



**Tijdschrift voor onderwijsresearch 1997 : een twee-
maandelijks tijdschrift voor onderzoekers van onderwijs,
gewijd aan theoretische, methodologische en professionele
problemen in de onderwijsresearch**

<https://hdl.handle.net/1874/214717>

Inhoud

Dit nummer bevat bijdragen met als speciaal thema:

META-ANALYSE

L.J.Th. van der Kamp en T.A.B. Snijders

Voorwoord / **pag. 2**

L.J.Th. van der Kamp en T.A.B. Snijders

De populaties van de meta-analyse / **pag. 3**

Tom A.B. Snijders

Meta-analyse via Multi-level modellen / **pag. 16**

J.J. Hox en E.D. de Leeuw

Een multiniveau meta-analyse naar sekse-verschillen in
leren / **pag. 29**

S. Severiens en G. ten Dam

Effecten van beeldende vakken op visueel-ruimtelijk
vermogen en esthetische waarneming: twee meta-
analyses / **pag. 42**

Folkert Haanstra

De relatie tussen leesprestaties en buitenschools lezen
in leerjaar 5-8 / **pag. 63**

M.E. Otter, R. Schoonen en K. de Gopper

Notities en Commentaren / **pag. 81**

Boekbesprekingen / **pag. 86**

Universiteit Utrecht
BIBLIOTHEEK CENTRUM UITHOF

BIBLIOTHEEK UNIVERSITEIT UTRECHT



2923 990 6

Voorwoord

L.J.Th. van der Kamp en T.A.B. Snijders

Vakgroep Methoden en Technieken Psychologie, Rijksuniversiteit Leiden

en Vakgroep Statistiek, Meettheorie en Informatica, Rijksuniversiteit Groningen

Over allerlei onderwerpen wordt door veel mensen in de wereld onderzoek gedaan. Voor wie in die onderwerpen geïnteresseerd is leidt dat tot allerlei informatie-problemen. Hoe houdt je het bij? In welke mate leiden verschillende onderzoeken tot dezelfde conclusies? Dit laatste is vaak maar in beperkte mate het geval. De recente discussies over de effecten van klassegrootte zijn hiervan een voorbeeld. In het ene onderzoek wordt wel, in het andere niet een significant effect gevonden. Nu zijn de resultaten van elk onderzoek op zich met onzekerheid behept: uit de onderzoeksverslagen zelf kunnen we niet aflezen, of er misschien fouten van de eerste of tweede soort zijn gemaakt, hoewel de statistiek ons vertelt dat dat om de haverklap moet gebeuren. Soms kunnen verschillen in uitkomsten van verschillende onderzoeken worden geweten aan het statistische toeval als bron van onzekerheid. Maar elk onderzoek is verschillend in toespitsing, operationalisatie en uitvoering, en ook daar kunnen verschillen in resultaten uit voortkomen.

Het is de bedoeling van meta-analyse om behulpzaam te zijn bij deze problematiek, en meer houvast te geven bij het maken van een synthese van eerder uitgevoerd onderzoek dan de traditionele soort van overzichtsartikelen. De cumulatie van onderzoeksbevindingen wordt op deze manier ondersteund. In een paar jaar tijd lijkt meta-analyse welhaast tot een rage te zijn geworden in de sociale, gedrags- en medische wetenschappen. Het aantal artikelen waarin over meta-analyses wordt gerapporteerd is groot, en de eerste meta-meta-analyses zijn niet uitgebleven. Dit themanummer van het *Tijdschrift voor Onderwijsresearch* kan dienen ter kennismaking voor wie nog weinig van meta-analyse weet, en tot een verder begrip voor wie de kennismaking al achter de rug heeft.

Van de vier artikelen zijn twee theoretisch van aard en leveren de andere twee interessante voorbeelden. Het artikel van T. Snijders, *De populaties van de meta-analyse*, behandelt de statistische motivering van de meta-analyse en enkele van de belangrijkste erbij gebruikte technieken. In de bijdrage van J. Hox en E. de Leeuw, *Multiniveau modellen voor meta-analyse*, wordt een recente aanpak van meta-analyse uit de doeken gedaan die wij zeer veelbelovend vinden. Deze aanpak wordt ook gebruikt in het eerste toepassingsartikel, waarin S. Severiens en G. ten Dam verslag doen over *Een meta-analyse naar sekseverschillen in leren in een multiniveau benadering*. Hierin blijkt dat, ondanks de verschillen tussen de 52 gebruikte studies, toch conclusies kunnen worden getrokken over verschillen in leerstijlen tussen mannen en vrouwen. Het andere toepassingsartikel, *Effecten van beeldende vakken op visueel-ruimtelijk vermogen en esthetische waarneming* van F. Haanstra, illustreert de meer traditionele aanpak van meta-analyse. Op grond van in totaal 69 studies concludeert dit artikel negatief wat betreft de aanwezigheid van een effect van beeldende vorming op visueel-ruimtelijke vermogens, maar positief ten aanzien van het effect op esthetische waarneming.

De populaties van de meta-analyse

Tom A.B. Snijders¹

ICS / Vakgroep Statistiek, Meettheorie en Informatica, Rijksuniversiteit Groningen

ABSTRACT

A conceptual introduction is given into ideas of meta-analysis. The focus is on the statistical assumptions made about the common denominator that justifies that a given set of studies is included in a meta-analysis. Five types of approach are distinguished: (A.) strictly parallel studies; (B.) studies testing a common null hypothesis; (C.) studies estimating a common parameter; (D.) a random sample from a population of studies; and (E.) a sample from a population of studies, structured by study characteristics that can serve as variables explaining outcome differences between studies. For practical purposes, approaches (B.) and (E.) are considered to be the most fruitful.

Trefwoorden: meta-analyse, replicatie, multiniveau-analyse, effectmaten, combinatie van toetsen.

1. META-ANALYSE

Meta-analyse is een verzamelterm voor statistische methoden voor de synthese van verspreide resultaten in de literatuur over een gemeenschappelijke vraagstelling. De grondgedachte daarachter is de volgende.

Rapportages van empirisch onderzoek in de vorm van statistische conclusies (uitkomsten van hypothesetoetsingen, parameterschattingen e.d.) zijn per definitie behept met onzekerheid. Bij hypothesetoetsing worden er fouten van de eerste en tweede soort gemaakt; aan parameterschattingen kleeft onnauwkeurigheid, wat door hun standaardfout wordt uitgedrukt. Wanneer meerdere onderzoeken over hetzelfde verschijnsel beschikbaar zijn, zouden de resultaten kunnen worden gecombineerd waarbij dan met deze onzekerheden rekening moet worden gehouden. Zo kunnen meerdere, niet steeds perfect overeenstemmende, conclusies worden getransformeerd tot één conclusie die met een kleinere mate van onzekerheid (dus kleinere kansen op fouten van de tweede soort, resp. kleinere standaardfouten) is behept dan de conclusies van de individuele onderzoeken.

Vragen die hierbij opkomen zijn de volgende:

1. Wanneer "mag" je verschillende onderzoeken samen nemen? Strikt replicatie-onderzoek is zeer zeldzaam, dus komt al gauw de vrees op dat appels en peren bij elkaar worden opgeteld, met als resultaat een onappetijtelijke vruchtenmoes.
2. Wanneer levert het statistisch instrumentarium iets op, dat je niet ook op meer eenvoudige manier had kunnen concluderen?
3. Hoe doe je het?

Aan een beantwoording van deze vragen is dit artikel gewijd. De nadruk zal meer liggen op conceptuele kwesties dan op hogeschooltechnieken. Voor de beantwoording van vraag 3 zal vooral naar de literatuur worden verwezen.

Populaties

De vraag over de zin van het samennemen van verschillende onderzoeken kan het best worden beantwoord door na te denken over wat er precies gemeenschappelijk kan zijn aan een aantal verschillende onderzoeken, en hoe dit gemeenschappelijke kan worden uitgedrukt in een statistisch model (men zou kunnen zeggen: een meta-model) dat al deze onderzoeken omvat. Zoals dat bij de statistiek behoort, maar vaak impliciet blijft, komen we hier ook terecht bij de vraag wat de populatie is waarop de statistische inferentie betrekking heeft. In oudere literatuur over meta-analyse, zoals bijv. Glass, McGaw en Smith (1981), wordt aan deze vraag volstrekt geen aandacht besteed. Het meer statistisch georiënteerde leerboek Hedges en Olkin (1985) noemt de populaties wel, maar tamelijk verstoppt in de wiskundige formulering van de gemaakte assumpties.

In dit artikel worden de volgende manieren onderscheiden om een aantal onderzoeken in een gemeenschappelijk kader te plaatsen.

- (A.) De onderzoeken zijn replicaties van elkaar. Vraagstelling en gebruikte instrumenten zijn identiek, de onderzochte populatie is steeds dezelfde, de onderzoeksdesigns sluiten naadloos op elkaar aan. Een zeldzame situatie! Als de oorspronkelijke data beschikbaar zijn, zouden deze kunnen worden samengevoegd; ook wanneer bijv. gemiddelden en covariantiematrices van elke studie kunnen worden gereconstrueerd, kan men een heel eind komen met het rechtstreeks samenvoegen van de studies en het analyseren ervan als één onderzoek. Strikt replicatie-onderzoek is zo zeldzaam dat ik er hier verder weinig woorden aan vul zal maken. Bovendien kunnen verschillende tijdstippen, verschillende onderzoeksplaatsen, enz. de veronderstelling van strikte replicaties hachelijk maken.
- (B.) De onderzoeken toetsen steeds dezelfde nulhypothese en er mag worden aangenomen dat ze onafhankelijk zijn. In dit geval kunnen methoden voor het combineren van onafhankelijke toetsingen worden toegepast. In paragraaf 2 wordt dit nader uitgewerkt.
- (C.) De onderzoeken zijn alle gericht op het schatten en toetsen van één en dezelfde parameter. Deze assumptie is weinig zwakker dan die in (A.), strikt replicatie-onderzoek. In de praktijk betekent dit ook dat er sprake is van één populatie van proefpersonen, die op wellicht verschillende manieren is onderzocht. Ook al is dit eveneens een zeldzame situatie, zal er toch in paragraaf 3 aandacht aan worden besteed.
- (D.) De onderzoeken kunnen worden beschouwd als een steekproef uit een populatie van studies. Verschillen tussen de onderzoeken wat betreft onderzochte populatie en onderzoeksmethode worden toegestaan, zolang er een basis van gemeenschappelijkheid is die kan worden uitgedrukt door het idee van een zinvol samenhangende populatie van studies. Hierbij is het noodzakelijk, aandacht te besteden aan de omschrijving van de populatie van studies, en de wijze waarop de onderzochte studies als "steekproef" uit deze populatie zijn verkregen. Eigenlijk is deze situatie vooral bedoeld als opstapje naar de volgende.
- (E.) De samengenomen studies kunnen worden beschouwd als een steekproef uit een populatie van studies, die bovendien gestructureerd zijn volgens kenmerken van de studies, die zelf als verklarende variabelen op studie-niveau kunnen worden gebruikt. Voor deze situatie is multiniveau-analyse een aantrekkelijke methode. Dit wordt behandeld in paragraaf 4.

Bij (B.) worden de minste assumpties gemaakt; hierbij is men dan ook beperkt tot vragen over het bestaan van een effect of samenhang, niet over de grootte van zo'n effect. Wat betreft assumpties staat (A.) aan het andere uiterste, met (C.) dicht erbij; hier zijn de assumpties zo zwaar dat het vaak twijfelachtig is, of er aan wordt voldaan. De situaties (D.) en vooral (E.) staan wat zwaarte van assumpties betreft in het midden en zijn voor de praktijk het meest vruchtbaar. Hierbij biedt (E.) een rijker structuur dan (D.), met ook ruimte voor onderzoek naar multivariate samenhangen.

Over de optredende populaties laat situatie (B.) zich niet verder uit, dan dat bij alle individuele studies populaties worden onderzocht. Bij situaties (A.) en (C.) berust het samenvoegen van de studies op de assumptie van een gemeenschappelijke populatie van proefpersonen, terwijl (D.) en (E.) uitgaan van een geneste situatie: een populatie van studies, waarbij elke studie betrekking heeft op een (zelfde of verschillende) populatie van proefpersonen.

De grote praktische problemen bij het verzamelen van materiaal voor meta-analyses, die in de literatuur (zie het eind van dit artikel, of ook de andere artikelen in het voor u liggende themanummer) worden genoemd, kunnen in dit kader worden beschouwd als voortvloeiend uit de nadere invulling en controle van de assumpties (A.) t/m (E.): wat is de omschrijving van de populatie van proefpersonen dan wel van studies, wat is de gemeenschappelijke nulhypothese of de gemeenschappelijke parameter, hoe kan men zorgen voor een verantwoorde verzameling ("steekproeftrekking") van studies, enz.

In het vervolg van dit artikel worden de vijf beschreven situaties, met de nadruk op (B.), (C.) en (E.), behandeld in de drie volgende paragrafen. Op de drie in het begin gestelde vragen kom ik terug in paragraaf 5.

2. COMBINEREN VAN HYPOTHESETOETSINGEN

Situatie (B.), waarin wordt gesteld dat in alle samen te voegen studies steeds dezelfde nulhypothese wordt getoetst, is een zeer terughoudende benadering van de meta-analyse. De gezamenlijke nulhypothese is vaak van zeer algemene aard, en in de individuele studies worden nadere specificaties ervan getoetst (bijv. voor specifieke uitkomstvariabelen, populaties, enz.). De alternatieve hypotheses mogen in de individuele studies op verschillende manieren zijn ingevuld; daarvoor wordt geen gemeenschappelijkheid geëist. De gemaakte assumptie over de gemeenschappelijkheid van de studies is niet zwaar, veelal kunnen simpele statistische methoden worden gebruikt voor de combinatie, en het resultaat is zelf ook weer beperkt tot een (gecombineerde) hypothesetoetsing. Voor het combineren van de toetsen is de enige verder gemaakte assumptie, dat de diverse onderzoeken onafhankelijk datamateriaal gebruiken.

De gecombineerde nulhypothese kan bijv. zijn: *klassegrootte (in het bereik van 20 tot 35 leerlingen) in het basisonderwijs heeft geen effect op leerprestaties*. Dit is een zeer algemeen geformuleerde nulhypothese, die in diverse studies op veel verschillende manieren zal zijn geconditioneerd, bijv. naar het soort leerprestaties, het leerjaar en naar verdere kenmerken van de onderwijssituatie, maar de gemeenschappelijke kern is de aangegeven nulhypothese. In het meta-onderzoek wordt nagegaan, of er in enige situatie een effect van klassegrootte is.

De meest terughoudende formulering van de meta-nulhypothese is als volgt. Er worden K studies gecombineerd; in de k -de studie wordt een nulhypothese $H_0^{(k)}$ getoetst. De gecombineerde nulhypothese is

$H_0^{(META)}$: alle nulhypotesen $H_0^{(1)}$ t/m $H_0^{(K)}$ zijn geldig.

Dit is een terughoudende formulering in de zin dat de gemeenschappelijkheid van de nulhypothese volgt uit de logische formulering ervan, en niet uit inhoudelijke overwegingen. Zo gauw tenminste één van de individuele nulhypotesen niet geldt, geldt ook de meta-nulhypothese niet, en is de wens dus om de meta-nulhypothese te verwerpen. In het voorbeeld: als er één conditie bestaat waaronder klassegrootte wel effect heeft, zou de meta-onderzoeker in zijn gecombineerde toetsing een significant resultaat wensen te vinden. Dit is een weliswaar beperkt doel, dat echter als screening van de resultaten betreffende een onderzoeksthema zeker zinvol kan zijn.

De klassieke methode voor deze gecombineerde nulhypothese is Fisher's methode voor het combineren van kansen (Fisher, 1932). Geef de overschrijdingskans (p -waarde) verkregen bij

de k -de toetsing aan met p_k . Dan is de gecombineerde toetsingsgrootheid

$$-2 \sum_{k=1}^K \ln(p_k) ;$$

deze wordt getoetst in de χ^2 verdeling met $2K$ vrijheidsgraden. Als één of enkele van de p -waarden zeer laag zijn, zal deze toetsingsgrootheid (door het minteken en de logaritmische transformatie) al een hoge waarde hebben. Dit leidt er toe, dat deze combinatiemethode een goed onderscheidingsvermogen heeft ook wanneer slechts één of enkele van de nulhypothese $H_0^{(k)}$ ongeldig zijn.

Er bestaan talloze andere methoden voor het combineren van onafhankelijke toetsingen. Een tweede veel gebruikte methode is de methode van Stouffer (Mosteller en Bush, 1954). Wanneer de k -de toetsing een resultaat geeft dat kan worden uitgedrukt als een Z -waarde Z_k , onder de nulhypothese standaard-normaal verdeeld, en alle toetsen zijn bijv. rechts-eenzijdig, dan is de gecombineerde toetsingsgrootheid

$$\frac{1}{\sqrt{K}} \sum_{k=1}^K Z_k ,$$

zelf ook weer te toetsen in een standaardnormale verdeling.

Wat betreft het vergelijken van deze twee combinatiemethoden moet allereerst worden opgemerkt dat als p_k bekend is, men Z_k kan uitrekenen en omgekeerd; bijv. met behulp van een tabel van de normale verdeling. Als de statistische effectgroottes (de werkelijke verwachtingswaarden van Z_k onder de alternatieve hypothese) voor alle onderzoeken ongeveer even groot zijn, is het onderscheidingsvermogen van Stouffers methode beter dan dat van Fishers methode; als de statistische effectgroottes van sommige onderzoeken heel laag zijn en van andere behoorlijk, heeft Fishers methode een hoger onderscheidingsvermogen (zie bijv. Marden, 1991).

Een voordeel van Stouffers methode is, dat deze eenvoudig kan worden gemodificeerd tot een methode waarin de onderzoeken verschillend worden gewogen, bijv. in overeenstemming met hun steekproefomvang, methodologische kwaliteit, of relevantie. Wanneer het gewicht voor de k -de studie wordt aangegeven met w_k , is de gecombineerde toetsingsgrootheid

$$\frac{\sum_{k=1}^K w_k Z_k}{\sqrt{\sum_{k=1}^K w_k^2}} .$$

Men kan bijv. nemen $w_k = \sqrt{n_k}$ de wortel uit de steekproefomvang. Maar eigenlijk is dit laatste een hybride benadering in het kader van het combineren van toetsen; deze gewogen versie van Stouffers methode is meer toepasbaar voor het combineren van parameterschattingen, dus als men de Z -waarde (bij benadering) beschouwt als een gestandaardiseerde schatting van een gemeenschappelijke parameter. Dit wordt toegelicht in paragraaf 3. Wanneer men strikt in het kader van het combineren van onafhankelijke toetsingsproblemen blijft, waarbij de effectgrootte van elk individueel onderzoek heel verschillend zou kunnen zijn, is op grond van het onderscheidingsvermogen de ongewogen methode van Stouffer te verkiezen boven deze gewogen methode (zie Schaafsma, 1966, par. 4.5).

Ter illustratie geef ik een voorbeeld uit de parapsychologie. Dit betreft onderzoek naar extra-sensorische informatie-overdracht in zogenaamd Ganzfeld-experimenten; zie Bem en Honorton (1994). (Een meer uitgebreide bespreking van het gebruik van meta-analyse ter evaluatie van parapsychologisch onderzoek wordt gegeven door Utts, 1991). Veronderstel dat de meta-onderzoeker geïnteresseerd is in de algemene nulhypothese, dat er geen sprake is van bui-

ten-zintuiglijke informatieoverdracht voor ongeoefende ("novice") proefpersonen. Van de door Bem en Honorton gerapporteerde resultaten gaan $K = 5$ onderzoeken over ongeoefende proefpersonen, met als (rechts-eenzijdige) Z_k -waarden $-0.30, 1.60, 0.67, 0.97$ en 2.20 . Dit is een typische rij uitkomsten voor meta-analyse: één uitkomst is (op zich genomen) significant, een andere heeft het "verkeerde" teken, en drie zijn wel positief maar niet significant. (Het aantal gecombineerde studies is wel aan de lage kant.) Fishers methode, waarvoor eerst de Z -waarden in p -waarden moeten worden omgezet, levert $\chi^2 = 21.67$, met $d.f. = 10, p = 0.017$. Stouffers methode levert $Z = 2.30$, met $p = 0.011$. Beide toetsen leveren dus de conclusie op, dat de nulhypothese wordt verworpen, oftewel dat niet in elk van de 5 onderzoeken de nulhypothese geldt. Er lijkt dus wel sprake te zijn van buiten-zintuiglijke informatieoverdracht.

3. COMBINEREN VAN PARAMETERSCHATTINGEN

Bij situatie (C.) wordt van de assumptie uitgegaan, dat in elk van de te combineren onderzoeken dezelfde parameter wordt geschat. Wiskundig gezien betekent dit dat de populatie-waarde van de parameter in alle onderzoeken hetzelfde getal is. Vanuit een inhoudelijke optiek is het zinvol te eisen, dat de betekenis van de parameter steeds hetzelfde is. Dit is een sterkere assumptie dan die van (B.). Uitgaande van (C.), zal men in elk deelonderzoek ook dezelfde hypothese toetsen, gewoonlijk dat deze parameter 0 is; dus (C.) impliceert (B.). Het omgekeerde is niet het geval: onderzoeken die alle gericht zijn op dezelfde nulhypothese van afwezigheid van een bepaald effect, behoeven niet noodzakelijk dit effect (als het wel bestaat, dus onder de alternatieve i.p.v. de nulhypothese) ook in dezelfde parameter uit te drukken. Verschillende operationaliseringen van het meetinstrument of de onderzoekspopulatie zullen over het algemeen leiden tot verschillende interpretaties van de geschatte parameter.

Assumptie (C.) is zo sterk dat men kan betwijfelen of er vaak aan wordt voldaan, maar heeft de aantrekkelijkheid van de conceptuele eenvoud. Deze benadering beschrijft het paradigma van veel meta-analyses. In de jaren '70 en '80 is bij de statistische uitwerking van de meta-analyse veel aandacht besteed aan het definiëren van effectmaten die vergelijkbaar zijn over diverse studies, en het ontwikkelen van zuivere schatters voor dergelijke effectmaten.

De meest basale effectmaat is wellicht Cohen's d voor het vergelijken van twee populaties. Deze is als populatie-maat gedefinieerd door

$$d = \frac{\mu_1 - \mu_0}{\sigma}$$

waarbij μ_1 en μ_0 de gemiddelden zijn in de twee populaties en σ de standaarddeviatie, waarvan wordt verondersteld dat hij in beide populaties dezelfde is. De reden voor deze laatste veronderstelling is vaak meer de statistische eenvoud dan het realiteitsgehalte ervan. Daarom heeft Glass voorgesteld het verschil van de gemiddelden te delen door de standaarddeviatie in de controlegroep, er van uitgaand dat er in alle onderzoeken sprake is van een vergelijkbare "onbehandelde" controlegroep. De formule voor de effectmaat van Glass is

$$\Delta = \frac{\mu_1 - \mu_0}{\sigma_0}$$

waarbij nu σ_0 de standaarddeviatie in de controlegroep is. In de meeste gevallen zal het in meta-analyses realistischer zijn om deze effectmaat te gebruiken, dan die van Cohen.

Een andere manier om te komen tot een gemeenschappelijke meetlat voor uitkomsten van verschillende studies wordt gebruikt door Rosenthal (1991). Hij stelt voor, om alle effectmaten te transformeren tot product-moment correlatiecoëfficiënten. Zo wordt bijv. een t -waarde voor het vergelijken van twee onafhankelijke steekproeven getransformeerd tot de corre-

latie tussen de uitkomstvariabele en de dummy-variabele die het groepslidmaatschap aangeeft. Naar mijn smaak leidt dit tot een tamelijk gewrongen aanpak: de correlaties liggen weliswaar op dezelfde schaal, maar ze zijn lang niet altijd een natuurlijke effectmaat. Andere gebruikte gemeenschappelijke maten zijn bijv., afhankelijk van het soort onderzoek, gestandaardiseerde regressie-coëfficiënten (Becker, 1992) en odds ratios. Nonparametrische effectmaten worden besproken door Hedges en Olkin (1984). Een overzicht over effectmaten met veel aandacht voor multivariate maten wordt gegeven door Tatsuoka (1993).

Doordat bij de effectmaten van Cohen en Glass gedeeld wordt door de standaarddeviatie verkrijgt men maten die dezelfde eenheid hebben: de eenheid is de standaarddeviatie van de uitkomstvariabele. Of er ook sprake is van dezelfde parameter, hangt af van de inhoud van de uitkomstvariabele en de omschrijving van de populaties. Stel bijvoorbeeld dat de meta-analyse betrekking heeft op het verschil in leerprestaties tussen kinderen op de basisschool in klassen met 32 leerlingen (beschouwd als controlegroep) en in klassen met 25 leerlingen. Hoe nauw moeten uitkomstmaat en populatie omschreven worden, opdat het aannemelijk is dat Glass' effectmaat in verschillende studies dezelfde parameter is, dus dezelfde inhoudelijke interpretatie en dezelfde getalwaarde heeft? Moet de uitkomstmaat worden beperkt tot prestaties op één vak? Moet de populatie worden omschreven als kinderen in dezelfde jaargroep, eventueel in klassen waar dezelfde onderwijsmethode wordt gevolgd? Moeten er restricties worden geplaatst op het soort buurt of de samenstelling van de klas wat betreft SES? Om overtuigend te kunnen stellen dat er sprake is van dezelfde parameter, dus dezelfde effectgrootte in alle studies, zullen er sterke beperkingen aan de op te nemen studies moeten worden gesteld.

Gebruikelijk is, om een twee-traps procedure te volgen voor de keuze van dit soort beperkingen. Allereerst wordt op grond van veldkennis een a priori inperking van de onderzoeksvraag (uitkomstmaat, populaties) gemaakt. Dit is de uiterst belangrijke eerste fase van de meta-analyse, die de criteria levert die worden gebruikt bij het verzamelen van studies. In het verlengde hiervan wordt een relevante parameter (effectmaat) gekozen, waarvan elke studie een schatting moet leveren; bijvoorbeeld een kans op succes indien de uitkomstvariabele dichotoom is, of Glass' effectmaat indien het gaat om het vergelijken van twee populaties met ongeveer normaal verdeelde criteria. Als tweede stap wordt na het verzamelen (volgens de gestelde criteria) van de studies, en het opzoeken of reconstrueren van de benodigde resultaten, statistisch getoetst of de studies inderdaad homogeen zijn, dus of de parameterwaarde wel steeds dezelfde is. Wanneer op deze manier is geconcludeerd dat een homogene verzameling studies is verzameld, worden de parameterschattingen ervan gecombineerd tot een gemeenschappelijke parameterschatting.

Het toetsen van homogeniteit is, begripsmatig gezien, een vreemde escapade. Anders dan gewoonlijk, is het hier namelijk de voorkeur van de onderzoeker om de nulhypothese (homogeniteit) niet te verwerpen! De onderzoeker heeft dus geen belang bij een toets met een hoog onderscheidingsvermogen... In andere woorden: het is de bedoeling om de nulhypothese, dat de studies homogeen zijn, te bewijzen. Maar hypothesetoetsing, met de bekende interpretatie dat een fout van de eerste soort ernstiger is dan een van de tweede soort, leidt nooit tot bewijs voor een nulhypothese. Een niet-significante homogeniteitstoets in een meta-analyse mag dus niet worden opgevat als een overtuigende ondersteuning van de assumptie van homogeniteit. De homogeniteitstoets is hoogstens een extra controle, maar geen vervanging van de inhoudelijk bepaalde inperking van de groep samengenomen onderzoekingen, als ondersteuning van de assumptie dat in alle studies een gemeenschappelijke parameter is onderzocht.

Om in de praktijk de schattingen te combineren, moeten eerst uit de gepubliceerde resultaten van de individuele studies de parameterschattingen en hun standaardfouten worden afgeleid. Diverse publicaties over meta-analyse (bv. Hedges en Olkin, 1985, Hunter en Schmidt, 1990, Rosenthal, 1991) geven hiervoor methoden; bijv. hoe uit *p*- of *t*-waarden, samen met andere (beschrijvende) kenmerken, zuivere schattingen van effectmaten zoals die van Cohen of Glass kunnen worden afgeleid. Vaak is het opsporen in de publicaties van de beno-

digde informatie een hels karwei; het blijkt niet zelden dat een artikel onvoldoende informatie bevat om de parameterschatting en standaardfout te berekenen. Als in de k -de studie een schatting $\hat{\theta}_k$ is gevonden met standaardfout SE_k , dan kan de bijdrage van deze studie worden gewogen met het gewicht $w_k = 1/SE_k^2$. Bijvoorbeeld wanneer de standaardfout omgekeerd evenredig is met de wortel uit de steekproefomvang, is deze weging w_k recht evenredig met de steekproefomvang. Deze gewichten leveren de kleinste variantie voor de gecombineerde schatting

$$\hat{\theta} = \frac{\sum_k w_k \hat{\theta}_k}{\sum_k w_k}$$

De standaardfout hiervan is, bij gebruik van de gewichten $w_k = 1/SE_k^2$,

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_k w_k}}$$

Dit komt overeen met de in de vorige paragraaf genoemde gewogen versie van Stouffers methode. De Z -waarde voor het toetsen van de nulhypothese θ in de k -de studie is namelijk

$$Z_k = \frac{\hat{\theta}}{SE_k}$$

Met de in deze en de vorige paragraaf aangegeven waarden voor de gewichten leveren de formules dat de gecombineerde Z -waarde uit par. 3 en de gecombineerde schatting uit deze paragraaf met elkaar overeenkomen volgens de formule

$$Z = \frac{\hat{\theta}}{SE(\hat{\theta})}$$

Als voorbeeld gebruik ik weer het bovengenoemde onderzoek van Bem en Honorton (1994). De vijf genoemde studies hadden dezelfde onderzoeksopzet, waarbij er een zender en een ontvanger van de buiten-zintuiglijke boodschap waren. De ontvanger moest kiezen tussen vier stimuli, waarvan één door de zender was uitgezonden. Het gaat hier per studie om het vergelijken van één populatie met een nul-model. Een geschikte effectmaat is de kans π op identificatie van de juiste stimulus; deze is 0.25 onder de nulhypothese. De laatste studie maakte gebruik van proefpersonen van een kunstacademie, terwijl de andere vier proefpersonen gebruikten die studenten waren van niet-artistieke richtingen. Omdat er theoretische vermoedens waren dat artistiek begaafde personen gevoeliger zijn dan anderen voor buiten-zintuiglijke waarneming, is het twijfelachtig dat de kans π voor de vijfde groep dezelfde is als voor de andere vier. Daarom worden hier voor het combineren van de parameterschattingen alleen de eerste vier studies samengenomen. De parameterschattingen zijn hier de vier geschatte kanssen (relatieve frekwenties), gegeven door 0.24, 0.36, 0.30 en 0.33, met steekproefomvang 50, 50, 50 en 36. Onder de assumptie van een gemeenschappelijke populatiekans zijn de standaardfouten omgekeerd evenredig met de wortel uit de steekproefomvang, en kan de steekproefomvang als gewicht worden gebruikt. Dit komt op hetzelfde neer als het beschouwen van alle 186 proefpersonen als een grote steekproef. (Zo zijn we hier dus toch weer terug in situatie (A.)! Dit illustreert dat assumptie (C.) bijna alleen overtuigend is wanneer sprake is van replicatie-onderzoek.) De gecombineerde schatting is nu $\hat{\pi} = 57/186 = 0.306$, met standaardfout (volgens de binomiale verdeling) 0.034. Het combineren van de schattingen levert een meer informatief resultaat op dan het combineren van de toetsingen: los van de significantie kan nu ook de relevantie worden beoordeeld. In vergelijking met de gokkans van 0.25 is de geschatte waarde 0.306 niet erg indrukwekkend te noemen. Bovendien is de schatting van 0.306 met een standaardfout van 0.034 niet significant hoger dan 0.25.

Tot slot van de bespreking van situatie (C.) is een samenvattende vergelijking met situatie (B.) op zijn plaats. In situatie (B.) is de nulhypothese steeds dezelfde, maar kunnen de alternatieve hypothesen betrekking hebben op verschillende parameters, en dus verschillend zijn. In onderzoek naar het effect van klassegrootte op leerprestaties kunnen in situatie (B.) studies met verschillende uitkomstmaten, onderwijsmethoden e.d. worden samengenomen omdat alleen de grove nulhypothese, "klassegrootte doet er niet toe", in de meta-analyse wordt onderzocht. In situatie (C.) wordt een veel sterkere homogeniteitseis gesteld. Deze aanscherping levert als statistische beloning allereerst meer informatie (de gecombineerde parameterschatting) maar ook een hoger onderscheidingsvermogen, dat wordt bereikt omdat er maar één parameter wordt getoetst in plaats van K verschillende.

4. GESTRUCTUREERDE POPULATIES VAN ONDERZOEKINGEN: MULTINIVEAU META-ANALYSE

In situatie (C.) wordt een assumptie gemaakt, die meestal kunstmatig is behalve wanneer er sprake is van replicatie-onderzoek. Vrijwel altijd is het realistischer om te erkennen dat er verschillen bestaan tussen de studies ten gevolge van verschillende specificaties van de onderzoeksvraagstelling en verschillende operationalisaties. Dan kan de onderzoeker de grootte van deze verschillen schatten en de mate van homogeniteit observeren, in plaats van bij voorbaat te postuleren dat de homogeniteit perfect is. Toch zal het volgens een of ander principe mogelijk moeten zijn, de diverse studies onder één noemer te brengen. Een vanuit de statistiek voor de hand liggende gedachte is geformuleerd als situatie (D.): de samen te voegen studies zijn een steekproef uit een populatie van studies. Een aselechte steekproef zal de meta-analyst vrijwel nooit trekken, maar het idee van een populatie van mogelijke studies (dat men als een metafoor mag opvatten) kan helderheid brengen in de vraag naar wat een verzameling studies in een meta-analyse gemeenschappelijk zou moeten hebben. Deze gedachte werd al voorgesteld door Cronbach (1980) in het kader van evaluatie-onderzoek.

Het opnemen van studies in de meta-analyse wordt nu gestuurd door het ideaal, dat de verzamelde studies als een steekproef uit een populatie van onderzoekingen kunnen worden beschouwd. De in (C.) genoemde eerste fase van de meta-analyse, waarin criteria voor het opnemen van studies worden vastgesteld, wordt nu beschouwd als (1) het omschrijven van de populatie van studies en (2) het vastleggen van de selectieprocedure (vergelijkbaar met de steekproefprocedure bij een "echte" steekproeftrekking). De homogeniteitseisen (inclusiecriteria voor studies) zullen iets zwakker mogen zijn dan in situatie (C.). Daar staat tegenover dat nog grotere zorgvuldigheid nodig is bij het zoeken naar studies die aan de inclusiecriteria voldoen. Door de grotere vaagheid van de inclusiecriteria is er immers ook meer ruimte voor selectiebias. Een van de meest besproken vormen van selectiebias is het bureaula-probleem ("file-drawer problem"): studies met een niet-significante uitkomst wat betreft de primaire onderzoeksvraag hebben een lagere publicatiekans dan studies met een wel significante uitkomst. (Bijv. vonden Lipsey en Wilson (1993) in een meta-analyse over meta-analyses dat in 92 meta-analyses de gemiddelde effect-grootte - volgens Glass of Cohen - in gepubliceerd onderzoek 0.53 was, en in ongepubliceerd onderzoek slechts 0.39.) Het kan daarom wenselijk zijn, de steekproef te laten bestaan uit niet alleen gepubliceerde artikelen maar ook ongepubliceerde rapporten. Een van de redenen, naast de noodzaak van goede omschrijving van de populatie van onderzoekingen, waarom een rapportage van een meta-analyse onvolledig is als hij niet zorgvuldig aangeeft hoe naar studies is gezocht, is dat zo kan worden gecontroleerd of er een risico bestaat van systematische overselectie van studies met een bepaald (bv. significant) resultaat.

Vanuit deze omschrijving van situatie (D.) is het maar een kleine stap naar situatie (E.). Gegeven dat verschillen tussen de studies worden toegelaten, is het wenselijk deze verschillen in de vingers te krijgen en ze door middel van kenmerken van de studies te verklaren. De

populatie van studies wordt dus gestructureerd aan de hand van verklarende variabelen. In de eerste fase van de meta-analyse moet de onderzoeker nu ook nagaan welke kenmerken (van de populatie, de uitkomstvariabele, de operationalisatie, het onderzoeksdesign, het jaartal van de studie, enz.) relevant zijn om als verklarende variabele voor de effectgrootte te gebruiken. Op het niveau van studies is nu sprake van een soort van regressie-analyse, waarbij de studies de cases zijn, de geschatte effectgrootte de afhankelijke variabele is, en de studie-kenmerken de verklarende variabelen zijn. Het aantal te gebruiken kenmerken zal, zoals bij regressie-analyse gebruikelijk is, in overeenstemming moeten zijn met het aantal verzamelde studies. Als er bijv. niet meer dan 10 of 20 studies worden samengenomen, zal het meestal gewenst zijn niet meer dan 1 tot 4 kenmerken te gebruiken.

In situatie (E.) is sprake van twee geneste niveaus van analyse-eenheden: proefpersonen binnen studies. Als sommige studies afkomstig zijn van dezelfde groep van onderzoekers ("laboratorium"), kan het zinvol zijn om drie niveaus te onderscheiden: proefpersonen zijn genest binnen studies, en deze weer binnen laboratoria. Dit is in principe net zo'n hiërarchisch geneste datastructuur als wordt geanalyseerd in de multiniveau-analyse (zie bijv. Bosker en Snijders (1990) voor een beknopte, Bryk en Raudenbush (1992), Goldstein (1995) of Hox (1994) voor een meer uitgebreide inleiding). Deze benadering noemt men ook wel de random effects aanpak van meta-analyse, omdat er toevalsverschillen tussen de studies in het model worden opgenomen. Het verschil met de gebruikelijke multiniveau-analyse is dat de meta-onderzoeker gewoonlijk niet beschikt over de data van de individuele proefpersonen, maar alleen over geaggregeerde (gepubliceerde) resultaten. Toch kunnen veel van de ideeën en technieken van de multiniveau-analyse hier worden toegepast, omdat situatie (E.) ook gemodelleerd kan worden met random effects-modellen. De verschillen in uitkomsten (geschatte effectgroottes) van de diverse studies worden in drie delen gesplitst: de steekproefvariantie binnen de studies, waarvan de omvang tot uiting komt in de binnen-studie standaardfouten; de door studie-kenmerken verklaarde verschillen tussen studies (gebruikt in een soort regressie-model); en de onverklaarde werkelijke verschillen tussen studies. Deze laatste zijn het gevolg van verschillen in specificatie van de onderzoeksvraag en in operationalisaties, die niet worden weerspiegeld door de gebruikte studie-kenmerken. In de multiniveau benadering bevat het statistisch model verschillende random effecten voor de steekproefvariabiliteit binnen de studies (met binnen de studies gemeten varianties) en voor de onverklaarde variabiliteit tussen de studies. Deze laatste wordt in benadering (C.) op 0 gesteld, en in benaderingen (D.) en (E.) uit de verzamelde data geschat. In hoofdstuk 9 en 10 van Hedges en Olkin (1985), hoofdstuk 7 van Bryk en Raudenbush (1992) en in de bijdrage van Hox en De Leeuw (1997; dit nummer) wordt het gebruik van multiniveau-analyse voor meta-analyse meer in detail behandeld. In Goldstein (1995, par. 3.7) staat een korte behandeling. Het voor u liggende tijdschriftnummer bevat verder in Severiens en Ten Dam (1997) een bijdrage waarin een voorbeeld hiervan wordt gegeven.

Deze benadering kan worden gecontrasteerd met de benadering van Glass, McGaw en Smith (1981), waarin regressie-analyse van onderzoeksresultaten wordt voorgesteld zonder dat er over populaties van onderzoeken wordt gesproken, en zonder het onderscheid te maken tussen werkelijke variatie en steekproefvariatie als bronnen van verschillen tussen studie-uitkomsten. Het wel maken van dit onderscheid, zoals gebeurt in de multiniveau-benadering, biedt meer helderheid.

In een originele bijdrage stelt Rubin (1992) dat het bepalen van een *gemiddelde* effectgrootte eigenlijk niet zo belangwekkend is. Wel relevant is de effectgrootte bij studies van *perfecte* methodologische kwaliteit, en hoe deze *afhangt* van inhoudelijke kenmerken van het onderzoek. In de praktijk variëren studies in methodologische kwaliteit, en worden er weinig of geen methodologisch perfecte studies uitgevoerd... Rubin bepleit om indicatoren van de methodologische kwaliteit van het onderzoek als studie-kenmerken mee te nemen, en te kijken naar effectgroottes die gelden wanneer de methodologische kwaliteit wordt geëxtrapoleerd tot het perfecte onderzoek. Als simpel voorbeeld zouden de steekproefomvang, de gemid-

delde betrouwbaarheid van de gebruikte meetinstrumenten en de non-responsfractie kunnen worden gebruikt als methodologische kwaliteitsindicatoren, waarbij dan de effectgrootte zou worden bestudeerd bij oneindige steekproefomvang, betrouwbaarheidscoëfficiënt 1, en non-responsfractie 0. Hier kan aan worden toegevoegd dat de methodologische kwaliteit niet alleen invloed kan hebben op de hoogte van de effectgrootte, maar ook tot een hogere onverklaarde tussen-studies variantie zou kunnen leiden. Dit kan in de multiniveau-analyse worden gerepresenteerd door de indicator voor methodologische kwaliteit niet alleen een vast maar ook een random effect te geven.

5. DISCUSSIE

Het grote voordeel van meta-analyse ten opzichte van het traditionele verbale literatuur-overzicht is natuurlijk de controleerbaarheid en de grotere precisie van de resultaten. Het belang van meta-analyse wordt geïllustreerd door Hedges, Laine en Greenwald (1994) die wél positieve verbanden vinden tussen financiële uitgaven op scholen en prestaties van leerlingen, in tegenstelling tot eerder (niet van meta-analyse gebruik makend) onderzoek. Een mega-voorbeeld is de meta-meta-analyse van Lipsey en Wilson (1993), waarin 302 meta-analyses worden gepresenteerd die tesamen een duidelijke ondersteuning bieden voor het effect van behandelingen op de gebieden van geestelijke gezondheid, organisaties, en onderwijs. Wie ervaring heeft met het uitvoeren van meta-analyse weet dat de door meta-analyse geleverde precisie ten koste gaat van grote inspanningen bij het reconstrueren van wat er in de verzamelde studies precies is gebeurd, en welke in getallen uitgedrukte resultaten er zijn behaald. Maar de aandacht voor meta-analyse levert meer op dan gepubliceerde meta-analyses.

De belangrijkste vrucht van de meer formele aanpak van literatuur-synthese zoals die in de meta-analyse ontwikkeld is, ligt misschien niet allereerst in alle ontwikkelde statistische technieken, maar vooral in het inzicht in de zorgvuldigheid die nodig is voor het samenvatten van resultaten uit de literatuur. Deze zorgvuldigheid komt tot uiting in vragen naar de in de verschillende studies gebruikte uitkomstvariabelen, de onderzochte populaties proefpersonen, en de samenvattende parameters; en naar de omschrijving van de in de meta-analyse onderzochte populatie van onderzoeken en de gebruikte methode van steekproeftrekking uit deze populatie. In de praktijk is het nagaan van de vergelijkbaarheid van studies op dergelijke criteria en het terugrekenen van gerapporteerde statistische resultaten naar in de meta-analyse bruikbare grootheden een zeer tijdrovend karwei. Soms zal de conclusie moeten worden getrokken dat de studies niet vergelijkbaar zijn, bijv. omdat te verschillende meetinstrumenten zijn gebruikt, of omdat de onderzoekspopulaties te verschillend zijn gedefinieerd; maar ook in zulke gevallen betekent het winst, dat men zorgvuldig heeft nagedacht over de overeenkomsten tussen de studies.

Ervaringen met meta-analyse, en met hoe moeilijk het is om uit een artikel dat te halen wat nodig is voor opname in de meta-analyse, leidt tot verder inzicht in eisen die aan onderzoeksrapportage (over "enkelvoudig" onderzoek) moeten worden gesteld. Opgenomen te worden in een meta-analyse is voor een onderzoeker immers een nobel streven, en vaak het hoogst bereikbare! Het alleen rapporteren van significante resultaten, of het alleen rapporteren van " $p \leq 0.05$ " of " $p > 0.05$ " is daarvoor onvoldoende. De onderzoeker zal meer nadruk moeten leggen op schattingen van effectgrootten en hun standaardfouten, dan op toetsingen; de volledige p -waarde moet worden gerapporteerd, of gemakkelijk te reconstrueren zijn; ook een goede beschrijving van de onderzoekspopzet is noodzakelijk voor opname in een meta-analyse.

Een andere vrucht van de aandacht voor meta-analyse is dat deze het besef bevordert, dat uitkomsten van onderzoek in hoge mate variabel zijn, wat veroorzaakt wordt door een combinatie van enerzijds toevalsfluctuaties en anderzijds verschillen in allerlei operationalisaties en (relevante of ondergeschikte) toespitsingen van het onderzoek. Toevalsfluctuaties kunnen ook verantwoordelijk zijn voor het feit dat het ene onderzoek wel, en het andere niet tot significantie leidt: het statistisch onderscheidingsvermogen (de "power") is

vaak zo matig, dat we doorlopend fouten van de tweede soort maken - helaas weten we niet, wanneer! Verschillen tussen onderzoeken wat betreft effectschattingen en wat betreft het al dan niet vinden van significantie behoeven dus niet als tegenstrijdigheden te worden geïnterpreteerd.

Van de in de inleidende paragraaf gestelde vragen is de eerste ("wanneer mag het?") de leidraad van dit artikel geweest. Verschillende gezichtspunten zijn mogelijk van waaruit diverse onderzoeken onder één noemer kunnen worden gebracht; deze zijn in dit artikel aan-gegeven met (A.) tot en met (E.). Voor praktisch gebruik pleit ik vooral voor de benaderingen (B.) en (E.), met name omdat bij deze benaderingen assumpties worden gemaakt die in allerlei situaties een realistisch gehalte hebben.

Bij (B.), het combineren van toetsen, worden geen andere veronderstellingen gemaakt dan dat het zinvol is om in de individuele studies getoetste nulhypothese $H_0^{(k)}$ samen te nemen en de gezamenlijke nulhypothese

$$H_0^{(META)} : \text{"alle nulhypothese } H_0^{(1)} \text{ t/m } H_0^{(K)} \text{ zijn geldig"}$$

te toetsen; en dat het datamateriaal voor de verschillende toetsingen onafhankelijk is. Dit is een zuinige benadering met een beperkt maar zinvol resultaat: gegeven dat sommige van de toetsingen wel (of "bijna") en andere niet een significant resultaat hebben opgeleverd, kan de meta-onderzoeker concluderen of de verzameling van alle onderzoeken een significante toets oplevert.

Bij (E.), de multiniveau benadering van meta-analyse, wordt bij voorbaat erkend dat er verschillen tussen de individuele studies bestaan, en blijkt pas uit de resultaten van de meta-analyse (dus achteraf) hoe groot de mate van gemeenschappelijkheid van de studies is. Natuurlijk ligt ook in deze benadering een cruciale rol bij het inzicht van de meta-analyst, dat van belang is bij de omschrijving van de populatie van studies, de keuze van effectmaten en de verzameling en bewerking van studies. Het toetsen van de homogeniteit van een aantal verzamelde studies, dat zo belangrijk is als controlemiddel bij benadering (C.) die er op mikt dat steeds een dezelfde effectparameter is geschat, en waaraan veel statistische aandacht is besteed, verdient hierbij zijn grote belang. De studies hoeven immers niet meer strikt homogeen te zijn. De rol van de homogeniteitstoets wordt teruggedrongen van een onmisbaar onderdeel tot een nuttige stap in een oriëntatie op het gemeenschappelijke karakter van de studies.

Het antwoord op de tweede in de inleiding gestelde vraag ("wanneer heb je er iets aan?"), is impliciet bevat in de voorafgaande paragrafen. Een noodzakelijke voorwaarde is dat een sluitende omschrijving is te geven van de criteria waaraan de samen te nemen studies moeten voldoen - volgens een van de benaderingen (A.) tot en met (E.), waarbij (B.) en (E.) de belangrijkste zullen zijn. Deze criteria liggen bij benadering (E.) enerzijds op het gebied van de populatie van studies (wat is het domein van deze meta-analyse?), anderzijds op dat van de steekproeftrekking van studies (hoe wordt er naar studies gezocht?). Het combineren van toetsen (B.) is al zinvol zo gauw er minstens twee studies gevonden zijn. Voor de multiniveau benadering (E.) moet de eis worden gesteld dat een behoorlijk aantal bruikbare studies beschikbaar is. Tien verzamelde studies is hiervoor al een heel laag aantal, en voor elk studie-kenmerk dat als verklarende variabele wordt gebruikt moet het aantal studies wat hoger worden (zoals gebruikelijk in een regressie-analyse).

Op de derde vraag, hoe je het doet, heb ik in het bestek van dit artikel nauwelijks in kunnen gaan. Enkele van de meest basale methoden voor het combineren van toetsen en van effectschattingen zijn in het bovenstaande genoemd. Wat betreft de multiniveau benadering zijn hoofdstuk 7 van Bryk en Raudenbush (1992) en Hox en De Leeuw (1997; dit nummer) goede inleidingen. Shadish (1996) geeft een toegankelijke behandeling van het gebruik van meta-analyse om causale processen op te sporen, waarin multiniveau-analyse met covariantie-structuur-analyse wordt gecombineerd. Een zeer beperkte keus van enkele andere publicaties die als inleiding kunnen worden gebruikt is de volgende.

Bangert-Drowns (1986): een niet-technische inleiding, waarin diverse benaderingen van meta-analyse kritisch met elkaar worden vergeleken.

Hedges (1992): een korte, meer statistisch-technische inleiding.

Cook et al. (1992): een waardevolle bundel die enkele algemene hoofdstukken over conceptuele en praktische aspecten van meta-analyse bevat, alsmede vier uitgebreide voorbeelden. Een van de voorbeeld-hoofdstukken gaat over het verschil tussen schoolprestaties van jongens en meisjes, en factoren die dit verschil beïnvloeden.

Rosenthal (1991): een beknopt en goed toegankelijk leerboek over meta-analyse. Nuttig omdat allerlei technieken aan de orde komen, maar met de manier waarop de auteur alles naar correlatie-coëfficiënten probeert toe te praten ben ik het niet eens, en ook niet steeds met zijn evaluatie van diverse technieken.

Hunter en Schmidt (1990): een uitgebreid leerboek, met nadruk op correlatie en artefact-correctie. Het boek bevat veel stellingnames en reacties op anderen, zowel op auteurs die meta-analyse bedrijven op een manier die H. & S. niet delen, als op auteurs die in het algemeen sterk kritisch staan ten opzichte van meta-analyse. Ik vind het jammer dat H. & S. in de bespreking (blz. 518-519) van verschillen tussen studies niet een populatie-begrip gebruiken.

Hedges en Olkin (1985): een goed, sterk wiskundig georiënteerd leerboek. Deze auteurs geven aan populaties van studies dezelfde plaats die er in het nu eindigend artikel aan wordt gegeven.

LITERATUUR

- Becker, B.J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics*, 17, 341-362.
- Bangert-Drowns, R.L. (1986). Review of developments on meta-analytic method. *Psychological Bulletin*, 99, 388-399.
- Bem, D.J., en Honorton, C. (1994). Does Psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, 115, 4-18.
- Bosker, R.J., Snijders, T.A.B. (1990). Statistische aspecten van multiniveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-329.
- Bryk, A.S., en Raudenbush, S.W. (1992). *Hierarchical linear models: applications and data analysis methods*. Newbury Park, CA: Sage.
- Cook, T.D., Cooper, H., Cordray, D.S., Hartmann, H., Hedges, L.V., Light, R.J., Louis, T.A., en Mosteller, F. (1992). *Meta-analysis for explanation. A casebook*. New York: Russel Sage Foundation.
- Cronbach, L.J. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Fisher, R.A. (1932). *Statistical methods for research workers* (4th ed.). Edinburgh: Oliver & Boyd.
- Glass, G.V., McGaw, B., en Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Goldstein, H. (1995). *Multilevel Statistical Models* (2nd ed.). London: Edward Arnold.
- Hedges, L.V. (1992). Meta-analysis. *Journal of Educational Statistics*, 17, 279-296.
- Hedges, L.V., Laine, R.D., en Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher* 23(3), 5-14. (Met een discussie erover in het volgende nummer, 23(4), 5-10.)
- Hedges, L.V., en Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, 96, 573-580.
- Hedges, L.V., en Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hox, J.J. (1994). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.
- Hox, J.J., en de Leeuw, E.D. (1997). Multiniveau modellen voor meta-analyse. *Tijdschrift voor Onderwijsresearch*. (Dit nummer.)
- Hunter, J.E. en Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park: Sage.
- Marden, J.I. (1991). Sensitive and sturdy p-values. *The Annals of Statistics*, 19, 918-934.
- Mosteller, F., en Bush, R.R. (1954). Selected quantitative techniques. In G. Lindzey (ed.), *Handbook of social psychology, Vol. 1. Theory and method* (p. 289-334). Cambridge, MA: Addison-Wesley.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park: Sage.

- Rubin, D.B. (1992). Meta-analysis: Literature synthesis or effect-size surface estimation? *Journal of Educational Statistics*, 17, 363-374.
- Schaafsma, W. (1966). *Hypothesis testing problems with the alternative restricted by a number of inequalities*. (Dissertatie R.U. Groningen). Groningen: Noordhoff.
- Severiens, S., en ten Dam, G. (1997). Een meta-analyse naar sekseverschillen in leren in een multi-niveau benadering. *Tijdschrift voor Onderwijsresearch*. (Dit nummer.)
- Shadish, W.R. (1996). Meta-analysis and the exploration of causal mediating processes: a primer of examples, methods, and issues. *Psychological Methods*, 1, 47-65.
- Tatsuoka, M. (1993). Effect Size. In G. Keren en C. Lewis (eds.), *Data Analysis in the Behavioral Sciences* (p. 461-479). Hillsdale, NJ: Lawrence Erlbaum Ass.
- Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, 6, 363-403.

1) Ik ben dank verschuldigd aan Ivo Molenaar, Leo van der Kamp en anonieme beoordelaars voor hun commentaar op een eerdere versie, dat heeft geleid tot verhelderingen in dit artikel.

Ontvangen 12-08-96
Definitief aanvaard 20-12-96

Meta-analyse via multiniveau-modellen

J.J. Hox en E.D. de Leeuw

Faculteit Pedagogische en Onderwijskundige Wetenschappen, Universiteit van Amsterdam¹

ABSTRACT

An important goal in meta analysis is to integrate the outcomes of several studies into one single result. This assumes that all studies are replications that differ by sampling fluctuation only; the studies are then said to be homogeneous. If the study outcomes differ by more than chance fluctuation, they are said to be heterogeneous, and the variation in outcomes should be further analyzed. An elegant method to analyze outcome variation is multilevel regression, because it enables us to include study characteristics as explanatory variables. Such study characteristics may be theoretically interesting variables, but also methodological characteristics to correct for artefacts. Various significance tests are available to assess the importance of study variables and to assess if the results are homogeneous or not.

Trefwoorden: meta-analyse, multiniveau-analyse, homogeniteitstoets, artefacten

INLEIDING

Kenmerkend voor meta-analyse is dat men een statistische analyse uitvoert op de uitkomsten van eerdere empirische onderzoeken. Het doel daarbij is het samenvatten van de resultaten van meerdere onafhankelijke studies die zich alle met dezelfde onderzoeksvraag bezighouden. Bijvoorbeeld: "Wat is het effect van een sociale vaardigheidstraining op sociaal-angstige kinderen?" Men verzamelt dan de verslagen van experimenten op dit terrein en vat de uitkomsten statistisch samen tot één 'superuitkomst' (zie ook het artikel van Snijders in dit nummer). Vaak is men echter ook geïnteresseerd in meer gedetailleerde vraagstellingen, zoals "Heeft de lengte van de training effect op de uitkomst?", of, "Heeft de samenstelling van de groep invloed op het effect van de training?". In dat geval codeert en analyseert men niet alleen de uitkomst van de studies, maar ook specifieke studiekekenmerken, zoals duur van de training, samenstelling van de experimentele en controle groep, sekse van de trainer, etcetera.

Een zeer algemeen model voor meta-analyse is het random effects model (Hedges & Olkin, 1985, 189 e.v.). Dit model gaat er van uit dat de uitkomsten (effecten) in de verschillende studies niet alleen van elkaar verschillen door toevallige steekproeffluctuaties, maar ook als gevolg van reële verschillen tussen de studies, bijvoorbeeld wat betreft het soort steekproef, de experimentele procedure, of de gebruikte instrumenten. Het doel van de meta-analyse is dan niet zozeer het combineren van alle uitkomsten tot één 'superuitkomst,' maar vooral het analyseren van de spreiding in de uitkomsten. Door de verschillen in de uitkomsten te relateren aan de gecodeerde studiekekenmerken kunnen we op het spoor komen van mogelijke moderator-variabelen, en zo een verklaring vinden voor verschillen tussen studies. In het random effects model wordt binnen de spreiding van de geobserveerde uitkomsten onderscheid gemaakt tussen toevallige steekproeffluctuaties rond de populatieparameter en systematische variatie van de populatieparameter zelf tussen de studies. Hedges en Olkin (1985, p. 194) geven formules om de variantie van de uitkomsten op te splitsen in ware parametervariantie en

toevallige steekproefvariantie, en om de significantie van de parametervariantie te toetsen. Wanneer de parametervariantie groot is, dan noemt men de uitkomsten *heterogeen*. In dat geval wordt meestal aangeraden de analyse voort te zetten door de studies te verdelen in clusters die onderling verschillen, maar intern *homogeen* zijn. De clustering van de studies kan vooraf gedaan worden op basis van de beschikbare studiekenmerken, bijvoorbeeld de duur van de gegeven training of de samenstelling van de groep, maar ook achteraf door een clusteranalyse op de uitkomsten. De verschillende clusters van studies worden vervolgens apart geanalyseerd met het random effecten model, in de hoop dat de gevormde clusters intern homogeen zijn. Men probeert zo inhoudelijk interessante studiekenmerken op het spoor te komen die de heterogeniteit van de uitkomsten verklaren.

Zoals ook Snijders (dit nummer) aangeeft, is er bij meta-analyse sprake van twee geneste niveaus van analyse-eenheden: personen binnen studies. Wanneer we zouden kunnen beschikken over de ruwe gegevens van alle verschillende studies dan zouden we de meta-analyse kunnen uitvoeren door middel van een 'gewone' multiniveau-analyse. In ons voorbeeld hebben we dan één afhankelijke variabele (een test voor sociale vaardigheid) en één onafhankelijke variabele (de indeling in de experimentele trainingsgroep of de controlegroep) op het individuele niveau, en een lineair regressiemodel voor het verband tussen deze beide. Het algemene multiniveau-regressiemodel stelt dat iedere studie zijn eigen regressiemodel heeft. Multiniveau-analyse wordt vervolgens gebruikt om het gemiddelde en de spreiding van de regressiecoëfficiënten over alle studies tezamen te schatten. Wanneer de spreiding van de regressiecoëfficiënten substantieel is, hebben we heterogene regressiehellingsen, en is het zinvol om groepskenmerken (bij meta-analyse studiekenmerken zoals de groepssamenstelling, maar ook bijvoorbeeld de methodologische kwaliteit van de betreffende studie) als verklarende variabelen in te voeren om daarmee de verschillen tussen de studies te verklaren. Het een en ander is dan een multiniveau-analyse die niet wezenlijk anders is dan de in onderwijskundig onderzoek gebruikelijke multiniveau-analyses op leerlingen binnen klassen. (Voor een inleiding in het multiniveau-regressiemodel zie bijv. Bosker en Snijders, 1990, en Hox, 1995; statistische details zijn te vinden in Bryk en Raudenbush, 1992 en Goldstein, 1995).

Bij meta-analyse kunnen we echter doorgaans *niet* beschikken over de ruwe gegevens van de individuen uit de oorspronkelijke studies, maar uitsluitend over de uitkomsten van de uitgevoerde analyses, zoals p-waarden, gemiddelden, standaardafwijkingen, of correlaties. In de klassieke meta-analyse zijn dan ook een groot aantal methoden ontwikkeld om uitkomsten van studies te combineren. Hunter en Schmidt (1990) geven een uitgebreide beschrijving van algoritmen en rekenregels, terwijl Hedges en Olkin (1985) een grondig overzicht geven van de statistische achtergronden. Voor een inleiding zie het artikel van Snijders in dit nummer.

Het is echter wel degelijk mogelijk om een multiniveau-regressieanalyse uit te voeren op de gebruikelijke meta-analyse gegevens. Raudenbush en Bryk (1985; Bryk and Raudenbush, 1992) beschrijven het random effects model als een speciaal geval van het algemene multiniveau-regressiemodel. Omdat de analyse niet wordt uitgevoerd op ruwe data, maar op daarvan afgeleide statistische grootheden, moet de analyse enigszins worden aangepast. Bij het programma HLM wordt daarvoor een speciale versie meegeleverd (VKHLM, Bryk, Raudenbush & Congdon, 1994), terwijl in MLn de benodigde aanpassing door de gebruiker zelf moet worden geprogrammeerd (Lambert & Abrams, 1995; zie ook de appendix bij dit artikel). Het voordeel van het toepassen van het multiniveau-regressiemodel en de daarvoor ontwikkelde software in plaats van de meer gebruikelijke meta-analytische methoden en software ligt in de flexibiliteit van de multiniveau-benadering. In het multiniveau-regressiemodel is het relatief eenvoudig om de gemiddelde uitkomst en de spreiding van de uitkomsten te schatten. Indien men *van tevoren* verklarende variabelen op studieniveau kan definiëren dan kunnen deze studiekenmerken gericht in het model worden opgenomen. Ook voor een post hoc verklaring van heterogene studieuitkomsten kunnen verklarende variabelen in het regressiemodel worden opgenomen. Bij het exploratief invoeren van verklarende variabelen in het regressiemodel is de mogelijkheid van kanskapitalisatie uiteraard altijd aanwezig.

HET MODEL

De verschillende studies die in een meta-analyse worden opgenomen gebruiken dikwijls sterk uiteenlopende instrumenten en statistische toetsen. Om de uitkomsten vergelijkbaar te maken worden daarom de resultaten zo goed mogelijk vertaald in een standaard effectmaat, zoals een correlatie of het gestandaardiseerde verschil tussen twee gemiddelden d . Het model voor de studie-uitkomsten is

$$d_j = \delta_j + e_j \quad (1)$$

waarin d_j de uitkomst is van studie j ($j=1, \dots, J$), δ_j de populatiewaarde van deze uitkomst, en e_j de steekproeffout van studie j . Verondersteld wordt dat de steekproeffout e_j een normale verdeling heeft waarvan de variantie σ_j^2 bekend is (op basis van de steekproefgrootte in studie j). Wanneer de steekproefgroottes van de afzonderlijke studies niet te klein zijn, dan is de steekproefverdeling van de uitkomsten in veel gevallen bij benadering normaal met een bekende variantie. Hedges en Olkin (1985, p. 175) suggereren een minimum van 20; Bryk en Raudenbush (1992, p. 157) houden het op 30. Deze veronderstelling is overigens ook van toepassing op andere gangbare meta-analyse technieken (vgl. Hedges & Olkin, 1985). Tabel 1 geeft een overzicht van enkele veel gebruikte uitkomstmaten en de variantie van de bijbehorende steekproefverdeling.

Tabel 1. Enige schatters van uitkomstmaten en hun steekproefvariantie.^{a)}

uitkomst	gebruikelijke schatter	normaliserende transformatie	steekproefvariantie schatter (na transformatie)
verschil gemiddelden spreiding (st.afw.)	$g = (\bar{Y}_E - \bar{Y}_C) / s$	onnodig	$(n_E + n_C) / (n_E n_C) + g^2 / (2(n_E + n_C))$
correlatie	r	$s^* = \ln(s) + 1 / (2df)$	$1 / (2df)$
proportie	p	$0.5 \ln((1+r)/(1-r))$	$1 / (n-3)$
		$\ln(p/(1-p))$	$1 / (np(1-p))$

a) Naar Bryk & Raudenbush (1992, p. 169). g is de uitkomstmaat voorgesteld door Glass (1976). s^* is een transformatie van de standaard afwijking s voorgesteld door Bryk en Raudenbush (1992). Dergelijke transformaties worden dikwijls toegepast om de steekproefverdeling te normaliseren. Bij de rapportage van de eindresultaten dient de inverse transformatie te worden toegepast. Voor andere schatters van uitkomstmaten en hun steekproefvariantie zie Hedges & Olkin (1985).

Uit de vergelijking voor model (1) blijkt dat de parameters δ_j (de uitkomsten) verondersteld worden te variëren over de studies. De variantie van de populatiewaarden δ_j (de parameter-variantie) wordt verklaard door het model

$$\delta_j = \gamma_0 + \gamma_1 Z_{1j} + \gamma_2 Z_{2j} + \dots + \gamma_p Z_{pj} + u_j \quad (2)$$

waarin $Z_1 \dots Z_p$ studiekenmerken zijn, $\gamma_0 \dots \gamma_p$ de regressiecoëfficiënten, en u_j de residuele voorspellingsfout waarvan we aannemen dat die normaal verdeeld is met variantie σ_u^2 . Wanneer we (2) substitueren in (1) dan krijgen we het volledige model

$$d_j = \gamma_0 + \gamma_1 Z_{1j} + \gamma_2 Z_{2j} + \dots + \gamma_p Z_{pj} + u_j + e_j \quad (3)$$

Wanneer er één studiekenmerken als predictoren in het model zijn opgenomen, dan wordt model (3)

$$d_j = \gamma_0 + u_j + e_j \quad (4)$$

Model (4), het zogenaamde 'intercept-only' model (cf. Bryk & Raudenbush, 1992), is equivalent aan het random effecten-model voor meta-analyse zoals beschreven door Hedges en Olkin (1985), zij het dat de gebruikelijke multiniveau-software een andere schattingsprocedure gebruikt dan de techniek die Hedges en Olkin beschrijven. (Hedges en Olkin stellen als schattingsmethode Weighted Least Squares voor in plaats van de bij multiniveau-analyse gebruikelijke Maximum Likelihood, maar beide methoden leiden doorgaans tot vergelijkbare schatters.)

In model (4) is de intercept γ_0 de schatting voor de gemiddelde uitkomst over de studies, de variantie van de steekproeffouten e_j oftewel σ_e^2 is de steekproeffluctuatie, en σ_u^2 is de variantie van de residuen u_j ofwel de parametervariantie. Daarmee is σ_u^2 een indicatie voor de mate van heterogeniteit van de uitkomsten, en de statistische toets van de nulhypothese dat σ_u^2 gelijk is aan nul equivalent aan de gebruikelijke homogeniteitstoetsen in de meta-analyse (voor een beschrijving van deze homogeniteitstoetsen zie Hedges & Olkin, 1985, p. 194 en Raudenbush, 1994). De proportie ware parametervariantie wordt gegeven door $\sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$.

Naast de formele toets kan de proportie ware parametervariantie gebruikt worden als een indicator voor de heterogeniteit van de uitkomsten. Dit is analoog aan de goede gewoonte om bij klassieke regressieanalyse niet alleen te kijken naar statistische significantie, maar ook naar verklaarde variantie. Binnen de meta-analyse wordt aangeraden de homogeniteitstoets voorzichtig te gebruiken. Zo wijzen Hunter en Schmidt (1990) er op dat de gebruikelijke chi-kwadraat-toets voor de homogeniteit bij het samenvatten van een groot aantal studies een groot onderscheidingsvermogen heeft. Met andere woorden, als de nulhypothese van homogeniteit verworpen wordt, dan wijst dit niet noodzakelijk op sterke heterogeniteit. Omgekeerd hoeft de afwezigheid van significantie niet altijd te betekenen dat de tussen-studies variatie verwaarloosbaar is. Hunter en Schmidt (1990) suggereren als vuistregel dat het zoeken naar interessante moderatorvariabelen pas zin heeft wanneer de proportie ware parametervariantie groter is dan 25 procent.

In het meer algemene model (3) zijn studiekenmerken Z_{pj} opgenomen om de verschillen in de uitkomsten te verklaren. In dit model is σ_u^2 de variantie van de residuen u_j nadat de studiekenmerken verdisconteerd zijn, ofwel de residuele parametervariantie. Ook deze kan op significantie worden getoetst. Wanneer de residuele parametervariantie significant is, dan zijn de ingebrachte studiekenmerken niet in staat alle variaties in de uitkomsten te verklaren. Het verschil tussen de parametervariantie in het intercept-only model (model 4) en de residuele parametervariantie in het model met verklarende studievariabelen kan worden opgevat als de hoeveelheid variantie die door de studievariabelen verklaard wordt.

Het multiniveau-model zoals beschreven in model (3) vertoont sterke overeenkomsten met het algemene lineaire model voor vaste effecten zoals besproken door Hedges en Olkin (1985, hoofdstuk 8). Hedges en Olkin gebruiken een afwijkende (matrix-)notatie, in de hier gebruikte notatie is hun model gelijk aan ons model (4) zonder de modelterm u_j . Het lineaire model voor vaste effecten blijkt daarmee een speciaal geval te zijn van hier beschreven multiniveau-model, waarin verondersteld wordt dat de residuele parametervariantie gelijk is aan nul.

De ervaring leert dat, wanneer er sprake is van heterogeniteit, deze doorgaans niet volledig kan worden wegverklaard door de beschikbare studie-variabelen; de residuele parametervariantie wordt zelden nul. Hunter en Schmidt (1990) stellen dat dit het gevolg is van een groot aantal artefacten, zoals o.a. de correctheid van de statistische assumpties in de oorspronkelijke studies, verschillen in betrouwbaarheid van de gebruikte instrumenten, coderingsproblemen en andere onvolkomenheden bij de meta-analyse zelf. Het is onwaarschijnlijk

lijk dat voor al deze verschillen verklarende studie-variabelen beschikbaar zijn; het uitgangsmateriaal (de rapporten en andere publikaties) is daarvoor doorgaans te summier. Heterogene resultaten zijn daarom eerder regel dan uitzondering, en het multiniveau-model is daarom in het algemeen te verkiezen boven het lineaire model met vaste effecten.

VOORBEELD EN VERGELIJKING MET KLASSIEKE META-ANALYSE

In deze paragraaf analyseren we een voorbeeld dataset met 'klassieke' meta-analysemethoden zoals geïmplementeerd in het meta-analyse programma van Schwarzer (1989). Dit programma is grotendeels gebaseerd op de methoden besproken door Rosenthal (1984), Hunter & Schmidt (1990) en Hedges & Olkin (1985). De dataset bestaat uit de (gefingeerde) resultaten van twintig studies waarin een experimentele groep en een controlegroep met elkaar vergeleken worden.

Een gebruikelijke effectmaat bij experimenteel onderzoek is het gestandaardiseerde verschil tussen de gemiddelden van de experimentele en de controlegroep, door Hedges en Olkin gedefinieerd als $g = (\bar{Y}_E - \bar{Y}_C) / s$, waarbij s de gepoolde standaarddeviatie is. Omdat g een niet geheel zuivere schatter is van het populatie-effect $\delta = (\mu_E - \mu_C) / \sigma$, geven Hedges en Olkin (1985, p. 81) de voorkeur aan een hiervan afgeleide maat d : $d = (1 - 3 / (4(n_E + n_C) - 9))g$. De steekproefvariantie van de effectgrootte d is $\sigma_d^2 = (n_E + n_C) / (n_E n_C) + \delta^2 / [2(n_E + n_C)]$ (Hedges & Olkin, 1985, p. 86). De steekproefvariantie is dus afhankelijk van de onbekende populatiewaarde δ . Doorgaans wordt hiervoor de steekproefwaarde d genomen, waarna σ_d^2 bekend wordt verondersteld. Analooq hieraan is de steekproefvariantie van g gelijk aan $\sigma_g^2 = (n_E + n_C) / (n_E n_C) + g^2 / [2(n_E + n_C)]$. In tabel 2 zijn van de twintig studies zowel de g als de d opgenomen; bij gebruikelijke steekproefgroottes blijken dergelijke correcties gelukkig in de praktijk weinig uit te maken.

Tabel 2. Gesimuleerde voorbeeldgegevens twintig studies

ID	weken	g	d	var(d)	p	n_{exp}	n_{con}	n_{tot}	betr.
1	3	-.268	-.264	.086	.810	23	24	47	.900
2	1	-.235	-.230	.106	.756	18	20	38	.750
3	2	.168	.166	.055	.243	33	41	74	.750
4	4	.176	.173	.084	.279	26	22	48	.900
5	3	.228	.225	.071	.204	29	28	57	.750
6	6	.295	.291	.078	.155	30	23	53	.750
7	7	.312	.309	.051	.093	37	43	80	.900
8	9	.442	.435	.093	.085	35	16	51	.900
9	3	.488	.476	.149	.116	22	10	32	.750
10	6	.628	.617	.095	.030	18	28	46	.750
11	6	.660	.651	.110	.032	44	12	56	.750
12	7	.725	.718	.054	.003	41	38	79	.900
13	9	.751	.740	.081	.009	22	33	55	.750
14	5	.756	.745	.084	.009	25	26	51	.900
15	6	.768	.758	.087	.010	42	17	59	.900
16	5	.938	.922	.103	.005	17	29	46	.900
17	5	.955	.938	.113	.006	14	31	45	.750
18	7	.976	.962	.083	.002	28	26	54	.900
19	9	1.541	1.522	.100	.0001	50	16	66	.900
20	9	1.877	1.844	.141	.00005	31	14	45	.750

De voorbeeld dataset bevat ook één theoretisch interessante verklarende variabele op het studie-niveau: het aantal weken dat de experimentele conditie heeft geduurd. Als de experimentele manipulatie effect heeft, dan mag verwacht worden dat het effect groter is naarmate de experimentele manipulatie langer heeft kunnen werken. Tabel 2 bevat voor elk van de twintig studies de tijdsduur van de experimentele manipulatie (in weken), de beide effectmaten (g en d), de steekproefvariantie van d , de eenzijdige overschrijdingskans van de betreffende studie gebaseerd op de gebruikelijke t -toets voor het verschil tussen twee gemiddelden, de steekproefgroottes van de experimentele en de controlegroep en van de studie als geheel, en de betrouwbaarheid van het gebruikte meetinstrument. Terwille van de leesbaarheid zijn in tabel 2 de studies geordend naar grootte van het gevonden effect.

Binnen de klassieke meta-analyse zijn verschillende benaderingen mogelijk, die vaak aanvullend op elkaar gebruikt worden. Een al heel oude benadering is het combineren van de *p*-waarden. De gecombineerde *p*-waarde (methode van Stouffer, zie Snijders in dit nummer) is in ons voorbeeld $p < 0.001$ ($Z = 7.73$). Doorgaans wordt dit geïnterpreteerd als een bewijs dat het gezochte effect bestaat. Strikt genomen hebben we hiermee echter alleen aangetoond dat tenminste een van de betrokken nulhypotheseën niet geldt (zie het artikel van Snijders in dit nummer). Meer informatief is het om voor elke studie een effectgrootte te berekenen en deze te combineren tot een enkele uitkomst. Wanneer we daarbij de mogelijkheid open willen houden dat de studies niet alle hetzelfde populatie-effect schatten dan kiezen we bij voorkeur voor het combineren van effectgrootten volgens het random effecten model. Een meta-analyse van de effectmaten d in Tabel 2 volgens dit model schat het gemiddelde effect $\delta = 0.58$, met een standaardfout van 0.11 ($p < 0.001$). Volgens het random effecten model verschilt het gemiddelde effect dus significant van nul. We moeten echter oppassen, want de homogeniteitstoets leidt tot de conclusie dat de uitkomsten heterogeen zijn ($\chi^2_{19} = 48.9$, $p < 0.001$).

De parametervariantie wordt geschat als 0.17, en de proportie parametervariantie wordt geschat op 0.65. Dat is aanmerkelijk groter dan de ondergrens van 0.25 die Hunter en Schmidt hanteren voor 'interessante' variantie. Het loont dus de moeite om te proberen de verschillen in de uitkomsten te verklaren. De gebruikelijke methode binnen klassieke meta-analyse volgens het random effects model is om te trachten de studies onder te verdelen in clusters die onderling wel homogeen zijn. Vaak wordt hiervoor een eendimensionale clusteranalyse op de uitkomsten gebruikt. Vervolgens tracht men op basis van de beschikbare studiekenmerken te verklaren waardoor de oorspronkelijke heterogeniteit is ontstaan. In ons voorbeeld levert een clusteranalyse op de effectmaat d drie clusters op, waarvan het eerste bestaat uit de studies 1 en 2, het tweede uit de studies 3 tot en met 18, en het derde uit de studies 19 en 20. Voor een interpretatie dienen we antwoord te geven op de vraag waarin de studies in het ene cluster zich onderscheiden van de studies in het andere cluster. Eén mogelijkheid is de duur van de experimentele ingreep. De gemiddelde duur van de manipulatie is 2 weken in het eerste, 6 weken in het tweede, en 9 weken in het derde cluster. Het lijkt er op dat de duur van de experimentele manipulatie de uitkomsten beïnvloedt.

Omdat we de hypothese hebben, dat de duur van de experimentele behandeling (het aantal weken) van invloed is, kunnen we ook direct een a-priori indeling maken gebaseerd op deze studievariabele. We maken twee clusters: het eerste cluster bestaat uit de negen studies van vijf weken of korter, het tweede cluster uit de elf studies van zes weken of langer. In het eerste cluster wordt het gemiddelde effect geschat op 0.33 ($se = 0.15$, $p = 0.01$), en in het tweede op 0.77 ($se = 0.15$, $p < 0.001$). In studies die langer duren is het effect kennelijk groter. De hypothese van homogeniteit wordt echter in beide clusters verworpen; de indeling in twee groepen van kortdurende en langer durende studies verklaart niet alle parametervariantie. In het eerste cluster is de geschatte proportie parametervariantie gelijk aan 0.61, in het tweede cluster is deze gelijk aan 0.69. Hoewel we geen formele toets hebben uitgevoerd voor het verschil in de gemiddelde uitkomst tussen beide groepen studies, lijkt de conclusie gerechtvaardigd dat de lengte van het experiment inderdaad een verklarende factor is voor de verschil-

len in de studieuitkomsten. De proportie parametervariantie lijkt echter vooral in het tweede cluster niet te verwaarlozen, de bevinding dat deze niet significant is zou ook het gevolg kunnen zijn van een zwakke statistische toets wegens het kleine aantal studies. Enige voorzichtigheid is op zijn plaats bij de gegeven interpretatie.

Een meta-analyse van de uitkomsten in Tabel 2 volgens het multiniveau 'intercept-only' model levert vrijwel dezelfde resultaten als de hierboven uitgevoerde klassieke meta-analyse volgens het random effecten model. Het gemiddeld effect (de intercept γ_0) wordt geschat als $\delta=0.58$, met een standaardfout van 0.11 ($p<0.001$). De nulhypothese van homogene uitkomsten wordt verworpen: de parametervariantie wordt echter met 0.14 iets lager geschat als bij het random effecten model ($\chi^2_{19}=49.8, p<0.001$). De proportie ware parametervariantie wordt geschat op 0.61.

De kracht van de multiniveau-benadering toont zich wanneer we de heterogeniteit aan een nader onderzoek onderwerpen. We kunnen het aantal weken dat het experiment heeft geduurd als verklarende variabele in de analyse betrekken, waarbij het bovendien niet langer nodig is deze variabele te dichotomiseren, zoals we hierboven in de klassieke meta-analyse deden. De resultaten staan in tabel 3.

Tabel 3. Resultaten random effecten model en multiniveau regressieanalyses op voorbeeldgegevens tabel 2.^{a)}

Model	random effecten	multiniveau nulmodel	multiniveau met predictor weken
delta/intercept weken behandeling	.58 (.11)	.58 (.11)	-.22 (.20) .14 (.03)
parameter variantie σ_u^2	.17	.14	.04
sig. via χ^2 toets ^a	$p<.001$	$p<.001$	$p=.09$

a) Tussen haakjes de standaardfout. Voor de bepaling van de significantie van de variantiecomponenten is deze niet gebruikt. De significantie van de parametervariantie is bepaald met een heterogeniteitstoets overeenkomend met de in de meta-analyse gebruikelijke Q-toets van Hedges & Olkin (1985). Zie de bespreking in de discussie.

Wanneer de tijdsduur van het experiment in het model wordt opgenomen (zie model 3), dan blijkt dat deze variabele een significant effect heeft. De bijbehorende regressiecoëfficiënt heeft een waarde van $\gamma_1=0.14$ ($p<0.001$), hetgeen inhoudt dat voor elke week dat het experiment duurt het effect gemiddeld met 0.14 toeneemt. In dit model is de intercept $\gamma_0=-0.22$ ($p=0.20$). Het intercept verschilt dus niet significant van nul. Inhoudelijk is dat goed te begrijpen. In een regressievergelijking is de intercept immers de geschatte waarde van het criterium wanneer alle predictoren gelijk zijn aan nul. In ons geval wordt geschat dat als de experimentele manipulatie nul weken heeft geduurd, de experimentele en de controlegroep niet significant van elkaar verschillen. De residuele parametervariantie wordt geschat op 0.04 ($p=0.09$). Van de oorspronkelijke parametervariantie van 0.14 wordt dus 71% verklaard door de duur van de experimentele manipulatie. De proportie residuele parametervariantie is 0.29; volgens de maatstaven van Hunter en Schmidt dus qua grootte niet geheel verwaarloosbaar.

Aangezien het effect groter wordt naarmate het experiment langer duurt, is het niet zinvol om alleen een gemiddelde effectgrootte te rapporteren. Eén mogelijkheid is om de geschatte effectgrootte te vermelden voor experimenten met een verschillende tijdsduur, of om te bepalen welke tijdsduur minimaal nodig is om een significant effect te bereiken. Het een en ander is eenvoudig uit te rekenen.

Wanneer we de conclusies van de klassieke meta-analyse en de multiniveau-benadering vergelijken, blijkt dat de resultaten in dezelfde richting wijzen. Dat is niet geheel verrassend, conceptueel lijken meta-analyse volgens het random effecten model en de multiniveau-benadering erg op elkaar. De multiniveau-benadering maakt het echter veel eenvoudiger om verklarende variabelen in de analyse op te nemen, en om te toetsen of na het opnemen van de verklarende variabelen nog significante tussen-studies variantie over is. In de volgende paragraaf wordt van de eerste mogelijkheid gebruik gemaakt om de meta-analyse te corrigeren voor twee mogelijke artefacten.

CORRECTIES VOOR ARTEFACTEN

In de meta-analyse wordt, met name door Hunter en Schmidt, dikwijls gepleit voor het corrigeren van effectmaten voor een aantal artefacten (Hunter en Schmidt, 1990, 1994). Een gebruikelijke correctie is de correctie voor attenuatie als gevolg van de onbetrouwbaarheid van het meetinstrument. Hiertoe wordt de effectmaat eenvoudig gedeeld door de wortel uit de betrouwbaarheidscoëfficiënt, waarna de verdere analyse op de gebruikelijke manier wordt uitgevoerd. Deze correctie is gelijk aan de in de psychometrie bekende correctie voor attenuatie. Hunter en Schmidt (1990) beschrijven nog een groot aantal andere correcties. Het probleem met deze benadering is dat de 'ware' effectgrootte die op deze manier geschat wordt vrijwel altijd groter is dan de geobserveerde effectgrootte. Wanneer bijvoorbeeld de oorspronkelijke meetinstrumenten een lage betrouwbaarheid hebben dan leidt de correctie voor onbetrouwbaarheid dus tot zeer grote effectgroottes. Omdat deze grote effecten niet daadwerkelijk geobserveerd zijn, zijn de door Hunter en Schmidt voorgestelde correcties dan ook enigszins omstreden. Schwarzer (1989, p. 56), die de correctie voor onbetrouwbaarheid standaard in zijn programma's heeft ingebouwd, raadt dan ook aan de gecorrigeerde resultaten alleen te rapporteren samen met de ongecorrigeerde.

Een nadeel van het vooraf corrigeren van de effectmaten voor allerlei artefacten, zoals onbetrouwbaarheid van de meetinstrumenten, is dat dergelijke correcties vrij grof zijn, en dat de gegevens a-priori worden aangepast. Wanneer de gebruikte, gerapporteerde betrouwbaarheid onnauwkeurig is, dan is de uitgevoerde correctie onjuist. Zo is de meest gebruikelijke maat voor de betrouwbaarheid, coëfficiënt alfa, een onderschatting van de betrouwbaarheid, zodat de voorgestelde correctie bij deze maat doorgaans een overcorrectie zal zijn. De correcties hebben bovendien invloed op de geschatte standaardfouten, zodat het niet inzichtelijk is wat het cumulatief effect van een aantal correcties is.

Een andere mogelijkheid om voor artefacten te 'corrigeren' is de betrouwbaarheid van de gebruikte meetinstrumenten als predictor in het multiniveau-model op te nemen, in het vaste of ook in het random-gedeelte van het model. Dit is niet optimaal, daar de theoretisch juiste correctie voor onbetrouwbaarheid formeel een multiplicatief model vereist in plaats van het additieve multiniveau-model. Het voordeel van deze werkwijze is echter dat het effect van de onbetrouwbaarheid van de metingen aan de hand van de voorliggende gegevens geschat wordt, zodat het probleem van de overcorrectie niet aan de orde is. Ook andere artefacten kunnen direct gecodeerd worden en als predictor in het model worden opgenomen.

Een specifieke variatie op het invoeren van artefact-variabelen als predictor is het onderzoeken van het effect van de totale steekproefgrootte. In de klassieke meta-analyse raden bijvoorbeeld Light en Pillemer (1984, zie ook Light et al., 1994) aan om de effectmaten in een plot af te zetten tegen de bijbehorende steekproefgroottes. Wanneer de verzameling studies

'well-behaved' is, dan hoort deze plot de vorm van een horizontaal liggende trechter te hebben: studies met een kleinere steekproefgrootte zijn meer variabel, maar schatten wel dezelfde populatieparameter. Wanneer vooral bij de kleinere studies veel onverwacht grote effecten gevonden worden, dan is dit een aanwijzing dat er wellicht selectief gepubliceerd is, en dat er ergens een denkbeeldige bureaulade met ongepubliceerde niet significante studies is. Dit staat bekend als het 'file-drawer' probleem. Behalve door inspectie van een trechterplot, kan deze hypothese ook formeel onderzocht worden door de totale steekproefgrootte als predictor op te nemen in een multiniveau-analyse.

We illustreren de werkwijze bij correctie voor onbetrouwbaarheid en steekproefgrootte weer aan de hand van de voorbeeldgegevens. Tabel 2 bevat de benodigde gegevens wat betreft de studiegrootte en de betrouwbaarheid van het gebruikte instrument.

Wanneer we volgens de methoden van de klassieke meta-analyse te werk gaan dan komen we tot de volgende resultaten. Na correctie voor onbetrouwbaarheid (standaard correctie voor attenuatie, zie Schwarzer, 1989) wordt de gecombineerde effectgrootte geschat op 0.64 in plaats van de ongecorrigeerde waarde van 0.58. De parametervariantie wordt geschat op 0.23 in plaats van de ongecorrigeerde waarde van 0.17. De correctie voor onbetrouwbaarheid leidt er dus toe dat de gecombineerde effectgrootte iets groter wordt, en de parametervariantie eveneens toeneemt. Het effect van de steekproefgrootte is in de klassieke meta-analyse moeilijk te onderzoeken anders dan door een trechterplot. Een trechterplot voor de gegevens uit Tabel 2 (hier niet opgenomen) laat op het eerste gezicht geen duidelijke trend zien.

Wanneer we een multiniveau-analyse toepassen en de verschillende studiekekenmerken elk apart opnemen als predictorvariabelen dan krijgen we de resultaten in tabel 4.

Tabel 4. Resultaten multiniveau regressieanalyses op voorbeeldgegevens tabel 3.^{a)}

Model:	(1)	(2)	(3)	(4)	(5)
Predictor:	intercept only	(1) plus N_{tot}	(1) plus betrouw. b.	(1) plus weken	(1) plus alle predictoren
intercept	.58 (.11)	.50 (.50)	.15 (1.23)	-.22 (.20)	.42 (.92)
N_{tot}		.001 (.01)			-.004 (.01)
betrouw. b.			.51 (1.47)		-.51 (1.18)
weken behandeling				.14 (.03)	.15 (.04)
parameter variantie σ_u^2	.14	.16	.16	.04	.05
sig. via χ^2 toets ^a	$p < .001$	$p < .001$	$p < .001$	$p = .09$	$p = .07$

a) Tussen haakjes de standaardfout. Voor de bepaling van de significantie van de variantiecomponenten is deze niet gebruikt. De significantie van de parametervariantie is bepaald met een heterogeniteitstoets overeenkomend met de in de meta-analyse gebruikelijke Q-toets van Hedges & Olkin (1985). Zie de bespreking in de discussie.

In Tabel 4 is het 'intercept only' model het basismodel dat de gemiddelde effectgrootte schat als 0.58 en de tussen-studies variantie als 0.14. Model 1 bevat de correctie voor de steekproefgrootte, model 2 de correctie voor de onbetrouwbaarheid, en model 3 het effect van de duur van de experimentele conditie als predictor. In model 4 zijn alle drie predictoren simul-

taan opgenomen. Zowel uit de afzonderlijke als uit de simultane analyses blijkt dat alléén de predictor 'duur van de experimentele behandeling' een significant effect heeft. Geconcludeerd kan worden dat verschillen in meetbetrouwbaarheid en in steekproefgrootte geen ernstige bedreiging van de validiteit van onze conclusies vormen. Het ontbreken van een significant verband tussen de studiegrootte en de effectgrootte geeft aan dat publikatie-selectie (file-drawer problem) vermoedelijk geen ernstige bedreiging is voor de validiteit van onze conclusie over het effect van de experimentele ingreep.

Het laatste model (model 4), waarin alle predictoren simultaan zijn opgenomen, is om andere redenen nog instructief. De schatting van de regressiecoëfficiënt voor de betrouwbaarheid (weliswaar niet significant) is daar negatief. Dat zou betekenen dat studies met een betrouwbaarder meetinstrument juist een kleiner effect rapporteren. Dat is wonderlijk, en ook tegengesteld aan de regressiecoëfficiënt in het model waar de betrouwbaarheid als enige predictor optreedt. Het is dan ook een suppressor-effect, veroorzaakt door matig positieve correlaties (van 0.25 tot 0.33) tussen de drie predictoren. Model (2), waarin alleen betrouwbaarheid als predictor is opgenomen, impliceert dat als de betrouwbaarheid stijgt van 0.75 naar 0.90, de effectgrootte met $(0.51) \times (0.15) = 0.08$ toeneemt. Dit lijkt aardig op de correctie van 0.06 zoals bepaald volgens de conventionele methode (zie hierboven). De grootte van de standaardfouten voor de regressiecoëfficiënten van de artefact-variabelen in tabel 4 (deze zijn in vergelijking met de grootte van de regressiecoëfficiënten zelf uitermate groot) maakt echter duidelijk dat de betreffende correctie kennelijk overbodig is, zo niet misleidend.

SAMENVATTING EN DISCUSSIE

Het gebruik van multiniveaumethoden bij meta-analyse heeft het grote voordeel dat het gemakkelijk is studiekenmerken in de analyse op te nemen als mogelijke verklarende variabelen voor verschillen in de uitkomsten. Deze studiekenmerken kunnen bedoeld zijn als een theoretisch interessante verklaring voor de geobserveerde verschillen, maar ze kunnen ook gebruikt worden om voor potentiële artefacten te corrigeren. Door significantietoetsen op de individuele regressiecoëfficiënten uit te voeren, en door te toetsen of na het opnemen van studievariabelen de residuele parametervariantie gelijk is aan nul, kan worden nagegaan welke variabelen mogelijk interessant zijn, en of de opgenomen variabelen alle verschillen verklaren, of slechts een deel.

Multiniveau-regressieanalyse veronderstelt dat de verbanden additief en lineair zijn. Verder wordt verondersteld dat de geanalyseerde effectmaten een normale verdeling volgen (dit geldt overigens ook voor de meeste conventionele schattingsmethoden in de klassieke meta-analyse, zie Hedges en Olkin, 1985). Voorzover aan deze aannamen niet voldaan is, kan soms door een transformatie een oplossing worden gevonden; zowel Hedges en Olkin (1985) als Bryk en Raudenbush (1992) bespreken er enige. Ook wordt verondersteld (opnieuw; dit geldt ook voor de conventionele schattingsmethoden) dat in de samengevatte studies zelf de statistische aannamen opgaan. Wanneer in de samengevatte studies parametrische methoden gebruikt worden, terwijl de gegevens niet normaal verdeeld zijn, dan introduceert dit 'ruis,' in de vorm van onverklaarbare parametervariantie. Onder andere om deze reden stellen Hunter en Schmidt (1990) voor om bij heterogeniteit de parametervariantie pas te analyseren wanneer deze groter is dan een kwart van de totale variantie.

Een interessante uitbreiding van het multiniveau model voor meta-analyse is het opnemen van meerdere uitkomstmaten per studie. Er is dan sprake van een multivariate meta-analyse. Een voorbeeld hiervan is als bij onderwijskundig onderzoek zowel de resultaten op een taaltest als een rekentest als afhankelijke variabelen gebruikt zijn. Doorgaans worden dan twee meta-analyses uitgevoerd, één voor de taaltest en één voor de rekentest. Wanneer het inhoudelijk zinvol is, dan ligt het voor de hand beide uitkomstmaten in een simultane analyse op te

nemen. Raudenbush en Bryk (1985) beschrijven reeds multiniveau modellen voor meerdere uitkomstmaten, waarbij er echter van uitgegaan wordt dat alle uitkomstmaten in *alle* studies beschikbaar zijn. Wanneer dit niet het geval is en er missing data optreden, kunnen meerdere uitkomstmaten nog steeds in één analyse worden opgenomen, maar wordt het model gecompliceerder. We verwijzen hiervoor naar Kalaian & Raudenbush (1996) en Goldstein (1995).

Voor de feitelijke analyse zijn momenteel twee programma's beschikbaar: een speciale versie van HLM (VKHLM, zie Bryk et al., 1994) die met de PC versie van het programma standaard wordt meegeleverd (voor een beschrijving zie de appendix), en MLn (Rasbash & Woodhouse, 1995). Bij MLn moet het model op een speciale manier worden gespecificeerd (voor een beschrijving zie de appendix). Een klein probleem is dat de verschillende programma's niet *precies* hetzelfde doen. VKHLM gebruikt als schattingsmethode Restricted Maximum Likelihood (REML), hetgeen iets andere uitkomsten geeft dan het conventionele random effecten-model, terwijl MLn de onderzoeker de keuze laat tussen Full en Restricted Maximum Likelihood (die in MLn resp. IGLS en RIGLS genoemd worden). De hier gerapporteerde resultaten zijn bij wijze van controle berekend met beide programma's; de verschillen tussen de parameterschattingen blijken minimaal.

Een groot verschil tussen VKHLM en MLn treedt echter op bij de toets op de significantie van de residuele parametervariantie. MLn berekent voor die variantie een standaardfout gebaseerd op de Maximum Likelihood (ML) methode, die gebruikt kan worden voor een Z-toets van die variantie, terwijl VKHLM een chi-kwadraat-toets uitvoert op de OLS-residuen. Bryk en Raudenbush (1992) beargumenteren dat de in HLM en VKHLM ingebouwde chi-kwadraat-toets superieur is aan de toetsingsmethode in MLn (en ook VARCL). Een toets gebaseerd op het delen van een schatter door de bijbehorende standaardfout veronderstelt namelijk dat deze schatter normaal verdeeld is, hetgeen voor varianties niet opgaat, met name niet wanneer deze klein zijn. Bovendien is de ML-toets asymptotisch, dat wil zeggen dat een goede werking bij kleine aantallen studies onzeker is. Een bijkomend voordeel van de chi-kwadraat-toets bij het uitvoeren van een meta-analyse is dat deze bij het intercept-only model identiek is aan de conventionele chi-kwadraat-toets van het conventionele random effecten-model. De hierboven gerapporteerde variantietoetsen zijn gebaseerd op de chi-kwadraat-toets (het is in MLn betrekkelijk eenvoudig om de chi-kwadraat-toets zelf te programmeren).

Op het terrein van de multiniveau-analyse in het algemeen bestaat een toenemende belangstelling voor Bayesiaanse methoden voor de schatting van effecten en standaardfouten. Hierbij wordt gebruik gemaakt van rekenintensieve methoden zoals Markov Chain Monte Carlo methoden (MCMC, een voorbeeld hiervan is de Gibbs sampler). Met name bij meta-analyse hebben MCMC methoden enkele voordelen (DuMouchel, 1990, 1994). Bayesiaanse schatters voor de varianties zijn minder gevoelig voor de problemen die ontstaan wanneer de varianties dicht bij nul liggen en/of geschat moeten worden op basis van een klein aantal studies. Verder is het bij MCMC methoden mogelijk om in het model een functie op te nemen die de kans op publikatie van het resultaat modelleert. In principe levert dit een elegante benadering op om het file-drawer probleem aan te pakken. Toepassing hiervan bij meta-analyse is bepaald nog niet standaard, en gebruikersvriendelijke software hiervoor ontbreekt vooralsnog. Een algemeen programma voor MCMC modellen is het programma BUGS (Spiegelhalter, 1995), dat voor meta-analyse gebruikt is door o.a. Tweedie et al. (1994).

REFERENTIES

- Bosker, R.J. & Snijders, T.A.B. (1990). Statistische aspecten van multiniveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-329.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Bryk, A.S., Raudenbush, S.W. & Congdon, R.T. (1994). *HLM 2/3. Hierarchical linear modeling with the HLM/2L and HLM/3L programs*. Chicago: Scientific Software International.

- DuMouchel, W.H. (1990). Bayesian meta-analysis. In: D. Berry (ed), *Statistical methodology in pharmaceutical sciences*. New York: Marcel Dekker.
- DuMouchel, W.H. (1994). *Hierarchical Bayesian linear models for meta-analysis*. Washington, DC: National Institute of Statistical Sciences.
- Glass, G.V. (1976). Primary, secondary and meta analysis of research. *Educational Researcher*, 10, 3-8.
- Gleser, L.J. & Olkin, I. (1994). Stochastically dependent effect sizes. In: H. Cooper en L.V. Hedges (eds.), *The handbook of research synthesis*. New York: Russel Sage Foundation.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hox, J.J. 1995. *Applied multilevel analysis. (Second edition)*. Amsterdam: TT-Publikaties.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Hunter, J.E. & Schmidt, F.L. (1994). Correcting for sources of artifact variation. In H. Cooper & L.V. Hedges (eds.), *The handbook of research synthesis*. New York: Russel Sage Foundation.
- Kalaian, H.A. & Raudenbush, S.W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, 1, 227-235.
- Lambert, P.C. & Abrams, K.R. (1995). Meta-analysis using multilevel models. *Multilevel Modelling Newsletter*, 7, 17-19.
- Light, R.J. & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Light, R.D., Singer, J.D. & Willet, J.B. (1994) The visual presentation and interpretation of meta-analyses. In: H. Cooper & L.V. Hedges (eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Rasbash, J. & Woodhouse, G. (1995). *MLn command guide*. London: Multilevel Models Project, University of London.
- Raudenbush, S.W. (1994). Random effects models. In: H. Cooper & L.V. Hedges (eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Raudenbush, S.W. & Bryk, A.S. (1985). Empirical Bayes meta analysis. *Journal of Educational Statistics*, 10, 75-98.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Schwarzer, R. (1989). *Meta-analysis programs. Program manual*. Berlijn: Institut für Psychologie, Freie Universität Berlin.
- Spiegelhalter, D. (1995). Bayesian analysis of complex random effects models. In T.A.B. Snijders, B. Engel, J.C. van Houwelingen, A. Keen, G.J. Stemerink & M. Verbeek (red). *Toeval zit overal: Programmatuur voor random coëfficiënt modellen*. Groningen: iec ProGamma.
- Tweedie, R.L., Scott, D.J., Biggerstaff, B.J. & Mengersen, K.L. (1994). Bayesian meta-analysis, with application to studies of ETS and lung cancer. Fort Collins, CO: Colorado State University, intern rapport.

Ontvangen 12-8-96

Definitief aanvaard 20-12-96

Appendix: Software voor multiniveau meta-analyse

Het eenvoudigste programma voor een multiniveau meta-analyse is het programma VKHLM (voor Variance Known HLM) dat bij het programma HLM (Bryk et al., 1994) wordt meegeleverd. VKHLM verwacht per studie een regel data waarin achtereenvolgens staan: de effectgrootte, de bijbehorende steekproefvariantie, en de verklarende studievariabelen. VKHLM werkt overeenkomstig aan HLM: het model wordt interactief gespecificeerd en VKHLM berekent vervolgens de parameters en standaardfouten, alsmede een groot aantal significantietoetsingen. Wanneer VKHLM wordt gebruikt om een 'intercept-only' model te specificeren wijken de resultaten nauwelijks af van die van het random effecten model in het programma META van Schwarzer (1989), mits er rekening mee gehouden wordt dat META de ingevoerde g 's intern transformeert naar d 's (de transformatie is eerder toegelicht in de tekst).

Ingewikkelder is het gebruik van het programma MLn. De datastructuur is analoog aan VKHLM: de effectgrootte, de bijbehorende standaardfout (de wortel uit de steekproefvariantie), de constante (VKHLM voegt die automatisch toe) en de verklarende studievariabelen.

belen. De opzet van de analyse is als volgt. We onderscheiden twee niveaus: niveau 1: de uitkomsten, en niveau 2: de studies. De predictor 'steekproeffout' wordt alleen opgenomen in het random deel op niveau 1 (de uitkomsten) met een coëfficiënt vastgezet op één (met RCONstrain), en niet in het gefixeerde deel. De constante wordt alleen opgenomen in het gefixeerde deel op niveau 1, en niet in het random deel op niveau 1. De constante wordt op het 2e niveau (de studies) normaal in het random deel opgenomen. Verklarende studievariabelen worden alleen opgenomen in het vaste deel. De toets op de variantie van de constante is bij MLn de asymptotische toets. De chi-kwadraat-toets uit Bryk en Raudenbush (1992) heeft als formule: $\chi^2 = \sum((d_j - \hat{d}_j) / \text{sed}_j)^2$, waarbij d_j het waargenomen effect en \hat{d}_j het voorspelde effect is ($d_j - \hat{d}_j$ is daarmee het residu) en sed_j de standaardfout van het effect in studie j . Het aantal vrijheidsgraden is $J - q - 1$, waarbij J het aantal studies is en q het aantal geschatte regressiecoëfficiënten. (De MLn code hiervoor is: PRED C20; CALC C20=((('d'-C20)/'sed')^2); SUM C20 to B1; CPRO B1 df. Deze code veronderstelt dat in de spreadsheet kolom C20 ongebruikt is, en dat de kolommen aangeduid met 'd' en 'sed' de d_j 's resp. de sed_j 's bevatten).

HLM en MLn zijn verkrijgbaar via het interuniversitaire expertisecentrum ProGamma (gamma.post@gamma.rug.nl). De meta-analyseprogramma's van Schwarzer zijn (inclusief uitvoerige documentatie) als freeware beschikbaar via Internet (http://fub46.zedat.fu-berlin.de:8080/~ahahn/meta_e.htm). Spiegelhalter's programma Bugs is eveneens verkrijgbaar via Internet (<http://www.mrc-bsu.cam.ac.uk/pub/methodology/bugs>). Alle genoemde programma's zijn beschikbaar voor MSDOS, enkele ook voor andere computerplatforms.

1) Faculteit POW, Universiteit van Amsterdam, Wibautstraat 4, 1091 GM Amsterdam, tel. 020-5251530, fax 020-5251200, E-mail HOX@EDUC.UVA.NL

Een multiniveau meta-analyse naar sekseverschillen in leren

S. Severiens en G. ten Dam

Instituut voor de lerarenopleiding, Universiteit van Amsterdam

ABSTRACT

To obtain a more coherent picture of the magnitude and direction of gender differences in learning, a meta-analysis was performed on 52 studies on this subject. The studies either used Kolb's Learning Style Inventory or Entwistle's Approaches to Studying Inventory. The results of these studies were analyzed in a multi-level approach with two levels: the respondents are nested within studies. Results show significant mean gender differences on a few dimensions of the two instruments. Men score higher on the Abstract Conceptualization scale of Kolb's instrument, Entwistle's ASI shows that women more often use the reproduction directed learning orientation and men more often a non-academic learning orientation. On a number of scales systematic variation in gender differences across studies is observed and this variation can be explained by the available study characteristics to a certain extent.

INTRODUCTIE

De afgelopen jaren is er nogal wat onderzoek verricht naar aan sekse gerelateerde aspecten van leerstijlen en leerstrategieën (zie Richardson & King, 1991; Severiens & ten Dam, 1994). De onderzoeksresultaten zijn echter allerminst eenduidig. Het gegeven dat in het onderzoeksdomein gebruik wordt gemaakt van uiteenlopende onderzoeksinstrumenten, maakt het moeilijk de resultaten zodanig met elkaar te vergelijken dat kan worden nagegaan welke aspecten van leren gerelateerd zijn aan sekse, en onder welke condities welke sekseverschillen optreden. Een samenvatting van het verrichte onderzoek kan hooguit op beschrijvende wijze plaatsvinden. In Severiens en ten Dam (1995; 1997) zijn we uitgebreid ingegaan op de theorie-ontwikkeling in het onderzoeksgebied sekse en leren. In het huidige artikel leggen we het accent meer op de methode met behulp waarvan de resultaten van eerder verricht onderzoek op statistische wijze kunnen worden samengevat.

Een geëigende manier om de resultaten uit verschillende onderzoeken te integreren, is het verrichten van een meta-analyse. De variëteit aan gebruikte onderzoeksinstrumenten vormt hiervoor echter een belemmering. Het is immers weinig zinvol resultaten samen te vatten die naar verschillende, onderling vaak onvergelijkbare, aspecten van leren verwijzen. Voor het verrichten van een meta-analyse dient het onderzoeksgebied dan ook te worden ingeperkt tot die onderzoeken waarin met eenzelfde onderzoeksinstrument is gewerkt. De *Learning Style Inventory* (LSI) van Kolb (1984) en de *Approaches to Studying Inventory* (ASI) van Entwistle (1981) lenen zich het best voor een dergelijke meta-analyse omdat deze instrumenten het meest gebruikt worden in het onderzoeksgebied.

Ten aanzien van Kolb's LSI concludeerden we op basis van een eerder verrichtte meta-analyse (Severiens & ten Dam, 1994), dat de leerstijl van mannen relatief meer gekenmerkt wordt door abstracte conceptualisering. De drie andere schalen lieten geen sekseverschillen zien. De resultaten waren evenwel heterogeen, wat betekent dat de sekseverschillen die ge-

vonden worden variëren tussen de studies. Uit de meta-analyse op de onderzoeken waarin gebruik is gemaakt van Entwistle's ASI, kwamen sekseverschillen ten aanzien van drie affectieve factoren naar voren: intrinsieke motivatie, faalangst en extrinsieke motivatie. De onderzoeksresultaten waren heterogeen op twaalf van de zestien schalen in de ASI. Wederom variëren de sekseverschillen dus tussen de studies. Met behulp van correlatie-analyses werd duidelijk dat de beschikbare achtergrondkenmerken van de 26 studies die in de meta-analyse waren opgenomen, de gevonden heterogeniteit niet konden verklaren. We concludeerden op basis van deze meta-analyse dan ook dat de verschillende onderzoeken met behulp van Kolb's LSI en Entwistles's ASI allesbehalve een helder beeld laten zien. De onderzoeksresultaten verschillen niet alleen, ze spreken elkaar soms ook tegen.

In dit artikel presenteren we de resultaten van een nieuwe meta-analyse. Deze verschilt op twee manieren van de eerder uitgevoerde meta-analyse. Ten eerste hebben we nu tevens de beschikking over de studies die in de afgelopen drie jaar verricht zijn, hetgeen een verdubbeling van het aantal studies betekent. Ten tweede worden de gegevens geanalyseerd in een multiniveau benadering (Bryk & Raudenbush, 1992; Hox & de Leeuw, dit nummer; Snijders, dit nummer). Een multiniveau benadering maakt het mogelijk in één analyse zowel sekseverschillen in leren te bekijken als de invloed van de factoren leeftijd, studierichting en land waarin de studie is verricht, na te gaan. De onderzoeksvragen luiden als volgt: In hoeverre verschillen de leerstijlen en leeroriëntaties van vrouwen en mannen van elkaar? En, hoe kan de eventueel gevonden heterogeniteit verklaard worden?

DE STUDIES

In verschillende databestanden (SSCI, ERIC, British Education Index, EUDISED-R&D, Dissertation Abstracts Online, Sociological Abstracts, Psychinfo, ADION/DION) is gezocht naar studies waarin gebruik is gemaakt van Kolb's Learning Style Inventory (LSI) of van Entwistle's Approaches to Studying Inventory (ASI) in de periode tussen 1980 en 1995. Dit resulteerde in 64 auteurs die in hun onderzoek gewerkt hebben met de LSI en 35 auteurs die hebben gewerkt met de ASI. Slechts enkele auteurs bleken in hun publikatie(s) gegevens vermeld te hebben ten aanzien van sekseverschillen. De overige auteurs zijn benaderd met het verzoek de benodigde gegevens op te sturen. Op dit verzoek reageerden uiteindelijk 54 auteurs: achttien auteurs stuurden de benodigde gegevens, de overigen hadden óf geen toegang tot het data-bestand óf hadden de sekse van hun respondenten niet gecodeerd. Aangezien sommige auteurs meer dan één steekproef beschikbaar hadden, hebben we voor dit artikel de beschikking over 30 studies waarin gebruik is gemaakt van de LSI en 22 waarin gebruik is gemaakt van de ASI (in de tabellen 1 en 4 worden de studies beschreven)¹.

Eén van de mogelijke gevaren in een meta-analyse betreft de onvoldoende kwaliteit van de betrokken studies (Light & Pillemer, 1984). De rapportage van de studies in de huidige meta-analyse is daarom nauwkeurig bekeken op mogelijke fouten. Een uitgebreide kwaliteitscontrole heeft bovendien plaatsgevonden doordat over de meeste studies is gepubliceerd in wetenschappelijke tijdschriften. Wij zijn er van uitgegaan dat de kwaliteit van de studies hierdoor voldoende is gewaarborgd.

META-ANALYSE EN MULTINIVEAU-ANALYSE

Het doel van een meta-analyse is het integreren van de resultaten uit verschillende onderzoeken naar hetzelfde onderwerp. Zowel Kolb's LSI als Entwistle's ASI bestaan uit verschillende dimensies die ieder een ander aspect van leren representeren. Een meta-analyse op elk van deze schalen laat zien in hoeverre en in welke richting de schalen sekse-sensitief zijn.

In Hox (1994, zie ook Hox & de Leeuw, dit nummer) wordt beschreven hoezeer het type databestand bij een meta-analyse lijkt op het hiërarchische databestand dat geanalyseerd wordt in multiniveau-analyses. Een meta-analyse waarbij de respondenten genest zijn binnen studies betreft in feite een zelfde situatie als een hiërarchisch databestand met twee niveaus. Het voordeel van een multiniveau benadering ten opzichte van de meer gebruikelijke benaderingen bij meta-analyse gegevens bestaat dan ook uit de mogelijkheid om bij het schatten van het gemiddelde effect en de variantie van dat effect, rekening te houden met zowel de variantie die toegeschreven kan worden aan de steekproef als de variantie die systematisch is en dus te maken heeft met kenmerken van de betreffende studies. Deze kenmerken kunnen vervolgens in de multiniveau modellen worden opgenomen om na te gaan in hoeverre de variantie van de effecten verklaard wordt door de verschillende kenmerken.

Voor een 'normale' multiniveau-analyse zijn de gegevens van alle afzonderlijke respondenten uit alle studies nodig. In een meta-analyse zijn deze echter meestal niet beschikbaar. Bryk en Raudenbush (1992) wijzen er in dit verband op dat bij redelijk grote steekproeven (30 of meer) aangenomen mag worden dat de statistische grootheden normaal verdeeld zijn en de variantie bekend is. Ze spreken in dit geval van het Level 1 Variance-known model. Als de variantie op niveau 2, het studie-niveau, significant is, zijn de effectgroottes in de afzonderlijke studies niet allemaal even groot. Deze verschillen kunnen dan niet toegeschreven worden aan steekproefvariantie; de variantie is systematisch. In dat geval wordt gesproken van heterogene resultaten. Met behulp van de kenmerken van de afzonderlijke studies, de achtergrondvariabelen, kan vervolgens geprobeerd worden deze heterogeniteit te verklaren. Dit is in een multiniveau benadering relatief eenvoudig. Als de variantie op het studie-niveau echter niet significant van nul afwijkt, en dus is toe te schrijven aan steekproefvariantie, dan wordt geconcludeerd dat de resultaten homogeen zijn. Het heeft in dat geval geen zin de achtergrondvariabelen ter verklaring op te voeren. Snijders (dit nummer) plaatst overigens een kanttekening bij de homogeniteitstoets: het verwerpen van de alternatieve hypothese (de resultaten zijn heterogeen) betekent nog niet dat de nulhypothese (de resultaten zijn homogeen) is bewezen. Er is dus enige voorzichtigheid geboden bij het interpreteren van de homogeniteitstoets en de conclusie dat de resultaten homogeen zijn.

Voor de meta-analyse naar sekseverschillen in leren betekent het bovenstaande dat bij heterogeniteit van de effectgroottes de sekseverschillen variëren tussen de studies. De resultaten kunnen bovendien onderling tegenstrijdig zijn. Zo is het mogelijk dat vrouwen in de ene studie bijvoorbeeld hoger scoren op de abstracte leerstijl, terwijl mannen in de andere studie hoger scoren op deze leerstijl. In het geval van heterogene resultaten is het zinnig om na te gaan in hoeverre de kenmerken van de studies voor de gevonden verschillen verantwoordelijk zijn. Scoren mannen bijvoorbeeld hoger op de abstracte leerstijl bij bepaalde vakken? Als er daarentegen sprake is van homogeniteit, nemen we aan dat de sekseverschillen even groot (of klein) zijn in alle studies. Variantie in effectgroottes schrijven we in dat geval toe aan steekproefvariantie.

In het programma van Bryk, Raudenbush, Seltzer en Congdon (1988) voor het analyseren van multiniveau gegevens (HLM) wordt een optie geboden voor het toepassen van het 'Level 1 Variance Known'-model ten behoeve van meta-analyse. Deze optie heet VKHLM (oftewel Variance Known Hierarchical Linear Models) (zie ook Hox & de Leeuw, dit nummer). Een typisch databestand voor meta-analyse bestaat uit een maat voor de effectgrootte per studie, de bijbehorende steekproefvariantie en de studiekarakteristieken die mogelijk van invloed zijn op de effectgrootte. Omdat in ons geval de steekproef in een aantal studies klein is, wordt als effectgrootte gebruik gemaakt van de verschilscore d (Hedges & Olkin, 1985, p. 80). Deze effectgrootte is het gestandaardiseerde verschil tussen twee groepen (vrouwen en mannen in dit onderzoek), gecorrigeerd voor steekproef-bias en gedefinieerd als

$$d = (1 - 3 / (4N - 9)) (M_{\text{mannen}} - M_{\text{vrouwen}}) / SD$$

waar SD de wortel is van het gewogen gemiddelde van de twee varianties en N het totaal aantal personen. De waarden van d-scores worden meestal geïnterpreteerd aan de hand van een vuistregel van Cohen (1977): een d-score van .2 geeft een klein verschil weer, een middelgrote d-score is ongeveer .5 en d-scores van .8 worden als groot geïnterpreteerd. De steekproefvariantie van d wordt geschat in de volgende formule (Hedges & Olkin, 1985, p. 86):

$$s^2(d) = (N_{\text{mannen}} + N_{\text{vrouwen}}) / (N_{\text{mannen}} N_{\text{vrouwen}}) + d^2 / 2(N)$$

We schatten in eerste instantie een model waarin geen verklarende variabelen zijn opgenomen. In dit zogeheten *intercept-only* model wordt de gemiddelde verschilscore voorspeld en wordt geschat hoe groot de variantie van de verschilcores in de populatie is. Er wordt een antwoord gezocht op de vraag in hoeverre er sprake is van systematische variantie tussen de verschilcores middels een chi-kwadraattoets voor homogeniteit. Vanwege het grote onderscheidingsvermogen van deze toets, wordt daarbij tevens gekeken naar de proportie van de ware parameter variantie (zie Hox & de Leeuw, dit nummer; Hunter & Schmidt, 1990). Als vuistregel wordt gehanteerd dat het alleen interessant is de heterogeniteit te verklaren, indien de proportie 25% of groter is.

In het geval van heterogeniteit nemen we de beschikbare achtergrondvariabelen in het model op: de gemiddelde leeftijd van de respondenten in de betreffende studie, de studierichting in de betreffende studie en het land waarin de studie verricht is. De achtergrondvariabele land is gekozen in navolging van Hanna, Kündiger en Larouche (1990) en Baker en Perkins Jones (1993): het is mogelijk dat sekseverschillen variëren al naar gelang het land. Van de variabele leeftijd weten we dat deze verband houdt met leerstijlen, en daarom mogelijkwijs ook met sekseverschillen in leren (zie bijv. Watkins & Hattie, 1981). Eerdere studies (zie bijv. Dipelholfer-Stiem, 1989) laten verbanden zien tussen leerstijlen, sekseverschillen en studierichting. Omdat studierichting en land nominaal verdeelde variabelen zijn, nemen we ze in het model op als dummy-variabelen. De variabelen studierichting en land worden elk beschreven met een aantal dummy-variabelen (het aantal categorieën min één) die telkens één van de categorieën vergelijken met de overige.

Als de homogeniteitstoets en de proportie parameter variantie aangeven dat de verschilcores niet heterogeen zijn, nemen we aan dat de sekseverschillen over de studies heen even groot (of klein) zijn. De achtergrondvariabelen worden dan logischerwijs niet in het model opgenomen.

Resultaten, Kolb's Learning Style Inventory

In het model van ervaringsleren van Kolb (1984) wordt leren gezien als een cyclus van vier fases: concreet ervaren, observeren en reflecteren, abstract conceptualiseren en actief experimenteren. Voor een inhoudelijke beschrijving van Kolb's leermodel en de daarbij behorende LSI, verwijzen we naar Vermunt (1992) of Boekaerts en Simons (1993). De betrouwbaarheden van de originele LSI zijn: Concrete Ervaring: $\alpha = .55$, Reflectieve Observatie: $\alpha = .62$, Abstracte Conceptualisering: $\alpha = .75$ en Actieve Experimentatie: $\alpha = .66$ (Kolb, 1976). Een herziene versie uit 1984 leverde iets hogere Cronbach α 's op (Pinto & Geiger, 1991; Veres, Sims & Locklear, 1991).

De LSI bestaat uit negen items die elk bestaan uit vier woorden. De woorden representeren één van de vier fases. Aan de respondenten wordt gevraagd deze woorden te rangschikken. Dit resulteert in aparte scores op elk van de vier fases uit de leeracyclus die aangeven in hoeverre de respondent een voorkeur heeft voor de betreffende fase. In tabel 1 worden de studies, de d-scores op de vier schalen en de beschikbare achtergrondgegevens vermeld. In vier aparte multiniveau-analyses wordt de gemiddelde d-score voorspeld².

Tabel 1 Studies die Kolb's LSI gebruiken en achtergrondvariabelen

nr onderzoekers	Nvrouwen	Nmannen	AC	RO	AE	CE	leeft	studie-richting	land
1 Cano-Garcia et al.	22	66	.15	-.07	.16	-.21	21	nat	Spanje
2 Cano-Garcia et al.	51	45	-.01	.02	.18	-.07	21	nat	Spanje
3 Cano-Garcia et al.	17	65	.05	.03	-.25	-.03	21	anders	Spanje
4 Cano-Garcia et al.	42	39	.00	.48	-.38	.03	21	med	Spanje
5 Cano-Garcia et al.	75	22	-.27	.18	-.14	.26	21	soc	Spanje
6 Cano-Garcia et al.	79	60	-.27	-.04	.07	.19	21	soc	Spanje
7 Cano-Garcia et al.	47	22	-.20	.27	-.18	.29	21	anders	Spanje
8 Cano-Garcia et al.	69	27	-.09	.22	.01	.24	21	let	Spanje
9 Cano-Garcia et al.	52	42	-.06	.27	-.31	-.01	21	soc	Spanje
10 Cano-Garcia et al.	95	22	.14	.48	-.21	-.00	21	let	Spanje
11 Algee et al.	36	12	.14	-.29	.21	.14	31	anders	USA
12 Choi	92	48	.39	-.26	-.16	.00	40	versch	USA
13 Katz	435	218	.14	.10	.04	-.15	20	versch	Israël
14 Vernon-Gerstenfeld	13	82	.35	-.01	-.54	.25	37	anders	USA
15 Hines et al.	202	134	.16	.20	-.04	-.26	20	soc	USA
16 Cornwell et al.	66	54	.44	-.34	.16	-.23	miss	soc	USA
17 Newland et al.	10	30	.04	.22	-.94	.66	25	med	USA
18 Hayden et al.	19	28	-.56	.08	-.11	.23	20	let	USA
19 Hayden et al.	16	12	.00	.95	.06	-.11	20	let	USA
20 Hayden et al.	19	17	-.26	-.44	.23	.81	20	nat	USA
21 Hayden et al.	19	28	-.18	-.09	.11	.18	20	let	USA
22 Hayden et al.	14	15	.09	-.20	.01	.46	20	let	USA
23 Hayden et al.	28	8	-.31	-.07	.26	.45	20	nat	USA
24 Hayden et al.	44	71	.03	.01	.28	-.24	20	let	USA
25 Rhodes	98	46	-.08	-.15	.00	.10	>30	versch	USA
26 Hudak et al.	54	43	.19	.00	.14	-.58	20	soc	USA
27 Bokoros et al.	125	62	.45	-.41	.16	-.34	31	soc	USA
28 Cordell	128	72	.25	.15	-.35	-.11	43	anders	USA
29 Logan	67	26	.11	.02	-.12	-.03	34	let	USA
30 Dorsey et al.	168	345	.19	.28	-.09	-.04	38	soc	USA

AC = verschildscore abstracte conceptualisering, RO = verschildscore reflectieve observatie, AE = verschildscore actieve experimentatie, CE = verschildscore concrete ervaring, leeft = gemiddelde leeftijd, land = land waar de studie is uitgevoerd, versch = verschillende, soc = sociale wetenschappen, med = medicijnen, let = letteren, nat = natuurwetenschappen, miss=ontbrekend N.b. In de publikaties van de studies 11, 12, 15 en 27 zijn de relevante statistische grootheden vermeld, statistische grootheden uit de overige studies zijn verkregen via de auteurs.

Tabel 2 Multiniveau modellen Kolb's Learning Style Inventory

		vast effect			random effect				
		coëfficiënt (s.e.)	t	p	variantie component	df	chi ²	p	proportie
Abstract Conceptualiseren	intercept	.091(.041)	2.24	.04	.009	29	34.23	.23	.15
	only model								
Reflectief Observeren	intercept	.053(.047)	1.12	.21	.022	29	45.71	.03	.24
	only model								
	intercept	-.021(.276)	-.08	.39					
	leeftijd	.005(.009)	.62	.32					
	nat	-.116(.213)	-.54	.34					
	let	.225(.174)	1.29	.17					
	soc	.085(.144)	.59	.33					
med	-.375(.264)	-1.42	.14						
USA	-.241(.127)	-1.90	.07						
Israël	.013(.237)	.06	.39						
Concreet Ervaren	intercept	-.029(.043)	-.66	.32	.014	29	42.25	.05	.26
	only model								
	intercept	.144(.249)	.58	.33					
	leeftijd	-.001(.008)	-.06	.39					
	nat	-.006(.199)	-.03	.39					
	let	-.017(.160)	-.10	.39					
	soc	-.200(.130)	-1.54	.12					
	med	-.118(.247)	-.48	.35					
USA	-.096(.117)	-.82	.28						
Israël	-.285(.205)	-1.39	.15						
Actief Experimenteren	intercept	-.039(.035)	-1.10	.22	.002	29	31.26	.35	.10
	only model								

Proportie = proportie van de ware parametervariantie, nat = natuurwetenschappen, let = letteren, soc = sociale wetenschappen, med = medicijnen. Nat, let, soc en med zijn dummy variabelen, in nat wordt nat met de overige studierichtingen vergeleken, in let wordt let met de overige studierichtingen vergeleken, etc. Hetzelfde geldt voor USA en Israël.

Allereerst bekijken we het *intercept only* model. Als het gaat om *abstract conceptualiseren*, blijken mannen enigszins hoger te scoren dan vrouwen: de verschillscore is weliswaar klein, maar significant (d -score = .09, zie tabel 2). Het resultaat is homogeen. We nemen aan dat de verschillen in de d -scores kunnen worden toegeschreven aan steekproefvariantie. Onafhankelijk van de situatie waarin de studies verricht zijn, hebben mannen gemiddeld genomen dus een iets grotere voorkeur dan vrouwen voor abstracte conceptualisering.

Ten aanzien van de leerfase *reflectief observeren* verschillen vrouwen en mannen niet van elkaar. De geschatte parameter is bijna gelijk aan nul (zie tabel 2). In dit geval betreft het echter een heterogeen resultaat: de d -scores variëren over de studies heen. De volgende stap bestaat dan ook uit het onderzoeken van het effect van de beschikbare achtergrondvariabelen. In ons geval zijn dat: de gemiddelde leeftijd per studie, de vier dummy-variabelen voor de studierichting van de respondenten en de twee dummy-variabelen die het land waar de studie verricht is weergeven (zie tabel 1). Uit tabel 2 blijkt dat geen van deze achtergrondvariabelen verband houdt met de grootte van de verschillscore³. Of er al dan niet sekseverschillen in reflectief observeren zijn, hangt blijkbaar met andere factoren samen dan die waarover wij de beschikking hebben.

Het *intercept only* model voor *concreet ervaren* resulteert eveneens in een schatting voor de verschillscore die niet significant afwijkt van nul (zie tabel 2): blijkbaar vinden mannen en vrouwen het opdoen van concrete ervaringen tijdens het leerproces even belangrijk. Tussen de studies doen zich echter nogal wat verschillen voor. De variantie op het studie-niveau is heterogeen. Ook nu gaan we, in een volgend model, het effect na van de leeftijd van de respondenten, het land van herkomst en de studierichting. Dit levert geen significante coëfficiënten op: de achtergrondvariabelen verklaren de verschillen tussen de studies niet.

De laatste schaal van de LSI betreft de fase van *actief experimenteren*. Er bestaan geen sekseverschillen in voorkeur voor deze fase (zie tabel 2). Dit geldt zowel over de studies heen, als voor de afzonderlijke studies: het resultaat is homogeen.

Samenvattend kunnen we stellen dat er, gemiddeld genomen, alleen een sekseverschil is bij de 'leerfase' abstract conceptualiseren. Mannen hebben een iets grotere voorkeur voor deze fase dan vrouwen. Op de schalen voor concreet ervaren en reflectief observeren, hebben we heterogene resultaten gevonden. De sekseverschillen op deze schalen variëren al naargelang de studies. Met behulp van de achtergrondvariabelen die wij tot onze beschikking hebben (leeftijd, studierichting en land), kunnen we deze heterogeniteit echter niet verklaren.

Resultaten. Entwistle's Approaches to Studying Inventory

De ASI (Entwistle, 1981) bestaat uit 64 items (5-punts items) met behulp waarvan studenten hun eigen leren kunnen beschrijven. Aan deze items liggen zestien schalen ten grondslag, die op hun beurt vier leeroriëntaties weergeven: de betekenisgerichte en de reproductiegerichte leeroriëntatie, de prestatie-oriëntatie en de niet-academische oriëntatie (zie tabel 3). Voor een uitgebreide beschrijving van de ASI verwijzen we naar Vermunt (1992). De betrouwbaarheden van de leeroriëntaties zijn redelijk; ze variëren tussen .59 en .79. Voor een discussie over de betrouwbaarheid van dit instrument verwijzen we naar Richardson (1990).

Door Meyer en Parsons (1989) zijn de factorstructuren van de ASI uit verschillende studies met elkaar vergeleken. Ze concluderen dat de betekenisgerichte oriëntatie vrij consistent naar voren komt, maar de overige oriëntaties beduidend minder. De manier waarop de leeroriëntaties zijn samengesteld uit schalen kan dus verschillen per studie. We besluiten daarom de composietcores op de leeroriëntaties, zoals gedefinieerd door de auteurs van de ASI zelf (Tait & Entwistle, 1996), te analyseren en daarnaast de scores van de afzonderlijke schalen te laten voorspellen in multiniveau modellen. In tabel 4 worden de studies, de d -scores op de vier leeroriëntaties en de achtergrondgegevens beschreven (zie noot 2)

Tabel 3 Dimensies van de Approaches to Studying Inventory (bron: Tait & Entwistle, 1996)

Subschaal	Korte beschrijving
<u>Betekenisgerichte oriëntatie</u>	
Diepte-aanpak	Actieve benadering gericht op het begrijpen van de betekenis.
Gebruiken van bewijs	Onderzoektoe bewijsmateriaal kritisch waarop conclusies zijn gebaseerd.
Relateren van ideeën	Stuurstof relateren met andere onderdelen van de cursus of met eigen ervaringen.
Begrip-leren	Gebruik van analogieën, nadruk op het opbouwen van een algemeen beeld over ideeën en verbanden ertussen.
Operationeel leren	Stapsgewijze, logisch-analytische werkwijze met de nadruk op details en feiten.
Intrinsieke motivatie	Geïnteresseerd in het leren om het leren zelf.
<u>Reproductiegerichte oriëntatie</u>	
Oppervlakte-aanpak	Nadruk op memoriseren, van buiten leren.
Cursus gebondenheid	Zoekt het leren tot door de docenten gedefinieerde taken en cursusmateriaal.
Zorgeloosheid	Aandacht overmatig gericht op details, niet geneigd relaties tussen ideeën te overdenken.
Faalangst	Pessimisme en angst over studieresultaten, gebrek aan zelfvertrouwen.
<u>Prestatiegerichte oriëntatie</u>	
Prestatie motivatie	Competitief en vol zelfvertrouwen.
Strategische aanpak	Zoekt actief naar informatie over examen-eisen en vragen; probeert indruk te maken op de staf.
<u>Niet-academische oriëntatie</u>	
Ongeorganiseerde studiemethoden	Niet in staat regelmatig en effectief te werken.
Negatieve studie-attitude	Gebrek aan interesse.
Globetrotten	Geneigd te snel conclusies te trekken en te generaliseren.
Extrinsieke motivatie	Geïnteresseerd in de studie vanwege de kwalificaties die het oplevert.

Tabel 4 Studies die Entwistle's ASI gebruiken en geselecteerde achtergrondvariabelen

nr onderzoekers	Nvrouwen	Nmannen	BO	RO	PO	NO	leeft	studie-richting	land
1 Cano-Garcia et al.	22	66	.08	.06	.02	.42	21	nat	Spanje
2 Cano-Garcia et al.	51	45	-.13	-.35	-.28	.08	21	nat	Spanje
3 Cano-Garcia et al.	17	65	-.14	-.38	-.47	.09	21	anders	Spanje
4 Cano-Garcia et al.	42	39	-.08	-.25	.10	.11	21	med	Spanje
5 Cano-Garcia et al.	75	22	.18	-.06	.21	.24	21	soc	Spanje
6 Cano-Garcia et al.	79	60	-.13	-.35	-.03	.20	21	soc	Spanje
7 Cano-Garcia et al.	47	22	-.19	-.36	.22	.01	21	let	Spanje
8 Cano-Garcia et al.	69	27	.05	-.65	-.09	-.08	21	let	Spanje
9 Cano-Garcia et al.	52	42	-.16	-.38	-.07	.15	21	soc	Spanje
10 Cano-Garcia et al.	95	22	-.41	-.31	-.38	.05	21	let	Spanje
11 Gunn*	85	70	-.18	-.17	.11	.10	34	versch	Gr.Britt
12 Meyer et al.	79	242	.14	-.33	.08	-.02	miss	nat	Z.Afrika
13 Meyer	128	370	-.18	-.14	.06	.11	≈18	versch	Gr.Britt
14 Meyer et al.	144	266	-.05	-.12	.00	-.02	≈18	nat	Gr.Britt
15 Solomonides*	389	40	-.02	-.01	.23	.24	21	nat	Gr.Britt
16 Miller et al.	650	465	-.08	-.23	.04	.21	miss	versch	USA
17 Gledhill et al.	41	135	-.03	.14	-.09	.32	≈24	med	Z.Afrika
18 Clarke	68	85	.03	-.19	.30	.05	miss	med	Australië
19 Watkins et al.	132	160	-.41	-.09	-.07	.15	21	versch	Australië
20 Richardson	27	68	.09	-.30	miss	miss	miss	soc	Gr.Britt
21 Richardson	113	114	.23	.00	miss	miss	miss	soc	USA
22 Coles	44	43	-.10	.17	-.12	miss	≈18	med	Gr.Britt

BO = verschilscore betekenis oriëntatie, RO = verschilscore reproductieve oriëntatie, PO = verschilscore prestatie oriëntatie, NO = verschilscore niet-academische oriëntatie, leeft = gemiddelde leeftijd, land = land waar de studie is uitgevoerd, miss = ontbrekend, soc = sociale wetenschappen, med = medicijnen, nat = natuur wetenschappen, let = letteren, versch = verschillende richtingen

N.b. In de publikaties van de studies 17 en 18 zijn de relevante statistische grootheden vermeld, statistische grootheden uit de overige studies zijn verkregen via de auteurs, in de studies 20 en 21 zijn de statistische grootheden berekend uit andere dan in de publikatie vermelde data.

* In twee studies in de meta-analyse is een herziene versie van de ASI gebruikt (Tait & Entwistle, in druk). In overleg met Tait is op schaalniveau gekeken naar de vergelijkbaarheid van de twee lijsten. Schalen uit de herziene versie die hetzelfde meten als de schalen uit de originele versie, zijn meegenomen in de verdere analyses.

Een model waarin de betekenisgerichte leeroriëntatie wordt voorspeld, laat zien dat de verschilscore over de studies heen niet significant verschilt van nul (zie tabel 5). Met andere woorden, vrouwen rapporteren deze leeroriëntatie even vaak als mannen. De variantie tussen de studies is niet significant en we concluderen dat het resultaat homogeen is. Het heeft ver-

der geen zin om de achtergrondvariabelen als verklarende variabelen op te voeren.

Als we de afzonderlijke schalen in ogenschouw nemen, zien we twee schalen die resulteren in een significant sekseverschil: vrouwen scoren hoger op het relateren van ideeën (d-score = -.18) en op operationeel leren (d-score = -.17). Opvallend is verder dat de verschillen op alle schalen uit deze leeroriëntatie heterogeen zijn. Blijkbaar is de variantie in sekseverschillen op de afzonderlijke schalen dus wél voor een deel systematisch. De achtergrondvariabelen: leeftijd en de dummy-variabelen voor land (vier) en studierichting (vier) zijn opgevoerd om de heterogeniteit te verklaren.

Op de schaal voor diepte-aanpak heeft de dummy-variabele die de natuurwetenschappen vergelijkt met de overige studierichtingen een significante coëfficiënt. De gemiddelde d-scores geven aan dat mannen iets hoger scoren op de schaal voor diepte-aanpak in de natuurwetenschappen, terwijl in de andere studierichtingen vrouwen hoger scoren op deze schaal. De variantie op studie-niveau blijft echter heterogeen, wat betekent dat er dus nog andere studiekenmerken zijn die zouden kunnen verklaren waarom sekseverschillen tussen de verschillende studies variëren.

Op de schaal voor intrinsieke motivatie hangen sekseverschillen samen met de dummy-variabelen voor studierichting: alle coëfficiënten verschillen significant van nul. De gemiddelde d-score bij natuurwetenschappen is 0, bij letteren .15, bij de sociale wetenschappen 0, bij medicijnen -.15 en bij de 'overige' categorie -.35. Mannen lijken dus intrinsieker gemotiveerd te zijn bij letteren, terwijl in de andere studierichtingen óf vrouwen meer intrinsiek gemotiveerd zijn óf geen verschil gevonden wordt. In het model waarin de achtergrondvariabelen zijn opgenomen is geen systematische parameter variantie meer. Het lijkt er dus op dat naast studierichting er geen andere achtergrondvariabelen meer zijn die verband houden met de geconstateerde variantie in d-scores.

De dummy-variabele die de Verenigde Staten met de overige landen vergelijkt, heeft een significante waarde op de schalen voor gebruiken van bewijs en begrip-leren. Als we de gemiddeldes betrekken bij het interpreteren van deze coëfficiënten, zien we dat de Verenigde Staten en Australië de landen zijn die het meest van elkaar verschillen. In de Verenigde Staten scoren mannen hoger op begrip-leren en gebruik van bewijs, terwijl in Australië vrouwen relatief meer gericht zijn op begrip-leren en het gebruiken van bewijs. Ook voor deze schalen geldt dat na het meenemen van de achtergrondvariabelen de heterogeniteit op studie-niveau verdwenen is.

De heterogeniteit van de andere schalen in de betekenisgerichte leeroriëntatie kunnen we niet verklaren met behulp van de beschikbare achtergrondvariabelen.

Tabel 5 Multiniveau modellen Entwistle's Approaches to Studying Inventory

intercept-only modellen	vast effect			random effect				
	coëfficiënt (s.e.) intercept	t	p	variantie component	df	chi ²	p	proportie
betekenisgerichte leeroriëntatie	-.053(.042)	-1.25	.18	.011	21	28.11	.14	.14
reproductiegerichte leeroriëntatie	-.180(.032)	-5.70	.00	.000	21	21.60	.42	.04
prestatiegerichte leeroriëntatie	.021(.032)	.63	.32	.000	19	16.93	<.50	.06
niet-academische leeroriëntatie	.132(.033)	3.97	.00	.000	18	10.07	<.50	.08

proportie = proportie van de ware parametervariantie

Op de *reproductiegerichte leeroriëntatie* scoren vrouwen gemiddeld hoger dan mannen (d-score = -.18, zie tabel 5). We vinden geen systematische verschillen tussen de afzonderlijke studies.

Een analyse van de afzonderlijke schalen laat zien dat, met uitzondering van de schaal voor cursus-gebondenheid, er op alle schalen sekseverschillen naar voren komen: in vergelijking met mannen zijn vrouwen vaker op details gericht, gebruiken ze vaker een oppervlakte-aanpak en zijn ze faalangstiger. De gevonden sekseverschillen op de schalen voor zorgeloosheid en faalangst zijn heterogeen en variëren dus tussen de studies. De beschikbare achtergrondvariabelen kunnen deze verschillen tussen de studies niet verhelderen.

Vrouwen en mannen verschillen niet van elkaar als het gaat om de *prestatiegerichte leeroriëntatie* (zie tabel 5). Dit geldt over alle studies heen. Opvallend is dat de afzonderlijke schalen wél significante verschillen opleveren. Vrouwen studeren strategischer (d-score=-.17) en mannen zijn meer prestatiegericht (d-score=.19). Deze resultaten zijn heterogeen. De achtergrondvariabelen verklaren de variantie alleen in het geval van de schaal voor strategische aanpak. De gemiddeldes geven aan dat mannen bij medicijnen iets strategischer leren dan vrouwen, terwijl dit bij de overige studierichtingen voor vrouwen geldt. Na het opnemen van de dummy-variabelen in het model, is de variantie op studie-niveau niet meer significant: de beschikbare studiekenmerken lijken de heterogeniteit te verklaren.

Op de *niet-academische leeroriëntatie* scoren mannen significant hoger dan vrouwen (d-score=.13, zie tabel 5). Wederom gaat het om een homogeen resultaat. De hogere score van mannen komt naar voren op de schalen voor een negatieve houding ten opzichte van de studie (d-score=.16) en extrinsieke motivatie (d-score=.29). De resultaten met betrekking tot de afzonderlijke schalen van deze leeroriëntatie zijn eveneens homogeen.

De resultaten van de meta-analyse op de ASI kunnen we als volgt samenvatten. Op twee leeroriëntaties zijn sekseverschillen geconstateerd. Vrouwen leren reproductiegerichter en mannen hebben vaker een niet-academische leeroriëntatie. De betekenisgerichte en prestatiegerichte oriëntatie komen even vaak bij vrouwen als bij mannen voor. De resultaten op het niveau van de leeroriëntaties zijn homogeen: de kenmerken van de afzonderlijke studies zijn niet van invloed op de grootte van het sekseverschil. Op schaalniveau daarentegen ontstaat een ander beeld. Op elf schalen zijn over de studies heen sekseverschillen geconstateerd en tien van de zestien schalen zijn heterogeen verdeeld. Weliswaar verklaren het land waarin de studie verricht is en de studierichting in enkele gevallen een deel van de gevonden heterogeniteit maar een duidelijk te interpreteren patroon komt niet naar voren.

Conclusies

In dit artikel is verslag gedaan van een meta-analyse op 52 onderzoeken waarin gebruik is gemaakt van Kolb's LSI of Entwistle's ASI. De vraagstelling luidde: In hoeverre verschillen de leerstijlen en leeroriëntaties van vrouwen en mannen van elkaar? En, hoe kan de eventueel gevonden heterogeniteit verklaard worden? Voordat we de resultaten samenvatten, besteden we eerst aandacht aan het gebruik van de multiniveau benadering voor het analyseren van meta-analyse gegevens.

We hebben gekozen voor een multiniveau benadering met het oog op het hiërarchische karakter van het databestand waarmee in een meta-analyse wordt gewerkt: de respondenten zijn genest binnen studies. In de multiniveau modellen die getoetst zijn, fungeerden de verschillen op de leerstijlen en leeroriëntaties als afhankelijke variabelen. De kenmerken van de afzonderlijke studies zijn opgevoerd ter verklaring van de eventuele heterogeniteit in de verschillen. In een multiniveau benadering is het mogelijk om rekening te houden met het onderscheid tussen steekproefvariantie en systematische variantie, terwijl de variantie in de verschillen wordt geschat. Dit is een duidelijk voordeel van deze benadering. De eenvoud waarmee in een volgende stap mogelijk verklarende studiekenmerken in het model opgenomen kunnen worden, maakt deze benadering extra aantrekkelijk. Heterogeniteit in meta-analyse gegevens is immers eerder regel dan uitzondering (Hox & de Leeuw, dit nummer).

In principe hadden de gegevens waarover we de beschikking hadden multivariaat geanalyseerd kunnen worden. Zowel in het geval van Kolb's LSI als in Entwistle's ASI hadden we namelijk te maken met meerdere schalen en dus met meerdere afhankelijke variabelen, die

tegelijktijd voorspeld zouden kunnen worden in een multivariate multiniveau-analyse. Een dergelijke multivariate analyse heeft als voordeel boven een univariate analyse dat het ook een indicatie geeft van sekseverschillen over de schalen heen. Met andere woorden, het geeft een beeld van sekseverschillen in leren in meer algemene zin. Maar gegeven de veelheid aan betekenis die schuilgaat onder het begrip leren, zou een dergelijk antwoord in feite weinig informatie geven. We hebben er daarom voor gekozen alleen die schalen samen te nemen die inhoudelijk gerelateerd zijn, dat wil zeggen die tesamen een leeroriëntatie of leerstijl vormen, en de analyses univariaat uit te voeren.

De resultaten van de meta-analyse op de 52 studies naar sekse en leren, laten een aantal consistente sekseverschillen zien. Mannen vinden, gemiddeld genomen, het abstract conceptualiseren van de leerstof iets belangrijker dan vrouwen. Op studie-niveau vinden we geen significante hoeveelheid variantie. Het 'omgekeerde' blijkt echter niet: vrouwen hebben in hun leerproces, gemiddeld genomen, geen sterkere voorkeur voor de fase van concrete ervaringen. In dit geval vinden we echter nogal wat verschillen tussen de studies onderling. De beschikbare achtergrondvariabelen konden deze verschillen niet verklaren. Met betrekking tot de schaal reflectief observeren kan worden geconcludeerd dat vrouwen en mannen over de studies heen niet significant van elkaar verschillen. De variantie op studie-niveau is ook hier heterogeen. Wederom konden de achtergrondvariabelen leeftijd, studierichting en land de gevonden heterogeniteit niet verklaren.

De resultaten van de in dit artikel gepresenteerde meta-analyse liggen in de lijn van de eerder uitgevoerde meta-analyse op een beperkter aantal studies (26 in totaal) (ten Dam & Severiens, 1995). Wel vinden we een aantal verschillen in de grootte van de verschillen. Zo heeft de toevoeging van de extra studies tot gevolg dat de verschillen op de schaal abstract conceptualiseren kleiner wordt. Kijken we naar tabel 1 (de studies 1 tot en met 11 zijn toegevoegd) dan lijken de studies uit Spanje 'verantwoordelijk' te zijn voor de daling van de verschillen. In de overige gevallen zijn de verschillen in de nieuwe meta-analyse ongeveer even groot gebleven.

Wat betreft de onderzoeken waarin gebruik is gemaakt van de ASI, laat de meta-analyse zien dat vrouwen hoger scoren op de reproductiegerichte leeroriëntatie; mannen scoren daarentegen niet hoger op een diepgaande verwerking van de leerstof (beteningsgerichte leeroriëntatie). De prestatiegerichte leeroriëntatie laat geen sekseverschillen zien. Tenslotte vinden we dat mannen, gemiddeld genomen, hoger scoren op de niet-academische leeroriëntatie. De homogeniteitstoets gaf op alle vier leeroriëntaties aan dat de resultaten niet heterogeen verdeeld zijn. Ze lijken dus onafhankelijk van de situatie waarin de studies zijn verricht. Deze resultaten kunnen niet worden vergeleken met de eerder uitgevoerde meta-analyse, aangezien de analyses destijds alleen op schaalniveau zijn verricht.

Kijken we ook nu naar de scores op schaalniveau, dan blijken er allereerst op verschillende schalen sekseverschillen te bestaan (bijv. relateren van ideeën, operationeel leren, faalangst, negatieve studiehouding, extrinsieke motivatie). Dit geldt ook daar waar er op het niveau van de leeroriëntatie geen verschillen tussen vrouwen en mannen zijn. Ten tweede zijn, terwijl de leeroriëntaties zoals gedefinieerd door Tait en Entwistle (1996) homogeen verdeeld lijken te zijn, de verschillen op de meeste schalen van de ASI heterogeen. Dit betekent dat de variantie in sekseverschillen op de schalen voor een deel systematisch is. De voor ons beschikbare achtergrondvariabelen (leeftijd, land en studierichting) konden deze systematiek in een klein aantal gevallen (gedeeltelijk) verklaren. De trend die we in de vorige meta-analyse constateerden, komt opnieuw naar voren. De grootste verschillen verschijnen op de affectieve factoren: een hogere score van mannen op de extrinsieke motivatie-schaal en van vrouwen op de faalangst-schaal. Het lijkt erop dat toevoeging van nog meer studies, en dit geldt voor zowel de ASI als de LSI, resulteert in dezelfde trends.

Deze bevinding roept de vraag op hoeveel studies nodig zijn in een meta-analyse, voordat de onderzoeksvraag afdoende beantwoord kan worden. De vraag lijkt zeker relevant in het

geval van homogeniteit, waarbij studiekekenmerken blijkbaar geen verband houden met de grootte van de verschillen. Een relatief klein aantal studies is dan wellicht voldoende om een steekhoudende conclusie te trekken met betrekking tot het sekseverschil. In het geval van heterogeniteit lijkt het daarentegen belangrijk voldoende studies in de meta-analyse op te nemen die variëren al naargelang relevante achtergrondkenmerken. Het aantal achtergrondkenmerken dat mogelijk samenhangt met sekseverschillen zal dan bepalend zijn voor het aantal studies in een meta-analyse dat nodig is om een betrouwbare verklaring te geven van de variantie in verschillen.

Samenvattend concluderen we dat de verschillen die naar voren zijn gekomen in de meta-analyse op leerstijl en leeroriëntatie niveau gemiddeld genomen klein of afwezig zijn, maar op schaalniveau (in het geval van de ASI) duidelijker naar voren komen. Een tweede conclusie betreft de gevonden heterogeniteit: sekseverschillen in leren kunnen in de context van elke studie andere vormen aannemen. Wordt er in de ene studie gevonden dat vrouwen een grotere voorkeur hebben dan mannen voor bijvoorbeeld reflectief observeren, in andere studies is dit niet het geval of zelfs precies het omgekeerde (zie tabel 1 en 4). De vraag die zich vervolgens aandient, is welke contextfactoren van invloed zijn op sekseverschillen in leeroriëntaties en leerstijlen. De ons bekende kenmerken van de studies verklaren de heterogeniteit maar in een klein aantal gevallen. Blijkbaar zijn er ook andere factoren 'verantwoordelijk' voor de grootte en richting van de gevonden verschillen tussen vrouwen en mannen. Vermunt (1992, p.27-29) geeft een overzicht van die factoren die een verklaring vormen voor de leerstijlen die mensen hebben. Hij onderscheidt persoonsgebonden invloeden als intelligentie, onderwijservaring en leeftijd, en contextgebonden invloeden als de aard van een leertaak, van het vakgebied, van de toetseisen en van de instructie in het algemeen. Voor onderzoek naar sekse en leren, gaat het er om die factoren te selecteren die mogelijk anders uitwerken ten aanzien van vrouwen dan van mannen. Zo mag leeftijd dan bijvoorbeeld een factor zijn die van invloed is op leerprocessen, voor vrouwen en mannen gaat er blijkbaar geen differentiërende werking van uit. Voor het verklaren van de gevonden heterogeniteit in sekseverschillen in leren, hebben we de beschikking nodig over meer studies naar aan sekse gerelateerde (elementen van) leren waarin mogelijk relevante persoons- en contextgebonden invloeden zijn opgenomen.

REFERENTIES

- Algee, A., & Bowers, W. (1993). *A comparison of learning styles between Asian and American Seminary students*. (ERIC No. ED 360 887).
- Baker, D.P., & Perkins Jones, D. (1993). Creating Gender Equality: Cross-national Gender Stratification and Mathematical Performance. *Sociology of Education*, 66, 91-103.
- Boekaerts, M., & Simons, P.R.J. (1993). *Leren en instructie. Psychologie van de leerling en het leerproces*. Assen: Dekker & van de Vegt.
- Bokoros, M.A., Goldstein, M.B., & Sweeney, M.M. et al. (1990). *Common cognitive personality factors in non-clinical measures*. Paper presented at the Annual Convention of the American Psychological Association, Boston, M.A. (ERIC No. ED 328 819).
- Bryk, A.S., Raudenbush, S.W., Seltzer, S.W., & Congdon, R.T. (1988). *An introduction to HLM. Computer Program and User's Guide*. Version 2.0.
- Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical Linear Models*. Newbury Park, CA: Sage.
- Cano-Garcia, F., & Justicia-Justicia, F. (1994). Learning strategies, styles and approaches: an analysis of their interrelationships. *Higher Education*, 27, 239-260.
- Choi, J.M. (1989). Learning styles of academic librarians. *College and research libraries*, 50(6), 691-699.
- Clarke, R.M. (1986). Students' approaches to learning in an innovative medical school: a cross-sectional study. *British Journal of Educational Psychology*, 56, 309-321.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Coles, C.R. (1985). Differences between conventional and problem-based curricula in their students' approaches to studying'. *Medical Education*, 19, 308-309.

- Cordell, B.J. (1991). A study of learning styles and computer-assisted instruction. *Computers and Education*, 16(2), 175-183.
- Cornwell, J.M., Manfredi, P.A., & Dunlap, W.P. (1991). Factor analysis of the 1985 revision of Kolb's Learning Style Inventory. *Educational and Psychological Measurement*, 51, 455-462.
- Dam, G. ten, & Severiens, S. (1995). Leerstijlen, leerstrategieën en leermotivatie. Een meta-analyse van het leren van vrouwen en mannen. *Pedagogische Studiën*, 72, 206-220.
- Dippelhofer-Stiem, B. (1989). The development of research-oriented learning in five European countries. *European Journal of Psychology of Education*, 4, 489-503.
- Dorsey, O.L., & Pierson, M.J. (1984). A descriptive study of adult learning styles in a non-traditional education program. *Lifelong Learning*, 8-11.
- Entwistle, N.J. (1981). *Styles of learning and teaching*. New York: John Wiley & Sons.
- Gledhill, R.F., & Van der Merwe, C.A. (1989). Gender as a factor in student learning: preliminary findings. *Medical Education* 23, 201-204.
- Gunn, L. (1994). *The State of the Postgraduate Swamp: modes of study and approaches to learning*. Master's Thesis, University of Aberdeen.
- Hanna, G., Kündiger, E., & Larouche C. (1990). Mathematical achievement of grade 12 girls in fifteen countries. In Burton, L. (Ed.), *Gender and Mathematics. An international perspective* (pp. 87-97). Strand: Cassell.
- Hayden, R.R., & Brown, M.S. (1985). Learning styles and correlates. *Psychological Reports*, 56, 243-246.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hines, S.J., & Seidman, S.A. (1988). *The effects of selected CAI design strategies on achievement, and an exploration of other related factors*. Proceedings of selected research papers presented at the Annual Meeting of the Association for Educational Communications and Technology. (ERIC No. ED 295 646).
- Hox, J.J. (1994). *Applied Multilevel Analysis*. Amsterdam: TT-Publicaties.
- Hox, J.J., & de Leeuw, E.D. (dit nummer). Multiniveau modellen voor meta analyse. *Tijdschrift voor Onderwijsresearch*.
- Hudak, M.A., & Anderson, D.E. (1990). Formal operations and learning style predict success in statistics and computer science courses. *Teaching of Psychology* 17(4), 231-234.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Katz, N. (1988). Individual learning style: Israeli norms and cross-cultural equivalence of Kolb's LSI. *Journal of Cross-Cultural Psychology*, 19(3), 361-379.
- Kolb, D.A. (1976). *Learning Style Inventory: Technical manual*. Boston: Mc Berand Company, revised 1978.
- Kolb, D.A. (1984). *Experiential Learning, experience as a source of learning and development*. Englewood Cliffs N.J.: Prentice-Hall.
- Light, R.J. & Pillemer, D.B. (1984). *Summing up; the science of reviewing research*. Cambridge MA: Harvard University Press.
- Logan, E. (1990). Cognitive styles and online behavior of novice searchers. *Information Processing and Management* 26(4), 503-510.
- Meyer, J.H.F. (1995). Gender group differences in the learning behaviour of entering first-year students. *Higher Education*, 29, 201-215.
- Meyer, J.H.F., & Parsons, P. (1989). Approaches to studying and course perceptions using the Lancaster Inventory - a comparative study. *Studies in Higher Education* 14(2), 137-153.
- Meyer, J.H.F., Dunne T.T., & Richardson, J.T.E. (1994). A gender comparison of contextualised study behaviour in higher education. *Higher Education*, 27, 469-485.
- Meyer, J.H.F., & Parsons, P. (1994). *An exploration of student learning in mathematics*. Paper presented at the 1994 HERDSA Conference, Australian National University, Canberra, July 6-10.
- Miller, C.D., Finley, J., & McKinley, D.L. (1990). Learning approaches and motives: Male and female differences and implications for learning assistance programs. *Journal of College Student Development*, 31(2), 147-154.
- Newland, J.R., & Woelfl, N.N. (1992). Learning style and academic performance within a group of sophomore medical students. *Academic Medicine*, 67(5), 349.
- Pinto, J.K., & Geiger, M.A. (1991). Changes in learning-style preferences: a prefatory report of longitudinal findings. *Psychological Reports*, 68, 195-201.
- Rhodes, R.W. (1990). Measurement of Navajo and Hopi brain dominance and learning styles. *Journal of American Indian Education*, 29(3), 29-40.
- Richardson, J.T.E. (1990). Reliability and replicability of the Approaches to Studying Questionnaire. *Studies in Higher Education*, 15(2), 155-168.

- Richardson, J.T.E., & King, E. (1991). Gender differences in the experience of higher education: Quantitative and qualitative approaches. *Educational Psychology, 11*(3-4), 363-382.
- Severiens, S.E., & Dam, G.T.M. ten (1994). Gender differences in learning styles: a narrative review and a quantitative meta-analysis. *Higher Education, 27*(4), 487-501.
- Severiens, S., & Dam, G., ten (1997). Gender and gender-identity differences in learning styles. *Educational Psychology, 7* (1/2), 79-93.
- Snijders, T. (dit nummer). De populaties van de meta-analyse. *Tijdschrift voor Onderwijsresearch*.
- Tait, H., & Entwistle, N. (1996). Identifying students at risk through ineffective study strategies. *Higher Education 31* (1), 97-116.
- Veres, J.G., Sims R.R., & Locklear, T.S. (1991). Improving the reliability of Kolb's revised Learning Style Inventory. *Educational and Psychological Measurement, 51*, 143-150.
- Vermunt, J.D.H.M. (1992). *Leerstijlen en sturen van leerprocessen in hoger onderwijs. Naar procesgerichte instructie en zelfstandig denken*. Amsterdam/Lisse: Swets & Zeitlinger.
- Vernon-Geerstenfeld, S. (1989). Serendipity? Are there gender differences in the adoption of computers. A case study. *Sex Roles: A Journal of Research, 21*(3-4), 161-173.
- Watkins, D., & Hattie, J. (1981). The learning processes of Australian university students: Investigations of contextual and personal factors. *British Journal of Educational Psychology, 51*(3), 384-393.

Ontvangen 12-8-96

Definitief aanvaard 20-12-96

1. In één geval (Cano-Garcia et al) had de auteur de beschikking over gegevens van zowel Entwistle's ASI als Kolb's LSI. In strikte zin betreft het hier geen afzonderlijke studies, waardoor de term onderzoeksgroepen in plaats van studies beter op zijn plaats zou zijn. Dit geldt ook voor Hayden et al. die in een onderzoek meerdere afzonderlijke groepen heeft meegenomen. Voor de duidelijkheid hebben we echter besloten de term studies te blijven gebruiken.
2. In feite geldt dat de studies genest zijn binnen onderzoekers en strikt genomen zou hier in de analyses rekening mee gehouden moeten worden. Een mogelijkheid is om in de analyses de studies die bij één onderzoeker horen te wegen, slechts één van de studies mee te nemen, of de studies te middelen. Om na te gaan in hoeverre de resultaten veranderen wanneer de studies van één onderzoeker worden 'samengevat', hebben we deze analyses tevens uitgevoerd. In het geval van Kolb's LSI veranderen de resultaten op de schaal concreet ervaren: vrouwen scoren nu gemiddeld hoger dan mannen. Bovendien is de variantie op de schaal abstract conceptualiseren nu significant. Op Entwistle's ASI is er een verandering op de schaal voor de betekenisgerichte leeroriëntatie: de variantie op het tweede niveau is significant en blijktbaar zijn deze d-scores nu dus heterogeen verdeeld. Het meenemen van de studies die genest zijn binnen onderzoekers als aparte studies (zoals beschreven in het huidige artikel) leidt dus niet tot significante effecten die bij het middelen 'verdwijnen'. Daarom hebben we besloten deze geneste studies als afzonderlijke studies te beschouwen.
3. Hierbij dient opgemerkt te worden dat er weinig variantie bestaat in de variabele leeftijd. De conclusies omtrent leeftijd zouden hierdoor kunnen zijn beïnvloed. Dit geldt zowel voor de studies waarin Kolb's LSI (zie tabel 1) gebruikt wordt als studies waarin Entwistle's ASI (zie tabel 4) gebruikt wordt.

Effecten van beeldende vakken op visueel-ruimtelijk vermogen en esthetische waarneming: twee meta-analyses

Folkert Haanstra

SCO-Kohnstamm Instituut van de Universiteit van Amsterdam

ABSTRACT

Perceptual learning has long been an important goal of art education. Empirical findings on this issue are scattered across individual studies and narrative reviews. This review uses meta-analytic procedures to integrate the research on the effects of art education on both visual-spatial ability and aesthetic perception. A literature search produced 30 relevant experimental studies concerning visual-spatial ability and 39 studies concerning aesthetic perception. The studies were published between 1960 and 1990. No significant education effect on visual-spatial ability was found, but a medium education effect on aesthetic perception exists in the studies that have been conducted. Highest average effects on visual-spatial ability were achieved with young children, while the opposite held for effects on aesthetic perception. A combination of the study of art works and studio art proved to be the most effective in educating the perception of aesthetic properties.

1. INLEIDING

In 1890 formuleerde de Nederlandse Vereniging voor Tekenonderwijs de 'ontwikkeling van het waarnemingsvermogen' als een belangrijke doelstelling van het tekenonderwijs, naast 'de vorming van den goeden smaak' (in: Van Rheeden, 1989, p.27). En de lijst van doelstellingen die een Amerikaanse groep tekendocenten in 1899 formuleerde bevatte eveneens de doelstellingen ontwikkeling van de waarneming en ontwikkeling van de appreciatie van visuele schoonheid (in Wilson, 1971, p.501).

Ontwikkeling van visuele vaardigheden (ook wel aangeduid als leerlingen "leren zien") is nog steeds een veel gehanteerde doelstelling van de beeldende vakken tekenen en handvaardigheid. En ook nu zijn er twee kanten aan te geven. Enerzijds een buiten de kunsten gelegen (instrumentele) kant van verbetering van de waarneming in algemene zin en anderzijds een kunstgerichte (intrinsieke) kant: het beter naar kunst leren kijken. Leerkrachten in het basisonderwijs vinden in meerderheid verbetering van de waarneming belangrijker dan het leren kijken naar kunst (Haanstra, van der Kamp en Oostwoud Wijdenes, 1992). In het algemeen worden aan kunstvakken vaak effecten toegeschreven die buiten het terrein van de kunsten vallen, zoals bevordering van cognitieve, sensomotorische en sociale vaardigheden, maar ook de verbetering van leerprestaties in andere vakken. De voortdurende druk op de kunstvakken om hun plaats in het curriculum te legitimeren versterkt de neiging dergelijke instrumentele doelen te beklemtonen. Het empirische gehalte van dergelijke doelstellingen is meestal onduidelijk. In dit artikel wordt nader ingegaan op de empirische onderbouwing van de doelstellingen t.a.v. visuele waarneming.

Welke nadere invullingen heeft deze doelstelling in recente stromingen in de beeldende

vakken gekregen? Binnen de stroming die het ontwikkelen van cognitieve vaardigheden centraal stelt, heeft 'leren zien' te maken met een toenemend visueel onderscheidingsvermogen, verbetering van cognitieve stijlen van waarneming, bevordering van visueel denken en mentaal voorstellingsvermogen. Visueel-ruimtelijk vermogen is het centrale concept dat deze invullingen gemeen hebben. De achterliggende gedachte is dat kunstactiviteiten nauw samenhangen met visueel-ruimtelijke vermogens en dat beeldende vakken kunnen leiden tot verhoogde vaardigheid op dit gebied.

In de meer vakinhoudelijk gerichte opvattingen staat het kijken naar kunst en vormgeving centraal. Rice (1988) omschrijft dit als het "weten wat te doen als je wordt geconfronteerd met een object dat is gemaakt of is tentoongesteld 'alleen om naar te kijken' zoals vaak gebeurt met kunstwerken in onze cultuur". Doelstellingen binnen deze invalshoek hebben formuleringen als het 'coderen en decoderen van betekenissen van kunstwerken', of het 'leren waarnemen van de zintuiglijke, formele en expressieve eigenschappen van esthetische afbeeldingen'.

Leren zien wordt in deze studie op twee manieren opgevat: enerzijds als het effectiever omgaan met visueel-ruimtelijke informatie (een instrumentele invulling van leren zien), anderzijds als het met meer kennis en inzicht kijken naar kunst (de intrinsieke kant van leren zien). Het eerstgenoemde doel kan worden aangeduid met het verbeteren van 'visueel-ruimtelijke vermogens', het laatste met verbetering van de 'esthetische waarneming'.

De hoofdvraagstelling luidt: *hoe groot zijn de effecten van beeldende vorming op visueel-ruimtelijke vermogens en op esthetische waarneming?* Om deze vraag te beantwoorden is eerder verricht onderzoek op dit gebied met behulp van meta-analyses samengevat. Alvorens de gehanteerde methode en de resultaten van de twee meta-analyses te presenteren gaan we eerst nader in op de begrippen visueel-ruimtelijk vermogen en esthetische waarneming, de gebruikte meetinstrumenten en resultaten van eerdere onderzoekssamenvattingen. Het artikel besluit met een discussie van de uitkomsten.

2. VISUEEL-RUIMTELIJK VERMOGEN EN ESTHETISCHE WAARNEMING

2.1 Visueel-ruimtelijk vermogen

Het visueel-ruimtelijk vermogen kan worden omschreven als een aantal verwante bekwaamheden, zoals het kunnen herkennen van vormen, het kunnen genereren, vasthouden en transformeren van visuele voorstellingen 'in de geest', het vermogen zich te oriënteren in de ruimte. Op een abstracter niveau gaat het om het hanteren van beelden bij probleem oplossen.

Vanuit de psychometrische benadering van de menselijke intelligentie is een groot aantal tests voor ruimtelijk inzicht ontwikkeld. Op basis van onderzoek met deze tests (o.a. McGee, 1979; Lohman, 1988; Carroll, 1993) kunnen diverse visueel-ruimtelijke factoren worden onderscheiden, bijvoorbeeld 'ruimtelijke visualisatie' en 'flexibiliteit van closure' (het vermogen een complexe Gestalt te doorbreken om daarin een eenvoudiger Gestalt te herkennen). Deze laatste factor wordt ook aangeduid met veldafhankelijkheid (Witkin & Goodenough, 1981). Auteurs zijn het niet eens over het aantal factoren dat kan worden onderscheiden en ook vergelijkbare deelvaardigheden worden niet altijd met dezelfde naam aangeduid. Wel is er een grote mate van overeenstemming over het gegeven dat er in de beschikbare visueel-ruimtelijke tests sprake is van twee soorten taken (Eliot, 1987):

- (1) herkenning, waarbij een figuur moet worden waargenomen, onthouden of getransformeerd;
- (2) manipulatie, waarbij een figuur mentaal moet worden verplaatst, geroteerd, gevouwen etc. Herkenningstaken vallen binnen het twee-dimensionale vlak, in tegenstelling tot de meer complexe manipulatie-taken.

Veel studies naar intelligentie gedurende de levensloop laten zien dat leeftijd van invloed

is op visueel-ruimtelijke vermogens. Vanaf de kindertijd nemen deze vermogens voortdurend toe en bereiken een piek in het midden of tegen het einde van de adolescentie (Lohman, 1988; Gardner, 1985). Minder zeker zijn de auteurs over de mate van vermogensverlies tijdens de volwassenheid. Volgens McGee (1979) is er in onderzoek enige ondersteuning voor de opvatting dat oefening en training tot verbetering van het visueel-ruimtelijk inzicht kunnen leiden. Ook een overzicht van Lean (1981) concludeert dat verschillende ruimtelijke vaardigheden, mits getraind met de juiste leertaken, aan te leren zijn. De beste resultaten worden geboekt met jonge kinderen; bij ouderen werkte oefening minder. Eliot (1987) concludeert dat wil training succesvol zijn, de inhoud ervan niet te ver verwijderd mag zijn van de te verbeteren visueel-ruimtelijke (deel) vaardigheden. Dit wordt bevestigd door de resultaten van een meta-analyse van Beanninger en Newcombe (1989).

Visueel-ruimtelijk vermogen is vereist in diverse alledaagse activiteiten, maar bijvoorbeeld ook in technische en medische beroepen en bepaalde schoolvakken. Men gaat ervan uit dat beeldende kunst en ruimtelijke vermogen met elkaar verbonden zijn, omdat beeldend kunstenaars gevoelig moeten zijn voor de formele aspecten van beelden, in staat moeten zijn composities mentaal te manipuleren, concepten moeten visualiseren, etc. (Winner & Pariser, 1985). De relatie tussen beeldende kunst en ruimtelijke vermogen beperkt zich echter niet tot productieve aspecten. Het kijken naar kunst hangt ook af van vaardigheden op het terrein van visueel-ruimtelijke waarneming. Zo zijn vragen over kunstwerken in kunstbeschouwelijke lessen vaak gebaseerd op de veronderstelling dat leerlingen in staat zijn bepaalde beeldaspecten of details binnen een context te identificeren en te isoleren (Bergland & MacGregor, 1988).

2.2 Onderzoek naar beeldende vorming en visueel-ruimtelijk vermogen

Correlatieel onderzoek (o.a. Hermelin & O'Connor, 1986; Eliot, Kuo-Hsuing Ho, Olson & Medoff, 1988; Winner & Casey, 1992) toont in bescheiden mate aan dat kunststudenten gemiddeld hoger scoren op bepaalde, maar niet alle visueel-ruimtelijke vaardigheden. Verschillende studies concluderen dat er verschillen zijn op herkenningsopdrachten, maar niet op manipulatie-opdrachten. Andere studies concluderen juist het tegenovergestelde. In het algemeen is het zo dat studenten wiskunde en natuurwetenschappen de kunststudenten overvleugelden bij visueel-ruimtelijke tests. Tevens laten onderzoeken zien dat kunststudenten lagere scores hebben op verbale tests en meer dan gemiddeld te kampen hebben met verbale problemen (Van Meel-Jansen, 1974; McWhinnie, 1991; Winner, Casey, Dasilva & Hayes, 1991). Volgens Winner e.a. zou dit kunnen betekenen dat individuen voor beeldende kunst kiezen omdat dit terrein van hen vraagt waar ze goed in zijn ('relevante vaardigheden-hypothese') en voor een deel omdat ze wellicht verbale problemen uit de weg gaan ('vermijdings-hypothese').

Overzichten van experimenteel onderzoek (McWhinnie, 1973; Witkin & Goodenough, 1981) geven geen uitsluitsel over trainingseffecten van beeldende vorming. Witkin en Goodenough (1981) vragen zich af of de gemiddeld hogere veldonafhankelijkheid van beeldende kunstenaars en musici het gevolg is van aanleg (de genoemde 'relevante vaardigheden-hypothese') of dat beeldende- en muzikale vorming deze vaardigheden versterken. Witkin en Goodenough achten dat laatste waarschijnlijk, omdat studenten leren beeldende kunst en muzikale composities te analyseren. De uitkomsten van een beperkt aantal experimentele onderzoeken op dit gebied zijn weliswaar niet volledig consistent, maar omdat ze overwegend steun aan de 'scholingshypothese' bieden worden de implicaties voor het onderwijs veelbelovend genoemd. McWhinnie kwam eerder tot een negatieve conclusie, hoewel hij zich grotendeels op dezelfde studies baseerde. In zijn samenvatting noemt McWhinnie (1973) de resultaten van onderzoek beperkt en op verschillende punten tegenstrijdig.

2.3 Esthetische waarneming

De eerste psychologische onderzoeken naar de esthetische waarneming dateren van de 19e eeuw. De Duitse psycholoog Fechner (1876) onderzocht al de relatie tussen de formele elementen van kunstwerken (lijnen, vormen, kleuren) en de reactie van de kijker. In de recente-

re experimentele esthetica is het belangrijkste kenmerk de hedonistische (genot gevende) waarde van de esthetische ervaring. Berlyne (1974, p.8) schrijft dat wanneer in zijn onderzoeken wordt geconcludeerd dat bepaalde kenmerken "positieve intrinsieke hedonistische waarde" hebben, bedoeld wordt dat het contact ermee in zichzelf prettig en vervullend is. Hij stelt verder dat dit begrip sterk verwant is met wat in de filosofische esthetica (Beardsley, 1981; Elias, 1993) het belangeloze karakter van esthetische ervaringen wordt genoemd.

De in de literatuur genoemde houdings- en gevoelsaspecten van de esthetische blik, zoals belangeloosheid of 'psychische distantie', verlies van ego en gevoelens van zelf-acceptatie, zijn niet hanteerbaar als onderwijsdoelen. In deze studie wordt esthetische waarneming teruggebracht tot zijn cognitieve infrastructuur (De Groot, 1986) en omschreven als: kijken met kennis van esthetische kenmerken van kunst, of zoals Dickie (1983, p. 69) het uitdrukt: 'the knowing perception of art'.

Sommigen menen dat esthetische waarneming de ontdekking inhoudt van wat eerst verborgen lijkt, maar wat 'op het tweede gezicht' wel visueel aanwezig is in een kunstwerk. Voor hen moet esthetische waarneming gericht zijn op de manier waarop de verschillende visuele elementen van een kunstwerk zijn geordend. Op basis van deze visie zijn er leermethoden ontwikkeld voor het leren zien van visueel-esthetische aspecten, met name de zintuiglijke, formele en expressieve kwaliteiten van kunstwerken.

Anderen zien esthetische waarneming op de eerste plaats als 'interpretatie'. Die interpretatie hoeft niet noodzakelijk plaats te vinden op basis van wat visueel 'aanwezig' is in een kunstwerk (Parsons, 1992).

Uit onderzoek naar leeftijdsgebonden verschillen in esthetische waarneming komt naar voren dat kinderen de voorkeur geven aan realistische kunstwerken en dat het afgebeelde onderwerp tevens een overheersende factor is. Pas op oudere leeftijd gaan andere esthetische kenmerken zoals expressie, vorm en stijl een rol spelen. Onderzoeken in de traditie van Piaget (Housen, 1983; Parsons, 1987) beschrijven esthetische ontwikkeling in termen van een afname van egocentrisme en een toename van zelfreflectie en het vermogen om vanuit het perspectief van anderen te kijken. De absolute oordelen (mooi of lelijk; goed of slecht) in de eerste levensfasen ontwikkelen zich tot een relativerender, meervoudige benadering waarbij argumenten pro en contra worden afgewogen.

2.4 Het meten van esthetische waarneming

Esthetische waarneming is vanuit verschillende visies geoperationaliseerd en omvat o.a. psychofysiologische metingen (zoals de EEG), gedragsmetingen (zoals het meten van de tijd die wordt besteed aan het kijken naar verschillende kunstwerken) en de semantische differentiaal. Voor het onderzoek zijn vooral het gebruik van voorkeursschalen en beoordelingschalen relevant alsmede inhoudsanalyses van verbale reacties op kunstwerken.

Bij esthetische voorkeur gaat het om de mate waarin iemands relatieve voorkeur voor verschillende stimuli correspondeert met een extern criterium. Bij beoordelingsschalen gaat het om de mate waarin iemands oordelen over de esthetische waarde van stimuli overeenstemmen met een externe standaard. Voorbeelden van oude tests om esthetisch beoordelingsvermogen te meten zijn de Meier Art Judgment Tests, (1940; 1963) en de Graves Design Judgment test (1948). Deze tests meten de mate waarin iemand universeel geachte principes van esthetische ordening, zoals proportionering, evenwicht en eenheid onderkent bij de keuze tussen meerdere afbeeldingen. Het externe criterium voor de esthetisch betere afbeelding is vastgesteld door experts zoals bij Graves, of is zoals bij Meier de afbeelding van een oorspronkelijk kunstwerk ten opzichte van modificaties daarvan (bijvoorbeeld een enigszins gewijzigde compositie). Vaker dan met deze gedateerde tests worden esthetische voorkeuren en oordelen gemeten met speciaal gemaakte tests waarbij voor- of afkeur moet worden uitgesproken voor een aantal geselecteerde kunstwerken.

Een in onderzoek veel gebruikt instrument zijn verbale of schriftelijke reacties op kunstwerken, die vervolgens worden gescoord en geanalyseerd met behulp van een aantal in-

houdelijke categorieën. Bijvoorbeeld de indeling van Acuff en Sieber-Suppes (1972) omvat beschrijvende, interpretatieve en evaluatieve categorieën.

2.5 Onderzoek naar effecten van beeldende vorming op esthetische waarneming

Verschillende besprekingen van hoofdzakelijk Amerikaans onderzoek naar effecten op esthetische waarneming zijn gepubliceerd binnen de context van meer algemene besprekingen van onderzoek naar beeldende vorming (o.a. Boyer, 1983; Eisner, 1982). Recente onderzoekssamenvattingen door Salome (1991a; 1991b) zijn bedoeld om docenten praktische aanwijzingen te bieden. Over het geheel bezien wijzen de conclusies in positieve richting. Bijvoorbeeld Boyer concludeert dat "uitkomsten significante leereffecten op begripsvorming en waarneming laten zien in de experimentele groepen" (p.37). Eisner merkt op dat slechts een klein aantal onderzoeken bewijs levert voor een relatie tussen kunsteducatie en esthetische waarneming. Hij geeft twee mogelijke redenen voor dit gebrek in zulke onderzoeksinspanningen: de schaarste aan valide meetmethoden voor esthetische waarneming en de 'vanzelfsprekendheid' van de vraag. "Maar weinig mensen vragen zich af of het volgen van onderwijs op het gebied van kunst effect heeft op de gevoeligheid voor kunst: ze nemen gewoon aan dat het zo is" (p.91). Eisner geeft aan dat niet alle onderzoeken in positieve richting wijzen en concludeert dat alleen leerlingen zelf werkstukken laten maken (productieve beeldende vorming) onvoldoende is en dat kunstbeschouwing en kunstgeschiedenis (reflectieve beeldende vorming) nodig zijn om effect te sorteren. Salome (1991b) komt tot een soortgelijke conclusie, maar anderen houden vast aan het overwegende belang van de productieve beeldende vorming, ook als het gaat om het vergroten van esthetische waarneming (Burton, 1992). Ook in Nederland bestaat een oude discussie over de problematische verhouding tussen 'schep-ten' en 'schouwen' in de beeldende vakken (Hermans & Schönau, 1988).

2.6 Conclusies t.a.v. eerdere onderzoekssamenvattingen

De bestaande samenvattingen van onderzoek naar effecten van beeldend onderwijs op 'leren zien' betreffen traditionele kwalitatieve samenvattingen waarbij de auteurs een beperkt aantal representatief geachte onderzoeken als uitgangspunt hebben gekozen. De samenvattingen met betrekking tot visueel-ruimtelijk vermogen zijn gedateerd en spreken elkaar tegen. De samenvattingen met betrekking tot esthetische waarneming ondersteunen weliswaar de stelling dat beeldend onderwijs op dit punt effectief kan zijn, maar laten verschillende vragen onbeantwoord. Met name de vraag of productieve beeldende vorming nodig en afdoende is voor het ontwikkelen van esthetische waarneming.

Uit onvrede met de traditionele manier van samenvatten en integreren van onderzoeksresultaten zijn alternatieven ontwikkeld, die de benaming 'meta-analyse' hebben gekregen (Glass, 1976). Meta-analyse is een verzameling methoden voor het statistisch analyseren en samenvatten van de uitkomsten van empirische onderzoeken over een bepaalde vraagstelling. In deze studie zijn meta-analyses uitgevoerd om beter antwoord te kunnen geven op de vraag in welke mate de beschikbare onderzoeken effecten op visueel-ruimtelijk vermogen en esthetische waarneming laten zien en welke factoren daarbij een belangrijke rol spelen.

3. METHODE VAN META-ANALYSE

Uitgaande van een bepaalde vraagstelling omvat een meta-analyse het verzamelen van studies, het coderen van kenmerken van deze studies (zoals treatment-kenmerken, proefpersoonkenmerken, methodische kenmerken), het berekenen van een of meer grootheden per studie (zoals een p-waarde, een effectgrootte of een correlatie) en het uitvoeren van een statistische analyse op deze gegevens. De onderzoeker die een meta-analyse wil uitvoeren staat voor de keuze welke methode hij zal toepassen. Er bestaan diverse benaderingen (Bangert-Drowns, 1986) die verschillen in doel (bijvoorbeeld het geven van een stand van zaken in de literatuur

of het schatten van het treatment effect in de populatie) en in statistische methoden (bijvoorbeeld de manier waarop variantie in effectgroottes wordt behandeld). Naast de 'klassieke' aanpakken van Glass (Glass, McGaw & Smith, 1981) en van Rosenthal (1991), zijn de recentere en meest toegepaste benaderingen die van Hedges en Olkin (1985) en Hunter en Schmidt (1990). Vergelijkingen van verschillende methodes op identieke datasets laten zien dat ze tot uiteenlopende resultaten kunnen leiden (Johnson, Mullen & Salas, 1995). Bangert-Drowns beveelt een eclectische aanpak aan waarbij afhankelijk van de vraagstelling elementen van verschillende benaderingen kunnen worden gecombineerd.

Dit onderzoek richt zich op de grootte van het (onderwijs)effect in de populatie en op verklaringen van mogelijke verschillen in effecten tussen studies. Zowel de benaderingen van Hedges en Olkin en van Hunter en Schmidt zijn hierbij toepasbaar en gekozen is voor een combinatie van beide benaderingen. Hieronder worden de opeenvolgende onderzoeksstappen beschreven.

In de eerste plaats zijn relevante onderzoeken verzameld. Daarbij golden de volgende criteria:

- de onderzoeken betreffen effecten van productieve of reflectieve beeldende vorming of combinaties hiervan, dat wil zeggen ze betreffen de vakgebieden tekenen en schilderen, handvaardigheid, film en fotografie en kunstgeschiedenis/kunstbeschouwing;
- er worden kwantitatieve resultaten voor effecten van beeldende vorming op visueel-ruimtelijke vermogens of esthetische waarneming gerapporteerd;
- de onderzoeken zijn gerapporteerd in de periode 1960 - 1990 in de Engelstalige of Nederlandse literatuur.

Als bronnen zijn gebruikt de gegevensbestanden van Dissertation Abstracts International, ERIC en PsychINFO, de Index of British Studies in Art and Design Education (Allison, 1986), de Bibliografie Onderzoek Kunstzinnige Vorming (Kooyman, 1989) en de tijdschriften Studies in Art Education en Visual Arts Research. Op basis van de minimale opbrengsten van eerdere literatuursearches zijn geen speciale Franse en Duitse bronnen geraadpleegd (Haanstra en Van Oijen, 1985).

De verzamelde onderzoeken zijn op verschillende kenmerken gecodeerd. Ter controle is de codering deels door twee beoordelaars gedaan. Per verzamelde studie zijn één of meer effectgroottes berekend: het gestandaardiseerde verschil tussen het gemiddelde van de experimentele en van de controlegroep. Als standaardisering is de gepoolde binnengroepenstandaardafwijking toegepast. Door de vergelijking van meer dan één treatment met de controlegroep, door het opdelen van proefpersonen in subgroepen (bijvoorbeeld naar sexe of leeftijd) of door het gebruik van meerdere operationalisering van de afhankelijke variabele kunnen op basis van een studie vaak meerdere effectgroottes worden berekend. Dit betekent dat de gegevens niet statistisch onafhankelijk zijn en dat studies met veel effectgroottes verhoudingsgewijs veel gewicht krijgen. Sommige auteurs (zoals Bangert-Drowns) wijzen het gebruik van meer dan één effectgrootte per studie volledig af, anderen (zoals Glass) vinden het een groter nadeel dat veel informatie verloren gaat wanneer men zich tot één effectgrootte beperkt. Besloten is om het aantal effectgroottes te beperken tot maximaal vier per studie. Het belangrijkste criterium om meer dan één effectgrootte te berekenen was of meerdere voor de vraagstelling relevante onderwijsmethodes met elkaar werden vergeleken.

Van de effectgroottes is het gewogen gemiddelde en de variantie berekend, waarbij is gecorrigeerd voor steekproeffouten. Dit houdt in dat de variantie veroorzaakt door steekproeffouten van de totale variantie is afgetrokken. In de terminologie van Hunter en Schmidt (1990) heet dit een 'bare bones' meta-analyse, omdat niet wordt gecorrigeerd voor andere mogelijke artefacten, zoals meetfouten van afhankelijke en onafhankelijke variabelen. Het niet of slechts gedeeltelijk corrigeren voor diverse artefacten leidt volgens Hunter en Schmidt in het algemeen tot een onderschatting van de gemiddelde effectgrootte. Behalve de 'bare bones' meta-analyse op alle berekende effectgroottes is ter controle ook een schatting van de gemiddelde effectgrootte gemaakt op basis van een gelijke weging van alle studies. Daarbij werden ef-

fectgroottes afkomstig van eenzelfde studie gemiddeld. Tenslotte is nagegaan in hoeverre de variantie in effectgroottes kan worden verklaard uit bepaalde studiekenmerken. Op basis van theorie en uitkomsten van eerder onderzoek zijn verwachtingen geformuleerd ten aanzien van relevante moderatorvariabelen. Behalve analyse van inhoudelijke variabelen is nagegaan of onderzoekskwaliteit en publicatievorm samenhangen met verschillen in effectgrootte. Dit is gebeurd met een zogeheten homogeniteitsanalyse. Deze methode is niet zonder kritiek gebleven en als alternatief zijn methoden in het kader van multiniveau-analyse voorgesteld (zie in dit themanummer o.m. de artikelen van Snijders en Hox en De Leeuw). Hoewel het aanbeveling verdient de beschikbare data met deze methode te heranalyseren, wordt in dit artikel volstaan met de homogeniteitsanalyse volgens Hedges en Olkin (1985). Voor de statistische analyses is gebruik gemaakt van het programma 'Meta' (Meijer en Kok, 1991).

De gehanteerde methode is te vergelijken met een variantie-analyse en kent een aantal opeenvolgende stappen. In de eerste plaats wordt aan de hand van een toetsingsgrootte Q_t (totaal) vastgesteld of de totale verzameling effectgroottes homogeen is, d.w.z. dat de effectgroottes niet méér van elkaar verschillen dan op basis van steekproefgrootte van de respectievelijke onderzoeken verwacht mag worden. Indien dit het geval is geeft de gemeenschappelijke effectgrootte antwoord op de onderzoeksvraag en zijn geen nadere analyses vereist. Indien de effectgroottes heterogeen zijn worden ze op basis van een potentiële moderatorvariabele in subgroepen verdeeld. Vervolgens worden toetsingsgrootheden berekend voor de homogeniteit tussen en binnen de subgroepen, respectievelijk Q_b (between) en Q_w (within). Q_w is de som van de homogeniteitsmaten binnen elke groep (Q_{w_i}) en de genoemde Q_t is de som van Q_b en Q_w . De toetsingsgrootheden zijn chi-kwadraat verdeeld als de nulhypothese geldt dat de effectgroottes homogeen zijn. Indien er homogene subgroepen worden geïdentificeerd en de effectgroottes tussen groepen verschillen significant dan stopt de procedure. Er kan dan, analoog aan variantie-analyse, post-hoc getoetst worden welke groepen van elkaar verschillen. Dergelijke contrasten tussen gemiddelden zijn echter moeilijk te interpreteren als de effectgroottes binnen groepen heterogeen zijn. Er kunnen dan andere moderators worden geanalyseerd of men kan subgroepen verder verdelen volgens een tweede moderator.

4. RESULTATEN

4.1 Effect van beeldende vakken op visueel-ruimtelijk vermogen

In totaal werden 40 relevante titels van onderzoeken verzameld. Een deel ervan werd niet gebruikt in de meta-analyse, omdat de onderzoeksverslagen niet binnen de gestelde onderzoekstermijn verkrijgbaar waren, of omdat de gerapporteerde kwantitatieve resultaten ontoereikend waren om effectgroottes te berekenen. Uiteindelijk bleven 30 onderzoeksverslagen over. Deze staan vermeld in bijlage 1.

Relevante kenmerken van de studies zijn gecodeerd. Bij een derde van de studies is dit dubbel gedaan. De inter-beoordelaarsbetrouwbaarheid was bevredigend (Cohen's kappa bedroeg .79).

4.1.1 Kenmerken van de verzamelde onderzoeken

De onderzoeken zijn op twee na afkomstig uit de Verenigde Staten. De verdeling over de afgelopen drie decennia is evenwichtig. Ruim de helft (16) zijn ongepubliceerde dissertaties. De meeste studies hebben een quasi-experimenteel design. De 30 onderzoeken leverden in totaal 54 effectgroottes op. Bij één studie werd het maximum van 4 effectgroottes berekend.

De grootte van de experimentele en controlegroepen varieert van 13 tot 865 personen. Dat laatste is een uitzondering, verder zijn alle onderzochte groepen beneden de 100 personen. Het totale aantal proefpersonen in alle onderzoeken is 3318. De gemiddelde leeftijden van de onderzochte groepen leerlingen liggen tussen de 4 en de 33 jaar.

Bij 31 effectgroottes is de inhoud van het onderzochte onderwijs productief: met name tekenen en schilderen en in een minderheid handenarbeid (boetsen, keramiek, draadcon-

structies). De rest betreft een combinatie van productief en reflectief, met name de combinatie van kunstbeschouwing en tekenen. Kunstbeschouwing moet hierbij ruim worden opgevat. Deels betreft het onderwijs dat niet of nauwelijks kunstinhoudelijk is en beter kan worden omschreven als waarnemingstraining. Voorbeelden hiervan zijn lessen voor 4- en 5-jarigen met als lesmateriaal geïllustreerde boekjes met "De avonturen van 'rechte lijn' 'hoek' en 'curve'" (Salome & Reeves, 1972) of oefeningen in het ontdekken van verborgen of onvolledige figuren (Szeto & Salome, 1977).

De meeste studies betreffen onderwijs dat een periode van 5 tot 15 weken beslaat. De tijds-aanduiding is niet exact, omdat vaak de informatie ontbrak om hoeveel uren het per week of in totaliteit ging. In enkele gevallen heeft de onderzoeker zelf een nieuwe test ontworpen, maar de meeste tests betreffen gestandaardiseerde psychologische tests of adaptaties ervan. Tot die laatste categorie horen bijvoorbeeld aanpassingen van Witkin's Embedded Figures Test voor jonge kinderen.

4.1.2 Resultaten van de meta-analyse

In tabel 1 worden de uitkomsten van de meta-analyse van leereffecten op visueel-ruimtelijke vermogens weergegeven.

Tabel 1. Meta-analyse van effecten van beeldende vakken op visueel-ruimtelijke vermogens

I 'Bare bones' meta-analyse gebaseerd op 54 effectgroottes	
Gewogen gemiddelde van de effectgroottes (\bar{d})	0.207
Ongecorrigeerde variantie	0.108
Variantie gecorrigeerd voor steekproeffout	0.065
90% betrouwbaarheidsinterval van de gemiddelde effectgrootte	$-0.118 < \bar{d} < 0.532$
II 'Bare bones' meta-analyse gebaseerd op één effectgrootte per studie	
Gewogen gemiddelde van de effectgroottes (\bar{d})	0.208
Ongecorrigeerde variantie	0.087
Variantie gecorrigeerd voor steekproeffout	0.063
90% betrouwbaarheidsinterval van de gemiddelde effectgrootte	$-0.114 < \bar{d} < 0.529$

Het gemiddelde 0.207 betekent dat de onderzochte lessen beeldende vorming leiden tot en verbetering van visueel-ruimtelijke vermogens met 0.207 standaarddeviatie eenheden.

Men kan de gemiddelde effectgrootte ook omzetten in een percentiel equivalent (Cohen, 1988). Het met 0.207 corresponderende percentiequivalent is 58. Dit houdt in dat 58% van de scores in de controlegroepen beneden de gemiddelde score in de experimentele groepen ligt. Op grond van de intervallenschatting van de gemiddelde effectgrootte (Hunter & Schmidt, 1990, p. 283) concluderen we dat het populatie-effect niet significant verschilt van 0.

Met behulp van een homogeniteitsanalyse is nagegaan of er sprake is van een gemeenschappelijke effectgrootte. De resultaten zijn weergegeven in tabel 2. Uit de significantie van Q_1 blijkt dat de effectgroottes heterogeen zijn en in subgroepen kunnen worden verdeeld. Van enkele variabelen werd op basis van theorie en eerder onderzoek verwacht dat ze kunnen bijdragen aan verschillen in effectgroottes. De verwachtingen ten aanzien van deze moderator-

variabelen zijn hieronder weergegeven. Tabel 2 laat zien of subgroepen ingedeeld volgens betreffende variabelen van elkaar verschillen.

1. Verwacht wordt dat effecten van beeldende vorming op visueel-ruimtelijke tests met herkenningstaken hoger zullen zijn dan effecten op visueel-ruimtelijke tests met manipulatie-taken. Deze verwachting was gebaseerd op de theorie van Eliot (1987). Onderzoeken die visueel-ruimtelijke vermogens meten met herkenningstaken lieten inderdaad meer effect zien dan onderzoeken met tests die gericht waren op de meer complexe manipulatie-taken, zoals het mentaal kunnen roteren van vormen.
2. Verwacht wordt dat beeldende lessen met specifieke waarnemingstraining hogere effecten opleveren dan meer algemene, kunstinhoudelijke lessen. Deze verwachting, gebaseerd op onderzoek van o.a. Beanninger en Newcombe (1989) werd niet bevestigd. De treatment-variabele gaf geen significante verschillen tussen subgroepen te zien.
3. Verwacht wordt dat effecten bij jonge kinderen het grootst zullen zijn. Deze verwachting, gebaseerd op onderzoek (Lean, 1981) en op ontwikkelingspsychologische theorieën, werd bevestigd. De variabele leeftijd laat significante verschillen tussen subgroepen zien en de hoogste effecten zijn gevonden bij de jongste groep, de 4- tot 6-jarigen.

Geen van de moderatorvariabelen alleen leidt tot homogene subgroepen. Formele post-hoc toetsing van contrastgroepen is om deze reden niet uitgevoerd. Het beste 'model' is de indeling naar leeftijd. Een mogelijk volgende stap in de analyse is om te kijken naar combinaties van kenmerken. Gezien het beperkte aantal effectgroottes is hier van afgezien.

Wel is ter controle nagegaan of de methodologische kwaliteit van de onderzoeken van invloed was op de effectgrootte. Hieraan werd de voorkeur gegeven boven een a priori verwerpen van onderzoeken op designkenmerken. In de eerste plaats is gekeken naar eventuele verschillen tussen zuiver experimenteel en quasi-experimenteel onderzoek. Volgens Wortman (1983) zijn er aanwijzingen dat gebrek aan experimentele controle tot inflatie van effectgroottes kan leiden. Lipsey en Wilson (1993) gingen in hun analyse van ruim 300 meta-analyses eveneens uit van deze veronderstelling. Zij vonden echter dat random designs gemiddeld een hogere effectgrootte laten zien. Ook in ons geval bleek niet dat quasi-experimentele opzetten gemiddeld hogere effectgroottes vertonen dan experimentele opzetten. Bij een deel van de onderzoeken fungeerde de onderzoeker zelf als docent in een of meer condities. De kennis van de onderzoeker/docent omtrent de verwachte uitkomsten zouden de onderzoeksresultaten kunnen vertekenen. De resultaten geven echter aan dat onderzoeken waarin de onderzoeker ook optrad als docent niet tot hogere effectgroottes leiden.

Tenslotte is nagegaan of er sprake is van een zogeheten 'file-drawer' effect (Rosenthal, 1979): het gevaar dat niet-significante onderzoeksuitkomsten een kleinere kans hebben om gepubliceerd te worden en eerder in de bureaula verdwijnen. De resultaten geven geen aanwijzing dat de gepubliceerde studies hogere effectgroottes vertonen. Overigens vonden Lipsey en Wilson (1993) in hun overzicht wel een bevestiging van het file drawer effect.

4.2 Effect van beeldende vakken op esthetische waarneming

In eerste instantie werden 68 relevant geachte titels verzameld. Om diverse redenen vielen 29 titels af. In enkele gevallen waren onderzoeksverslagen niet op tijd verkrijgbaar en soms waren de gerapporteerde kwantitatieve resultaten ontoereikend om een effectgrootte te berekenen. Ook vielen bepaalde soorten onderwijs en onderzochte effecten (de operationalisering ervan in meetinstrumenten) buiten de criteria. Het ging met name om instructie die te ver afstond van de reguliere beeldende vorming. Het betrof dan zeer korte trainingssessies in laboratoriumachtige situaties. De onderwijsmethode hield daarbij vaak geen verbale of schriftelijke uitleg in, maar nonverbale methoden als het belonen van juiste keuzes. Uiteindelijk bleven 39 onderzoeksverslagen over. Deze staan vermeld in bijlage 2. Wederom werd een derde van de studies dubbel gecodeerd en was de gemiddelde inter-beoordelaarsbetrouwbaarheid bevredigend (Cohen's kappa .80).

Tabel 2. Homogeniteitsanalyse van effectgroottes m.b.t. visueel ruimtelijk inzicht

Kenmerk	N	\bar{d}	sd	Q_{wi}	Q_w	Q_b	Q_t
Totale groep	54	.21	.03	--	--	--	119.77 *
1. Test taak:							
herkenning	42	.26	.04	104.87*			
manipulatie	12	.06	.06	7.45	112.32*	7.44	
2. Treatment experimentele groep:							
algemene beeldende vorming	28	.21	.04	48.79			
gerichte waarnemingstraining	11	.12	.07	25.44	116.95*	2.81	
combinatie	15	.30	.07	42.71*			
3. Leeftijd leerlingen:							
4 - 6 jr	10	.54	.09	15.29			
7 - 12 jr	20	.19	.04	39.93*	102.61*	17.16*	
13 - 20 jr	15	.08	.07	30.86			
21 jr of ouder	9	.22	.11	16.51			
4.1 Design:							
quasi-experimenteel	21	.15	.04	37.88			
quasi-experimenteel met match.	18	.16	.07	51.40*	102.83*	7.32	
experimenteel	12	.38	.08	13.56			
4.2 Onderzoeker als docent?							
nee	27	.23	.04	68.32*			
ja	21	.08	.06	21.71	90.03*	4.44	
5. Publicatie:							
ongepubliceerde dissertatie	29	.22	.05	65.99*			
gepubl. artikel/boek/ERIC doc.	25	.21	.04	53.76*	119.75*	0.02	

* $p < .005$ N = aantal effectgroottes; \bar{d} = gemiddelde effectgrootte; sd = standaarddeviatie van effectgroottesVoor Q_{wi} , df = aantal effectgroottes binnen elke groep minus één;Voor Q_w , df = totale aantal effectgroottes (N) minus het aantal betrokken groepen;Voor Q_b , df = het aantal groepen minus één;Voor Q_t , df = het totale aantal effectgroottes minus éénNB: Q_t is alleen in de eerste rij gerapporteerd, omdat deze identiek is voor alle vergelijkingen.

4.2.1 Beschrijving van de onderzoeken

De onderzoeken zijn op vijf na afkomstig uit de Verenigde Staten. De meeste onderzoeken zijn verricht in de laatste twee decennia. Bijna de helft (18) zijn ongepubliceerde dissertaties. De meeste studies zijn quasi-experimenteel van opzet. De 39 onderzoeken leverden in totaal 79 effectgroottes op. Bij vier studies werd het maximum van vier effectgroottes berekend. De grootte van de experimentele en controlegroepen varieert van 10 tot 135 personen. Het totale aantal proefpersonen in alle onderzoeken is 3987. De gemiddelde leeftijden van de onderzochte groepen leerlingen liggen tussen de 4 en de 65 jaar. De meeste onderzoekssituaties betreffen het reguliere onderwijs (basisonderwijs en voortgezet onderwijs). In een klein aantal onderzoeken gaat het om buitenschoolse beeldende vorming, zoals museumlessen en een cursus voor ouderen.

De onderzochte lessen betreffen zowel productieve beeldende vorming (in meerderheid tekenen en schilderen), reflectieve vorming (kunstbeschouwing en kunstgeschiedenis, soms aangevuld met museumbezoek), als combinaties hiervan. Reflectieve vorming omvatte in enkele gevallen de traditionele 'praatje-bij-een plaatje' lessen (docent vertelt bij dia's) maar vaker betrof het kunstbeschouwing/kunstgeschiedenis uitgaande van een bepaald besprekingsmodel, zoals van Mittler (1980) of Broudy (1987). Daarbij worden kunstwerken getoond en volgens een bepaald stramen behandeld, zoals: beschrijving, analyse (vormgericht), interpretatie (betekenisgericht) en evaluatie (eindoordeel).

De meeste studies onderzoeken onderwijs dat een periode van 5 tot 15 weken beslaat. Evenals in de andere analyse is de tijdsaanduiding noodgedwongen globaal, omdat exacte informatie in veel verslagen ontbrak. De gebruikte instrumenten betreffen voorkeursschalen, beoordelingsschalen en verbale reacties. In ongeveer 60% van de gevallen heeft de onderzoeker zelf een nieuw meetinstrument ontworpen of heeft een bestaande test aangepast.

In tabel 3 worden de uitkomsten van de meta-analyse van leereffecten op esthetische waarneming weergegeven.

Tabel 3. Meta-analyse van effecten van beeldende vorming op esthetische waarneming

I 'Bare bones' meta-analyse gebaseerd op 79 effectgroottes

Gewogen gemiddelde van effectgroottes (\bar{d})	0.533
Ongecorrigeerde variantie	0.215
Variantie gecorrigeerd voor steekproeffout	0.151
90% betrouwbaarheidsinterval van de gemiddelde effectgrootte	$0.035 < \bar{d} < 1.030$

II 'Bare bones' meta-analyse gebaseerd op één effectgrootte per studie

Gewogen gemiddelde van de effectgroottes (\bar{d})	0.536
Ongecorrigeerde variantie	0.164
Variantie gecorrigeerd voor steekproeffout	0.132
90% betrouwbaarheidsinterval van de gemiddelde effectgrootte	$0.071 < \bar{d} < 1.00$

Het gemiddelde 0.533 betekent dat de onderzochte lessen beeldende vorming leiden tot een verbetering van esthetische waarneming met 0.533 standaarddeviatie eenheden. Het corresponderende percentieequivalent is 70. Dit houdt in dat 70 % van de scores in de controle-

groepen beneden de gemiddelde score in de experimentele groepen ligt. Op grond van de intervaleschatting concluderen we dat het populatie-effect significant verschilt van 0.

Op de verzamelde effectgroottes is een homogeniteitsanalyse uitgevoerd. De effectgroottes zijn heterogeen (zie tabel 4). Van een beperkt aantal relevant geachte moderatorvariabelen is nagegaan of ze bijdragen aan verschillen in effectgroottes.

1. In navolging van de eerste meta-analyse is nagegaan of het gebruik van verschillende testtaken (voorkeur, oordeel en verbale reacties) verband hield met de grootte van de effecten. Dat bleek het geval: de gemiddeld laagste effecten werden gemeten met voorkeursschalen, de hoogste effecten met analyses van verbale reacties. Effecten op beoordelingsschalen zitten daar tussenin. Een mogelijke verklaring is de mate van overeenkomst tussen de leersituatie en de toetsingssituatie. Het geven van verbale reacties wordt met name in reflectieve vorming direct onderwezen: de leerlingen wordt een vocabulair bijgebracht waarmee ze beter onder woorden leren brengen wat er aan kunstwerken te zien is. Voorkeursschalen zijn een meer afgeleide toetsvorm. Ze hebben bovendien een grotere houdingscomponent, die minder beïnvloedbaar is.
2. Verwacht wordt dat lessen reflectieve beeldende vorming gemiddeld hogere effecten op esthetische waarneming zullen opleveren dan lessen productieve beeldende vorming. Deze verwachting was gebaseerd op eerder onderzoek (o.a. Wilson, 1981; Salome, 1991) en werd bevestigd. Echter de hoogste effecten, dus ook hoger dan de reflectieve lessen, werden gevonden in de studies die een combinatie van productieve en reflectieve lessen onderzochten. In enkele onderzoeken werden productieve lessen als controlegroep gebruikt en vergeleken met reflectieve aanpak of een combinatie. Deze onderzoeken bevestigden het beeld van de grotere effectiviteit op esthetische waarneming van reflectieve lessen in vergelijking tot productieve.
3. Verwacht wordt dat effecten bij jonge kinderen het laagst zullen zijn. Deze verwachting was gebaseerd op theorieën over ontwikkelingsstadia van esthetische waarneming, waarin wordt gesteld dat beschouwers pas op latere leeftijd oog (kunnen) krijgen voor esthetische kenmerken met betrekking tot vorm, symboliek, etc. De leeftijdsvariabele draagt bij aan verschillen in effecten. De laagste effecten zijn gevonden bij de jongste groep, de 4 tot 6 jarigen en de hoogste bij de volwassen studenten.

Geen van de moderatorvariabelen alleen leidt tot homogene subgroepen. Formele post-hoc toetsing van contrasten is daarom achterwege gelaten. Gezien het beperkte aantal effectgroottes is eveneens afgezien van het verder opsplitsen van de verzameling effectgroottes door het combineren van moderatorvariabelen.

Evenals in de meta-analyse van effecten op visueel-ruimtelijk vermogen is ter controle nagegaan of de methodologische kwaliteit en het al of niet gepubliceerd zijn van de onderzoeken van invloed was op de effectgrootte. Het onderscheid tussen zuiver experimenteel en quasi-experimenteel onderzoek was niet verbonden met variaties in effecten. Echter onderzoek waarin onderzoeker en docent dezelfde persoon waren had gemiddeld hogere effecten. Dat wijst erop dat in die groep studies de impliciete verwachtingen van de docent/onderzoeker de uitkomsten hebben vertekend. Zo'n sociale beïnvloeding is waarschijnlijk eerder mogelijk bij het meten van een waardegebonden concept als esthetische waarneming dan bij een cognitieve vaardigheid als visueel-ruimtelijk inzicht. Uit de resultaten bleek dat er geen sprake was van een file-drawer effect.

Tabel 4. Homogeniteitsanalyse van effectgroottes m.b.t. esthetische waarneming

Kenmerk	N	\bar{d}	sd	Q_{wi}	Q_w	Q_b	Q_t
Totale groep	78	.50	.03	--	--	--	245.14 *
<i>1. Test taak:</i>							
esthetische oordeelsschaal	34	.51	.05	81.40*			
esthetische voorkeurschaal	24	.31	.04	73.90*	184.08*	61.05*	
verbale beschrijving	20	.89	.06	28.78			
<i>2. Treatment experimentele groep:</i>							
productief	17	.24	.06	14.00			
reflectief	24	.44	.05	77.12*	200.72*	44.41*	
combinatie prod. en reflect.	37	.72	.04	109.60*			
<i>3. Leeftijd leerlingen:</i>							
4 - 6 jr	6	.26	.13	6.46			
7 - 12 jr	29	.48	.04	97.33*	231.76*	13.38*	
13 - 20 jr	32	.50	.05	96.71*			
21 jr of ouder	11	.79	.09	31.25*			
<i>4.1 Design:</i>							
correlatieel	4	.56	.03	4.90			
quasi-experimenteel	48	.53	.03	198.16*	239.55*	5.58	
quasi-experimenteel met match.	4	.57	.17	3.28			
experimenteel	18	.36	.07	33.21			
<i>4.2 Onderzoeker als docent?</i>							
nee	36	.39	.04	74.05*	226.33*	18.81*	
ja	42	.64	.04	152.28*			
<i>5. Publicatie:</i>							
ongepubliceerde dissertatie	39	.48	.04	124.18*	244.52*	.62	
gepubl. artikel/boek/ERIC doc.	39	.52	.04	120.34*			

* $p < .005$ N = aantal effectgroottes; \bar{d} = gemiddelde effectgrootte; sd = standaarddeviatie van effectgroottesVoor Q_{wi} , df = aantal effectgroottes binnen elke groep minus één;Voor Q_w , df = totale aantal effectgroottes (N) minus het aantal betrokken groepen;Voor Q_b , df = het aantal groepen minus één;Voor Q_t , df = het totale aantal effectgroottes minus éénNB: Q_t is alleen in de eerste rij gerapporteerd, omdat deze identiek is voor alle vergelijkingen.

5. DISCUSSIE

5.1 Onderwijsrelevantie en vervolgonderzoek

De algemene conclusie van de twee meta-analyses luidt dat de beschikbare onderzoeksresultaten geen steun geven aan de verwachting dat beeldende vorming direct bijdraagt aan verbetering van visueel-ruimtelijke vermogens. Een uitzondering vormen de resultaten met jonge kinderen. Wel draagt beeldende vorming bij tot verbetering van esthetische waarneming.

Visueel-ruimtelijke vermogens en esthetische waarneming zijn in deze studie opgevat als respectievelijk de instrumentele en de intrinsieke component van de doelstelling 'leren zien'. Er zijn belangrijke verschillen in de aard van de beide onderzochte leereffecten. Esthetische waarneming zoals geoperationaliseerd in de verzamelde studies, is 'kennis-intensief': het doet een direct beroep op de in het onderzochte onderwijs verworven domeinspecifieke kennis. Dit geldt in de hoge mate voor de verbale reacties op kunstwerken. Visueel-ruimtelijke taken zijn 'proces-intensief' (Lohman, 1989) en vereisen algemene probleemoplossingsvaardigheden, die samenhangen met algemene intelligentiefactoren.

Zoals aangegeven werd door verschillende auteurs de samenhang tussen visueel-ruimtelijke en beeldende vermogens als theoretisch plausibel beschouwd. Het gaat echter waarschijnlijk om specifiekere relaties dan die in de samengevatte studies zijn onderzocht. Er zal daarom nauwkeuriger geanalyseerd moeten worden, welke visueel-ruimtelijke processen een rol spelen in de verschillende vormen van beeldende kunst (o.a. twee-dimensionaal versus drie-dimensionaal werken) en in de verschillende fasen van het beeldende proces. Het meten van visueel-ruimtelijke vaardigheden zal bovendien niet uitsluitend met de gebruikelijke "paper-and-pencil" tests moeten gebeuren. Kosslyn (in Haanstra, 1994) ziet beperkte mogelijkheden van beïnvloeding van bepaalde deelvaardigheden van het mentale voorstellingsvermogen ('mental imagery') door beeldende vorming. Naar zijn mening is de overdracht van trainingseffecten naar andere taken en contexten beperkt, omdat de specifieke inhoud van de taak en specifieke oplossingsstrategieën een belangrijke rol spelen.

De verscheidenheid aan visueel-ruimtelijke vaardigheden, het belang van taakinhoudelijke aspecten en de beperkte overdracht van training naar andere gebieden duiden op complexe relaties van deze vaardigheden met specifieke beeldende vaardigheden en/of bepaalde fasen in het beeldend proces. Dit is voorlopig meer een interessant gebied voor fundamenteel onderzoek dan voor toegepast onderwijskundig onderzoek met directe implicaties voor de onderbouwing en inrichting van het beeldend onderwijs.

Bij veel van de verzamelde studies naar esthetische waarneming ligt de nadruk van het onderwijs op het leren visueel onderscheiden van formele esthetische kenmerken. De gebruikte instrumenten als beoordelings- en voorkeursschalen, maar ook de inhoudsanalyses van verbale reacties zijn eveneens sterk op vormaspecten gericht. Van verschillende kanten (o.a. Van Rheeden, 1993) is gewezen op de gevaren van een steriel aanleren van kijk- en besprekingschema's van kunst (het 'beeldend boekhouden') en op het normatieve en tijdgebonden karakter van bepaalde esthetische vormprincipes. Het oordeel over hedendaagse beeldende kunst hangt minder af van het oog hebben voor kleurcontrasten, lijnvoering, vormen en contra-vormen, en veel meer van theoretische inzichten, interpretaties en herinterpretaties. De uitdrukking 'leren zien' is in deze opvatting nauwelijks nog relevant te noemen. In de praktijk van het beeldend onderwijs is een dergelijke 'postmoderne' benadering echter niet gemeengoed geworden en is de doelstelling leren zien nog actueel, zij het dat het sterk normatieve karakter van de oudere tests voor esthetische waarneming door de meerderheid wordt afgewezen.

Voor de onderwijspraktijk is het verband tussen onderwijsinhoud en de hoogte van effecten op esthetische waarneming van belang. Onderwijsprogramma's die zowel reflectieve als productieve beeldende vorming omvatten lieten de grootste effecten op esthetische waarneming zien. In verschillende curriculumontwikkelingsprojecten in de beeldende vakken is ook als uitgangspunt gekozen dat 'scheppen' en 'schouwen' volledig geïntegreerd dienen te worden. Bij een deel van de studies in de meta-analyse is inderdaad sprake van zo'n

integratie, bijvoorbeeld doordat een bepaalde techniek, een bepaald thema of een kunststroming zowel productief als reflectief worden behandeld. Omdat de onderzoeksrapportages soms vrij summier zijn weten we van andere studies alleen dat beide aspecten aangeboden zijn, zonder dat duidelijk is in hoeverre ze in het onderwijs op elkaar betrokken worden. Vervolgonderzoek zou zich kunnen richten op de vraag in hoeverre een strikt geïntegreerde aanpak noodzakelijk is of dat een veel losser verband tussen productieve en reflectieve elementen al voldoende is om wenselijke effecten, waaronder verbetering van esthetische waarneming te bewerkstelligen.

5.2 Legitimering, empirisch onderzoek en meta-analyse

Zowel in de Verenigde Staten als in Nederland moet de positie van beeldende vakken in het onderwijs steeds weer bevochten worden. In de inleiding is al gesteld dat vaak instrumentele doelen worden aangevoerd om het nut van deze vakken te beklemtonen. In Nederland is dat bijvoorbeeld te zien bij de discussie over de positie van de beeldende vakken in de tweede fase van het voortgezet onderwijs. Daarbij wordt sterk de nadruk gelegd op de mogelijke bijdrage van beeldende vakken aan algemene vaardigheden (Van den Akker, Arnold & Teule, 1994). En in de huidige experimenten in het basisonderwijs met de verlengde schooldag en de zogeheten magneetscholen aan de kunstvakken een rol toegedacht bij het bevorderen van leerprestaties (Oostwoud Wijdenes, Oud & Haanstra, 1994)

Het legitimeringsdebat in de beeldende vorming wordt in de eerste plaats gevoerd vanuit verschillende waardenorientaties en theoretische invalshoeken, maar dient ons inziens zo veel mogelijk te worden onderbouwd met empirische gegevens. Doordat meta-analyses een grote hoeveelheid uiteenlopende onderzoeksuitkomsten op systematische en controleerbare manier hanteerbaar maken, kunnen ze een goede rol vervullen bij het legitimeringsdebat. De meta-analyses in deze studie zijn gebaseerd op voornamelijk Amerikaans onderzoek, maar gezien de parallellen tussen het Amerikaanse en het Nederlandse legitimeringsdebat en de vele gemeenschappelijke inhoudelijke aspecten van het beeldende onderwijs zijn de uitkomsten relevant voor de Nederlandse situatie.

Die uitkomsten ondersteunen degenen die zich sterk maken voor een vakgerichte aanpak met intrinsieke (kunsthoudelijke) doelen, en manen tot voorzichtigheid bij het hanteren van bepaalde instrumentele doelen, zoals bijvoorbeeld het bevorderen van cognitieve vaardigheden.

Bij veel kunstzinnige vormers bestaat overigens scepsis over de rol van empirisch onderzoek in het legitimeringsdebat. Het vergt vaak langdurig, intensief (en dus kostbaar) onderzoek om effecten aan te kunnen tonen. Meta-analyses vereisen bovendien gekwantificeerde onderzoeksresultaten. Kwantitatief onderzoek gericht op meetbaar gedrag is zeker met betrekking tot de evaluatie van beeldende vorming (en van kunsteducatie in het algemeen) vaak bekritiseerd (bijvoorbeeld Eisner, 1985). Mede door deze kritiek is de omvang van (quasi-) experimenteel onderzoek naar effecten van beeldend onderwijs beperkt gebleven.

Met betrekking tot sommige doelstellingen zullen de meeste betrokkenen geen moeite hebben met een kwantitatieve aanpak. Bijvoorbeeld om na te gaan in hoeverre de stelling waar gemaakt kan worden dat de beeldende vakken kunnen bijdragen aan betere schoolprestaties in andere leergebieden. In deze studie ging het om leereffecten met betrekking tot complexere en meer waardegebonden begrippen en was de vertaling van deze begrippen in meetinstrumenten niet eenduidig en op een aantal punten discutabel. Toch is meer inzicht in de haalbaarheid van deze leereffecten verkregen en zijn verbanden tussen effecten en kenmerken van onderwijs en leerlingen opgespoord die relevant zijn voor de onderwijspraktijk. Gepleit wordt daarom voor voortzetting van (quasi-)experimenteel onderzoek ook met betrekking tot de moeilijk grijpbare leereffecten en leerervaringen. Wel zal het tot ver in de 21ste eeuw duren voor er zinvolle meta-analyses kunnen worden uitgevoerd op uitsluitend Nederlands onderzoek op het gebied van de kunsteducatie.

LITERATUUR

- Acuff, B.C. & Sieber-Supes J. (1972). *A manual for coding descriptions, interpretations, and evaluations of visual art forms*. Palo Alto: Stanford University, Stanford Center for Research and Development in Teaching.
- Akker, L. van den, Arnold F. & Teule B. (1993). Het belang van de beeldende vakken voor een vervolgstudie op HBO en WO. *Beeldaspecten*, Nr.7/8, 2-6.
- Allison, B. (1986). *Index of British studies in art and design education*. Aldershot, Gower Publishing Group.
- Bangert-Drowns, R.L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, 99(3), 388-399.
- Beanninger, M. & Newcombe, N. (1989). The role of experience in spatial test performance: a meta-analysis. *Sex Roles*, 20 (5/6), 327-344.
- Beardsley, M.C. (1981). *Aesthetics: Problems in the philosophy of criticism*. Indianapolis, Hackett.
- Bergland, D.R. & MacGregor, R.N. (1988). Strategies adopted by elementary school students in shape manipulation and shape recognition tasks. *Visual Arts Research*, 14(2), 51-56.
- Berlyne D.E. (Ed.) (1974). *Studies in the new experimental aesthetics*. New York: John Wiley and Sons.
- Boyer, A.B. (1983). An examination of experimental aesthetic research from 1970 to 1980 related to the visual perception of paintings with implications for art education. *Visual Arts Research*, 9(1), 34-41.
- Broudy, H.S. (1987). *The role of imagery in learning*. Los Angeles: The Getty Center for Education in the Arts.
- Burton, J.M. (1992). Art education and the plight of the culture: A status report. *Art Education*, 45(4), 7-18.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Dickie, G. (1983). Perceiving art. *Visual Arts Research*, 9(2), 66-70.
- Eisner, E.W. (1982). Aesthetic education. In H.W. Mitzel, J.H. Best & W. Rabinowitz (Eds.), *Encyclopedia of educational research* (Vol 1, pp. 87-94). New York: MacMillan.
- Eisner, E.W. (1985). *The art of educational evaluation: A personal view*. London & Philadelphia: The Falmer Press.
- Elias, W. (1993). *Tekens aan de wand: Hedendaagse stromingen in de kunsttheorie*. Antwerpen-Baarn: Uitgeverij Hadewijch.
- Eliot, J. (1987). *Models of psychological space: Psychometric, developmental, and experimental approaches*. New York: Springer Verlag.
- Eliot, J., Kuo-Hsuing Ho, Olson D. & Medoff D. (1988). *The MICA study: the relationship between spatial abilities and art achievement*. University of Maryland.
- Fechner, G. (1876). *Vorschule der Aesthetik*. Leipzig: Breitkopf & Härtel.
- Gardner, H. (1985). *Frames of mind: The theory of multiple intelligences*. London: Paladin Books, Granada Publishing Ltd.
- Glass, G.V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glass, G.V., McGaw B. & Smith M.L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.
- Groot, A.D. de (1986). *Begrip van evalueren*. s'-Gravenhage: VUGA.
- Graves, M. (1948). *Design judgement test*. New York: The Psychological Cooperation.
- Haanstra, F. (1994). *Effects of art education on visual-spatial ability and aesthetic perception: Two meta-analyses*. (Academisch proefschrift) Amsterdam: Thesis Publishers.
- Haanstra, F. & Oijen, L. van (1985). *Leereffecten van kunstzinnige vorming. Een inventarisatie van onderzoek*. Amsterdam: SCO.
- Haanstra, F., Van der Kamp, M. & Oostwoud Wijdenes, J. (1992). *De steunfunctie kunstzinnige vorming in de provincie Gelderland*. Amsterdam: SCO Rapport 282.
- Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hermans, P. & Schönau, D. (1988). Reflectieve beeldende vorming tussen schouwen en scheppen. *Kunsten & Educatie*, 1(2), 45-52.
- Hermelin, B., & O'Connor, N. (1986). Spatial representations in mathematically and artistically gifted children. *British Journal of Educational Psychology*, 56, 150-157.
- Housen, A. (1983). *The eye of the beholder: measuring aesthetic development*. Unpublished dissertation. Harvard University.
- Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park: Sage Publications.
- Johnson, B.T., Mullen, B. & Salas, E. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology*, 80(1), 94-106

- Kooyman, R. (ed.) (1989). *Bibliografie Onderzoek Kunstzinnige Vorming*. Utrecht: LOKV.
- Lean, G.A. (1981). *Training of spatial abilities: A bibliography*. Papua New Guinea: Mathematics Education Centre, University of Technology.
- Lindauer, M.S. (1981) Aesthetic experience: a neglected topic in the psychology of the arts. In D. O'Hare (Ed.) *Psychology and the arts*. Sussex: The Harvester Press.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48(12), 1181-1209.
- Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181-248). Hillsdale, N.J.: Erlbaum.
- Lohman, D. F. (1989). Human intelligence: an introduction to advances in theory and research. *Review of Educational Research*, 59(4), 333-373.
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86(5), 889-918.
- McWhinnie, H. (1973). Perceptual learning in art, possible or impossible? *Scientia Paedagogica Experimentalis*, 10, 101-109.
- McWhinnie, H.J. (1991). *The identification of spatial types one*. Final Report. University of Maryland.
- Meel-Jansen, A. van (1974). *Kreativiteit en cognitieve stijl*. Den Haag: Mouton.
- Meier, N. (1940) *The Meier art tests: I Art judgment*. Iowa City: Bureau of Educational Research and Service, Iowa.
- Meier, N. (1963). *Meier art tests, II Aesthetic perception*. Iowa City: Bureau of Educational Research and Service, Iowa.
- Meijer, J. & Kok, F. (1991). *Program Meta*. Amsterdam: Center for Educational Research, University of Amsterdam.
- Mittler, G.A. (1980). Learning to look/looking to learn: a proposed approach to art appreciation at the secondary school level. *Art Education*, 33, 17-21.
- Oostwoud Wijdenes, J.D., Oud, W. & Haanstra, F. (1994). *Kunstmagneetscholen: een literatuurstudie*, Utrecht, LOKV.
- Parsons, M.J. (1987). *How we understand art: A cognitive developmental account of aesthetic experience*. Cambridge: Cambridge University Press.
- Parsons, M.J. (1992). Cognition as interpretation in art education. In B. Reimer & R. A. Smith (Eds.) *The arts, education and aesthetic knowing* (Ninety-First Yearbook of the National Society for the Study of Education, Part II). Chicago: The University of Chicago Press.
- Rheeden, H. (1989). *Om de vorm: Een eeuw teken-handenarbeid-en kunstnijverheidsonderwijs in Nederland*. Amsterdam: SUA.
- Rheeden, H. van (1993). Wat er op lijkt, maar het niet is. *Maandblad voor de Beeldende Vakken*, 6, p.22
- Rice, D. (1988). Vision and culture: the role of museums in visual literacy. *The Journal of Museum Education*. Fall, 13-17.
- Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. (Rev. ed.). Beverly Hills: Sage Publications.
- Salome, R.A. (1991a). Research on drawing. *Translations from theory to practice*, 1 (1), NAEA.
- Salome, R.A. (1991b). Research on perceiving and responding to art. *Translations from theory to practice*, 1(2), NAEA.
- Salome, R.A. & Reeves, D. (1972). Two pilot investigations of perceptual training of four and five year old children. *Studies in Art Education*, 13(2), 3-9.
- Szeto, J.W. & Salome, R.A. (1977). The effects of scanning practice and perceptual training upon several visual functions: implications for drawing. *Studies in Art Education*, 19(1), 45-51.
- Wilson, B. (1971). Evaluation of learning in art education. In: B.S. Bloom, J.T. Hastings & G.F. Madaus (eds.), *Handbook on formative and summative evaluation of student learning* (pp. 500-558), New York: McGraw-Hill.
- Wilson, B. (1981). The triumph of American culture over the art educational establishment: findings from the U.S. National Assessment of Educational Progress in Art. In H. van Merwijk (Ed.) *The Product of a Process*; a selection of papers presented at the 24th INSEA World Congress: Rotterdam.
- Winner, E., Casey, M., Dasilva, D. & Hayes, R. (1991). Spatial abilities and reading deficits in visual art students. *Empirical Studies of the Arts*, 9(1), 51-63.
- Winner, E. & Casey M. (1992). Cognitive profiles of artists. In G. Cupchik & J. Laszlo (Eds.), *Emerging visions of the aesthetic process in psychology, semiology, and philosophy*. Cambridge: University Press.

- Winner E. & Pariser, D. (1985). Giftedness in the visual arts. *Items*, December, 65-69.
- Witkin, H.A. & Goodenough, D.R. (1981). *Cognitive styles: Essence and origins*. Field dependence and field independence. New York: International Universities Press.
- Wortman, P.M. (1983). Evaluation research: A methodological perspective. *Annotated Review of Psychology*, 34, 223-260.

BIJLAGE

Onderzoeken opgenomen in de meta-analyses

1. De 30 onderzoeken naar effecten op visueel-ruimtelijke vermogens

* = onderzoek ook opgenomen in meta-analyse esthetische waarneming

1. Comet, J.J.K. (1986). *The effect of a perceptual drawing course on other perceptual tasks*. Unpublished dissertation. Yeshiva University.
2. Correro, G.C. (1986). *The effect of visual art training upon visual perception measures of certain first grade children*. Unpublished dissertation, University of Alabama.
3. Dorethy, R.E. (1973). Motion parallax as a factor in the differential spatial abilities of young children. *Studies in Art Education*, 14(2), 15-27.
4. Efland, A. (1965). *The effect of perceptual training upon the differentiation of form in children's drawing*. Unpublished dissertation. Stanford University.
5. Gair, S.B. (1975). An art based remediation program for children with learning disabilities. *Studies in Art Education*, 17(1), 55-67.
6. Janov, D.R. (1986). *The effects of structured criticism upon the perceptual differentiation and studio compositional skills displayed by college students in an elementary art education course*. Unpublished dissertation. Ball State University.
7. Johnson, E.C. (1976). *A comparison of the effects of two programs on the development of visual perceptions and reading achievement: Art, A perceptual approach and the Frostig program for the development of visual perception*. Unpublished dissertation. Indiana University.
8. Kannegieter, R.B. (1971). The effects of a learning program in activity upon the visual perception of shape. *Studies in Art Education*, 12(2), 18-27.
9. Kensler, G.L. (1964). *Perceptual training and utilization of space cues in perspective drawing*. Unpublished dissertation. Stanford University.
10. McCloskey, P. (1979). The facilitation of spatial ability and problem solving in adolescent pupils through learning in design. *Educational Review*, 31(3), 259-267.
11. McCord, M.M. (1973). *Comparative effects of selected art experiences upon factors of visual perception*. Unpublished dissertation. Indiana University.
- 12.* McGaughey, R.G. (1976). *An assessment of perceptual maturity, creativity, aesthetic judgement, self-concept development and social adjustment differences between sixth-grade students who have participated in selected 1-6 art programs and those who have not*. Unpublished dissertation. University of South Dakota.
- 13.* McWhinnie, H.J. (1967). The effects of a learning experience upon preference for complexity and asymmetry in fourth grade children. *California Journal of Educational Research*, 18(5), 219-229.
- 14.* McWhinnie, H.J. (1968). Effects of a learning experience upon preference for complexity and asymmetry in fifth grade children. *California Journal of Educational Research*, 19(4), 183-189.
- 15.* McWhinnie, H.J. (1969). *The effects of a learning experience upon preference for complexity and asymmetry; variables of field independence, and the ability to handle visual information*. Final report. Office of Education.
16. Pflaumer, L. (1980). *An assessment of the effects of a visual perception training program*. Unpublished dissertation. Illinois State University.

17. Reeves, D.J. (1971). *The assessment of altered differentiation as affected by experimental treatments*. Unpublished dissertation. Illinois State University.
 18. Rennels, M.R. (1969). Two methods of teaching spatial tasks to disadvantaged negroes. *Studies in Art Education*, 11(1), 44-51.
 19. Richards, A.G. (1988). Perceptual training in drawing among students from two countries. *Studies in Art Education*, 29(3), 302-308.
 20. Rosenfield, D.W. (1985). *Spatial relations ability skill training: the effect of a classroom exercise upon the skill level of spatial visualization of selected vocational education high school students*. Unpublished dissertation. University of Cincinnati.
 21. Royal, D.C. (1986). *A comparative study of the effects of a personalized system of instruction and conventional methods of instruction upon student achievement in the elements of design at the junior high level*. Unpublished dissertation. Ball State University.
 22. Salome, R.A.; Reeves (1972). Two pilot investigations of perceptual training of four and five year old children. *Studies in Art Education*, 13(2), 3-9.
 - 23.* Silverman, R.H. (1962). *Comparing the effects of two versus three-dimensional art activity upon spatial visualization, aesthetic judgement and art interest*. Unpublished dissertation. Stanford University.
 24. Silverman, R.H.; Hoepfner, R. (1969). *Developing and evaluating art curricula designed or disadvantaged youth*. Final report. Los Angeles, California.
 25. Smith, S.A. (1984). *The effects of a cognitive-affective structured arts curriculum upon the visual perceptual skills of preschool disadvantaged children using a visual training approach*. Unpublished dissertation. Southern Illinois University.
 26. Stone, H.E.F. (1982). *Visualization effects of three-dimensional art forms on high school students*. Unpublished dissertation. Georgia State University.
 27. Szeto, J.W.; Salome, R.A. (1977). The effects of scanning practice and perceptual training upon several visual functions: implications for drawing. *Studies in Art Education*, 19(1), 45-51.
 28. Tidhar, C.E. (1984). Children communicating in cinematic codes: Effects on cognitive skills. *Journal of Educational Psychology*, 76(5), 957-65.
 29. Trotty, S.A. (1977). *An examination of the relationship of imagery instruction to the spatial visualization and manipulation, art information and attitude and qualitative visual art performance of elementary education majors*. Unpublished dissertation. Purdue University.
 30. Wiley, S.E. (1984). *The relationship between amount of experience in art, visual perception, and picture memory*. Unpublished dissertation. Ball State University.
2. De 39 onderzoeken naar effecten op esthetische waarneming
1. Acuff B.C. (1974). *Aesthetic inquiry and the creation of meaning: a program of instruction in the visual arts for the elementary school*. Unpublished dissertation. Stanford University.
 2. Anderson, R.R. (1974). *A comparison of three methodologies as affectors of aesthetic sensitivity and judgement: discovery of concept and principle, art history and studio activity*. Unpublished dissertation. University of Minnesota.
 3. Atkins, S.S. (1986). *The effectiveness of a studio-based art appreciation curriculum*. Unpublished dissertation. University of Georgia.
 4. Bergamo, D.J. (1978). *The effect of environment-centered art instruction on the development of aesthetic and creative responses in high school art students*. Unpublished dissertation. Ohio State University.
 5. Bramlett, B.J. (1983). *A comparison study of two methods for teaching art appreciation to seventh and eighth graders*. Unpublished dissertation. University of South Carolina.
 6. Burg, K.E. (1983). *Effects of two teaching methods on visual aesthetic preference and language use*. Unpublished dissertation. Arizona State University.

7. Cole, E.S. (1985). *The effect of a cognitively oriented aesthetic curriculum on the responses of 4, 6 and 8 year olds enrolled in an art museum program*. Unpublished dissertation. University of Toledo.
8. Covington, S. (1987). *The effect of art instruction upon children's aesthetic preferences and their graphic representation*. Unpublished dissertation. University of Maryland at College Park.
9. Day, M.D. (1976). Effects of instruction on high school students art preferences and art judgments. *Studies in Art Education*, 18(1), 25-40.
10. Fitzner, D.H. (1980). The effects of combined art teaching approaches on the development of aesthetic sensitivity among selected elderly adults. *Studies in Art Education*, 21(2), 22-27.
11. Gilliat, M.T. (1980). The effects of habituation, the Feldman-Mittler methodology and studio activities on expanding art preferences of elementary students. *Studies in Art Education*, 21(2), 43-49.
12. Haanstra, F. (1980). De evaluatie van een educatieve tentoonstelling. *Muson*, 3, 23-26.
13. Holland, J.B. (1972). *Differences in aesthetic sensitivity through teaching selected elements of design through painting in contrasting groups of fifth grade students*. Unpublished dissertation. University of Minnesota.
14. Hollingsworth, P.L. (1983). The combined effect of mere exposure, counterattitudinal advocacy and art criticism methodology on upper elementary and junior high students' affect toward art works. *Studies in Art Education*, 24(2), 101-109.
15. Hysell, D.M. (1973). Testing an advance organizer model in the development of aesthetic perception. *Studies in Art Education*, 14(3), 9-17.
16. Kelsall, R.T. (1983). Towards critical study in the primary school. *Journal of Art & Design Education*, 2(1), 49-60.
17. Larkins, E.C. (1977). *An experimental study of the effects of a learning laboratory strategy for teaching art history-criticism for college freshmen*. Unpublished dissertation. Florida State University.
18. Lee Chi Chong (1976). *An experimental study to determine the influence of a cross cultural learning experience*. Unpublished dissertation. University of Minnesota.
19. McGaughey, R.G. (1976). *An assessment of perceptual maturity, creativity, aesthetic judgment, self-concept development and social adjustment differences between sixth grade students who have participated in selected 1-6 art programs and those who have not*. Unpublished dissertation. University of South Dakota.
20. McWhinnie, H.J. (1966). Effects of a learning experience on preference for complexity and asymmetry. *Perceptual & Motor Skills*, 23, 119-122.
21. McWhinnie, H.J. (1967). The effects of a learning experience upon preference for complexity and asymmetry in fourth grade children. *California Journal of Educational Research*, 18(5), 219-229.
22. McWhinnie, H.J. (1968). The effects of a learning experience upon preference for complexity and asymmetry in fifth grade children. *California Journal of Educational Research*, 19(4), 183-189.
23. McWhinnie, H.J. (1969). *The effects of a learning experience upon preference for complexity and asymmetry; variables of field independence, and the ability to handle visual information*. Final report, Office of Education.
24. Mittler, G. (1972). Efforts to secure congruent and incongruent modifications of attitude toward works of art. *Studies in Art Education*, 13(2), 58-69.
25. Mittler, G. (1976). An instructional strategy designed to overcome the adverse effects of established student attitudes toward works of art. *Studies in Art Education*, 17(3), 13-31.
26. Neperud, R.W. (1986). The relationship of art training and sex differences to aesthetic valuing. *Visual arts research*, 12(2), 1-10.
27. Palland J.T. & Jong, J. De (1962). *Schoolkinderen in het museum. Een onderzoek naar*

- de invloed van museumlessen op de receptieve kunstzin van kinderen uit de zesde klasse van scholen voor gewoon lager onderwijs. Universiteit van Amsterdam.
28. Phares, S.W. (1984). *Student's affective awareness of aesthetic form and symmetry through a directed reading lesson approach*. Unpublished dissertation. University of Miami.
 29. Sakatani, K.K. (1988). *A study comparing the effects of logo turtle graphics and paint graphics on the response of students to color, line and shape*. Unpublished dissertation. Stanford University.
 30. Seefeldt, C. (1979). The effects of a program to increase young children's perception of texture. *Studies in Art Education*, 20(2), 40-43.
 31. Silverman, R.H. (1962). *Comparing the effects of two versus three-dimensional art activity upon spatial visualization, aesthetic judgment and art interest*. Unpublished dissertation. Stanford University.
 32. Stevanov, Z. (1982). *A study of three instructional procedures in the teaching of art survey to non-art majors on the college level*. Unpublished dissertation. University of Kansas.
 33. Suggs, M.S. (1976). *Phenomenological analysis, art studio production and task order effects on student aesthetic attitude, art studio production and aspective perception of paintings*. Unpublished dissertation. University of Maryland at College Park.
 34. Taylor, A.P. and Trujillo, J.L. (1973). The effects of selected stimuli on the art products, concept formation and aesthetic judgmental decisions of four-year-old children. *Studies in Art Education*, 14(2), 57-66.
 35. Temme, J.E.V. (1983). *Over smaak valt te twisten, sociaal-psychologische beïnvloedingsprocessen van esthetische waardering*. Utrecht: Elinkwijk B.V..
 36. Warren, C.B. (1979). *A study of the effects of certain methods of teaching art appreciation on the development of aesthetic judgment and art vocabulary in college students*. Unpublished dissertation. University of New Jersey.
 37. White, D.C. (1980). *A pilot study investigating the relationship between an art programme and the development of aesthetic judgmental skills in children aged 13-14*. Unpublished dissertation. Birmingham Polytechnic.
 38. Wilson, B. (1966). An experimental study designed to alter fifth and sixth grade students' perceptions of paintings. *Studies in Art Education*, 8(1), 33-42.
 39. Wilson, B. (1972). The relationship between years of art training and the use of aesthetic judgmental criteria among high school students. *Studies in Art Education*, 13(2), 34-43.

Ontvangen 12-8-96

Definitief aanvaard 20-12-96

De relatie tussen leesprestaties en buitenschools lezen in leerjaar 5-8

Martha E. Otter, Rob Schoonen en Kees de Glopper¹
SCO-Kohnstamm Instituut, Universiteit van Amsterdam

ABSTRACT

The purpose of the study is to investigate the effects of leisure time reading on reading performances. It is a four-year longitudinal study carried out in grade 3, 4, 5 and 6 of primary school, with a total of 736 students. Reading achievement is measured five times: at the beginning and end of grade 3 and at the end of grade 4, 5 and 6 during eight, twelve and ten weeks, respectively. To investigate the relationship leisure time reading and reading performances covariance structural analysis via LISREL is used. The hierarchical structure of the data (students in classes) is taken into account by means of the multilevel procedure proposed by Muthén. The study results showed that leisure-time reading had no effect on reading achievement at school.

INLEIDING

Begrijpend lezen is een basisvaardigheid die van groot belang is voor het maatschappelijk en sociaal functioneren. De onderwijsloopbaan, participatie op de arbeidsmarkt en deelname aan de cultuur zijn gebaat bij een goed ontwikkelde leesvaardigheid. Niet voor niets is leesonderwijs van oudsher een centrale taak van de school.

Over de opbrengsten van het lesonderwijs op Nederlandse basisscholen bestaat reeds een geruim aantal jaren zorg (Wesdorp, 1985; Commissie Evaluatie Basisonderwijs, 1994). De leesprestaties halverwege en aan het einde van de basisschool blijven achter bij de verwachtingen (Sijstra, 1992; Zwarts, 1990) en ook in internationaal opzicht blijven de prestaties van Nederlandse basisschoolleerlingen achter bij die van leeftijdgenoten uit veel andere landen (De Glopper & Otter, 1993; Elley, 1994).

De tegenvallende resultaten worden in de eerste plaats toegeschreven aan tekortkomingen in de aanpak van het onderwijs in aanvankelijk en begrijpend lezen. De kern van de diagnose is dat er niet of nauwelijks sprake is van instructie in aanvankelijk en begrijpend lezen (Weterings & Aarnoutse, 1986; Aarnoutse & Weterings, 1991). Met een andere inhoud en aanpak van het lesonderwijs kunnen naar de overtuiging van velen de leesprestaties belangrijk verbeterd worden. Met name van instructie in tekstbegripstrategieën wordt veel winst verwacht (Paris, Wasik & Turner, 1991; Pearson & Fielding, 1991).

Verbetering van de leesprestaties wordt in de tweede plaats verwacht van stimulering van taal- en leesactiviteiten buiten school. De achterliggende gedachte is dat niet alleen lesonderwijs, maar ook taal- en leesactiviteiten buiten school een belangrijke bijdrage leveren aan de leesontwikkeling van kinderen. Door rijmspelletjes en liedjes, typische taalactiviteiten voor ouders en hun kinderen, kunnen jonge kinderen allerlei vaardigheden ontwikkelen die in de eerste fase van de leesontwikkeling van groot belang zijn (Bradley & Bryant, 1983; Lundberg, Frost & Petersen, 1988). Ook van voorlezen wordt algemeen aangenomen dat er een positieve invloed uitgaat op de taal- en leesvaardigheid (Heath, 1982; Snow, 1983). Recent zijn twee reviews gepubliceerd waarin de empirische evidentie voor de effecten van voorle-

zen op de leesvaardigheidsontwikkeling wordt samengevat (Bus, IJzendoorn & Pellegrini, 1995; Scarborough & Dobrich, 1994). In beide publikaties wordt de conclusie getrokken dat voorlezen effect heeft op de taal- en leesontwikkeling. In beide onderzoeken wordt ongeveer 8% van de individuele verschillen in taal- of leesvaardigheid toegeschreven aan verschillen in de hoeveelheid voorlezen die kinderen genoten hebben. Het effect van voorlezen is niet zeer groot te noemen, maar ook niet te verwaarlozen.

Het effect van voorlezen wordt in de regel verklaard vanuit meer dan het voorlezen alleen. Voorlezen is niet alleen leerzaam omdat kinderen met schrijftaal, tekst en verhalen of prenten in aanraking komen, maar ook omdat kinderen vragen kunnen stellen en volwassenen instructie- en voorbeeldgedrag vertonen.

Zelfstandige leesactiviteiten van kinderen, beginnend vanaf het moment dat kinderen op eigen kracht kunnen lezen, worden veelal genoemd als volgende bron voor de groei van de leesvaardigheid. De plausibele aanname is dat oefening kunst baart. Het veronderstelde mechanisme is daarbij als volgt. Door zelfstandig te lezen doen leerlingen op allerlei terreinen nieuwe kennis op (woordenschat, kennis van de wereld, tekststructuren, etc.) hetgeen een positieve invloed heeft op hun leesvaardigheid. Bovendien oefenen kinderen door zelfstandig lezen eerder verworven vaardigheden (herkennen van woordbeelden, verhogen van de leesnelheid, toepassen van begripsstrategieën, etc.).

De hypothese dat oefening kunst baart, mag nog zo plausibel klinken, de empirische evidentie voor de stelling dat leerlingen door zelfstandig lezen beter leren lezen is gering. Waar Bus, IJzendoorn en Pellegrini (1995) en Scarborough en Dobrich (1994) nog enige tientallen studies bespreken, waaronder tal van experimenten, is het aantal serieuze studies naar de effecten van buitenschools lezen op de leesvaardigheid zeer gering. Dat is des te opmerkelijker, gezien het feit dat zelfstandig lezen in de vrije tijd genoemd wordt als belangrijke factor bij het ontstaan en de groei van individuele verschillen in leesvaardigheid (Stanovich, 1986).

Het aantal waardevolle studies naar de effecten van buitenschools lezen op de leesprestaties noemen we zeer beperkt omdat veel onderzoek gebruik maakt van een 'posttest-only' onderzoeksonderwerp (Greaney, 1980; Greaney & Hegarty, 1987; Waldberg & Tsai, 1984; Cunningham & Stanovich, 1990; Rowe, 1991; Allen, Cipielewski, Stanovich, 1992; Elley, 1994; Kuhlemeier, 1996). Dit betekent dat de leesprestaties in de genoemde studies slechts éénmaal gemeten zijn waardoor eventuele verbanden tussen het buitenschoolse lezen en de leesprestaties op geen enkele wijze causaal geïnterpreteerd kunnen worden. Een eventuele samenhang tussen het buitenschoolse lezen en de leesprestaties kan namelijk duiden op zowel een effect van het buitenschoolse lezen op de leesprestatie als een effect van de leesprestatie op het buitenschoolse lezen. Ten slotte kan het verband tussen de leesprestatie en het buitenschoolse lezen ook spurieus zijn en verklaard worden door variabelen die niet in het onderzoek zijn opgenomen (bijvoorbeeld door intelligentie of sociaal milieu).

Daarbij is in het merendeel van de hierboven genoemde studies het buitenschoolse lezen gemeten met een vragenlijst, een onderzoeksmethode die niet zonder problemen is (zie bijvoorbeeld Kuhlemeier, 1996).

Een onderzoekopzet die een beter inzicht geeft in mogelijke effecten van het buitenschools lezen op de leesprestaties is de zogenaamde 'pretest-posttest' onderzoekopzet. Deze onderzoekopzet is bijvoorbeeld gebruikt in het onderzoek van Anderson, Wilson en Fielding, (1988), Taylor, Frye en Maruyama (1990), Koolstra (1993) en Carver en Leibert (1995). Naast een adequatere onderzoekopzet is in drie van de vier zojuist genoemde studies het buitenschoolse lezen gemeten met de zogenaamde dagboekmethode. Deze methode voor het meten van het buitenschoolse lezen is minder aan twijfel onderhevig is dan de eenmalige vragenlijst. Hieronder worden de studies van Anderson e.a. (1988), Taylor e.a. (1990), Koolstra (1993) en Carver en Leibert (1995) besproken.

Anderson, Wilson en Fielding (1988) hebben de relatie bestudeerd tussen buitenschools lezen en leesprestaties bij 155 leerlingen in leerjaar 7. In dit onderzoek hielden de leerlingen middels een dagboekje bij hoelang zij thuis in boeken, strips of bijvoorbeeld de krant lezen.

Uit het onderzoek van Anderson c.s. komen twee opmerkelijke zaken naar voren. In de eerste plaats lijken de onderzoeksresultaten erop te duiden dat veel 11-jarige Amerikaanse kinderen niet of zeer weinig lezen. Gemiddeld leest een typische leerling in leerjaar 7 ongeveer acht tot twaalf minuten per dag, waarvan maar vier à vijf minuten in boeken (Anderson e.a., 1988, p. 299). Dat Amerikaanse jongeren weinig lezen wordt nog eens benadrukt indien we bedenken dat de gerapporteerde leestijden van Anderson c.s. betrekking hebben op zowel het lezen voor het plezier als het lezen voor school: "(...) ours included all reading, whether done for enjoyment or not" (p. 299).

Een andere belangrijke conclusie uit het onderzoek is dat buitenschools lezen samenhangt met de schoolse leesprestaties: 3.4% van de verschillen in de leesprestaties kunnen worden verklaard door de tijd die leerlingen besteden aan buitenschools lezen.

De resultaten uit het onderzoek van Anderson c.s. moeten op grond van een tweetal argumenten echter in twijfel worden getrokken. Zoals hierboven reeds is vermeld was het oorspronkelijke onderzoeksplan een voormeting-nametingsonderzoeksontwerp: een meting van de leesprestaties, gedurende het schooljaar metingen van het buitenschoolse lezen en aan het einde van het schooljaar wederom een meting van de leesprestaties. Deze onderzoeksopzet is uiteindelijk niet gerealiseerd omdat de tweede prestatiemeting (de posttest) volgens de onderzoekers niet bruikbaar was ('de leerlingen hadden door het warme weer hun best niet gedaan'). Om het onderzoeksdesign te handhaven hebben de onderzoekers leesprestatiegegevens uit leerjaar 4 gebruikt als eerste leesprestatiemeting (pretest) en hebben zij de oorspronkelijke voormeting gebruikt als nameting (posttest). Dit betekent dat de onderzoekers de metingen van het buitenschoolse lezen in feite verricht hebben ná de posttest van de leerlingen in begrijpend lezen. De oorzakelijkheid van de gevonden relatie is daarmee eigenlijk uitgesloten. De onderzoekers stellen weliswaar dat het buitenschoolse lezen een stabiel gedragspatroon is, zodat de leestijd ná de halverwege leerjaar 7 gelijkgesteld kan worden aan de (niet gemeten) leestijd in leerjaar 4, 5 en 6, maar uit onderzoek van Otter (1995a) blijkt echter naar voren dat het leesgedrag van jonge kinderen lang niet zo stabiel is als Anderson c.s. veronderstellen. Er zijn kinderen die het lezen als vrije-tijdsbesteding pas in de loop van leerjaar 5, 6 of 7 ontdekken en er zijn ook kinderen die het lezen in de loop van de leerjaren inruilen voor andere activiteiten zoals bijvoorbeeld muziek maken of sport.

In de tweede plaats blijkt bij nadere beschouwing van het onderzoek van Anderson e.a. (1988) dat de gevonden verbanden mogelijk het gevolg zijn van hun analysemethode. Zij hebben namelijk de geringe verschillen die er waren tussen de kinderen -ze lazen namelijk bijna allemaal zeer weinig- met behulp van een logaritmische transformatie uitvergroot waardoor het correlatieve verband met de leesprestaties flink omhoog ging ("went up considerably" (p. 290)). Dit betekent dat de resultaten van Anderson c.s. impliceren dat het onderscheid tussen héél weinig lezen (bijvoorbeeld één minuut per dag) en 'gewoon' weinig lezen (bijvoorbeeld twee minuten per dag) verband houdt met de leesprestaties. Het is echter zeer de vraag wat de betekenis van zulke minimale verschillen in de leestijd kan zijn. Naar onze mening is het niet aannemelijk dat de leesprestaties van kinderen profijt hebben van bijvoorbeeld ééns per twee weken tien minuten thuis in boeken lezen.

Taylor, Frye en Maruyama (1990) hebben onderzoek verricht naar de relatie tussen buitenschools lezen en de leesprestaties in leerjaar 7 en 8. In januari en mei van hun onderzoeksjaar werden aan 195 leerlingen leesvaardigheidstoetsen afgenomen en in de tussenliggende periode hielden de leerlingen middels een dagboekje bij hoelang zij thuis in boeken lazen. Overeenkomstig de resultaten van het onderzoek van Anderson e.a. (1988) komt ook uit het onderzoek van Taylor e.a. (1990) naar voren dat veel kinderen niet of zeer weinig thuis lezen. Daarom hebben ook Taylor c.s. de verschillen tussen de kinderen met een logaritmische transformatie uitvergroot. Echter, anders dan in de studie van Anderson c.s. vinden Taylor c.s. géén effect van het buitenschoolse lezen op leesprestaties. Als mogelijke verklaring voor het ontbreken van een verband wordt door de onderzoekers de veronderstelling geopperd dat de leerlingen de tijdbesteding aan het buitenschoolse lezen misschien niet goed in hun dag-

boekjes hebben bijgehouden. Kennelijk sluiten de auteurs de mogelijkheid uit dat er daadwerkelijk geen verband is tussen het weinige buitenschools lezen en de leesprestaties.

Ook het onderzoek van Koolstra is hier van belang (1993). In een grootschalig longitudinaal panelonderzoek zijn de effecten van het buitenschools lezen (en televisie-kijken) op de leesprestaties onderzocht. Het onderzoek vond plaats in leerjaar 4 tot en met 8 met in totaal 828 leerlingen. De leesprestaties werden in het begin van elk leerjaar gemeten alsmede het buitenschoolse lezen. In tegenstelling tot Anderson e.a. (1988) en Taylor e.a. (1990) is in deze studie het buitenschoolse lezen met een eenmalige vragenlijst gemeten. De effecten die Koolstra rapporteert, zijn teleurstellend. In leerjaar 5 en 7 vindt Koolstra effecten van het buitenschoolse lezen op de leesprestaties. In leerjaar 6 en 8 zijn deze effecten afwezig. De gerapporteerde effecten in leerjaar 5 en 7 zijn zéér klein: nog geen half procent van de verschillen in de leesprestaties kunnen in leerjaar 5 en 7 verklaard worden door verschillen in de frequentie waarmee kinderen buiten schooltijd voor hun plezier in boeken lezen. Daarbij moet aangetekend worden dat het meten van het buitenschoolse lezen met een vragenlijst niet zonder problemen is (zie bijvoorbeeld Kuhlemeier, 1996).

Ten slotte mag het onderzoek van Carver en Leibert (1995) niet onvermeld blijven. Deze onderzoekers lieten een groep van 43 kinderen gedurende 30 dagen twee uur per dag in boeken lezen (in totaal dus 60 uur). Uit de analyses blijkt dat het lezen van boeken géén effect heeft gehad op de ontwikkeling van de leesprestaties van de kinderen.

De bovenstaande bespreking van de vier onderzoeken maakt duidelijk dat er maar zeer weinig empirische steun is voor de veronderstelling dat buitenschools lezen een positief effect heeft op de schoolse leesvaardigheid. In twee van de vier studies worden geen effecten aangetoond en in de andere studies slechts zeer kleine effecten, die op goede gronden in twijfel kunnen worden getrokken.

Het is gezien het voorafgaande al met al een open vraag in hoeverre buitenschools lezen² effect heeft op de schoolse leesprestaties. In het onderstaande beschrijven we een longitudinaal onderzoek dat de veronderstelde effecten van buitenschools lezen op de schoolse leesprestaties kritisch tracht te toetsen.

METHODE

Scholen en leerlingen

Bij aanvang van het onderzoek waren de deelnemende respondenten 9-jarige leerlingen in het basisonderwijs (leerjaar 5). Voor de steekproefopzet is Nederland in tien regio's ingedeeld. Een paar moeilijk bereikbare gebieden gezien vanuit de regio's - met name Zuid-Limburg, Zeeuws-Vlaanderen en de Waddeneilanden - zijn niet bij een van de regio's ingedeeld.

Vervolgens zijn per regio aselekt twintig scholen getrokken. Van deze twintig scholen zijn er steeds vier benaderd voor deelname aan het project. De keuze van de vier scholen was niet aselekt, omdat de proefleider voor een bepaalde regio de scholen die het dichtst bij hem of haar in de buurt waren als eerste benaderd heeft voor deelname. We verwachten echter dat deze 'selecte' keuzes niet samenhangen met relevante onderzoeksvariabelen, zodat er weinig reden is om aan te nemen dat die keuzes de representativiteit van de steekproef nadelig beïnvloed hebben.

In het schooljaar 1990/1991 deden per regio vier scholen mee, met in totaal 42 leerkrachten en 973 leerlingen. Om verschillende redenen hebben vier scholen bij aanvang van het schooljaar 1991/1992 zich uit het project teruggetrokken. De belangrijkste redenen waren de komst van nieuwe onervaren leerkrachten en schoolfusies.

Door het ontbreken van financiële middelen aan het einde van het tweede onderzoeksjaar was het niet duidelijk of het project voortgezet zou kunnen worden in leerjaar 7 en 8 (schooljaar 1992/1993 en 1993/1994). Toen in januari 1993 voortzetting van het project werd gewaarborgd, bleek dat een zestal scholen verplichtingen waren aangegaan in andere onder-

zoeksprojecten. Om de school niet verder te belasten, zagen deze scholen af van verdere deelname aan het onderhavige (arbeidsintensieve) project. In Tabel 1 wordt een overzicht gegeven van de deelnemende scholen, klassen en leerlingen in de vier onderzoeksjaren.

Tabel 1: Overzicht van het aantal scholen, klassen en leerlingen per onderzoeksjaar

Schooljaar	Leerjaar	Scholen	Klassen	Leerlingen
1990/1991	5	40	42	973
1991/1992	6	36	41	958
1992/1993	7	30	34	807
1993/1994	8	30	30	779

Uit Tabel 1 blijkt dat ten opzichte van leerjaar 5, in leerjaar 6 het aantal scholen met vier verminderde en het aantal klassen met slechts één. De oorzaak hiervan is dat in het zesde leerjaar de leerlingen over meer combinatieleerjaren verdeeld waren. Het aantal leerlingen nam in leerjaar zes met slechts 15 leerlingen af. De verklaring hiervoor is dat de dagboekjes in drie combinatieleerjaren door *alle* leerlingen zijn ingevuld. De resultaten die wij zullen presenteren hebben echter *uitsluitend* betrekking op de leerlingen die in schooljaar 1990/1991, 1991/1992, 1992/1993 en 1993/1994 in respectievelijk leerjaar 5, 6, 7 en 8 van de steekproef zaten.

Onderzoeksontwerp

In leerjaar 5 zijn in oktober 1990 en mei 1991 leestoetsen afgenomen (zie Tabel 2: T1 en T2). In leerjaar 6, 7 en 8 zijn leestoetsen in juni afgenomen (T3, T4 en T5). Gedurende het vierjarige longitudinale onderzoek hebben de leerlingen hun leesgedrag in dagboeken bijgehouden (zie Tabel 2: D01 t/m D36: in schooljaar 1990/1991 gedurende acht weken, in de schooljaren 1991/1992, 1992/1993 en 1993/1994 gedurende respectievelijk twaalf, zes en tien weken. De acht dagboekweken in het schooljaar 1990/1991 lagen in drie (aaneengesloten) perioden: (1) 5-26 november 1990, (2) 14 januari - 4 februari en (3) 4-17 maart 1991. De twaalf dagboekweken in het schooljaar 1991/1992 (leerjaar 6) waren verspreid over het najaar, de winter en het voorjaar. De weeknummers waren in 1991 achtereenvolgens 43, 45, 47 en 50 en in 1992 2, 5, 8, 11, 14, 20, 22 en 25.

Door het ontbreken van financiële middelen aan het begin van het derde onderzoeksjaar (1992/1993) konden de dagboekmetingen in leerjaar 7 pas in de winter aanvangen. Tijdens de winter en het voorjaar is het leesgedrag gedurende zes weken gemeten (zie Tabel 2). De weeknummers waren achtereenvolgens 10, 13, 16, 19, 23 en 25. In schooljaar 1993/1994 zijn de metingen weer gespreid over het najaar, de winter en het voorjaar. De weeknummers waren in 1993 achtereenvolgens 40, 43, 46 en 49 en in 1994 2, 5, 12, 15, 22 en 25.

Tabel 2: Afname leestoetsen (T1 t/m T5), dagboekmetingen van het leesgedrag (D01 t/m D36)

Schooljaar	leerjaar	oktober-november	januari-juni	mei - juni
1990/1991	5	T1 D01 - D03	D04 - D08	T2
1991/1992	6	D09 - D12	D13 - D20	T3
1992/1993	7		D21 - D26	T4
1993/1994	8	D27 - D30	D31 - D36	T5

Dataverzameling

Proefleiders verzorgden de dataverzameling. De proefleiders waren zeer ervaren onderzoeksassistenten en allen afkomstig uit het onderwijs.

Aan het begin van elk onderzoeksjaar hebben de proefleiders persoonlijk contact gezocht met de leerkrachten. In dit contact werd het draaiboek van het onderzoek uitvoerig besproken. Tevens werden de dagboeken van de leerlingen mondeling toegelicht. Voorafgaande aan deze bespreking hadden de leerlingen een aantal dagen de dagboeken bijgehouden, waardoor eventuele problemen besproken konden worden. De proefleiders hadden voor en na afloop van elke dagboekweek telefonisch contact met de leerkrachten. Hierdoor konden problemen snel gesignaleerd en opgelost worden.

Om de toetsafnames te standaardiseren (over proefleiders en in de tijd) zijn alle afnamen aan de hand van een draaiboek afgenomen.

INSTRUMENTARIUM

Leestoetsen

Aan het begin en einde van leerjaar 5 is dezelfde leestoets afgenomen. De toets is ontwikkeld onder auspiciën van de International Association for Educational Achievement (IEA) in het kader van een internationaal onderzoek naar de opbrengsten in begrijpend lezen (Elley, 1994; De Gloppe en Otter, 1993).

De leesvaardigheid van de leerlingen aan het einde van leerjaar 6 is gemeten met een selectie van teksten en vragen uit de IEA leesvaardigheidstoets en met behulp van een door het Cito ontwikkelde toets voor begrijpend lezen voor leerjaar 6, toets E4 (Cito, 1980). De leesvaardigheid aan het einde van leerjaar 7 is gemeten met behulp van de hierboven al genoemde Cito-toets E4 en met Cito-toets E5. Aan het einde van leerjaar 8 is de leesvaardigheid gemeten met Cito-toets E5 (zie boven) en met een selectie van teksten en vragen uit een toets ontwikkeld door de International Association for Educational Achievement (IEA) voor leerlingen uit het tweede leerjaar van het voortgezet onderwijs (Elley, 1994; De Gloppe & Otter, 1993). In Tabel 3 wordt een overzicht gegeven van de oorsprong van de toetsen, het aantal teksten en het aantal bijbehorende vragen.

Tabel 3: Leestoetsen in leerjaar 5 tot en met 8: bron, aantal teksten en vragen

Toetsmoment	Oorsprong van de toets	Teksten	Vragen
Begin/einde leerjaar 5	IEA-toets 9-jarigen	21	66
Einde leerjaar 6	IEA-toets	7	33
	Cito-toets E4	6	25
Einde leerjaar 7	Cito-toets E4	6	25
	Cito-toets E5	6	25
Einde leerjaar 8	Cito-toets E5	6	25
	Selectie uit IEA-toets 14-jarigen	6	40

Buitenschools lezen

Hieronder wordt uiteengezet hoe het buitenschoolse lezen met behulp van dagboeken is gemeten. Ook komt de scoring aan de orde en wordt aangegeven hoe het probleem van ontbrekende waarnemingen is opgelost.

Dataverzameling

Alle leerlingen kregen voor elke dagboekweek in leerjaar 5 tot en met 8 één dagboekje. Voor elke dag was er één bladzijde in een dagboek gereserveerd. Elk dagboekje begon op dinsdag met vragen over hun leesgedrag op maandag en eindigde één week later op maandag met vragen over hun leesgedrag op vrijdag, zaterdag en zondag. De dagboekjes werden aan het begin van de ochtend ingevuld en het invullen nam zes à zeven minuten in beslag.

In de eerste plaats werd gevraagd of de leerlingen (gisteren) gelezen hadden (ja/nee). Om meer specifieke informatie te verkrijgen over het buitenschools lezen werd verder geïnformeerd naar de titel van het boek, de naam van de auteur en het aantal gelezen bladzijden.

Verschillen tussen de dagboeken in leerjaar 5 en de overige leerjaren

In leerjaar 5 was in het dagboekje elke schooldag onderverdeeld in drie dagdelen: (1) vóór schooltijd (2) ná schooltijd maar vóór het avondeten en (3) ná het avondeten (zie Otter, 1995a). Na inspectie van de gegevens van de acht dagboekweken in leerjaar 5 werd de conclusie getrokken dat er voor schooltijd zéér infrequent gelezen werd. Op grond hiervan is besloten om het 'voorschoolse' dagdeel in leerjaar 6 tot en met 8 te laten vervallen. In *alle* onderzoeksjaren waren de weekeinddagen onderverdeeld in twee dagdelen: (1) vóór het avondeten en (2) na het avondeten (zie Otter, 1995a).

Scoring

De dagboeken van de leerlingen zijn als volgt gescoord: bij de vraag of kinderen gelezen hebben, is aan elk 'ja' antwoord de score 1 toegekend, aan elk 'nee' antwoord de score 0. Voor het lezen van boeken zijn dag-, week- en gemiddelde weekcores per leerjaar berekend. De maximale dagscore was voor schooldagen in leerjaar 5 drie en in leerjaar 6 tot en met 8 twee. De maximale dagscore in het weekeinde was in alle onderzoeksjaren gelijk, namelijk twee. De maximale weekscore bedroeg in leerjaar 5 negentien ($5 \cdot 3 + 2 \cdot 2$) en in leerjaar 6 tot en met 8 veertien ($7 \cdot 2$).

Ontbrekende waarnemingen

De ontbrekende waarnemingen van de leerlingen zijn in alle onderzoeksjaren eerst op weekniveau 'behandeld': indien méér dan de helft van de negentien (leerjaar 5) of veertien scores (leerjaar 6 tot en met 8) ontbraken, hebben deze leerlingen een code meegekregen die stond voor 'ontbrekende weekscore'. De ontbrekende scores van de overige kinderen zijn aangevuld met hun persoonlijke gemiddelde score berekend over de overige dagen.

Vervolgens zijn de ontbrekende waarnemingen op jaarniveau behandeld (leerjaar 5: acht weken; leerjaar 6 twaalf weken; leerjaar 7 zes weken; leerjaar 8 tien weken). Indien per leerjaar meer dan 25% van de dagboekweken ontbraken, ten gevolge van bijvoorbeeld ziekte of schoolreisjes, zijn de leerlingen uit het bestand verwijderd (bijna tien procent van de leerlingen). De ontbrekende weekcores van de overige leerlingen zijn aangevuld met hun persoonlijke gemiddelde score berekend over de overige weken.

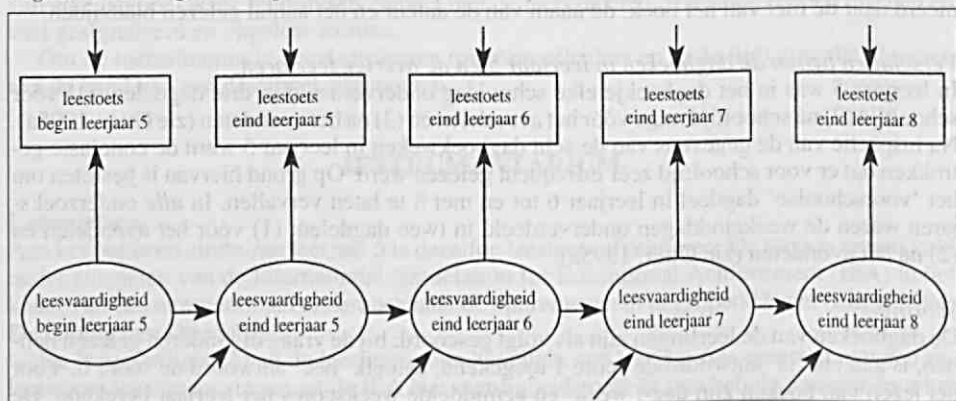
ANALYSES

Van alle instrumenten (leestoetsen en dagboekmetingen) worden beschrijvende gegevens gerapporteerd zoals gemiddelden en standaarddeviaties. Tevens wordt met behulp van de Kolmogorov-Smirnov Goodness of Fit Test nagegaan of aangenomen mag worden dat de scores uit een normaal verdeelde populatie komen.

Ook komt de psychometrische kwaliteit van het instrumentarium aan de orde. Per instrument zullen gegevens gerapporteerd worden over de passing van het één-factormodel volgens Hagglund (1982), de interne consistentie volgens Fleishman en Benson (1987) en Cronbach. Ook wordt een 95%-betrouwbaarheidsinterval rond Cronbachs α gerapporteerd.

Vervolgens zal worden nagegaan of op de scores van de leestoetsen en de gemiddelde leesfrequentiescores een quasi-simplexstructuur van toepassing is. Uit de literatuur is namelijk bekend dat herhaalde metingen in longitudinaal onderzoek, een quasi-simplexstructuur te zien geven (Jöreskog, 1970), d.w.z. dat nabij gelegen metingen sterker met elkaar correleren dan verder uit elkaar liggende metingen. Deze structuur kan gemodelleerd worden als in Figuur 1.

Figuur 1: Een Quasi-simplex model



Uit Figuur 1 blijkt dat de leesscores worden opgevat als indicatoren van de latente variabele leesvaardigheid. Verder komt uit de figuur naar voren dat meetfouten in de geobserveerde variabelen zijn toegestaan. Voor de dagboekmetingen (leesfrequentie) wordt een overeenkomstig model gepostuleerd. Indien uit de passing van de quasi-simplexmodellen naar voren komt dat zij het databestand goed representeren, wordt van deze structuur uitgegaan bij het toetsen van modellen over de invloed van het buitenschools lezen op de leesprestaties.

Voor de passing van de modellen zijn verschillende schattingsprocedures beschikbaar. Wij hebben ervoor gekozen de passing van de gepostuleerde modellen op de data te toetsen met behulp van de 'maximale-waarschijnlijkheidsmethode' ('maximum likelihood', ML). De ML-methode stelt ons namelijk in staat om concurrerende modellen te evalueren met behulp van de chi-kwadraat-verschiltest van Bollen (1989)³.

Omdat het databestand hiërarchisch van aard is (leerlingen zijn onderdeel van klassen en scholen) is er voor gekozen om de relatie tussen buitenschools lezen en de leesprestaties te analyseren met covariantiestructuuranalyses (Jöreskog & Sörbom, 1988) waarin met de hiërarchische structuur van de gegevens rekening gehouden wordt. Dit kan door middel van de multiniveauprocedure die is voorgesteld door Muthén (Muthén, 1989, 1990; Hox, 1993). Deze benadering houdt in de onderhavige studie in dat, *voorafgaand* aan de modeltoetsingen, met behulp van covariantiestructuuranalyses de variantie op klasniveau volledig wordt uitgepartialiseerd.

RESULTATEN

Hieronder worden de resultaten gepresenteerd: eerst de beschrijvende en psychometrische gegevens van de leestoetsen en het buitenschools lezen, vervolgens de resultaten van de analyse van de structuur van het databestand. Ten slotte worden de resultaten gepresenteerd van de analyses van de invloed van het buitenschools lezen op de leesprestaties.

Leestoetsen

In Tabel 4 staan de beschrijvende gegevens van de leesprestaties.

Tabel 4: Beschrijvende gegevens voor de leesprestaties in leerjaar 5 tot en met 8: aantal respondenten (N), aantal vragen (k), geobserveerd minimum en maximum (min-max), gemiddelde (M) en standaarddeviatie (sd)

Leerjaar	N	k	min- max	M	sd
5 begin	965	66	6 - 62	35.7	11.3
5 eind	971	66	11 - 63	43.4	10.7
6	885	58	10 - 57	41.8	9.5
7	749	50	6 - 46	34.6	7.6
8	712	65	12 - 64	43.4	11.7

Bij inspectie van de tabel komt naar voren dat de opgaven aan het begin van groep 5 het moeilijkst zijn geweest. De gemiddelde p-waarde van de toetsvragen in groep 5 bedraagt namelijk .54 (35.7/66). De leesvaardigheidstoets aan het einde van groep 6 had de laagste moeilijkheidsgraad. De gemiddelde p-waarde van de toetsvragen bedraagt in dat leerjaar .72. De gemiddelde p-waarden in de overige jaren ontlopen elkaar niet veel. Deze zijn voor het einde van leerjaar 5, 7 en 8 respectievelijk .66, .69 en .67.

Verder is uit Tabel 4 af te lezen dat in geen enkel leerjaar de kinderen de minimale of de maximale score behalen hetgeen een aanwijzing is dat zowel bodem- als plafondeffecten uitgebleven zijn.

Met behulp van de Kolmogorov-Smirnov Goodness of Fit test is nagegaan of aangenomen mag worden dat de verschillende leesscores uit een normaal verdeelde populatie komen. Voor alle leestoetsen moet deze aanname op statistische gronden verworpen worden ($p < .01$). Echter, de histogrammen van de scoreverdelingen, de gepiektheids- en de scheefheidscoëfficiënten geven aan dat de schendingen niet ernstig zijn. De significantie van de schendingen is waarschijnlijk terug te voeren op het grote aantal observaties.

In Tabel 5 worden enkele psychometrische gegevens van de leestoetsen gerapporteerd.

Tabel 5: Psychometrische gegevens van de leesvaardigheidstoetsen in leerjaar 5 tot en met 8: aantal observaties (N), passing van het één-factormodel (gfi), betrouwbaarheid volgens Fleishman & Benson (fb), α en een 95%-betrouwbaarheidsinterval rond α (int. α)

Leerjaar	N	gfi	fb	α	int. α
5 begin	965	.95	.91	.90	.90 - .91
5 eind	971	.96	.91	.90	.89 - .91
6	885	.96	.90	.90	.89 - .91
7	749	.99	.89	.89	.88 - .90
8	712	.97	.92	.92	.91 - .93

Uit Tabel 5 kan worden opgemaakt dat de passing van het één-factormodel op alle meetmomenten ruim voldoende tot zeer goed te noemen is: de 'goodness-of-fit' (gfi) is minimaal .95 en maximaal .99. De conclusie dat de leestoetsen in leerjaar 5 tot en met 8 min of meer één factor lijken te meten, mag getrokken worden. Dit laatste duidt erop dat de leestoetsen die in leerjaar 6, 7 en 8 uit verschillende deelttoetsen bestaan min of meer hetzelfde meten.

Verder kan uit Tabel 5 worden afgelezen dat de betrouwbaarheid volgens Cronbachs α en Fleishman & Benson (1987) goed genoemd mag worden: minimaal zijn deze coëfficiënten .89 en maximaal .92.

Leesfrequentie

In Tabel 6 worden per leerjaar de beschrijvende gegevens van de gemiddelde leesfrequentiescores (gemiddelde frequentie per week) gerapporteerd. Opgemerkt wordt dat in leerjaar 5 de weekfrequentiescore maximaal negentien is (zie instrumentatie), in leerjaar 6 tot en met 8 is deze maximaal veertien.

Tabel 6: Beschrijvende gegevens voor de leesfrequentiedata in leerjaar 5 tot en met 8: aantal respondenten (N), theoretisch minimum en maximum (tmin-tmax), geobserveerd minimum en maximum (gmin-gmax), gemiddelde (M) en standaarddeviatie (sd)

Leerjaar	N	tmin-tmax	gmin-gmax	M	sd
5	903	0 - 19	0 - 19	3.4	3.4
6	903	0 - 14	0 - 14	1.8	1.9
7	775	0 - 14	0 - 14	1.3	1.8
8	733	0 - 14	0 - 13	1.1	1.4

Bij inspectie van de Tabel 6 blijkt dat de kinderen naarmate zij ouder worden steeds minder gaan lezen. Lezen zij in leerjaar 5 nog op gemiddeld 18% van de dagdelen (3.4/19), in leerjaar zes is dit percentage nog slechts 8% (1.1/14). Uit Otter en Schoonen (1996) blijkt dat het aantal kinderen dat gemiddeld minder dan één keer per week thuis in een boek leest in de loop van de jaren fors stijgt: van 29 procent van de kinderen in leerjaar 5 naar bijna 63 procent van de kinderen in leerjaar 8.

Met behulp van de Kolmogorov-Smirnov Goodness of Fit test is nagegaan of aangenomen mag worden dat de leesfrequentiescores uit een normaal verdeelde populatie komen. Deze aanname moet voor alle leerjaren verworpen worden ($p < .01$). Uit de histogrammen van de scoreverdelingen, de gepiektheids- en de scheefheidscoëfficiënten blijkt dat de verdelingen sterk asymmetrisch zijn: de linkerkant is sterk gepiekt en de verdelingen lopen naar rechts uit. Anders gezegd: er zijn veel kinderen die aangeven zeer weinig te lezen en maar zeer weinig kinderen die aangeven veel te lezen. Daarbij wordt de scheefheid en de gepiektheid sterker naarmate de kinderen ouder worden.

In Tabel 7 worden enkele psychometrische gegevens van de leesfrequentiegegevens gepresenteerd.

Tabel 7: Psychometrische gegevens van de leesfrequentiegegevens in leerjaar 5 tot en met 8: aantal observaties (N), passing van het één-factor model (gfi), betrouwbaarheid volgens Fleishman & Benson (fb), α en een 95%-betrouwbaarheidsinterval rond α (int. α)

Leerjaar	N	gfi	fb	α	int. α
5	903	.97	.91	.91	.90 - .92
6	903	.99	.92	.92	.91 - .93
7	775	1.00	.88	.88	.86 - .89
8	733	1.00	.81	.81	.79 - .83

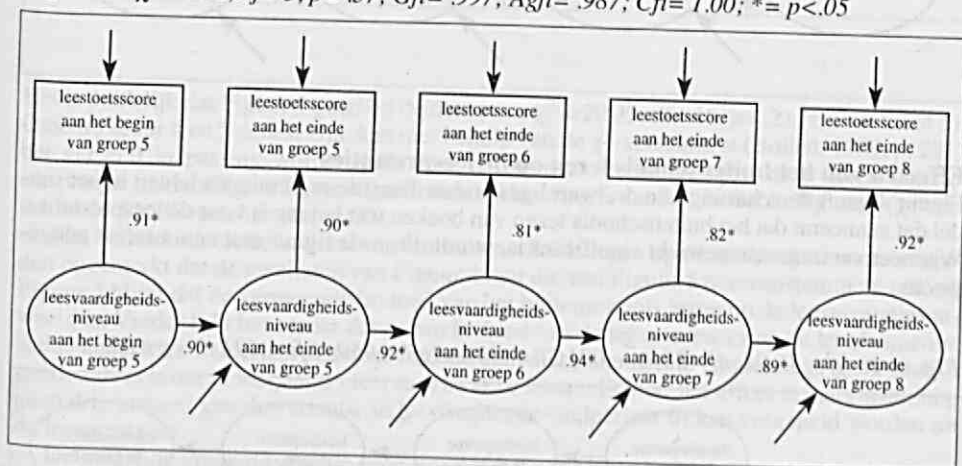
Uit de tabel blijkt dat de passing van het één-factormodel voor de weekfrequentiescores zeer goed te noemen is: de 'goodness-of-fit' (gfi) is .97 of hoger. De conclusie dat op de leesfrequentiescores het één-factormodel van toepassing lijkt, is dan ook gerechtvaardigd.

Uit Tabel 7 komt verder naar voren dat de betrouwbaarheid volgens Cronbachs α en Fleishman & Benson (1987) in leerjaar 5 tot en met 7 goed genoemd mag worden. In leerjaar 8 is de betrouwbaarheid nog steeds redelijk doch wel beduidend lager dan in de overige leerjaren (zie verder Otter en Schoonen, 1996).

Structuur van het data-bestand

In Figuur 2 worden de resultaten weergegeven van de quasi-simplexmodeltoetsing (vgl. Figuur 1) van de leesprestaties. Wanneer een gestandaardiseerd regressiegewicht significant is, wordt dit in het model met een asterisk (*) aangegeven.

Figuur 2: $\chi^2 = 3.18$, $df = 3$, $p = .37$; $Gfi = .997$; $Agfi = .987$; $Cfi = 1.00$; $* = p < .05$



In de eerste plaats kan uit Figuur 2 worden opgemaakt dat alle passingsmaten erop duiden, dat het quasi-simplexmodel goed bij de data past. Statistisch gezien wordt het model aanvaard ($\chi^2 = 3.18$, $df = 3$, $p = .37$). Ook de beschrijvende maten duiden op een goede passing; alle passingsmaten naderen de 1 (gfi, agfi en cfi).

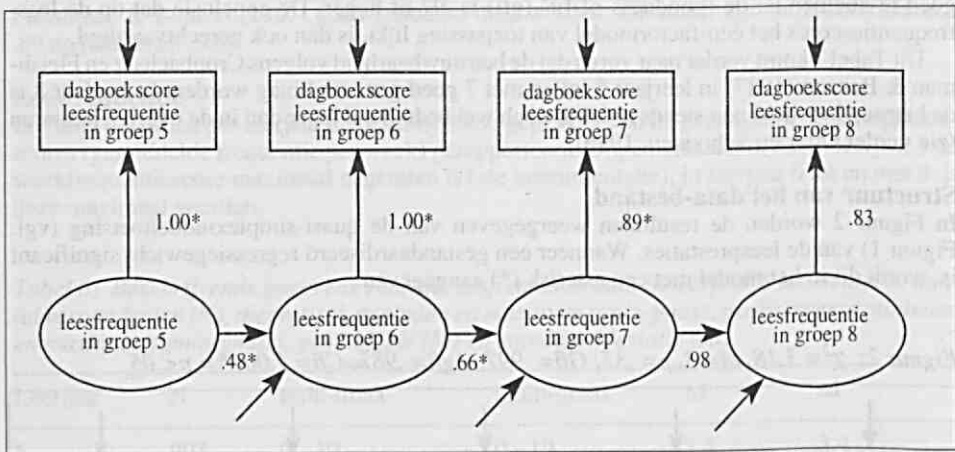
Verder blijkt uit de resultaten dat leesvaardigheid als een redelijke stabiele trek kan worden opgevat; de regressiecoëfficiënten zijn zeer hoog: minimaal; .89 (leerjaar 7 naar 8) en maximaal .94 (leerjaar 6 naar 7).

In Figuur 3 (pag. 74) worden de resultaten weergegeven van de quasi-simplexmodeltoetsing van de leesfrequentiescores.

Uit Figuur 3 kan worden opgemaakt dat het quasi-simplexmodel de data van de leesfrequentie goed beschrijft. Het model wordt statistisch gezien aanvaard ($\chi^2 = .92$, $df = 1$, $p = .34$) en ook de beschrijvende passingsmaten duiden op een goede passing (gfi, agfi en cfi naderen de 1).

Verder blijkt dat de leesfrequentie zich in de loop van de jaren *ontwikkelt* tot een stabiele trek. Zo is de regressiecoëfficiënt van leerjaar 5 naar 6 .48; van leerjaar 6 naar 7 .66 en van leerjaar 7 naar 8 .98; hoewel deze laatste coëfficiënt niet significant is⁴.

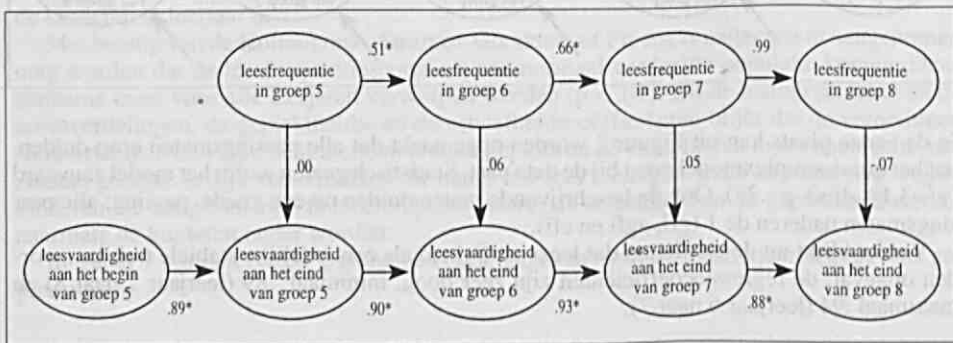
Figuur 3: $\chi^2 = .92$, $df = 1$, $p = .34$; $Gfi = .999$; $Agfi = .992$; $Cfi = 1.00$; $* = p < .05$



Effecten van het buitenschools lezen op de leesprestaties

Figuur 4 geeft de schatting van de 'ware' gestandaardiseerde regressiegewichten in het model dat aanneemt dat het buitenschools lezen van boeken van belang is voor de leesprestaties. Wanneer een regressiegewicht significant is, wordt dit in de figuur met een asterisk aangegeven.

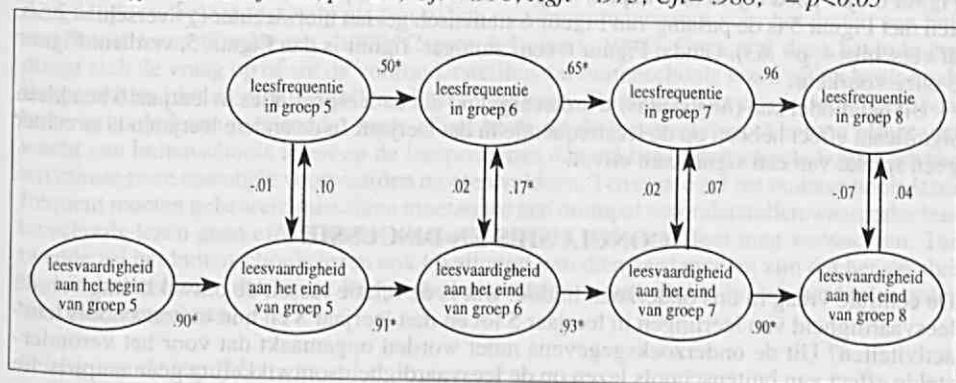
Figuur 4: $\chi^2 = 36.94$, $df = 20$, $p = .00$; $Gfi = .983$; $Agfi = .959$; $Cfi = .991$; $* = p < .05$



In de eerste plaats wordt opgemerkt dat, hoewel het model op statistische gronden verworpen moet worden ($\chi^2 = 36.94$, $df = 20$, $p = .00$), de passing redelijk is. De beschrijvende passingsmaten zijn namelijk behoorlijk te noemen ($gfi = .983$; $agfi = .959$; $cfi = .991$). Verder blijkt uit de regressies dat de leesfrequentie *geen* invloed heeft op de leesprestaties van de leerlingen. In leerjaar 5 is het regressiegewicht .00, in leerjaar 6 en 7 wel positief, respectievelijk .06 en .05, maar niet significant en in leerjaar 8 licht negatief (-.07), maar evenmin significant⁵.

De modificatie-indices van Figuur 4 suggereren dat de passing van het model aanzienlijk verbeterd zou worden als er ook paden gepostuleerd worden van de leesprestaties naar de leesfrequentie. In Figuur 5 wordt dit model weergegeven.

Figuur 5: $\chi^2 = 20.33$, $df = 16$, $p = .21$; $Gfi = .990$; $Agfi = .973$; $Cfi = .988$; $* = p < 0.05$

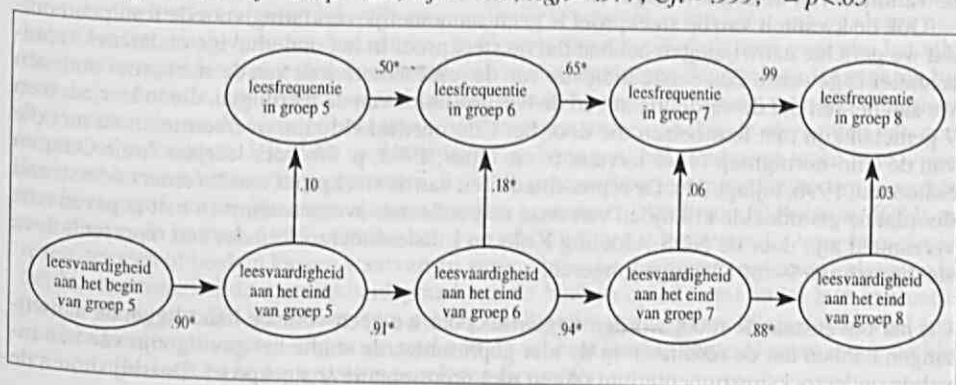


Het is duidelijk dat Figuur 5 goed bij de data past: $\chi^2 = 20.33$, $df = 16$, $p = .21$ (zie voetnoot 5). Omdat Figuur 4 en 5 genest zijn, kan met behulp van de χ^2 -verschilstest (Bollen, 1989, p. 291) het verschil in passing worden geëvalueerd. Uit deze evaluatie komt naar voren dat het reciproque model (Figuur 5) een significante verbetering is van het model in Figuur 4 (χ^2 verschil = 16.61, df verschil = 4, $p = .00$).

Met betrekking tot de invloed van het buitenschools lezen op de leesprestaties moet worden opgemerkt dat de resultaten van Figuur 4 met die van Figuur 5 overeenstemmen: ook uit Figuur 5 blijkt dat de vermeende invloed van het buitenschools lezen op de leesprestaties afwezig is. Verder lijkt het eerder dat er een invloed van de leesprestaties op de leesfrequentie is dan andersom. In leerjaar 6 is deze invloed statistisch significant (regressiegewicht is .17): goede lezers lezen in leerjaar 6 meer dan zwakke lezers. Echter, het effect is marginaal: nog geen drie procent van de variantie in leesfrequentie (in leerjaar 6) kan voorspeld worden uit de leesprestaties.

De volgende vraag die aan de orde is, is of het reciproque model (Figuur 5) kan worden versimpeld. In Figuur 6 worden de schattingen gegeven van de gestandaardiseerde regressiecoëfficiënten van een model dat aanneemt dat uitsluitend de leesprestaties een effect hebben op de leesfrequentie en niet vice versa.

Figuur 6: $\chi^2 = 22.89$, $df = 20$, $p = .29$; $Gfi = .990$; $Agfi = .977$; $Cfi = .999$; $* = p < .05$



Figuur 6 lijkt goed bij de data te passen ($\chi^2 = 22.89$, $df = 20$, $p = .29$; zie voetnoot 5). Vergelijken met Figuur 5 is de passing van Figuur 6 statistisch gezien niet slechter (χ^2 verschil = 2.56, df verschil = 4, $p = .63$). Omdat Figuur 6 een 'zuiniger' figuur is dan Figuur 5, verdient Figuur 6 onze voorkeur.

Uit het model kan (nog steeds) worden afgeleid dat de leesprestaties in leerjaar 6 een klein significant effect hebben op de leesfrequentie in dat leerjaar. In de andere leerjaren is er echter géén sprake van een significant effect.

CONCLUSIES EN DISCUSSIE

De centrale vraag in ons onderzoek luidde: wat is de relatie tussen de ontwikkeling van de leesvaardigheid van leerlingen in leerjaar 5 tot en met leerjaar 8 en hun buitenschools leesactiviteiten? Uit de onderzoeksgegevens moet worden opgemaakt dat voor het veronderstelde effect van buitenschools lezen op de leesvaardigheidsontwikkeling géén empirische steun gevonden is. Dit roept allerlei vragen op. Is het niet mogelijk dat de resultaten het gevolg zijn van een matige kwaliteit van het onderzoeksinstrumentarium? Zijn de resultaten mogelijk veroorzaakt omdat de steekproef niet representatief is geweest? Of is er misschien sprake van een niet geheel correcte vooronderstelling? Op elk van de vragen zal hieronder uitvoerig worden ingegaan.

De verklaring dat de onderzoeksresultaten veroorzaakt zijn door gebruik van een ondeugdelijk meetinstrumentarium, is gezien de psychometrische karakteristieken van de meetinstrumenten (leestoetsen en dagboeken) niet aannemelijk. Bovendien zijn er enkele positieve indicaties betreffende de *validiteit* van het meetinstrument voor het buitenschools lezen. In de eerste plaats is de validiteit van de dagboeken onderzocht door zowel de leerlingen als hun ouders de dagboekjes te laten invullen (zie Otter, 1993a). De overeenstemming tussen de gegevens van de ouders en de leerlingen over het lezen van boeken bedroeg, na correctie voor attenuatie, .79. Omdat zowel ouders als leerlingen aangemerkt kunnen worden als bron van invaliditeit, achten we deze samenhang alleszins redelijk. De validiteit van de dagboeken voor het meten van het buitenschools lezen wordt verder ondersteund door het gegeven dat 73% van de titels die leerlingen noteerden in hun dagboekjes direct (dus zonder hulp van de auteursnaam) met behulp van een computercatalogus geïdentificeerd konden worden (De Leeuw, 1992). Ook de analyses van Otter en Schoonen (1996) ondersteunen de validiteit van de dagboekmetingen. Zij zijn nagegaan of de gegevens verzameld met de dagboeken overeenstemmen met gegevens uit onderzoek waarin het buitenschools lezen of televisie kijken gemeten is met een zogenaamd criteriuminstrument, bijvoorbeeld integraal tijdsbudgetonderzoek (lees- en kijktijd) of de knoppenmethode (kijktijd). Zij vinden geen aanleiding om te twifelen aan de validiteit van de dagboekgegevens (Otter en Schoonen, 1996).

Ook de kwaliteit van de steekproef is geen aannemelijk verklaring voor de resultaten omdat we gerichte aanwijzingen hebben dat de steekproef in het onderhavige onderzoek representatief is geweest. Een eerste indicatie van de representativiteit van de steekproef ontleen we aan het feit dat de verdelingen van de leesprestaties van de leerlingen, die in leerjaar 6 en 7 gemeten zijn met leestoetsen die door het Cito ontwikkeld zijn, overeenstemmen met die van de Cito-normgroep (voor leerjaar 6: zie Otter, 1993, p. 96; voor leerjaar 7: zie Otter en Schoonen, 1996, bijlage 21). De representativiteit van de steekproef wordt verder ondersteund doordat de gemiddelde kijktijden van onze respondenten overeenstemmen met gegevens die verzameld zijn door de NOS-Afdeling Kijk- en Luisteronderzoek onder een representatieve steekproef van 9- tot 12-jarige jongeren.

Uit het bovenstaande moet worden opgemaakt dat we geen voor de hand liggende aanwijzingen hebben dat de resultaten in de hier gepresenteerde studie het gevolg zijn van een invalide onderzoeksinstrumentarium of een niet-representatieve steekproef. Daarbij sluiten de

onderzoeksresultaten nauw aan bij ander onderzoek waarin gebruik is gemaakt van zowel een deugdelijk onderzoeksontwerp als een valide meetinstrumentarium voor het meten van buitenschools lezen (Taylor e.a., 1990; Carver & Leibert, 1995). Gezien al deze bevindingen dringt zich de vraag op of we de vooronderstelling dat buitenschools lezen effect heeft op de leesprestaties niet moeten bijstellen of herformuleren. Duidelijk lijkt ons in ieder geval dat buitenschools lezen niet zonder meer effect heeft op de leesprestaties. Als men effect verwacht van buitenschools lezen op de leesprestaties dan zal het buitenschools lezen toch aan ten minste twee essentiële voorwaarden moeten voldoen. Ten eerste zal het buitenschools lezen frequent moeten gebeuren; misschien moeten we een drempel veronderstellen waaronder buitenschools lezen geen effect heeft en waarboven men wel een effect mag verwachten. Ten tweede zal het buitenschools lezen ook kwalitatief van dien aard moeten zijn dat het aansluit bij het 'soort lezen' dat bij de leesprestaties gemeten wordt. Beide 'voorwaarden' worden hieronder besproken.

Belang van de leestijd

In de eerste plaats verwachten we dat het buitenschools lezen pas effect kan sorteren op de ontwikkeling van de leesprestaties als de tijd die aan lezen besteed wordt substantieel is. Uit de gegevens van Anderson e.a (1988), Taylor e.a. (1990) en Otter en Schoonen (1996) komt namelijk naar voren dat leerlingen in het basisonderwijs in hun vrije tijd zelden in boeken lezen. Gemiddeld genomen lezen Nederlandse leerlingen in leerjaar 6 zes minuten per dag in een leesboek, in leerjaar 7 minder dan vijf minuten en in leerjaar 8 nog geen vier minuten per dag! Wil het veronderstelde effect zich kunnen tonen, dan is een eerste vereiste dat kinderen voldoende thuis lezen.

Toch lijkt uitsluitend véél lezen niet voldoende om een betere lezer te worden. Het onderzoek van Carver en Leibert (1995) maakt duidelijk dat een ruime dosis vrij lezen misschien wel een noodzakelijk, maar zeker geen voldoende voorwaarde is om een betere lezer te worden. Van 3600 minuten (!) lezen in bibliotheekboeken verspreid over 30 dagen ging immers géén effect uit op de leesprestaties.

Uit de onderzoeksresultaten dat véél lezen niet zonder meer effect sorteert op de leesprestaties, leiden wij af dat ook de *kwaliteit* van het leesproces een factor van belang is.

Belang van de kwaliteit van het leesproces

Onze veronderstelling dat niet alleen de kwantiteit maar ook de kwaliteit van het leesproces van belang is om een betere lezer te worden, sluit nauw aan bij Carvers theorie (1977, 1990) over het *alledaags* lezen. Carvers theorie over lezen maakt heel expliciet onderscheid tussen 'alledaags' en 'studerend' lezen. Tijdens het alledaags lezen leest men (a) op een constante snelheid waarbij men (b) géén gebruik maakt van leer- of herinneringsprocessen. Het alledaags lezen is volgens Carvers theorie *niet* toereikend om iets te *leren* tijdens het lezen. Daarvoor dienen kinderen (maar ook volwassenen) over te schakelen van het *alledaags* lezen naar leesprocessen zoals *leren* en *herinneren*. In Carvers theorie schakelt men over van het alledaags lezen naar leesprocessen zoals *leren* en *herinneren* als men daartoe wordt aangezet, bijvoorbeeld door de moeilijkheidsgraad van een tekst of de context waarin men leest. In moeilijke teksten komt de lezer onbekende woorden tegen en/of onbekende elementen van concepten of procedures. Hierdoor wordt de lezer 'gedwongen' het alledaagse lezen in te wisselen voor leer- en herinneringsprocessen die in de schematheorie centraal staan om te verklaren hoe men *leert* en *herinnert* als men teksten *bestudeert* (Anderson & Pearson, 1984; zie ook Carver & Leibert, 1995). Omdat het waarschijnlijk is dat kinderen die voor hun *plezier* lezen, teksten of boeken kiezen waarvan zij verwachten dat die geen problemen opleveren bij het lezen ervan, zal de moeilijkheidsgraad van de boeken op of (iets) onder hun instructieniveau zal liggen. Dit laatste heeft tot gevolg dat het niet erg aannemelijk is dat zij veel activiteiten hoeven of moeten ondernemen die in de schema theorie centraal staan om te verklaren hoe men leert en herinnert als men teksten bestudeert.

Dat 'gewoon' lezen, of in Carvers termen (1979, 1990) het alledaags lezen, niet voldoende is om een betere lezer te worden, óók niet in de schoolse situatie, lijkt ook naar voren te komen in het overzicht van Pearson en Fielding (1991) van de effecten van instructiestrategieën op het begrijpendleesproces. Genoemde auteurs komen tot de conclusie dat, onder de juiste condities, bijna elk interventieprogramma 'werkt'. De programma's worden werkzaam zodra leerlingen actief met de inhoud van het verhaal bezig zijn: de leerlingen vatten het samen, formuleren en beantwoorden vragen, of maken bijvoorbeeld inferenties over de afloop. Genoemde auteurs zijn dan ook van mening dat het lezen 'als activiteit op zich' de leesprestaties niet per definitie zal doen verbeteren.

Deze redenering volgend zal het buitenschools lezen pas dan een positief effect hebben op de leesprestaties van de kinderen als zij thuis actief lezen. Maar is de aanname dat kinderen thuis actief bij het lezen betrokken zijn een aannemelijke? Met andere woorden is het plausibel te veronderstellen dat kinderen thuis, lekker lui in het bed of op de bank, activiteiten zullen gaan ondernemen zoals het samenvatten van het verhaal of het formuleren en beantwoorden van vragen naar aanleiding van het gelezene? Wij zetten vraagtekens bij de actieve houding van de lezer thuis en wel op grond van twee argumenten.

In de eerste plaats komt uit de onderwijspraktijk naar voren dat er nog maar nauwelijks onderricht wordt gegeven in begrijpendleesstrategieën (Durkin, 1979; Weterings & Aarnoutse 1986 en Aarnoutse en Weterings 1991). Dit betekent dat het aannemelijk is dat de meeste kinderen over géén of weinig effectieve strategieën beschikken waardoor zij deze activiteiten thuis niet eens kunnen toepassen. Daarbij komt dat de training van dit soort (cognitieve) activiteiten krachtig moet zijn, willen deze enigszins (zelfs in schoolse leessituaties) beklijven (Pearson & Fielding 1991). Blijkbaar doen leerlingen het niet 'uit zichzelf'.

In de tweede plaats is het buitenschools lezen is een activiteit die het kind voor z'n plezier onderneemt. Het is onaannemelijk dat bij het lezen thuis inspanningen worden geleverd waar de kinderen in de schoolse situatie slechts na veel moeite en aandacht toe overgehaald kunnen worden.

REFERENTIES

- Aarnoutse, C.A.J., & Boland, Th. (1987). *Schalen voor het meten van de leesattitude en de leespreferentie bestemd voor de hoogste groep van het basisonderwijs*. Nijmegen: Berkhout Nijmegen B.V.
- Aarnoutse, C.A.J., & Weterings, A.C.E.M. (1991). *Onderwijs en begrijpend lezen*. Nijmegen: Vakgroep Onderwijskunde.
- Allen, L., Cipielewski, J., & Stanovich, K.E. (1992). Multiple indicators of children's reading habits and attitudes: construct validity and cognitive correlates. *Journal of Educational Psychology*, 84, 4, 489-503.
- Anderson, R.C., & Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. 1, pp. 255-291). White Plains, NY: Longman.
- Anderson, R.C., Wilson, P.T., & Fielding, L.G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23, 285-303.
- Bock, H. de (1985). Het epos van het kijkonderzoek. *Massacommunicatie*, 8, 209-214.
- Bradley, L. & Bryant, P.E. (1983). Categorizing sounds and learning to read - a causal connection. *Nature*, 301, 419-421.
- Bus, A.G., IJzendoorn, M.H. & Pellegrini, A.D. (1995). Joint Book Reading Makes for Success in Learning to Read: A Meta-Analysis on Intergenerational Transmission of Literacy. *Review of Educational Research*, 65, 1, 1-21.
- Carver, R.P. (1977). Toward a theory of reading and rauding. *Reading Research Quaterly*, 13, 863-870.
- Carver, R.P. (1990). *Reading rate: A review of research and theory*. New York: Academic Press.
- Carver, R.P., & Leibert, R.E. (1995). The effects of reading library books at different levels of difficulty upon gain in reading ability. *Reading Research Quarterly*, 30, 1, 26-48.
- Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit. Evaluatie van het basisonderwijs*. De Meern: Inspectie van het Onderwijs.

- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.
- Cunningham, A.E., & Stanovich, K.E. (1990). Assessing print exposure and orthographic processing skill in children: a quick measure of reading experience. *Journal of Educational Psychology*, 82, 4, 733-740.
- Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 15, 481-533.
- Elley, W.B. (1994). *The IEA Study of Reading Literacy*. Oxford: Elsevier Science.
- Fleishman, I., & Benson, J. (1987). Using LISREL to evaluate measurement models and scaling reliability. *Educational and Psychological Measurement*, 47, 925-939.
- Glopper, K. de, & Otter M.E. (1993). *Nederlandse leesprestaties in internationaal perspectief*. Amsterdam: SCO.
- Greaney, V. (1980). Factors related to amount and types of leisure-time reading. *Reading Research Quarterly*, 15, 337-357.
- Greaney, V., & Hegarty, M. (1987). Correlations of leisure-time reading. *Journal of Research in Reading*, 10, 3-20.
- Hagglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika*, 47, 209-222.
- Heath, S.B. (1982). What no bedtime story means: narrative skills at home and school. *Language Sociology*, 11, 49-76.
- Hox, J.J. (1993). Factor analysis of multilevel data. Gauging the Muthén model. In J.H. Oud & R. Vogelsang (Eds.), *Advances in longitudinal multivariate analysis in the behavioral sciences*. Leiden: DSWO Press.
- Jöreskog, K.G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS Inc.
- Jöreskog, K.G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, 23, 121-145.
- Koolstra, C.M. (1993). *Television and children's reading: A three-year panel study*. Leiden: University of Leiden (Doctoral Dissertation).
- Kuhlemeier, J.B. (1996). *Taalvaardigheid, taalactiviteiten en taalattituden. Een validatiestudie*. Arnhem: Cito (Academisch proefschrift).
- Lundberg, L., Frost, J. & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, 23, 263-284.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Presidential address to the psychometric society, July 1989. *Psychometrika*, 54, 557-585.
- Muthén, B. (1990). *Mean and covariance structure analysis of hierarchical data*. Paper presented at the psychometric society meeting in Princeton, NJ, June 1990. UCLA Statistics Series #62. Accepted for publication in *Journal of Educational Statistics*.
- Otter, M.E. (1993). *Leesvaardigheid, leesonderwijs en buitenschools lezen: instrumentatie en effecten*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam. (Academisch proefschrift).
- Otter, M.E. (1995a). *Buitenschools lezen effectief voor de schoolse leesvaardigheid? Een vierjarig longitudinaal onderzoek*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam.
- Otter, M.E. (1995b). *TV kijken en de leesprestaties: een replicatie van het onderzoek van Koolstra*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam.
- Otter, M.E., & Schoonen, R. (1996). *Aap, Noot, Niets ... Het spook van de ontleding in het basisonderwijs*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam.
- Paris, S.G., Wasik, B.A., & Turner, J.C. (1991). The development of strategic readers. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds.), *Handbook of Reading Research*. (Volume II, pp. 609-640). New York: Longman.
- Pearson, P.D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M.L. Kamil, P. Mosenthal & P.D. Pearson (Eds.), *Handbook of Reading Research*. (Volume II, pp. 815-860). New York: Longman.
- Roller, K.J. (1991). The influence of reading activity at home on students' attitudes toward reading, classroom attentiveness and reading achievement: an application of structural equation modelling. *British Journal of Educational Psychology*, 61, 19-35.
- Scarborough, H.S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, 14, 245-302.
- Snow, C.E. (1983). Language and literacy: relationships during the preschool years. *Harvard Educational Review*, 53, 2, 165-189.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.
- Sijtsma, J. (1992). *Balans van het taalonderwijs halverwege de basisschool. Uitkomsten van de eerste taalpeiling medio basisonderwijs*. Arnhem: Cito.

- Taylor, B.M., Frye, B.J., & Maruyama, G.M. (1990). Time spent on reading and reading growth. *American Educational Research Journal*, 27, 2, 351-362.
- Waldberg, H.J., & Tsai, S. (1984). Reading achievement and diminishing returns to time. *Journal of Educational Psychology*, 76, 442-451.
- Wesdorp, H. (1985). *Goed onderwijs, wat is dat? Voorstudie Periodieke Peiling van het Onderwijsniveau*. 's-Gravenhage: SVO.
- Weterings, A.C.E.M., & Aarnoutse, C.A.J. (1986). De praktijk van het onderwijs in begrijpend lezen. *Pedagogische Studiën*, 63, 387-400.
- Zwarts, M.A. (1990). *Balans van het taalonderwijs aan het einde van de basisschool*. Arnhem: Cito.

Ontvangen 15-5-96

Definitief aanvaard 16-1-97

-
1. De auteurs willen graag G.J. Mellenbergh en J. Hox danken voor hun uitvoerige commentaar op eerdere versies van dit artikel.
 2. In het algemeen wordt aangenomen dat, indien we iets weten over het leesgedrag van kinderen, we ook informatie hebben over de leesattitude. De aanname is namelijk dat de leesattitude een richtinggevende invloed heeft op het leesgedrag. Dit impliceert dat kennis van de leesattitude een belangrijk gegeven is om inzicht te krijgen in het leesgedrag. Hoewel deze aanname zeer plausibel klinkt, is empirische evidentie voor deze assumptie mager (Aarnoutse & Boland, 1987, p. 42) of afwezig (Otter, 1995b, p. 54-56). Daarom zal in het onderhavige onderzoek het leesgedrag als variabele worden opgenomen en niet een mogelijke *indicator* van het leesgedrag, de leesattitude.
 3. De analyses zijn ook uitgevoerd met behulp van de ongewogen kleinste-kwadratenmethode ('unweighted least squares': ULS) omdat de ULS-methode geen enkele aanname doet over de verdeling van de variabelen. Indien de resultaten op basis van de ULS-methode afwijken van die op basis van de ML-methode, zal dit gerapporteerd worden.
 4. De resultaten van de 'ongewogen kleinste kwadraten-methode' (ULS) zijn nagenoeg overeenkomstig met die van de ML-methode. Het enige verschil tussen de resultaten is dat de *niet* significante regressiecoëfficiënten (zie Figuur 3) bij de ULS-methode wel significant zijn. De hoogte en de richting van de regressiecoëfficiënten zijn vrijwel overeenkomstig met die in Figuur 3. Ook de passing van het model is goed te noemen (zowel gfi , $agfi$ en $cf1$ zijn .999).
 5. De resultaten van de 'ongewogen kleinste kwadraten-methode' (ULS) zijn nagenoeg overeenkomstig met die van de ML-methode. Het enige verschil tussen de resultaten is dat de *niet* significante regressiecoëfficiënt tussen de latente variabelen leesfrequentie in leerjaar 7 en 8 bij de ULS-methode wel significant is. De hoogte en de richting van de regressiecoëfficiënten zijn in alle gevallen vrijwel overeenkomstig. Ook de passing van het model is goed te noemen (zowel gfi , $agfi$ en $cf1$ zijn .999).

Notities en Commentaren

EEN ALGEMEEN MODEL VOOR LOTING EN SELECTIE BIJ NUMERUS CLAUSUS¹⁾

Willem K.B. Hofstee en Henk A.L. Kiers
Heymans Instituut, R.U. Groningen

Bij het uiten van onvrede over 'de loting' bij de toelating tot studierichtingen met een numerus clausus is men zich niet steeds bewust dat het systeem van gewogen loting alle kanten op kan. In de jaren zeventig bedacht door A.J. Vermaat, Anti-Revolutionair parlementariër en wiskundige, mengt de gewogen loting twee principes: gelijke rechten op toelating voor houders van een diploma (leidend tot loting) en recht op grond van relatieve prestatie (leidend tot vergelijkende selectie). Een rationele discussie over dit toelatingssysteem wordt vergemakkelijkt als men zich realiseert dat zuivere loting en zuivere selectie als speciale (nl. uiterste) gevallen kunnen worden begrepen van het systeem van Vermaat. Immers, dat inziende wordt duidelijk dat de discussie gaat over de lotingsgewichten en niet over het systeem van gewogen loting.

We presenteren hier een formalisering van het systeem van gewogen loting. Met behulp van dat instrument kan op een meer systematische manier een keuze worden gemaakt voor een stel lotingsgewichten. Hier is niet aan de orde de wenselijkheid of onwenselijkheid van de numerus clausus. De discussie daarover is waarschijnlijk belangrijker dan de onderhavige. Evenmin is expliciet aan de orde de vraag of de gewogen loting, bij handhaving van de numerus clausus, beter kan worden vervangen door een wachtlijst; echter, wanneer de volgorde op zo'n wachtlijst door gewogen loting tot stand komt, is het onderstaande mutatis mutandis van toepassing. Al helemaal niet aan de orde is de recente discussie over vervanging van eindexamencijfers door een ander meetmiddel (voorzover een toelatingsgesprek die betiteling verdient). We beperken ons tot de vraag hoe het systeem van gewogen loting kan worden gevarieerd.

De concrete gewichten per cijferklasse, dat wil zeggen afgerond gemiddeld eindexamencijfer, zijn momenteel als volgt:

klasse:	6	6 $\frac{1}{2}$	7*	7 $\frac{1}{2}$	8	8 $\frac{1}{2}$
gewicht:	.67	.8	1	1.25	1.5	2

*Bij de zevens rekenen we ook de 'overige opleidingen', die hetzelfde lotingsgewicht ontvangen. Verder betekent 6 een cijfer lager dan 6 $\frac{1}{2}$, enzovoort.

De gegadigden worden op grond van hun gemiddeld eindexamencijfer ingedeeld in deze klassen, en per klasse wordt een inlotingskans bepaald. Bij een lage toelatingsratio, dus wanneer er veel aanmeldingen zijn in verhouding tot het aantal plaatsen, verhouden de inlotingskansen zich inderdaad conform de gewichten. Bij een hoge ratio echter kloppen de gewichten niet meer. Immers, gegadigden met een 7 zouden slechts een toelatingskans van 50% mogen hebben, want dat is het maximum wanneer de toelatingskans voor de 8 $\frac{1}{2}$ -en twee keer zo groot zou moeten zijn. Op die manier zouden er plaatsen leeg blijven. In de praktijk begint men in zo'n geval met de gegadigden uit de hoogste lotingsklassen rechtstreeks toe te laten, dus een kans van 1 te geven. De feitelijke kansverhoudingen hangen dus mede af van de toelatingsratio.

Door alle gewichten identiek te stellen verkrijgt men zuivere loting. Door, afhankelijk van de toelatingsratio, de laagste cijferklassen een gewicht 0 te geven, verkrijgt men zuivere selec-

tie. Door de gewichten steiler te laten verlopen verkrijgt men *ceteris paribus* een selectiever systeem, enzovoorts.

Een algemeen model

Op te merken valt dat de Vermaat-gewichten een versnelde functie zijn van het gemiddeld eindexamencijfer. We nemen aan dat daarvoor goede redenen zijn en kiezen voor een exponentieel model. Wellicht zijn ook andere functies (lineaire, vertraagde, s-vormige) verdedigbaar; de modellen daarvoor kunnen eventueel analoog worden geconstrueerd.

Een eis aan het model is dat zuivere loting en zuivere selectie er grensgevallen van zijn. Dat wordt bereikt door in het model een hellingsparameter h op te nemen; h is een beleidsparameter voor de mate van selectiviteit van het systeem. In het grensgeval van zuivere selectie is er vervolgens een bepaald gemiddeld eindexamencijfer $X = X_g$ waarbeneden het lotingsgewicht 0 wordt; die grens X_g vloeit voort uit de frekwentieverdeling van X en het aantal beschikbare plaatsen. X_g is dus een empirische parameter waarvan de waarde jaarlijks en over numerus-clausus-studierichtingen varieert. Tot slot voeren we een constante c in welke helpt de ijking van het model aan de Vermaat-gewichten te optimaliseren. Het model luidt:

$$w_x = e^{h(X-X_g) + c} \quad (1)$$

of:

$$\ln(w_x) = hX - hX_g + c \quad (2)$$

met w_x het gewicht dat hoort bij klasse X . De log-gewichten zijn dus een lineaire functie van X , zodat de Vermaat-gewichten bij een gegeven grenswaarde X_g kunnen worden gereconstrueerd via lineaire regressie.

Voor $h = 0$ zijn de gewichten dezelfde, nl. e^c , voor iedere waarde van het eindexamencijfer X , en treedt dus zuivere loting op. Voor $h = \infty$ krijgen alle $X < X_g$ het gewicht 0 (en alle $X > X_g$ hetzelfde, oneindig grote gewicht), zodat zuivere selectie optreedt. Tussenliggende waarden $0 < h < \infty$ leveren gewogen loting op in strikte zin, dus met uitsluiting van de grensgevallen.

De beleidsparameter h kan worden geïnterpreteerd als een selectiviteitsparameter S/L , waar bij S het gewicht is dat aan selectie, en L het gewicht dat aan loting wordt toegekend; bij $S=0$ volgt zuiver loting, bij $L=0$ volgt zuivere selectie. Bij de interpretatie van tussenliggende waarden van h echter geldt als waarschuwing dat h afhankelijk is van het gekozen grondtal van de exponent. Als men in plaats van e een grondtal a zou kiezen, resulteert een andere hellingsparameter $h' = h/\ln(a)$. Men kan dus h (of S/L) slechts in vergelijkende, niet in absolute zin interpreteren.

In plaats van h direct te kiezen, verdient het dan ook voorkeur het model indirect toe te passen: men leidt de waarde van de beleidsparameter af uit de bestaande praktijk én overweegt vervolgens of die geïmpliceerde waarde overeenkomt met de bedoelingen. Eventueel verhoogt of verlaagt men de waarde en overweegt of het effect daarvan acceptabeler is, enzovoort.

Uitdrukking (1) kan worden herschreven als

$$w_x = e^{hX+c} e^{-hX_g} \quad (3)$$

zodat duidelijk is dat bij verschuiven van de toelatingsratio, en daarmee van X_g , de gewichten proportioneel blijven aan de oorspronkelijke. De gewichtsverhoudingen blijven dus identiek. Een eigenschap van het model is derhalve dat de selectiviteit van de gekozen oplossing ongevoelig is voor verschuivingen in de toelatingsratio. In de toepassing echter zal verlaging van X_g (als gevolg van verlaging van het aantal aanmeldingen) tot gevolg hebben dat bij de hogere klassen eerder het plafond wordt bereikt van een toelatingkans van 1.

Een toepassing

We hebben het model toegepast op de aanmeldingsgegevens voor Medicijnen in 1995. Deze aanmeldingen waren bij benadering als volgt verdeeld:

klasse	6	6 $\frac{1}{2}$	7*	7 $\frac{1}{2}$	8	8 $\frac{1}{2}$
aantal	1271	924	1429	396	254	87

*Wederom inclusief 'overige opleidingen'.

Het aantal beschikbare plaatsen voor deze gegadigden was 1686, zodat bij zuivere selectie de grens X_g tussen 7 en 7 $\frac{1}{2}$ zou zijn komen te liggen. Bij gebrek aan gedifferentieerdere gegevens interpoleren we $X_g = 7.168$; in de praktijk zou men over gemiddelde eindexamencijfers in decimalen nauwkeurig moeten beschikken om X_g te vinden. Regressie van de log-gewichten op de cijferklassen levert bij deze grenswaarde $h = .434$ en $c = -.029$ op. De teruggeschatte lotingsgewichten zijn als volgt:

klasse	6	6 $\frac{1}{2}$	7	7 $\frac{1}{2}$	8	8 $\frac{1}{2}$
gewicht	.65	.81	1.00	1.25	1.55	1.93

Zoals opgemerkt, kan men de vraag stellen of lotingsgewichten overeenkomen met de politieke inzichten van dit moment. Vervolgens kan men de selectiviteitsparameter variëren en een evaluatie ondernemen van het effect daarvan. We illustreren dit proces aan de hand van de waarde $h = 1$, onder constant houden van c . We herhalen de waarschuwing dat deze waarde bij een ander grondtal anders zou uitvallen, en dus geen bijzondere betekenis heeft. Deze oplossing levert de volgende gewichtsverhoudingen voor de casus:

klasse	6	6 $\frac{1}{2}$	7	7 $\frac{1}{2}$	8	8 $\frac{1}{2}$
gewicht	.39	.64	1.05	1.74	2.87	4.73

In Figuur 1 (pag. 84) zijn de vigerende gewichten, de terugschatting ervan, en de alternatieve gewichten afgebeeld. De lijnen snijden elkaar bij de grenswaarde van 7.168.

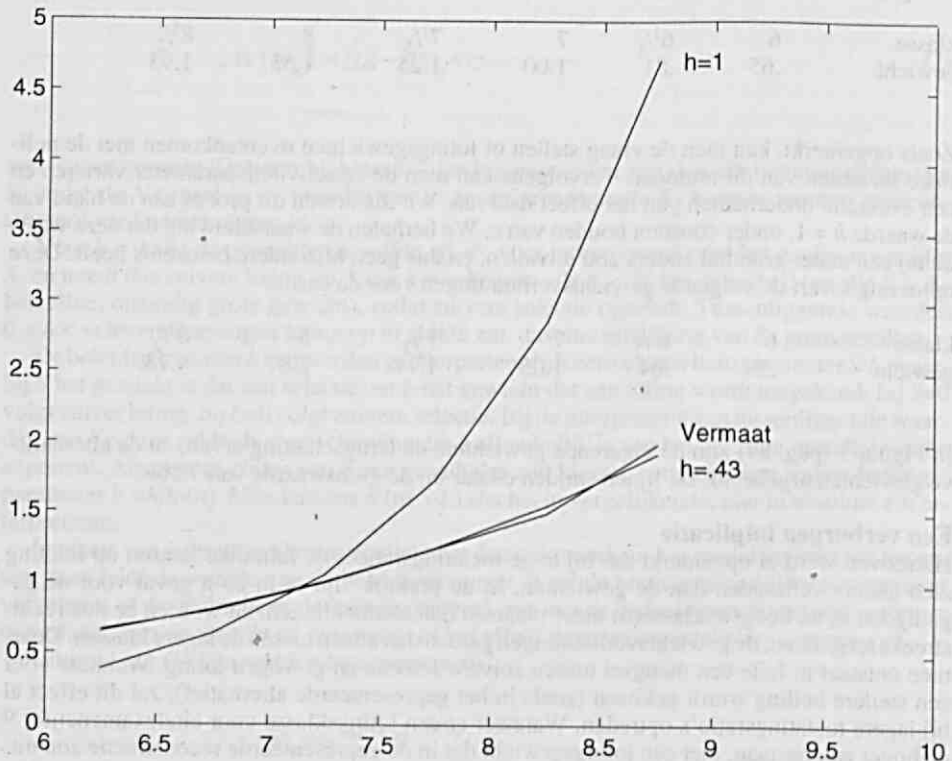
Een verborgen implicatie

Hierboven werd al opgemerkt dat bij hoge toelatingsratio's de feitelijke kansen op inloting zich anders verhouden dan de gewichten. In de praktijk zijn er in zo'n geval voor de gegadigden in de hoogste klasse(n) meer plaatsen dan aanmeldingen, en worden ze dus rechtstreeks toegelaten; de gewichtsverhoudingen gelden dan alleen tussen de lagere klassen. Daarmee ontstaat in feite een mengsel tussen zuivere selectie en gewogen loting. Wanneer voor een steilere helling wordt gekozen (zoals in het gepresenteerde alternatief), zal dit effect al bij lagere toelatingsratio's optreden. Wanneer er een lotingsklasse voor eindexamencijfer 9 en hoger zou bestaan, met een lotingsgewicht dat in de gepresenteerde reconstructie zou uitkomen op 2.4, zou wellicht een gegadigde voor de medicijnenstudie met zo'n cijfer recht-

streeks zijn geplaatst (en zou de politieke discussie over het systeem misschien niet opnieuw zijn begonnen).

Men kan tegen deze implicatie verschillend aankijken. Enerzijds kan men redeneren dat het niet in de geest is van het politiek compromis dat personen die niet onder een hardheidsclausule vallen toch in feite rechtstreeks worden geplaatst. Een oplossing zou dan zijn de lotingsgewichten anders te gebruiken: geef iedere gegadigde een aantal loten pro rata van zijn of haar lotingsgewicht, en verwijder alle loten van een persoon van wie een lot is getrokken uit de urn. Die werkwijze voorkomt rechtstreekse plaatsing: iemand met zeer veel loten kan buiten de prijzen vallen. Hij zou er echter toe leiden dat in geen enkel geval de inlotingskansen zich verhouden conform de gewichten. Men kan voorts als bezwaar hebben dat de toelatingskanscurve een scherpe knik vertoont, namelijk daar waar de exponentiële kromme, bij de inlotingskans $p=1$, overgaat in een horizontale lijn. Daarmee verkrijgt het systeem een onwenselijke eigenschap die destijds heeft geleid tot afschaffing van de "7 $\frac{1}{2}$ -regel" (vanaf 7 $\frac{1}{2}$ rechtstreeks toelaten, daarbeneden ongewogen loten): bij zo'n knik maakt het erg veel verschil of men net onder of boven de drempel scoort. Een oplossing zou bestaan uit het kiezen van een ander model, bijvoorbeeld een S-curve die de feitelijke praktijk beter weergeeft maar geen scherpe knik vertoont. Anderzijds echter kan men van mening zijn dat de gesignaleerde implicatie nog niet zo gek is. Men zou dan in de voorlichtingsbrochures kunnen expliciteren dat de gegadigden in de hoogste klasse(n) in de praktijk rechtstreeks worden toegelaten.

Figuur 1



Conclusie

We hebben een model gepresenteerd bestaande uit een klasse van functies waarvan zuivere selectie op eindexamencijfers, de vigerende gewogen loting, en zuivere loting bijzondere gevallen zijn. De kern van het model bestaat uit een selectiviteitsparameter, die is op te vatten als een gewichtsverhouding met betrekking tot het selectie- versus het lotingsprincipe. Met behulp van het model kunnen de gewichten op systematische wijze worden gevarieerd.

Gesignaleerd werd vervolgens dat bij hoge toelatingsratio's, bij hoge waarden van de selectiviteitsparameter, en bij toevoegen van een hogere lotingsklasse (cijfer 9 en hoger) in feite rechtstreekse toelating plaatsvindt voor de hoogste klasse(n). Daarmee is het systeem van gewogen loting feitelijk een mengsel van rechtstreekse toelating en gewogen loting.

Naschrift

Inmiddels is bekend dat de Commissie-Drenth aanbeveelt 50% van de plaatsen te reserveren voor de aanmelders met de hoogste eindexamencijfers - waarbij de grens voor Medicijnen bij 7.3 of 7.4 zou komen te liggen - en beneden die grens ongewogen te loten, met een kans die zou neerkomen op plm. 25%. Deze oplossing lijkt sterk op de regel die in het begin van de jaren 70 werd gehanteerd, en die op goede gronden is vervangen door de huidige gewogen loting. De voorspelbare onwenselijke neveneffecten van de 'Drenth-regel', in vergelijking met de oplossing die hier werd aangegeven, zijn (a) een gevoel van onredelijke behandeling bij honderden uitgelote studenten per jaar met een eindexamengemiddelde net beneden de drempel; (b) versterkte onttrekking van intellectueel potentieel aan technische studierichtingen. Een 'mixed blessing' is dat aanmelders met een verwacht gemiddeld cijfer in de buurt van de grens harder zullen werken voor hun eindexamencijfer, en degenen die daar duidelijk boven of onder zitten minder hard dan onder het huidig regiem. Uiteraard is de Drenth-regel eenvoudiger dan andere vormen van gewogen loting. Aangezien hier echter werd aangetoond dat rechtstreekse toelating ook kan voortvloeien uit varianten van de huidige gewogen loting, weegt het voordeel niet op tegen de nadelen.

1. De auteurs zijn prof. dr. T.A.B. Snijders en prof. dr. T.J. Wansbeek erkentelijk voor hun commentaren op een eerdere versie.

Boekbesprekingen

Lowyck, J., Verloop, N. (red.)

Onderwijskunde: een kennisbasis voor professionals

Groning: Wolters Noordhoff. ISBN 90 01 54531 9

De doelgroep

Het boek wordt in de Inleiding gepresenteerd als een 'opvolger' van het "bekende standaardwerk *Beknopte didaxologie*", dat verscheen in 1972. "Beknopte" beleefde drie herdrukken en werd vertaald in het Duits en Frans. De laatste, grondig gewijzigde editie verscheen in 1981. Het werd dus hoog tijd voor een opvolger.

Vergelijking van de twee boeken leert dat er uiteraard enige overeenkomsten zijn, maar er zijn vooral veel verschillen. Overeenkomstig is onder meer dat beide redacties van Nederlandse en Vlaamse herkomst zijn. Dat leidt er toe dat in "Onderwijskunde" op een aantal plaatsen ook aandacht besteed wordt aan verschillen in onderwijsbestel en -beleid tussen de twee landen.

In beide boeken is hetzelfde doel geformuleerd: 'een baken te zijn in de steeds groeiende stroom van sociaal-wetenschappelijke literatuur betreffende onderwijzen en leren'. Een baken is van belang om een koers te bepalen. De vraag is welke doelgroepen steun van het baken zullen ondervinden. Zullen dat de a.s. leraren zijn, of de beginnende studenten onderwijskunde?

In "Onderwijskunde" streeft men naar een wetenschappelijk gehalte zonder daarbij de leesbaarheid uit het oog te verliezen. Het uitgangspunt voor de behandeling is gevonden bij de docent. De informatie is geselecteerd op mogelijke relevantie voor de (aanstaande) leraar. Niettemin is ook rekening gehouden met wetenschappelijk geïnteresseerde lezers (bijvoorbeeld aanstaande onderwijskundigen) doordat een groot aantal literatuurverwijzingen is opgenomen.

Het beroep op professionaliteit van de leraar is in beide boeken even groot. In "Beknopte" vormen de componenten uit een model voor het didactisch handelen de basis voor de structuur van het geheel. In "Onderwijskunde" is het perspectief van de docent het uitgangspunt.

De verschillen tussen de twee boeken zijn belangwekkend en relevant. Voor het onderwijzen is het prescriptieve karakter van het didactisch model verlaten. In "Onderwijskunde" is voor het leren een toenemend belang gehecht aan het kennisconstructie proces, dat zich voltrekt wanneer de leerling zelf actief met leerstof bezig is. Dit laatste is alleszins een verbetering vergeleken met de behandeling van leerprocessen in "Beknopte".

De inhoud

Het boek bestaat uit 9 hoofdstukken. Voor elk van die hoofdstukken zijn afzonderlijke auteurs aangetrokken. De eerste drie hoofdstukken betreffen de 'gegevenheden' die een leraar in en rondom zijn werk aantreft, te weten de leerlingkenmerken (P.R.J. Simons), de school (E.C.H. Marx, O.A. de Vries) en de klas (S.A.M. Veenman, P. Slegers). Over de leerlingkenmerken bevat het boek een overzicht van onderzoeksgegevens over de relaties die gevonden zijn tussen de verschillende typen kenmerken van leerlingen, hun leerprocessen en hun leerprestaties. In hoofdstuk 2 komt eerst de school als organisatie in beeld. Aandacht wordt besteed aan het beleidsvoerend vermogen en de beleidscyclus. Over schoolcultuur en schoolontwikkeling wordt helaas niet gesproken. Daarna behandelt men 'de feitelijkheid van de klas', de werkplek voor de leraar. Hier komen thema's aan de orde als klasklimaat, groepsindeling en groepsgrootte.

In hoofdstuk 3 krijgen de maatschappelijke context en het onderwijsbeleid een deskundige

behandeling (A.M.P.Knoers). Verschillende trends worden getypeerd die het gevolg zijn van modernisering zoals individualisering, differentiatie en technologisering van het onderwijs.

De leraar staat centraal in hoofdstuk 4 (N.Verloop). Rond het begrip 'professionaliteit' wordt uiteengezet welke eisen nu en in de toekomst voor dit beroep gelden. Recente veranderingen in het denken over het leraarsberoep blijven niet onbesproken. Hierna volgen vier hoofdstukken over gebieden waarop de leraar actief kan zijn. In hoofdstuk 5 is dat het beïnvloeden van leerprocessen (L. Verschaffel). Basisprincipes voor het ontwerpen van krachtige leeromgevingen worden geschetst. In hoofdstuk 6 (W.J.Nijhof) behandelt de auteur de vraag hoe belangrijk een curriculum is voor de leraar. Het gaat met name om de rol van gebruiker en van ontwikkelaar. De keuze van didactische werkvormen en media wordt behandeld in hoofdstuk 7 (J.Lowyck). Naast traditionele audiovisuele leermiddelen komt daarin ook de computer aan de orde. Het belang van didactische evaluatie wordt in hoofdstuk 8 uiteengezet (N.Verloop, F. Van der Schoot). Een aantal situaties wordt beschreven waarin evaluatie een rol kan vervullen bij het nemen van beslissingen, zoals selectie, resultaatbepaling, diagnose en sturing van het onderwijsleerproces. Het slothoofdstuk (W.Th.J.G. Hoeben) bevat een goed gestructureerde analyse van kwaliteitszorg in het onderwijs. De basisgedachte van dit hoofdstuk is dat de onderwijskwaliteit het beste gediend is met een verbetering van het werk van de man of vrouw voor de klas. Integrale kwaliteitszorg betekent hier onderling afstemming van het werk van leraren, scholen, externe ondersteuning en overheid.

Evaluatie

Het boek bevat in ruime mate wat de ondertitel suggereert: een kennisbasis voor professionals. Het betreft een systematisch geordend boek, dat duidelijk vanuit de optiek van een leerboek voor a.s. leraren (voor het voortgezet onderwijs) is opgesteld. Bij elk hoofdstuk is een inleidend overzichtsschema geplaatst en ter afsluiting een samenvatting. Dat bevordert het bestuderen. De teksten zijn helder en goed leesbaar geschreven.

Op verschillend plaatsen illustreren of ondersteunen voorbeelden en 'excursies' de teksten. Een excursie is een uitweiding over meer complexe en specialistische zaken, die niet essentieel zijn voor de opbouw van het betoog. Er zijn verschillende uitstekend geslaagde excursies opgenomen.

De verschillen met "Beknopte didaxologie" zijn vergaand. In de meeste gevallen vind ik dit nieuwe boek een vooruitgang. Ik blijf met de vraag zitten of dit nu een adequate keuze is uit het kennisgebied van de onderwijskunde voor aanstaande leraren en tegelijkertijd voor aanstaande onderwijskundigen. Het maakt een voornamelijk theoretische indruk. De verbinding van de inhoud met de alledaagse praktijk van het leren en onderwijzen zal niet eenvoudig tot stand te brengen zijn.

Het boek wil een baken zijn, maar de vraag is of leraren op grond van de inhoud een eigen koers kunnen bepalen. Daarbij ervan uitgaande dat hun handelen voor een deel gebaseerd kan zijn op de inhoud van dit boek. Er lijkt mij na het passeren van dit baken nog een hele weg te gaan. Ik neem me wel voor het boek te gaan gebruiken in de inleidende kolleges voor eerstejaars studenten onderwijskunde.

Terecht staat in dit boek de leraar centraal, maar deze professional wordt wel gezien in relatie tot de school als geheel. Die school op haar beurt krijgt haar opgave en mogelijkheden in een maatschappelijke en onderwijspolitieke context. Het is een verdienste van deze uitgave dat die relatie als de crux van het geheel gezien wordt.

Als achtergrond informatie voor het complexe werk van leraren in scholen is het een geslaagde uitgave. Het is door ter zake deskundige mensen geschreven. De samenhang is, ondanks het grote aantal auteurs, duidelijk in het boek nagestreefd en gerealiseerd. Dat is een grote verdienste op naam van de redacteurs.

Een detail: een beetje vreemd vond ik het schema van Dwyer uit 1967 over het nut van soorten visualiseringen aan te treffen. Dat komt ook al voor in "Beknopte". Het is nu tijd ook dit materiaal te vervangen door meer recente voorbeelden, die beslist wel beschikbaar zijn.

Een interessante kwestie, die ook in "Beknopte" alleen maar is aangestipt is de vraag hoe dit boek zich nu verhoudt tot de vakdidactiek. Naar mijn waarneming is deze relatie met vakdidactiek nog steeds onhelder. Zullen vakdidactici dit werk als basis erkennen waarop zij kunnen voortbouwen?

N.A.J. Lagerweij,
Vakgroep Onderwijskunde Universiteit Utrecht
Heidelberglaan 2, 3584 CS Utrecht

Gonny Schellings

*An educational approach to selecting main points in texts
Motives, methods, tasks and individual differences*
Amsterdam: s.n., 1995, IX-123 pp.

Een belangrijke stap die aan het verwerken van o.m. tekstuele informatie voorafgaat is het selecteren. Het belang en de impact op het verdere verwerkingsproces van het selecteren is reeds uitvoerig onderzocht. Veelal wordt in dit onderzoek een grote variatie in geselecteerde informatie-elementen vastgesteld. Met andere woorden, mensen verschillen onderling in wat ze meer en minder belangrijk achten. Deze grote variatie kan in onderwijssituaties problematisch zijn. Herhaaldelijk werd vastgesteld dat docenten onderling verschillen voor wat ze belangrijk vinden en als kerngedeelte in een tekst aanduiden. Docenten wijken bovendien in hun selectie geregeld af van wat op basis van een analyse van de tekststructuur als belangrijk naar voren komt. Leerlingen of studenten kunnen zich dan ook niet altijd op de tekststructuur baseren om te achterhalen wat de docent belangrijk acht. Ook tussen leerlingen of studenten worden grote onderlinge verschillen aangetroffen. Dergelijke variaties blijken samen te hangen met onder meer verschillen qua voorkennis en interesse.

In het onderwijskundig onderzoek naar het selecteren van kerngedeelten neemt het door B. Van Hout-Wolters geïntroduceerde onderscheid tussen een linguïstische, een cognitief-psychologische en een onderwijskundige benaderingswijze bij het selecteren een voorname plaats in. Dit onderscheid vormt tevens het conceptueel raamwerk van het proefschrift van G. Schellings waarin deze als een diamantslijper die telkens een nieuw facet van de ruwe steen tot leven brengt, op zeer systematische wijze en grondig gedocumenteerd drie onderzoekingen naar het selecteren van kerngedeelten rapporteert.

In een eerste onderzoek werd de overeenkomst onderzocht tussen de door leerlingen en de door hun docent als belangrijk gemarkeerde tekstfragmenten. Net als in vroeger onderzoek werd een grote diversiteit in gemarkeerde kerngedeelten aangetroffen. De selectie door de leerlingen vertoont meer of minder overeenkomst met de door hun docent geselecteerde gedeelten. Nieuw in dit onderzoek is dat aan de leerlingen eveneens werd gevraagd hun keuze te beargumenteren. Deze keuzen worden telkens gecategoriseerd binnen een van de drie selectiebenaderingen. Vastgesteld werd dat leerlingen wiens selectie een sterke overeenkomst vertoont met de selectie van hun docent, frequenter onderwijskundige redenen vermelden om hun selectie te beargumenteren. Het sterker profileren van de leertaak zou dan ook kunnen bijdragen tot een grotere overeenkomst tussen de selectie van leerlingen en die van hun docent. Aan deze implicatie hangt evenwel de in het proefschrift niet behandelde conditie vast dat er een overeenkomst is tussen de door docenten naar voren gebrachte leerdoelen en de door hen uitgevoerde selectie.

Het tweede onderzoek is meer onderzoekstechnisch van aard. Nagegaan werd wat de invloed is van verschillende selectiemethoden (vrij en gelimiteerd onderstrepen, beoordelen van be-

lang, en maken van aantekeningen) op de aard en de mate van geselecteerde delen. In alle condities werd opnieuw veel variatie in geselecteerde delen vastgesteld. De selectiemethode beïnvloedt wel de hoeveelheid geselecteerde delen. Zo voeren leerlingen die aantekeningen maken een verdergaande selectie door dan leerlingen die vrij mogen onderstrepen of het belang beoordelen. Deze reductie van gemarkeerde kerngedeelten heeft evenwel niet tot gevolg dat een bepaald type tekstgedeelten meer of minder wordt geselecteerd. Deze onderzoeksuitkomst vormt een grote geruststelling omdat in het onderzoek naar het selecteren van kerngedeelten omwille van praktische redenen vaak het markeren als selectiemethode werd gehanteerd. Hier werd vastgesteld dat de onderzoeksresultaten geen artefact zijn van de selectiemethode.

Het derde onderzoek vormt de kern van het proefschrift. Het wordt in twee afzonderlijke hoofdstukken besproken. Vooreerst werd nagegaan of selectietaken de aard en de mate van geselecteerde tekstfragmenten beïnvloeden. Hiertoe werd vooreerst elk gedeelte van de teksten als onderwijskundig, linguïstisch dan wel als interessant gecategoriseerd. Voor de drie verschillende biologieteksten werd aan de leerlingen werd gevraagd de tekstgedeelten te markeren die (a) door een docent interessant vinden, (b) de auteur belangrijk acht en (c) de leerlingen zelf het meest interessant vinden. Elke taak resulteerde weerom in een grote variëteit aan geselecteerde kerngedeelten. Wel bleken als 'onderwijskundig' gecategoriseerde tekstgedeelten het meest te worden geselecteerd in de conditie met een onderwijskundige taak en de linguïstische het meest in de conditie met een linguïstische taak. In de interesseconditie werden alle soorten tekstfragmenten geselecteerd. Schellings concludeert dat veel leerlingen heel wat moeilijkheden ervaren hun selectie op de precieze taak af te stemmen. De geringe taak-specificiteit van het selectie-gedrag van leerlingen doet heel wat vragen rijzen. Enerzijds blijken leerlingen onvoldoende in staat hun selectie op de taken af te stemmen, anderzijds kan ook worden beargumenteerd dat de taken zelf onvoldoende duidelijk werden geformuleerd of dat leerlingen geleerd hebben taakomschrijvingen niet te vertrouwen. Schellings gaat vooral in op de eerste mogelijke verklaring en pleit daarom, in navolging van J. Vermunt, voor een vorm van procesgerichte instructie. In vervolgonderzoek zou het zeker ook nuttig zijn de andere verklaring te toetsen. De taken zouden kunnen worden geëxpliciteerd door voorbeeldtestitems aan te reiken. Tevens zou hiermee de band gelegd zijn met het onderzoek naar ingebouwde vragen.

In het tweede deel van het onderzoek worden verschillen onderzocht tussen leerlingen die zich wel en die zich niet aan de verschillende leertaken aanpassen. Hiertoe werden vooreerst op basis van het door hen gehanteerde selectie-patroon vijf groepen geïdentificeerd. Twee groepen worden gekenmerkt doordat ze steeds weer hetzelfde (linguïstische of onderwijskundige) type informatie in de teksten selecteren. Een derde groep selecteert, ongeacht de taak, steeds weer zeer veel kerngedeelten en een vierde weinig kerngedeelten. De vijfde groep is een zogenaamde 'aanpasgroep', de door deze leerlingen gemaakte selectie is taakspecifiek. Vervolgens werd nagegaan of deze aanpasgroep verschilt van de andere groepen op een aantal leerlingkenmerken. Vastgesteld werd dat de aanpasgroep in vergelijking met de andere groepen (behalve de beperkende groep) minder kerngedeelten selecteert. De aanpasgroep verschilt ook van de andere groepen voor wat het rapportcijfer voor biologie betreft en van een aantal groepen voor wat gehanteerde cognitieve strategieën en leerconcepties betreft. Over het algemeen valt het evenwel moeilijk hierin een eenduidig patroon te herkennen dat mede als uitgangspunt voor het ontwerpen van een procesgerichte instructie zou kunnen worden genomen.

Ter afronding van het proefschrift wordt het nut vastgesteld van het door Van Hout-Wolters geïntroduceerde theoretisch kader in verband met de selectie van kerngedeelten. Wel wordt gepleit voor een andere benaming van de verschillende benaderingswijzen. Tevens wordt ingegaan op de nood aan procesgerichte instructie die leerlingen beter in staat zou moeten stel-

len meer taakrelevante selectie te maken. Het ontwerpen van een dergelijke procesgerichte instructie en met name het bepalen van wat het selectieproces juist inhoudt en vooronderstelt lijkt nog geen evidentie te zijn. Mede daarom wordt ook door de auteur gewezen op de belangrijke taak die docenten hebben in verband met het profileren van het belang van informatie-elementen in teksten.

Het proefschrift analyseert systematisch een aantal belangrijke vraagstukken in verband met het selecteren van kerngedeelten. Tevens reveleert het een aantal nog resterende problemen en roept het nieuwe vragen op. Zo is het onderzoek sterk gericht op schriftelijk materiaal en gaat het uit van een situatie waarin leerdoelen in de eerste plaats door een docent worden bepaald. De vraag dient gesteld hoe het selecteren van kerngedeelten verloopt wanneer de leerling een grotere verantwoordelijkheid heeft voor het bepalen van de leerdoelen. Meer bepaald zal voor meer open leeromgevingen moeten worden nagegaan of leerlingen in staat zijn informatie te selecteren die hen in staat stelt de door henzelf geformuleerde leerdoelen te realiseren. Daarnaast roept het proefschrift ook nieuwe vragen op inzake het theoretisch kader dat als uitgangspunt werd genomen. De naamswijziging voor de verschillende benaderingswijzen die Schellings voorstelt, suggereert reeds moeilijkheden met het conceptueel kader. Het onderzoek naar het selecteren van kerngedeelten kan winnen bij een verdere explicitering van het criterium dat wordt gehanteerd om de verschillende benaderingswijzen van elkaar te onderscheiden. Schellings stelt dat de selectiebron het indelingscriterium vormt. Op basis hiervan kan men descriptief vaststellen welke gedeelten meer of minder belangrijk geacht worden door een auteur, een docent of een leerling. Moeilijker wordt het wanneer door een onafhankelijke onderzoeker voor tekstgedeelten wordt bepaald welke gedeelten door deze drie groepen worden geselecteerd. Een dergelijk (predictief) gebruik van het conceptueel kader lijkt te impliceren dat de drie benaderingswijzen niet alleen qua selectiebron verschillen maar tevens procesmatig onderscheidbaar zijn. Twee assumpties die vervat zitten in het predictief gebruik van het conceptueel kader vereisen eveneens verder onderzoek. Het betreft de assumpties van correspondentie en stabiliteit. Een predictief gebruik van het conceptueel kader gaat er van uit dat auteurs er in slagen teksten zo te structureren dat een linguïstische analyse van de tekststructuur de voor de auteur belangrijkste gedeelten van de tekst reveleert. Er wordt tevens van uitgegaan dat wat een auteur belangrijk vindt tijdens het schrijfproces ongewijzigd blijft. Hetzelfde geldt voor de docent, ook hier wordt stabiliteit van leerdoelen verondersteld en een perfecte correspondentie tussen deze doelen en de geselecteerde kerngedeelten. Deze correspondentie is van groot belang in een onderwijscontext, omdat leerlingen vaak alleen over de doelen beschikken om kerngedeelten te selecteren. Verschillen tussen docenten in de geselecteerde kerngedeelten zijn evenwel een indicatie dat doelen heel verschillend worden geïnterpreteerd. Ook bij studenten wordt correspondentie verondersteld met name tussen wat ze interessant vinden en selecteren. Bovendien wordt er van uitgegaan dat wat iemand interessant vindt a priori kan worden bepaald. Het feit dat in het derde onderzoek juist in de interesse-taak een grote variëteit aan kerngedeelten wordt geselecteerd, doet vragen rijzen over deze mogelijkheid.

Het proefschrift van Schellings levert een belangrijke bijdrage tot het onderzoek naar kerngedeelten omdat het binnen een rijke onderzoekstraditie nieuwe inzichten verschaft over een aantal belangrijke onderzoeksvragen maar ook en vooral omdat het nieuwe vragen doet stellen die in vervolgonderzoek kunnen worden opgenomen.

*Jan Elen,
Centrum voor Instructiepsychologie en -technologie,
K.U. Leuven*

RECTIFICATIE

Noot van de redactie:

In het artikel van Severiens en Ten Dam, Verschillen in leerstijl, volume 21, nummer 4, p. 301 zijn in tabel 5 de verkeerde kopjes boven de kolommen geplaatst. Hieronder vindt u de gecorrigeerde tabel.

Tabel 5. Correlaties tussen de ILS-VWO en NSRV

	vrouwelijke eigenschappen	vrouwelijk gedrag	mannelijke eigenschappen	mannelijk gedrag
de betekenisgerichte leerstijl	-.04 (p=.71)	-.06 (p=.59)	.15 (p=.13)	.29 (p=.00)
de ongerichte leerstijl	.23 (p=.03)	.04 (p=.69)	-.10 (p=.31)	-.11 (p=.30)
de reproductiegerichte leerstijl	-.04 (p=.72)	-.12 (p=.23)	-.04 (p=.72)	-.12 (p=.23)
de certificaatgerichte leerstijl	.16 (p=.11)	.03 (p=.79)	.09 (p=.39)	.01 (p=.93)

Noot. In alle gevallen is de N 97; afwijkend van de totale N vanwege ontbrekende gegevens.

Inhoud

Curriculumorganisatie en studievoortgang

E.P.W.A. Jansen

Hoofdredacteur van *De Lerende Organisatie*, voorzitter van de ILS, lid van de CUROC

ABSTRACT

Curriculumorganisatie en studievoortgang / pag. 3

E. Jansen

Leerstijl in de onderbouw van het voortgezet onderwijs.

Een onderzoek naar de bruikbaarheid van de ILS / pag. 15

M. Boekaerts, R. Otten en R. Simons.

Constructie en evaluatie van een tekstbegripstoets voor de brugklas. Een instrument voor signaleren en typeren van leesproblemen / pag. 37

H. Hacquebord

Multi-level modellen voor de analyse van leerwinst vergeleken / pag. 54

H. van den Bergh en K. de Glopper

Televisie kijken en lezen.

Het effect van tv-kijken op de leesfrequentie en de leerprestaties in leerjaar 5 tot en met 8 van het basisonderwijs / pag. 76

M. Otter, R. Schoonen en K. de Klopper

Boekbesprekingen / pag. 96

Curriculumorganisatie en studievoortgang

E.P.W.A. Jansen¹

Rijksuniversiteit Groningen, onderzoeksinstituut GION, afdeling COWOG

ABSTRACT

The last decades the demand for university education in the Netherlands has increased and until some years ago each year more students attended university. But not all of these students graduate and the students who do, use more time than the four years of the official program. The policy of the Minister of Education aimed at an increase of completion rates and at a reduction of the time to graduate. The last decade in research as well as in policy on university education factors within the curriculum organization of university education have gained more attention. This is in contrast with the sixties and seventies when student-related factors were more prominent. This article focuses on the relation of the curriculum organization and the study success in the first year of university education. Study success is defined as the numerical return after the first year, that is a fraction with enrolling student as the numerator and graduated students (for the propedeutic examination) as the denominator. In this research project we investigated the study success of five cohorts of students (enrollment years 1987-1991) within six different departments at the University of Groningen. Aspects of the curriculum organization that contributed positively to study success were the spread of study load by spreading examinations and programming less subjects parallel, whereas resits should not be spread over the year.

INLEIDING

Sinds de invoering van de Wet Tweefasenstructuur Wetenschappelijk Onderwijs (1981) is het numerieke rendement van opleidingen sterk onder de aandacht gekomen. Deze wet kan gezien worden als een eerste grootschalige ingreep van de overheid in het kader van de efficiëntieverhoging van het Wetenschappelijk Onderwijs. In de jaren daarna zijn een reeks van maatregelen afgekondigd die de financiering van het Wetenschappelijk Onderwijs beheersbaar moet houden. Deze maatregelen liggen zowel op het terrein van de studenten, de studiefinanciering, als op het terrein van de bekostiging van de instellingen zoals regelingen waarbij instellingen afgerekend worden op het aantal behaalde diploma's en niet meer op het aantal studenten dat instroomt.

Zowel de overheid als de instellingen en niet in de laatste plaats de studenten zelf hebben er belang bij dat studenten niet onnodig vertraagd het studieprogramma doorlopen. Met het instellen van de commissie-Wijnen in 1992 en het beschikbaar stellen van gelden om de studeerbaarheid van programma's te verhogen in 1995 lijkt de overheid het uitgangspunt te huldigen dat er in het onderwijs manipuleerbare factoren te onderkennen zijn die van invloed zijn op de studievoortgang van studenten.

De vraag naar manipuleerbare factoren in het wetenschappelijk onderwijs die van invloed zijn op de studievoortgang van studenten staat in het onderzoek (Jansen, 1996) dat hierna beschreven wordt centraal.

FACTOREN VAN INVLOED OP STUDIEVOORTGANG

In de Verenigde Staten is een rijke onderzoekstraditie naar (studie)loopbanen in het Hoger Onderwijs (o.a. Spady, 1970; Tinto, 1975, 1987; Pascarella & Terenzini, 1991). Bekend is het model ter verklaring van studiestaking van Tinto (1975), dat ook in verschillende schoolloopbaanonderzoeken in Nederland is gebruikt (o.a. De Jong e.a., 1991). Een uitgebreid overzicht van de verschillende benaderingen en modellen uit de angelsaksische literatuur geeft Bijleveld (1993) in zijn proefschrift "Numeriek rendement en studiestaking". Voor ons onderzoek naar manipuleerbare factoren in het onderwijs die van invloed zijn op de studievoortgang van studenten zijn deze modellen minder geschikt. Ten eerste spelen factoren in de curriculumorganisatie een geringe rol in deze modellen. Variabelen zijn gemeten op studentniveau, zowel direct aan de student gekoppelde eigenschappen (bijvoorbeeld leeftijd of persoonskenmerken), omgevingsfactoren (bijvoorbeeld sociaal economische status, woon-situatie, financiering) als meningen van studenten over kenmerken van de opleiding of instelling (bijvoorbeeld waardering contact met docenten en medestudenten). Daar speelt nog bij dat de situatie binnen het hoger onderwijs in de Verenigde Staten vrij sterk verschilt van de situatie in het Nederlandse universitaire onderwijs.

Het onderzoek in Nederland naar factoren die van invloed zijn op studiesucces of studievoortgang hebben wij in drie categorieën onderverdeeld: onderzoek naar studentgebonden factoren, onderzoek naar programmeringsprincipes en onderzoek naar effectieve instructie. Het meeste onderzoek is uitgevoerd binnen één studierichting of binnen één cohort aan één onderwijsinstelling.

Studentgebonden factoren

In Nederland is vanaf de jaren zeventig veel onderzoek gedaan naar de relatie tussen persoonsgebonden variabelen en studiesucces of studiestaken. Baneke (1987) en Crombag e.a. (1975) vonden een zwak verband tussen prestatie-motivatie en studiesucces. Van andere persoonlijkheidskenmerken konden geen relaties met studiesucces worden aangetoond. Bijleveld (1993) constateert op grond van angelsaksisch onderzoek eveneens dat er niet zoiets bestaat als een 'drop-out' persoonlijkheid. Studiesucces kan volgens Lacante (1983) het best omschreven worden als een functie van intelligentie en inzet.

Algemene studievaardigheden en wijze van leerstofbestudering bleken eveneens geen relatie te vertonen met studiesucces (Crombag, 1975; Meerum Terwogt-Kouwenhoven, 1990). Inzet, planningsgedrag en tijdbesteding zijn daarentegen wel van invloed op studiesucces (Schouwenburg, 1994; Meerum Terwogt-Kouwenhoven, 1990; Van Os en Van Strien, 1975). Dat de tijd die studenten besteden aan hun studie een belangrijke factor is, behoeft geen betoog. Uit het onderwijsleermodel van Carroll (1963) was het belang van de effectief bestede studietijd reeds lang duidelijk. De vraag is echter op welke wijze studenten aangezet kunnen worden tot een optimaal gebruik van hun tijd. Carroll gaat ervan uit dat een optimale inrichting van de leeromgeving, het onderwijsprogramma of onderwijssysteem hiertoe bijdraagt onder twee randvoorwaarden, namelijk 'quality of instruction' en 'opportunity to learn'. Dit model ligt ook ten grondslag aan het model voor effectief onderwijs dat Creemers (1994) ontwikkeld heeft. Op grond van deze theorieën en onderzoeken kunnen we stellen dat maatregelen die ingrijpen op het planningsgedrag en tijdbesteding van studenten van belang zijn om de studievoortgang te bevorderen.

De laatste jaren wordt ook de selectiediscussie weer aangewakkerd. Uit verschillende onderzoeken blijkt dat het moeilijk is criteria te vinden waarop geselecteerd zou moeten worden. Van Dyck (1995) heeft er weinig vertrouwen in dat welke vorm van selectie dan ook meer talent doet overblijven. Van de instroomkenmerken van studenten blijkt het gemiddelde vooropleidingscijfer, het geslacht en de leeftijd nog enige samenhang te vertonen met studiesucces of studievoortgang (Van Overwalle, 1985; Roeleveld e.a. 1994; Jansen, 1995).

Effectiviteit van instructie

Onderwijs kan in ruime zin opgevat worden als de wijze waarop de opleiding is georganiseerd, maar ook in beperkte zin als het daadwerkelijke onderwijzen, de kennisoverdracht van de docent aan de studenten. Hiervoor gebruiken we in navolging van Jochems (1990) de term instructie. Het bevat de gehele cursusopzet waarbinnen de docent één van de elementen is. 'Quality of instruction' is volgens Carroll één van de randvoorwaarden voor de optimale inrichting van de leeromgeving. Goede instructie omvat vier functies, namelijk oriënteren, oefenen, controleren en terugkoppelen (Smuling e.a., 1990; Pilot e.a. 1992). Met name het belang van het mechanisme van controle en terugkoppeling voor het verbeteren van studieprestaties van studenten wordt door verschillende auteurs benadrukt (Buis, 1978; Mettes en Pilot, 1980; Been, 1981, 1992; Roossink, 1990).

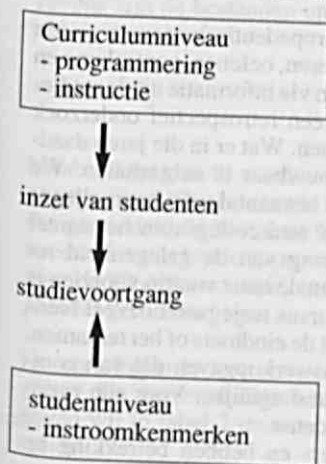
Programmering

De organisatie van het curriculum kan eveneens bijdragen aan het verhogen van de studieinzet van studenten en daarmee aan de studievoortgang. Onder inrichting of organisatie van het curriculum verstaan we onder andere de gehanteerde periode-indeling (bijvoorbeeld trimesters of blokonderwijs), aantal vakken dat geprogrammeerd is en tentamenspreiding. Crombag e.a. (1985) veronderstelden dat de Teaching-Learning Paradox van Dubin en Taveggia (1968) doorbroken kon worden door het werkgedrag van studenten als intermediaire variabele op te nemen bij het leggen van verbanden tussen onderwijzen en leren. Op grond van onderzoek concludeerden zij dat studenten hun werkgedrag aanpassen aan de inrichting van het curriculum. Dit is later in andere studies specifiek aangetoond door Van der Drift en Vos (1987). Studenten laten zich leiden door de nabijheid van tentamens. Onderwijs met weinig concurrentie tussen vakken en tussen vakken en tentamens zal leiden tot een optimale besteding van de studietijd. Blokonderwijs, dat wil zeggen onderwijs waarbij slechts één vak tegelijk geprogrammeerd is dat direct wordt afgesloten met een tentamen, leidt tot een verbeterde studievoortgang (Jansen, 1993; Joostens, 1990; Schoonen en Joostens, 1993, Vaughan en Carlson, 1992).

Model

Op grond van de hierboven beschreven bevindingen komen wij tot het volgende model ter verklaring van verschillen in studievoortgang van studenten die studeren in verschillende curriculumorganisaties in het wetenschappelijk onderwijs.

Figuur 1: Theoretisch model van kenmerken van invloed op studievoortgang



In het onderzoek gaan we op grond van dit model na welke aspecten van de curriculumorganisatie bijdragen aan de studievoortgang van studenten. De hypothesen met betrekking tot de programmering zijn:

1. Naarmate in een curriculum de toetsen meer gespreid zijn, zal de studievoortgang van studenten hoger zijn (minimalisering uitstelgedrag).
2. Naarmate in een curriculum minder vakken parallel geroosterd zijn, zal de studievoortgang van studenten hoger zijn (minder concurrentie tussen vakken).
3. Naarmate in een curriculum vaker in één week twee tentamens geroosterd zijn, zal de studievoortgang van studenten lager zijn (concurrentie tussen tentamens).
4. Naarmate in een curriculum de herkansingen meer verspreid zijn over het jaar zal de studievoortgang van studenten lager zijn (concurrentie met het reguliere onderwijsaanbod en eerste-kanstentamens).
5. Naarmate in een curriculum vaker herkansingen geprogrammeerd zijn in een week waarin ook een regulier tentamen geroosterd is, zal de studievoortgang van studenten lager zijn (concurrentie tentamen - herkansing).

Met betrekking tot de instructiekenmerken zijn de volgende hypothesen geformuleerd:

6. Naarmate in een curriculum/studieprogramma meer aandacht gegeven wordt aan het geven van terugkoppeling, zal de studievoortgang van studenten hoger zijn (stimuleren tot regelmatig studiegedrag).
7. Naarmate in een curriculum/studieprogramma meer gelegenheid gegeven wordt tot oefenen, zal de studievoortgang van studenten groter zijn (stimuleren tot actief studiegedrag).
8. Naarmate in een curriculum/studieprogramma meer aandacht gegeven wordt aan oriëntatie, zal de studievoortgang van studenten lager zijn (te veel aandacht voor oriëntatie leidt tot minder actief studiegedrag en uitstelgedrag).

METHODE

Het onderzoek is gebaseerd op de gegevens van de propedeuses van zes studierichtingen aan de Rijkuniversiteit Groningen. Van deze studierichtingen (bedrijfskunde, farmacie, geschiedenis, geneeskunde, pedagogische en andragogische wetenschappen (PAO) en psychologie) zijn van vijf cohorten (1987-1991) de studiegidsen, roosters, studentgegevens en studievoortgangsgegevens geanalyseerd.

Operationalisering

Studievoortgang is geoperationaliseerd als het behalen van het propedeutisch examen na één jaar. De instructiekenmerken volgen de onderwijsfuncties oriënteren, oefenen, controleren en feedback geven. De meting van de variabelen is tot stand gekomen via informatie uit de studiegidsen. Deze meting kan alleen maar globaal zijn, omdat het een retrospectief onderzoek betreft en we aangewezen zijn op de informatie uit de studiegidsen. Wat er in die jaren daadwerkelijk is gebeurd binnen het onderwijs is niet meer betrouwbaar te achterhalen. We hebben de instructiekenmerken beperkt tot vier maten, te weten het aantal uren hoorcolleges als operationalisering van de oriëntatiefunctie, het aantal uren werkcolleges en het aantal onderdelen met extra oefenmogelijkheden als operationalisering van de gelegenheid tot oefenen, het aantal feedbackmomenten als operationalisering van de mate waarin voorzien is in het geven van feedback. Voor deze laatste variabele is per cursus nagegaan hoeveel feedbackmomenten expliciet in de studiegids staan aangegeven naast de eindtoets of het tentamen. Het gaat hierbij bijvoorbeeld om tussentijdse toetsen, om huiswerkopgaven die van commentaar worden voorzien, tussentijdse verslagen of referaten en dergelijke. Voor alle variabelen geldt dat zij zijn gemeten als totaal over de gehele propedeuse.

De programmeringskenmerken zijn afgeleid uit de roosters en hebben betrekking op

de spreiding van vakken en tentamens. We hebben vijf variabelen: de spreiding van toetsen (aantal toetsperiodes), de spreiding van herkansingen (aantal herkansingsperiodes), het gemiddelde aantal vakken dat parallel is geroosterd, het aantal malen dat er twee tentamens in één week geprogrammeerd zijn en het aantal malen dat er een herkansing in een week met een regulier tentamen is geprogrammeerd. De spreiding van toetsen is bepaald als het aantal periodes waarin toetsen geprogrammeerd zijn. Een blokprogrammering met één tentamen na afloop van elke cursus krijgt zo de hoogste score. Telt een dergelijk studieprogramma acht cursussen, dan zijn er acht onderscheiden toetsmomenten en krijgt de variabele 'spreiding toetsing' een acht. In een trimestersysteem worden de tentamens na afloop van elk trimester in een beperkte tijdsperiode afgenomen. In de roostering zien we dan duidelijk drie tentamenperiodes. Op de variabele spreiding toetsing heeft een studierichting met een dergelijke tentamenroostering een drie. De operationalisering van de variabele spreiding herkansingen is hieraan analoog.

Als instroomkenmerken zijn het gemiddeld eindexamencijfer, het geslacht en de leeftijd van de studenten als controlevariabelen in de analyses opgenomen.

Analysemodel

Uitgangspunt voor de analyse is het model zoals weergegeven in figuur 1. In dit model zien we de verschillende niveaus waarbij het niveau van de studierichting niet expliciet is opgenomen, om het figuur leesbaar te houden. Inzet van studenten beschouwen we als een black box. In onze analyse gaan we na welke effecten de instroomkenmerken en de instructie- en programmeringskenmerken hebben op de studievoortgang van studenten. Dit tweede niveau noemen we het curriculumniveau. Het bestaat uit gegevens van cohort * studierichting. Studierichting is in ons model als derde niveau opgenomen, echter zonder variabelen op dat niveau. In de analyses maken we gebruik van het random coëfficiënt model (zie onder andere Kreft en De Leeuw, 1987; Bosker en Snijders, 1990) om de curriculumeffecten te schatten. Dit random coëfficiënt model wordt gebruikt voor data waarin meer nestingsniveaus aanwezig zijn. In ons geval gaat het om studenten die studeren binnen een curriculum binnen een studierichting. Het curriculumniveau is gekoppeld aan een cohort binnen een studierichting. De curriculumvariabelen zijn de variabelen waar we uitspraken over willen doen. De drie-niveau analyse is uitgevoerd met VARCL (Longford, 1988).

In deze bijdrage worden de analyses voor studenten met een VWO-diploma beschreven. De bestanden zijn opgeschoond voor 'no-show' studenten, dit zijn studenten die wel bij een studierichting staan ingeschreven maar aan geen enkel tentamen hebben deelgenomen en dus eigenlijk niet als serieuze student voor die studierichting kunnen worden aangemerkt. Tevens zijn de bestanden opgeschoond voor studenten met meer dan 840 uur vrijstellingen in het eerste jaar.

RESULTATEN

In tabel 1 (pagina 8) worden voor de VWO-groep de verdelingskenmerken van de variabelen weergegeven. Over de totale analysegroep (5151 studenten met een VWO-diploma) zien we dat gemiddeld 34% van de studenten de propedeuse binnen één jaar heeft behaald. Op het tweede niveau bij de instructie- en programmeringskenmerken zien we bijvoorbeeld dat er over de 30 cellen studierichting*cohort 4.9 toetsperiodes bestaan en gemiddeld 3.9 vakken parallel geroosterd zijn.

De verschillen in instructie- en programmeringskenmerken *tussen* de studierichtingen zijn groter dan de verschillen *binnen* één studierichting tussen de vijf cohorten. Als voorbeeld geven wij in tabel 2 de waarden op de variabele aantal toetsperiodes per studierichting en cohort.

Tabel 1: Verdelingskenmerken van variabelen voor de VWO-groep

	gemiddelde/ proportie	geobserveerde minimum/maximum
propedeuse in één jaar	.34	
studentvariabelen		
leeftijd	19.4	17 - 33
gemiddeld eindexamencijfer - 6	.78	-.25 - 3.20
geslacht (vrouw = 0)	.50	
spreidingsvariabelen		
spreiding toetsen	4.9	3 - 8
spreiding herkansingen	3.4	1 - 8
gemiddeld aantal vakken	3.9	1 - 6
twee tentamens in 1 week	2.4	0 - 7
herkansing in tentamenweek	1.1	0 - 5
instructievariabelen		
uren hoorcollege	275.6	162 - 516
uren werkcollege	155.6	51 - 392
extra oefening	4.8	1 - 8
feedbackmomenten	21.3	13 - 33

Tabel 2: Aantal toetsperioden per studierichting en cohort

cohort	87	88	89	90	91
studierichting					
bedrijfskunde	4	4	4	4	6
farmacie	8	8	8	8	8
geschiedenis	5	4	4	3	3
geneeskunde	3	5	5	5	5
PAO	7	8	6	6	6
psychologie	5	5	5	6	5

Bij het opbouwen van het drie-niveaumodel zijn de variabelen groepsgewijs ingevoerd. Eerst het lege model met alleen een intercept zonder predictoren, daarna de studentvariabelen, de programmeringsvariabelen en de instructievariabelen. De programmeringsvariabelen zijn eerst ingevoerd omdat de kwaliteit van die variabelen hoger is dan die van de instructievariabelen. Vanuit de roosters en studiegidsen konden wij vrij nauwkeurig vaststellen welke waarden de programmeringsvariabelen aannamen bij de verschillende studierichtingen en jaren. Dit in tegenstelling tot de instructievariabelen, waarbij de operationalisatie noodgedwongen globaal bleef, alleen gebaseerd op studiegidsen, die uiteraard een soort 'ideaalbeeld' laten zien van wat er in het studie-onderdeel gebeurt.

Propedeuse binnen één jaar

De resultaten van de VARCL-analyse op de variabele propedeuse behaald in één jaar zien we in tabel 3. Omdat de variabele dichotoom is, is binnen VARCL gekozen voor het logistisch regressiemodel.

Tabel 3: Resultaten van de logistische meerniveau-analyse afhankelijke variabele propedeuse binnen één jaar, 5151 VWO-studenten

Boven de streep geschatte effecten met standaardfout tussen haakjes; onder de streep enkele andere modelkenmerken

	leeg	+student	+program.	+instructie
intercept	-.412	-.493	-1.969	-1.670
leeftijd		-.035 _{ns}	-.039 _{ns}	-.039 _{ns}
		(.020)	(.021)	(.022)
geslacht (vrouw=0)		-.197	-.222	-.254
		(.065)	(.068)	(.074)
gemiddeld eindex.cijfer -6		1.086	1.231	1.375
		(.055)	(.058)	(.064)
spreiding toetsen			.410	.222 _{ns}
			(.104)	(.114)
spreiding herkansingen			-.228	-.238
			(.092)	(.052)
gemiddeld aantal vakken			-.031 _{ns}	-.333
			(.114)	(.114)
twee tentamens 1 week			.042 _{ns}	-.016 _{ns}
			(.048)	(.054)
herkansing in tentamenweek			.007 _{ns}	-.147 _{ns}
			(.081)	(.104)
uren hoorcollege				.003
				(.001)
uren werkcollege				-.005
				(.002)
aantal feedbackmomenten				-.026 _{ns}
				(.041)
aantal onderd. extra oef.				.444
				(.075)
variantiecomponent				
niveau 2	.177	.199	.088	.101
	(.078)	(.083)	(.066)	(.056)
niveau 3	.424	.533	.283	.000
	(.207)	(.229)	(.167)	(.084)
deviance	5949.65	5402.81	5232.03	5147.19
d.f.		3	5	4
p-waarde **		<.001	<.001	<.001

ns: effect niet significant ($\alpha = .05$, tweezijdig)

** toets steeds ten opzichte van het voorgaande model

De tabel is als volgt opgebouwd. In de kolommen staan de verschillende modellen. We beginnen met het lege model, dat is het model met alleen een intercept, zonder predictoren. In het volgende model worden de studentvariabelen erbij opgenomen. In het daarop volgende model nemen we de variabelen rond de spreiding van de studielast erbij op, enzovoort. Het toevoegen van de verschillende groepen variabelen laat steeds een betere passing van het model zien. In het onderste deel van de tabel zien we dat de deviance afneemt van 5949.65 in het

lege model tot 5147.19 in het meest uitgebreide model met student-, spreidings- en instructievariabelen. Kijken we naar dit meest uitgebreide model dan zien we op studentniveau dat geslacht en gemiddeld eindexamencijfer² significante effecten vertonen. Vrouwen en studenten met een hoger gemiddeld eindexamencijfer hebben meer kans de propedeuse binnen één jaar te behalen. Bij de programmeringskenmerken zien we significante effecten van de spreiding van de herkansingen en het aantal vakken dat parallel geroosterd is. Beide effecten zijn negatief, dat wil zeggen dat minder spreiding van de herkansingen en minder vakken parallel de kans om de propedeuse binnen één jaar te behalen verhogen. Opvallend is dat in het model zonder de instructievariabelen juist spreiding van de toetsen een positief effect laat zien in plaats van het gemiddeld aantal vakken parallel. Dit is mogelijk een multicollineariteitsprobleem gezien de hoge correlatie tussen beide variabelen (-.80). Tussen de overige curriculumvariabelen zijn de correlaties laag (variërend tussen de .01 en .40). Bij de instructievariabelen zien we significante effecten van uren hoorcolleges, uren werkcolleges en aantal onderdelen met extra oefenmogelijkheden.

Na invoering van de instructievariabelen is de variantie op het derde niveau tot nul gereduceerd. Op het tweede niveau is de variantiecomponent na invoering van de instructievariabele .10. De toename van de variantiecomponent op het tweede niveau na invoering van de predictorvariabelen is een verschijnsel dat vaker voorkomt bij hiërarchische lineaire modellen (zie Snijders en Bosker, 1994). De variantiecomponent op studentniveau is niet vermeld in de tabel omdat deze door het programma VARCL op 1 wordt gesteld. Dat is slechts een wijze van systematisering en geen geschatte parameter vanuit de data.

Om een indruk te geven van het verwachte effect op de slaagkans om de propedeuse binnen één jaar te behalen, is in tabel 4 per variabele in de verschillende modellen aangegeven wat hun bijdrage zou zijn aan de verhoging of verlaging van de slaagkans bij een 'gemiddelde' student. Dat is een student met een slaagkans van .34 binnen de groep als geheel. Verder is het uitgangspunt dat alle andere omstandigheden gelijk blijven. De verschillende effecten kunnen dan ook niet bij elkaar opgeteld worden. Dit komt voort uit de gebruikte statistische techniek. Bij de berekening van het effect van iedere predictor wordt ervan uitgegaan dat de waarden op de overige predictoren nul zijn.

Tabel 4: Resultaten in termen van geschatte toe- of afname van de slaagkans voor de propedeuse binnen één jaar, ceteris paribus, 5151 VWO-studenten

	+ student	+ program.	+ instructie
basiskans	.34	.34	.34
leeftijd	ns	ns	ns
geslacht (vrouw=0)	-.04	-.05	-.05
gemiddeld eindex.cijfer -6	+.26	+.30	+.33
spreiding toetsen		+.10	ns
spreiding herkansingen		-.05	-.05
gemiddeld aantal vakken		ns	-.07
twee tentamens 1 week		ns	ns
herkansing in tentamenweek		ns	ns
uren hoorcollege (per 100 uur)			+.07
uren werkcollege (per 100 uur)			-.10
aantal feedbackmomenten			ns
aantal onderd. extra oef.			+.11

ns: niet significante effecten

De tabel moet als volgt gelezen worden. In het hypothetische geval van een gemiddelde student binnen de totale groep met een slaagkans van .34, zou deze student indien deze een vol punt meer scoort op de variabele gemiddeld eindexamencijfer zijn/haar kans om te slagen voor de propedeuse binnen één jaar verhogen met 33%, alle andere omstandigheden gelijk verondersteld. Het effect van het gemiddelde eindexamencijfer is groot, maar we moeten er wel bij bedenken dat de variabele herberekend was. Eén punt meer op deze variabele betekent dat een student 7.78 als gemiddeld eindexamencijfer zou hebben. Daarmee behoort deze student bij de bovenste 10% van de onderzochte populatie. De variabele gemiddeld aantal vakken parallel geroosterd heeft de waarden 1 tot en met 6. Indien het gemiddeld aantal vakken met één verminderd wordt van bijvoorbeeld 4 naar 3 heeft dat een verwacht positief effect op de slaagkans van 7%, uiteraard ceteris paribus.

Spreading van de herkansingen en het gemiddeld aantal vakken dat parallel geroosterd is laten significante effecten zien. Deze effecten wijzen in de richting van de geformuleerde hypothesen. Van de instructiekenmerken zien we significante effecten bij de variabelen uren hoorcollege, uren werkcollege en aantal onderdelen met extra oefenmogelijkheden. De eerste twee effecten zijn tegengesteld aan de verwachtingen. Het laatste effect is wel conform de verwachting.

DISCUSSIE

In dit onderzoek hebben we de invloed van curriculumkenmerken op de studievoortgang van studenten onderzocht. In tegenstelling tot een aantal recente onderzoeken waarin slechts binnen één studierichting onderzoek is verricht (bijvoorbeeld Meerum Terwogt-Kouwenhoven, 1990; Cohen-Schotanus, 1994) of onderzoek waarin geen empirische ondersteuning is gezocht voor verklaringen van studievoortgang (bijvoorbeeld Bijleveld, 1993), hebben we de propedeuses van zes studierichtingen aan de Rijksuniversiteit Groningen vijf cohorten studenten in ons onderzoek betrokken. In aansluiting op eerdere bevindingen (Van der Drift en Vos, 1987; Jansen, 1992; 1993; Joostens en Schoonen, 1987) en in aansluiting op de verwachtingen vanuit het model van Caroll (1963) en Creemers (1994) blijken maatregelen die ingrijpen op het planningsgedrag van studenten en die aanzetten tot regelmatig studiegedrag, de studievoortgang van studenten te bevorderen. In dit onderzoek is nogmaals aangetoond dat minder vakken parallel roosteren bijdraagt aan een verhoging van de studievoortgang. In dit artikel is studievoortgang als dichotome variabele - wel of niet geslaagd voor de propedeuse binnen één jaar - opgenomen. De effecten van de spreidingsvariabelen zijn ook aangetoond voor de continue variabele 'aantal gerealiseerde studiebelastinguren na één jaar' (zie Jansen, 1996). Opvallend is het gegeven dat het beter is herkansingen minder te spreiden over het jaar. Over de plaats van de herkansingen wordt namelijk verschillend geredeneerd. Hertentamens zouden juist zo dicht mogelijk na de eerste tentamenmogelijkheid moeten plaatsvinden, omdat dan de reeds opgedane kennis nog niet te ver is weggezaakt. Herkansingen die geprogrammeerd worden in een periode dat het 'reguliere' onderwijs weer start, werken echter concurrentie tussen dat onderwijs en de voorbereiding voor het hertentamen in de hand, met gevolg dat de student één van beide of beide onderwijsactiviteiten niet goed kan voorbereiden. Onze onderzoeksresultaten lijken te wijzen op de laatstgenoemde redenering.

Hiermee zijn de hypothesen 2 en 4 ondersteund door het onderzoek. Hypothese 1 over de spreading van het aantal toetsen werd wel ondersteund in het model waarin alleen de programmeringskenmerken waren opgenomen. De verschuiving in het model met instructiekenmerken naar het gemiddeld aantal vakken parallel doet het vermoeden versterken dat deze twee kenmerken eenzelfde zijde van de medaille weergeven. Gezien de hoge correlatie tussen beide kenmerken lijkt dit een logische verklaring. De hypothesen over de concurrentie tussen tentamens, of tussen tentamens en herkansingen konden geen ondersteuning vinden. Dit zijn kenmerken die blijkbaar veel minder van belang zijn voor de (belemmering van) de studievoortgang van studenten.

De instructiekenmerken laten een minder eenduidig beeld zien. We zien een positief effect van het aantal onderdelen met extra oefenmogelijkheden. Dat duidt op een ondersteuning van hypothese 7. We hadden verwacht dat ook het aantal uren werkcolleges een bijdrage zou leveren aan de ondersteuning van deze hypothese. Het aantal uren werkcollege geeft juist een tegengesteld effect, evenals het aantal uren hoorcollege, dat gezien kan worden als een operationalisering van de aandacht voor de functie oriëntatie. De meest voor de hand liggende verklaring is de kwaliteit van de operationalisaties. De instructiekenmerken zijn operationaliseerd aan de hand van de informatie uit studieguiden. Hieraan kleefte het bezwaar dat onduidelijk is wat er precies binnen het onderwijs gebeurt is. Hoeveel hoor- en werkcolleges zijn daadwerkelijk door de studenten bezocht? Wat is er gedaan binnen een werkcollege, zowel door de studenten als door de docenten? Zijn alle feedbackmomenten zoals uit de studieguiden gedestilleerd ook effectief als feedbackmechanisme gebruikt? Daarnaast zijn er nog aanwijzingen (zie Jansen, 1996) dat er bij de instructiekenmerken sprake is van studierichtings-effecten.

In dit onderzoek is gefocussed op die kenmerken van het curriculum die door de studierichting te beïnvloeden zijn. Studentkenmerken, externe omstandigheden zoals bijvoorbeeld studiefinanciering en wettelijke maatregelen rond de beperking van de inschrijvingsduur zijn van grote invloed, waarschijnlijk zelfs van grotere invloed op het (werk)gedrag van studenten en daarmee op de studievoortgang. Dat was echter niet de insteek van dit onderzoek. Wij waren geïnteresseerd in de bijdrage van beïnvloedbare factoren in de curriculumorganisatie aan de studievoortgang van studenten. Het uitgangspunt daarbij was dat in deze tijd van beperkingen in studiefinanciering en inschrijvingsduur de instelling/studierichting de plicht heeft zorg te dragen voor een curriculum dat optimaal ingericht is om studenten in de hen toegekende tijd verantwoord door de studie te loodsen. Het onderzoek heeft over verschillende studierichtingen en cohorten studenten aangetoond dat in ieder geval in de propedeuse het spreiden van vakken en tentamens en het minder spreiden van herkansingen een positieve bijdrage kunnen leveren aan de studievoortgang van studenten. Een verbetering van de studeerbaarheid van studieprogramma's zal daarom in de eerste plaats in die richting gezocht moeten worden. Nader onderzoek naar instructiekenmerken en met name in het gebruik van de instructie door studenten en docenten, in relatie tot de studievoortgang van studenten is noodzakelijk.

LITERATUUR

- Baneke, J.J. (1987), *Studiesucces, persoonlijkheid en stress bij geneeskundestudenten vóór en na de invoering van de Tweefasestructuur*. (proefschrift) Amsterdam: Universiteit van Amsterdam.
- Been, P.H. (1981), *De invloed van aanleg en gebruik van studiehandleiding op tijdbesteding en studieresultaat*. Groningen: COWOG/Rijksuniversiteit Groningen.
- Been, P. (1992), Feedback. in: *ABC van minder docentafhankelijk onderwijs/uitgave t.g.v. 25 jaar COWOG*. Groningen: COWOG/Rijksuniversiteit.
- Bijleveld, R.J. (1993), *Numeriek rendement en studiestaking. Een theoretische analyse van factoren die samenhangen met rendement en studiestaking in het wetenschappelijk onderwijs*. Utrecht: Lemma.
- Bosker, R.J. & T.A.B. Sniijders (1990), Statistische aspecten van multiniveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 5, p. 317 - 329.
- Buis, P. (1978), *Het functioneren van terugkoppeling in het wetenschappelijk onderwijs: twee voorafgaande voorwaarden*. Lisse: Swets & Zeitlinger.
- Carrol, J.B. (1963), A model of school learning. *Teachers College Record*, 64, 8, p. 723 - 733.
- Cohen-Schotanus, J. (1994), *Effecten van curriculumveranderingen. Studiewaardering, studeergedrag, kennis en studiedoorstroom in een veranderend medisch curriculum*. (dissertatie) Groningen: Rijksuniversiteit Groningen.
- Creemers, B.P.M. (1994), *The effective classroom*. London: Cassell, School development series.
- Crombag, H.F.M., J.G. Gaff & T.M. Chang (1975), Study behavior and academic performance. *Tijdschrift voor Onderwijsresearch*, 1, 1, p. 3 - 14.

- Crombag, H.F.M., K.D.J.M. van der Drift & P. Vos (1985), De inrichting van curricula en het werkgedrag van studenten. *Universiteit en Hogeschool*, vol. 31, p. 234 - 247.
- Drift, K.D.J.M. van der & P. Vos (1987), *Anatomie van een leeromgeving. Een onderwijseconomische analyse van het universitaire onderwijs*. Lisse: Swets & Zeitlinger.
- Dubin, R. & T.C. Taveggia (1968), *The teaching-learning paradox: a comparative analysis of college teaching methods*. Eugene Oregon: University of Oregon Press.
- Dyck, M. van (1995), *Selectie voor Hoger Onderwijs*. Utrecht: Adviesraad voor het onderwijs.
- Jansen, E.P.W.A. (1992), Kan onderwijsprogrammering studiesnelheid beïnvloeden? *Onderzoek van Onderwijs*, 21, 4, p. 54-55.
- Jansen, E. (1993), Educational programming related to study-progress. In: J.K. Koppen en W.D. Webler (red.), *Strategies for increasing access and performance in higher education*. Amsterdam: Thesis publishers.
- Jansen, E. (1995), De kracht van de vernieuwde tweede fase VWO. *Onderzoek van Onderwijs*, 24, 1, p. 9-11.
- Jansen, E.P.W.A. (1996), *Curriculumorganisatie en studievoortgang. Een onderzoek onder zes studierichtingen aan de Rijksuniversiteit Groningen*. Groningen: GION -Studies over opvoeding en onderwijs; 1. Proefschrift Rijksuniversiteit Groningen.
- Jochems, W. (1990), *Productiever onderwijs*. Delftse Universitaire Pers.
- Jong, U. de, P. Koopman en J. Roeleveld (1991), *Snelwegen en slingerpaden in en om het Hoger Onderwijs*. Eindrapport project 'Studieloopbanen in het Hoger Onderwijs'. Den Haag: Ministerie Onderwijs en Wetenschappen: Achtergrond Studies Hoger Onderwijs en Wetenschappelijk Onderzoek, 10.
- Joostens, Th.H. (1990), De studie in de farmacie te Groningen. Resultaten na de invoering van blokkonderwijs. In: Jochems, W.M.G. e.a. (red.) *Onderwijsverbetering, voorbeelden uit het Wetenschappelijk Onderwijs*, p. 29 - 38. Delft: Delftse Universitaire Pers.
- Joostens, Th.H. & A.J.M. Schoonen (1987), Blokken helpt: de pharmacie-opleiding aan de R.U. Groningen. In: H.J.M. van Berkel, A.E. Bax, H.M.C. Schellekens (red.) *Differentiatie in het Hoger Onderwijs*. Amsterdam: Versluys.
- Kreft, G.G. & J. de Leeuw (1987), Het schatten van schooleffecten. Problemen bij de keuze van het model en de techniek. *Tijdschrift voor Onderwijsresearch*, 12, 2, p. 75 - 85.
- Lacante, M.F.E. (1983), Van intelligentie, persoonlijkheid, studiestrategie en studeergedrag naar studieresultaat. *Pedagogische Studiën*, 60, p. 289 - 299.
- Longford, N.T. (1988), *VARCL3 - Full covariance structure modelling*. Princeton, N.J., Educational Testing Service.
- Meerum Terwogt-Kouwenhoven, K. (1990), *Niet gewogen, toch te licht bevonden. Analyse van de rendementsproblematiek aan de Universiteit*. (proefschrift) Amsterdam. Kampen: Mondiss.
- Ministerie van Onderwijs en Wetenschappen (1981), *Beleidskader invoering tweefasenstructuur*. Den Haag: Ministerie van Onderwijs en Wetenschappen.
- Os, W. van & H.J.M. van Strien (1975), *Vallen en opstaan*. Amsterdam: Vrije Universiteit afdeling Onderwijsresearch.
- Overwalle, F. van (1985), Slagen en mislukken in de eerste kandidatuur: Op zoek naar oorzaken. *Tijdschrift voor Onderwijsresearch*, 10, 3, p. 107 - 129.
- Pascarella, E.T. & P.T. Terenzini (1991), *How college affects students. Findings and insight from twenty years of research*. San Francisco, Oxford: Jossey-Bass Publishers.
- Pilot, A., B. van Hout-Wolters en H. Kramer-Pals (1992), *Schriftelijk studiemateriaal*. Groningen: Wolters-Noordhoff.
- Roeleveld, J., U. de Jong, H. Oosterbeek en H.D. Webbink (1994), *Verder studeren, deel 3. Studievoortgang en onderwijskeuzen 1992*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Roosink, H. (1990), *Terugkoppelen in het natuurwetenschappelijk onderwijs, een model voor de docent*. Enschede: academisch proefschrift.
- Schoonen, B.J.M. & Th.H. Joostens (1993), Modulaire Curriculum Organization and Study Completion. The Case of Pharmacy at the University of Groningen. In: Joostens, Th.J.; G.W.H. Heijnen & A.J. Heevel (eds) *Doability of Curricula*. Lisse: Swets & Zeitlinger.
- Schouwenburg, H.C. (1994), *Uitstelgedrag bij studenten*. (proefschrift) Groningen: Rijksuniversiteit Groningen.
- Smuling, E.B., J. Brants & A. Pilot (1990), *Oriëntatie op leren en onderwijs*. Groningen: Wolters Noordhoff, Hoger Onderwijs Reeks.
- Snijders, T.A.B. en R.J. Bosker (1994), Modeled variance in two-level models. *Sociological Methods and research*, 22, p. 342-363.
- Spady, W. (1970), Drop-outs from higher education: an interdisciplinary review and synthesis. *Interchange*, 1, p. 64-85.

Staatsblad nr. 137 (1981), *Wet tweefasenstructuur wetenschappelijk onderwijs*.

Tinto, V. (1975), Drop-out from Higher Education: a theoretical synthesis of recent research. *Review of educational research*, 45, 1, p. 89-125.

Tinto, V. (1987), *Leaving College: rethinking causes and cures of student attrition*. Chicago: The university of Chicago.

Vaughan, Ch. & Chr. Carlson (1992), Teaching and learning. One-course-at-a-time. *Innovative Higher Education*, 16, 4, p. 263-276.

1. E.P.W.A. Jansen, GION/afdeling COWOG, Rijksuniversiteit Groningen, Postbus 335, 9700 AH Groningen, e-mail: e.p.w.a.jansen@ppsw.rug.nl

2. De variabele gemiddeld eindexamencijfer is herberekend door er zes punten van af te trekken. Hiermee moet rekening gehouden worden bij de berekening van de geschatte effecten van deze variabele. De variabele is herberekend zowel om te voorkomen dat er een al te hoge correlatie tussen intercept en helling ontstaat als om te voorkomen dat de geschatte random effects varianties kunstmatig laag lijken te zijn, omdat het gemiddelde van een variabele gekwadeerd terugkomt in zijn random effect variantie.

Ontvangen 11-11-96

Definitief aanvaard 16-1-97

Een onderzoek naar de bruikbaarheid van de ILS

Leerstijlen in de onderbouw van het voortgezet onderwijs

Monique Boekaerts, Roelie Otten en Robert-Jan Simons
Vakgroep Onderwijsstudies, Rijksuniversiteit Leiden
Vakgroep Onderwijskunde, Katholieke Universiteit Nijmegen

SAMENVATTING

Vermunt's Inventaris LeerStijlen (ILS) is een zelfrapportage instrument waarmee de leerstijl van studenten in het (open) hoger onderwijs kan worden gediagnosticeerd. Het begrip leerstijl wordt in de ILS opgevat als een karakteristieke combinatie van cognitieve verwerkingsactiviteiten en metacognitieve sturingsactiviteiten, die gebaseerd zijn op daarmee samenhangende studie-opvattingen (leermodellen) en studiemotieven (leeroriëntaties). Er worden vier leerstijlen onderscheiden: de ongerichte, reproductiegerichte, betekenisgerichte en toepassingsgerichte leerstijl. De vraag is in welke mate Vermunt's voor studenten ontwikkelde leerstijleninstrument bruikbaar is om de leerstijl van leerlingen in de onderbouw van het voortgezet onderwijs in kaart te brengen.

Secundaire analyses op data van een viertal onderzoeken wezen op een aantal problemen bij het gebruik van de ILS in de onderbouw van het voortgezet onderwijs, maar ook een aantal mogelijkheden om het instrument voor deze doelgroep aan te passen. De problemen zijn het grootst bij de leermodellen en leeroriëntaties. In tegenstelling tot Vermunt's resultaten bij universiteitsstudenten, is er bij leerlingen in de onderbouw van het voortgezet onderwijs geen relatie tussen, aan de ene kant, modellen met betrekking tot leren en leeroriëntaties en, aan de andere kant, verwerkings- en sturingsstrategieën. Van brede leerstijlen zoals die gevonden zijn bij volwassenen, is in de onderbouw van het voortgezet onderwijs geen sprake. Vermunt's verwerkings- en sturingsstrategieën zijn wel bruikbaar voor de onderbouw van het VO, maar dienen, onzes inziens, te worden aangepast aan de belevingswereld van leerlingen in dit type onderwijs. Hierbij is een aantal bevindingen van belang. In de onderbouw van het voortgezet onderwijs is er slechts sprake van drie leerstijlen; de toepassingsgerichte leerstijl is niet apart terug te vinden. De cognitieve strategie 'kritisch verwerken' is niet herkenbaar aanwezig. Bovendien is het onderscheid tussen interne en externe sturing niet eenduidig. Sturing door de docent hangt in de ogen van de leerlingen samen met interne sturing van het leerproces. Een laatste punt van aandacht is, dat de schaal stuurloosheid een duidelijker verband (in negatieve zin) heeft met schoolprestaties dan de overige schalen.

INLEIDING

In de huidige vernieuwingen in het onderwijs, de basisvorming en de recent ontwikkelde studiehuisaanpak staat 'leren-leren' centraal (ARO, 1994; WRR, 1986). Deze nieuwe visie op onderwijs houdt in dat de school leerlingen niet alleen vakinhoudelijke kennis en vaardigheden aanleert, maar ook tot taak heeft het zogenaamde leervermogen (Shuell, 1988), het autonoom leren (De Klerk, Lodewijks & Simons, 1981), of zelfregulerend leren (Boekaerts, 1997; Brown, 1987; Kluwe 1982, 1987; Simons, 1991) te ontwikkelen. De achterliggende gedachte hierbij is, dat de kwaliteit van de kennis en vaardigheden die leer-

lingen in het onderwijs verwerven, afhankelijk is van de (denk)activiteiten die de leerlingen bij het uitvoeren van leertaken verrichten (bijv. Wang, Haertel & Walberg, 1990). De lerende is een actieve deelnemer aan het proces van kennisverwerving. Alleen wanneer het onderwijs de leerlingen ook de vaardigheden leert om zich de aangereikte kennis daadwerkelijk 'eigen te maken', d.w.z. in hun individuele kennisbestand en gedragsrepertoire te integreren en zelfstandig te 'updaten', is het onderwijs effectief.

Verschillende auteurs wijzen erop, dat er bij zelfregulatie meer aan de orde is dan alleen metacognitieve vaardigheden. Volgens Corno (1986) dienen ook de affectieve componenten van het leerproces bewaakt en gestuurd te worden. Cross en Paris (1988) zijn van mening dat metacognitieve kennis verweven is met affectieve en motivationele (zelf)kennis. Zimmerman tenslotte stelt dat de leerling meer zelfregulerend is naarmate hij en metacognitief en gemotiveerd is en actief deelneemt aan zijn eigen leerproces. Leerlingen met initiatief, intrinsieke motivatie en persoonlijke verantwoordelijkheid, behalen meer academische successen dan leerlingen die deze kenmerken niet vertonen (Zimmerman, 1989; Zimmerman & Martinez-Pons, 1988). Boekaerts en Simons (1993) stellen dan ook dat zelfregulerend leren betekent dat: (1) leerlingen gekarakteriseerd worden door een actieve, op diepte-verwerking gerichte, cognitieve aanpak van de leerstof, (2) over meta-cognitieve kennis en vaardigheden beschikken om cognitieve strategieën adequaat en flexibel toe te passen (relatief onafhankelijk van de sturing van anderen en/of leerboeken), en (3) over motivationele gedragscontrole vaardigheden beschikken om in staat te zijn zelfstandig leerdoelen te formuleren en na te streven, relatief onafhankelijk van de motiverende, c.q. demotiverende invloed van anderen en/of specifieke omstandigheden. Zie voor meer informatie over dit motivationele aspect van zelfregulatie Boekaerts (1996).

Om de effectiviteit van het onderwijs op dit nieuwe taakgebied te kunnen evalueren en optimaliseren zijn meetinstrumenten nodig die het mogelijk maken de positie van leerlingen op het gebied van deze 'hogere-orde' leer- en denkvaardigheden in kaart te brengen en eventuele problemen te diagnostiseren. Op dit moment staat de Inventaris LeerStijlen (ILS) van Vermunt (1992) sterk in de belangstelling. Het aantrekkelijke van Vermunt's instrument is dat het belangrijke theoretische inzichten op het gebied van 'denk- en studievaardigheden' combineert en ordent. Veel instrumenten zijn slechts gericht op een beperkt aspect van leer- en studievaardigheden. Te denken valt aan vragenlijsten die zich met name richten op het opsporen van motivatieproblemen, zoals de Vragenlijst StudeerProblemen VSP (Hermans, 1977; Schouwenburg & Vugteveen, 1995), de Studeergedragvragenlijst (Rogier, 1986), en de Studie-Management en Academische Resultaten Test SMART (Klein, Topman & Van der Ploeg, 1994). Daarnaast zijn er vragenlijsten die wel een breder scala aan studievaardigheden bevatten, maar eerder gebaseerd zijn op een taakanalytische benadering dan op een theoretische visie op leren en instructie, zoals bijvoorbeeld de Vragenlijst Leren en Studeren VLS (Vorst, 1993). Vermunt's ILS, daarentegen, is gebaseerd op een theorie van leren en instructie waarin het begrip leerstijl een kernconstruct vormt. Van effectief onderwijs, of effectieve instructie is, aldus Vermunt (1992), sprake wanneer het onderwijs voldoende rekening houdt met de leerstijl van de lerende, zodat er sprake is van 'constructieve fricties' en niet van 'destructieve fricties'. Leren is meer dan kennis verzamelen; het is een proces van gedragsadaptatie, waarbij een 'optimale' discrepantie tussen huidig en gewenst gedrag een belangrijke voorwaarde is voor succes (vgl. Boekaerts, 1993). Anders gesteld: effectief onderwijs houdt in dat de docent zijn of haar sturende en begeleidende rol in het leerproces geleidelijk afbouwt, rekening houdend met de specifieke leerkenmerken van de leerling (Simons & Vermunt, 1986).

Vermunt beschrijft de kenmerken van de lerenden in termen van leerstijlen. In het algemeen worden leerstijlen opgevat als cognitieve vaardigheden of voorkeuren, bijvoorbeeld gekarakteriseerd als diepte- tegenover oppervlakte-aanpak (Marton & Säljö, 1984), serialistische tegenover holistische stijl (Pask, 1988), of een min of meer vaste voorkeur voor leren aan de hand van concrete ervaringen, het reflecteren op ervaringen, abstract

conceptualiseren, dan wel actief experimenteren (Kolb, 1984). Zich aansluitend bij de gangbare visie, beschouwt Vermunt (1992) "de neiging om uit eigen beweging van dezelfde leerstrategieën gebruik te maken in verschillende leersituaties." (p. 20) als een belangrijk aspect van het leerstijlbegrip. Tegelijkertijd breidt Vermunt de inhoud van het begrip leerstijl uit. Hij beargumenteert dat er sprake is van karakteristieke combinaties van cognitieve verwerkingsstrategieën en metacognitieve sturingsstrategieën, die gebaseerd zijn op specifieke studie-opvattingen (mentale leermodellen) en studiemotieven (leeroriëntaties). "Met een leerstijl wordt dan een samenhangend geheel bedoeld van leeractiviteiten die studenten gewoonlijk ontplooiën, hun leeroriëntatie en hun mentale leermodel, een geheel dat kenmerkend is voor hen in een bepaalde periode." (p. 21). Voor het eerst wordt de leerstijl opgevat als een combinatie van cognitieve strategieën, metacognitieve leermodellen en motivationele kenmerken van de lerende. Het opnemen van zowel affectieve als cognitieve aspecten in de leerstijl doet onzes inziens meer recht aan de complexiteit van leerprocessen, hoewel er bij het uitwerken van het affectieve aspect in studiemotieven nog wel de nodige vraagtekens zijn te plaatsen.

Vermunt maakt gebruik van een drietal niveaus om het leergedrag te beschrijven. Op het meest basale niveau worden specifieke cognitieve verwerkingsactiviteiten (relateren & structureren, kritisch verwerken, memoriseren & herhalen, analyseren en concrete verwerking), metacognitieve sturingsactiviteiten (interne sturing leerproces & leerresultaten, interne sturing leerinhoud, externe sturing leerproces, externe sturing leerresultaten, en stuurlaas leergedrag), opvattingen met betrekking tot leren en de taak van het onderwijs (opbouw van kennis, opname van kennis, gebruik van kennis, stimulerend onderwijs, en samen studeren), en studie-motieven oftewel leeroriëntaties (persoonlijk geïnteresseerd, certificaatgericht, testgericht, beroepsgericht, en ambivalent) onderscheiden. Dit niveau wordt door ons in het vervolg aangeduid met leerstrategieën. *Leerstrategieën* in een brede betekenis worden opgevat, als zowel cognitieve, affectieve als metacognitieve leeractiviteiten (zie bijv. Boekaerts & Simons, 1993).

Op het middenniveau worden de onderscheiden cognitieve verwerkingsactiviteiten gecombineerd en gekarakteriseerd in termen van een drietal verwerkingsstijlen: stapsgewijze, concrete en diepte-verwerking. De sturingsactiviteiten deelt Vermunt op in: sturing verricht door de leerling zelf (zelfsturing of interne sturing leerinhoud en leerproces), sturing door externen, de docent, het leerboek, het onderwijsprogramma (externe sturing), en stuurlaasheid. Wat betreft de leeroriëntaties en mentale leermodellen wordt er geen middenniveau onderscheiden. Het middenniveau verwijst onzes inziens naar de traditionele leerstijl, of zoals door Vermunt verwoord, naar de *smalle leerstijl*. Het combineren van leerstrategieën in cognitieve verwerkings- en metacognitieve sturingsstijlen staat centraal.

Op het hoogste niveau ten slotte, wordt het begrip *brede leerstijl* gepostuleerd. Het is de voorkeur voor een bepaalde wijze van leren die tot uiting komt, in zowel het leermodel en de leeroriëntaties die de lerenden voorstaan, als in de cognitieve en metacognitieve stijlen die hun leren kenmerken.

Vermunt vond, met behulp van principale componenten analyses, bij studenten in het wetenschappelijk onderwijs (Open universiteit en reguliere universiteit) zodanige samenhangen tussen de cognitieve verwerkingsstrategieën, metacognitieve strategieën, leeroriëntaties, en leermodellen, dat er inderdaad sprake is van duidelijke patronen in studiegedrag in de zin van de door hem veronderstelde brede leerstijlen. Er waren vier brede leerstijlen te onderscheiden: de reproductiegerichte, betekenisgerichte, toepassingsgerichte, en ongerichte leerstijl.

De drie niveaus worden in Figuur 1 op pagina 18 nader toegelicht.

LEERSTRATEGIËN

Verwerkingsstrategieën
relateren & structureren
kritisch verwerken
memoriseren & behalen
analyseren
concrete verwerking

Sturingsstrategieën
interne sturing leerproces & leerresultaten
interne sturing leerinhoud
externe sturing leerproces
externe sturing leerresultaten
stuurloos leergedrag

Mentale leermodellen
opbouw van kennis
gebruik van kennis
opname van kennis
stimulerend onderwijs
samen studeren

Leerorientaties
persoonlijk geïnteresseerd
certificaalgericht
testgericht
beroepsgericht
ambivalent

SMALLE LEERSTIJL

Verwerkingsstijlen
diepte-verwerking
stapsgewijze verwerking
concrete verwerking

Sturingsstijlen
interne sturing
externe sturing
stuurloosheid

opbouw van kennis
opname van kennis
samen studeren

persoonlijk geïnteresseerd
certificaalgericht
testgericht
beroepsgericht
ambivalent

BREDE LEERSTIJL

Betekenisgerichte
diepte-verwerking

stapsgewijze verwerking

concrete verwerking

interne sturing

externe sturing

stuurloosheid

opbouw van kennis

opname van kennis

stimulerend onderwijs
samen studeren

persoonlijk geïnteresseerd

certificaalgericht
testgericht

beroepsgericht

ambivalent

Toepassingsgerichte

Ongerichte

Vermunt's (1992) onderzoek bood verdere ondersteuning voor deze vier brede leerstijlen door de gevonden verschillen in de relatie met studieprestaties. De betekenisgerichte leerstrategie 'relateren en structureren' was in positieve zin gerelateerd aan studieresultaten in het universitair onderwijs. De verwerkingsstrategieën die tot de reproductiegerichte en toepassingsgerichte leerstijl horen, waren niet of nauwelijks gerelateerd aan studieresultaten in het universitair onderwijs. Terwijl een tweetal kenmerken van de ongerichte leerstijl, die ook wel als 'problematisch' wordt gekenschetst, negatief gerelateerd waren aan studieresultaten, te weten: 'stuurloosheid' en 'ambivalente leeroriëntatie'.

Vermunt's ILS is in meerdere onderzoeken op zijn bruikbaarheid getoetst. Replicatiestudies in het wetenschappelijk onderwijs resulteerden in nagenoeg dezelfde factoren met hun respectievelijke factorladingen (Busato, Prins, Hamaker & Visser, 1995; Schouwenburg, 1996). Vergelijkend onderzoek naar leerstijlen in het wetenschappelijk (WO) en hoger beroeps-onderwijs (HBO), leverde ook in het HBO sterk vergelijkbare factorpatronen op (Roosendaal, 1994; Vermunt, 1995). In het HBO bleken de reproductiegerichte en de toepassingsgerichte leerstijlen relatief vaak voor te komen, terwijl bij WO-studenten de betekenisgerichte en ongerichte leerstijlen de overhand hadden. In een onderzoek naar leerstijlen, studieprestaties en ervaren studiebelasting bij verschillende HBO-opleidingen, vond Weersink (1996) eveneens vergelijkbare factorpatronen. De schaal concrete verwerking laadde in tegenstelling tot de verwachting in haar onderzoek echter niet op de toepassingsgerichte factor. Zij typeerde deze factor daarom als beroepsgericht.

Naast replicatie van de factorpatronen zijn er ook een aantal onderzoeksresultaten die de validiteit van Vermunt's leerstijlen ondersteunen. Prins, Busato, Hamaker en Visser (1996), bijvoorbeeld, vonden (via hardop-denken) relaties tussen de op de ILS gerapporteerde leerstijl en het gedrag van eerstejaars psychologiestudenten bij het bestuderen van een tekst over sociale psychologie. Studenten met een reproductiegerichte leerstijl memoriseerden meer dan studenten met een ongerichte leerstijl. Studenten met een betekenisgerichte leerstijl reflecteerden meer op hun eigen leerproces dan studenten met een ongerichte leerstijl: ze constateerden vaker dat ze iets niet begrepen (onbegrip) en verwezen meer naar de door hen gevolgde werkwijze. Er werden geen verschillen gevonden met betrekking tot de activiteiten 'structureren en relateren' en 'kritisch verwerken'. De auteurs suggereren hiervoor een aantal mogelijke verklaringen waaronder de relatief kleine steekproef (27 studenten). Schouwenburg (1996) bestudeerde de relaties tussen Vermunt's leerstijlen en een aantal andere studievaardigheidsinstrumenten. Het betrof de Studie-Management en Academische Resultaten Test SMART (Klein, et al., 1994), de Vragenlijst StudeerProblemen VSP (Schouwenburg & Vugteveen, 1995), de Nederlandse vertaling van Lay's Procastination Scale LPS (Schouwenburg, 1994) die de neiging tot uitstellen meet, en negen zelfbeoordelingsschalen betrekking hebbend op cognitieve verwerking en studeergedrag, ontleend aan de Vragenlijst Leren & Studeren VLS (Vorst, 1993). In de principale componentenanalyse op de schaalcores van de verschillende instrumenten bleek dat Vermunt's leerstijlen betrekkelijk onafhankelijk zijn van time-management of omgevings-management-achtige variabelen en werkdiscipline. Wel werden er factoren gevonden die samenhangen tussen leerstijl, studeergedrag en motivationele oriëntatie tijdens het leren vertegenwoordigden. De betekenisgerichte leerstijl hing samen met op diepte-verwerking gericht studeergedrag, te weten 'argumenten analyseren' en 'verbanden leggen' uit de Vragenlijst Leren & Studeren. De reproductiegerichte en ongerichte leerstijl waren gecombineerd met faalangst en tentamencompetentie. De toepassingsgerichte leerstijl ten slotte hing samen met academische competentie en desinteresse in de studie. Volgens Schouwenburg betekenen deze resultaten dat er eigenlijk slechts sprake is van twee leerstijlen. De reproductiegerichte en toepassingsgerichte leerstijl verwijzen in zijn terminologie niet naar leerstijlen, maar respectievelijk naar emotionele stabiliteit en studiemotivatie. De kern van Vermunt's leerstijlconcept is echter juist dat de aard (kwaliteit) van de leeractiviteiten die de lerende tijdens het leerproces verricht zowel samenhangt met (meta)cognitieve als motivationele kenmerken van de lerende. Schouwenburg's resultaten vormen dus eerder een

bevestiging dan een weerlegging van Vermunt's leerstijlen. Vermetten, Vermunt en Lodewijks (1995) ten slotte, vonden effecten van studentgerichte instructie (expliciteren van leerdoelen, werkwijze, enz.) op de ILS-resultaten. Studentgerichte instructie bij eerstejaars rechtenstudenten resulteerde in het eerste semester in een sterkere gerichtheid op een 'kritische verwerking' van de stof en in het tweede semester in een afname van de ervaren 'stuurloosheid'. Bovendien toonden ze aan dat Vermunt's smalle leerstijlen een belangrijke stabiele persoonsgebonden component bevatten. Ze vergeleken de verwerkings- en sturingsstijlen van studenten op domeinspecifieke versies van de oorspronkelijk algemeen gestelde ILS. Er werd een hoge mate van overeenstemming (correlaties tussen de .60 en .70) gevonden in de verwerkings- en sturingsstrategieën die rechtenstudenten bij verschillende onderdelen van hun studie zeggen te gebruiken.

De bruikbaarheid van de ILS is eveneens onderzocht voor onderwijstypen die wat verder van het universitair onderwijs afliggen. Slaats, Van der Sanden en Lodewijks (1994) en Noij, Slaats en Lodewijks (1995) vonden bij leerlingen in het MBO alleen de ongerichte leerstijl en de reproductiegerichte leerstijl in zuivere vorm terug. De betekenisgerichte leerstijl hing, of samen met toepassingsgerichte activiteiten, te weten 'concrete verwerking', of samen met reproductieve activiteiten. In een vervolgonderzoek van Slaats, Van der Sanden en Lodewijks (1996) werden vergelijkbare resultaten gevonden. Zij benoemen de door hen gevonden leerstijlen respectievelijk als: passieve, reproductieve, constructieve, en veelzijdige leerstijl. Recentelijk is de ILS, na tekstuele aanpassingen, ook bij *brugklassers* in het voortgezet onderwijs gebruikt. Koerts (1995) gebruikte de ILS om de effecten van een nieuwe studievaardigheidsmethode (de HAVLOT) op het studiegedrag en de studieprestaties te onderzoeken. Hij vond de factorstructuur van Vermunt niet terug. In plaats daarvan concludeerde hij dat een driefactor oplossing de data het best ordende. Hij benoemde deze als de ondernemende, uitvoerende en stuurloze leerstijl. Klatter (1996) onderzocht de kwaliteiten van de ILS bij brugklassers in relatie tot een door haar ontwikkeld domeinspecifiek leerstijlinstrument voor tekstverklaren (de VILT). Ze vond dat de domeinspecifieke meting van verwerkingsstijlen in de VILT een betere voorspeller was voor leerprestaties dan de meer algemene verwerkingsstijlen (smalle leerstijlen) van de ILS. Opgemerkt dient echter wel te worden dat de VILT items bevat die sterke overeenkomst vertonen met de schoolse wijze waarop tekstverklaren wordt getoetst. Aan de hand van interviews stelde ze verder vast, dat brugklassers een groot aantal items uit de ILS niet goed begrijpen. Kortom, Vermunt's leerstijlen vertonen een hoge mate van generaliseerbaarheid in het hoger onderwijs. De resultaten in het middelbaar beroepsonderwijs en het voortgezet onderwijs zijn minder duidelijk.

PROBLEEMSTELLING

Vermunt's leerstijlconcept gaat ervan uit dat de lerenden de beschikking hebben over een scala aan cognitieve verwerkingsactiviteiten en dat zij in staat zijn te reflecteren op het eigen leren. Hij veronderstelt dat lerenden een visie op leren en op de taak van het onderwijs hebben ontwikkeld en op grond van deze visie leeractiviteiten kiezen. De vraag is of men deze brede leerstijlen ook bij leerlingen in het begin van het voortgezet onderwijs kan verwachten, temeer daar het karakter van het hoger en voortgezet onderwijs verschillend is. Het hoger onderwijs is gebaseerd op individuele keuze van studenten, terwijl er in het voortgezet onderwijs sprake is van leerplicht. Bovendien is er in het hoger onderwijs sprake van selectie in die zin dat, afgezien van dispensatie-regelingen, alleen leerlingen uit de hogere vormen van het voortgezet onderwijs toegang hebben tot deze vorm van vervolgonderwijs. Tot slot hebben studenten in het hoger onderwijs een langduriger ervaring met het volgen van (verschillende typen) onderwijs. Dit maakt dat aan studenten striktere eisen kunnen worden gesteld wat betreft de kwaliteit van hun studieactiviteiten en de zelfreflectie op hun studieactiviteiten, dan aan leerlingen in het voortgezet onderwijs. In het voortgezet onderwijs dient rekening te

worden gehouden met de beperkte capaciteiten van een deel van de leerlingen, met de taak van het onderwijs om leerlingen voor dit verplichte onderwijs te motiveren en met een beperkte onderwijservaring van een deel van de leerlingen.

In dit artikel wordt nagegaan in welke mate Vermunt's leerstijlconcept bruikbaar is in de onderbouw van het voortgezet onderwijs (VO). Deze vraag wordt nader gespecificeerd als:

- (1) Zijn de verschillende leerstijlschalen van Vermunt's ILS, d.w.z. de leerstrategieën en smalle leerstijlen, ook intern consistent en onderscheidbaar aanwezig in de onderbouw van het VO?
- (2) Zijn de onderlinge relaties tussen de leerstrategieën vergelijkbaar met de brede leerstijlen van volwassenen?
- (3) En, tot slot, zijn de gevonden relaties met prestaties in het hoger onderwijs ook aanwezig in de onderbouw van het VO?

METHODE

Data

Voor de analyses is gebruik gemaakt van data van een viertal onderzoeken die uitgevoerd zijn in het kader van het strategisch onderzoeksprogramma 'Motivatie en Zelfregulatie als determinanten van Onderwijseffecten' (zie voor een beschrijving Boekaerts & Simons, 1990, 1992). In deze onderzoeken werd de ILS in verschillende leerjaren van de onderbouw van het voortgezet onderwijs afgenomen. Er is gebruik gemaakt van de data van een longitudinaal onderzoek in de onderbouw van het VO (N=873), naar 'De invloed van de kwaliteit van instructie op de motivationele oriëntatie en de schoolcarrière van leerlingen' (Van Amelsvoort, in voorbereiding), data van een tweetal studies (N=196; N=242) uit het project 'Zelfregulatie vaardigheden en schoolsucces' (Kluvers & Simons, in voorbereiding), en data van het tweede leerjaar (N=712) van het project 'Motivationale oriëntatie en meta-motivatie als determinanten van schoolsucces' (Otten & Boekaerts, 1994). In Tabel 1 zijn in het kort de relevante kenmerken van de betreffende onderzoeken samengevat. Door het geheel of gedeeltelijk ontbreken van ILS-gegevens wijken de aantallen leerlingen in de hierna gepresenteerde analyses enigszins af van de hier genoemde aantallen.

Tabel 1: *Overzicht studies*

ILS	Leerjaar	Tijd	scholen	N	Karakterisering	Onderzoek
deels	1	Kerst April	1 AVO Zuid Nederland	196	pretest-posttest-control group design*	Simons-1
totaal	1, 2 en 3	Kerst	20 AVO-scholen Zuid-Oost Nederland	878	longitudinaal	Bergen
totaal	1, 2 en 3	Kerst	1 havo/ath/gym school	242	cross-sectioneel	Simons-2
deels	2	Kerst	10 AVO-scholen Randstad	712	eenmalige meting	Boekaerts

* Het betreft een training van studievaardigheden met behulp van de HAVLOT-methode in studielessen, al dan niet gecombineerd met een training tijdens de vaklessen en wel of niet gecombineerd met het trainen van metacognitieve vaardigheden.

ILS-bk: Inventaris Leerstijlen voor de brugklas

Vermunt's Inventaris LeerStijlen is door Van Rijswijk, Vermunt, De Jong en Kluvers aangepast voor leerlingen uit de onderbouw van het voortgezet onderwijs, en met name voor brugklasleerlingen (zie voor een beschrijving, Kluvers & Simons, 1992). De aanpassing is vooral een tekstuele aanpassing, de inhoud van de oorspronkelijke ILS-items is zo veel mogelijk

gehandhaafd. De eerste versie van de ILS-bk is in een proefafname bij 30 volwassenen (cursisten Centrum voor Vakopleiding) aangescherpt. Deze verbeterde versie is vervolgens in Bergen's longitudinale onderzoek (zie Tabel 1) bij 626 brugklasleerlingen gebruikt. Op grond van efficiëntie-overwegingen is rekening houdend met de psychometrische gegevens uit deze studie, ook een verkorte versie van 102 items samengesteld. In deze versie zijn in het algemeen 1 of 2 items per schaal geschrapt, met uitzondering van de mentale leermodellen. Hier is het aantal items per schaal systematisch met 3 verminderd. In de hierna te bespreken studies is zowel gebruik gemaakt van de versie met 120 items als van deze zogenaamde verkorte versie. Een overzicht van de items in de oorspronkelijke en verkorte versie wordt in Tabel 2 gegeven. In deze Tabel staat tevens aangegeven in welke onderzoeken de verkorte versie is gebruikt.

De ILS-bk meet een viertal aspecten, te weten: cognitieve verwerkingsstijl, sturingsstijl, mentale leermodellen en leeroriëntaties. De *verwerkingsstijl* staat voor de wijze van omgaan met de informatiestroom in het leerproces. Diepte-verwerking verwijst naar de mate waarin leerlingen relaties leggen met bijvoorbeeld voorkennis, de informatie structureren, en/of de leerstof kritisch verwerken. Stapsgewijze verwerking verwijst naar het afzonderlijk en achtereenvolgens verwerken en memoriseren van onderdelen van de leerstof. Concrete verwerking houdt in dat leerlingen proberen de stof te begrijpen door er zich concrete voorstellingen van te maken en zich af te vragen hoe ze de stof kunnen gebruiken. De *sturingsstijl* heeft betrekking op de manier waarop het leren wordt gereguleerd en gecontroleerd. Interne sturing verwijst naar sturing door de leerling zelf. Externe sturing duidt op sturing verricht door het onderwijs. Stuurloosheid verwijst naar het ontbreken van sturing. De mentale *leermodellen* verwijzen naar opvattingen ten aanzien van leren, zoals opbouw van kennis, gebruik van kennis en opname van kennis, ofwel naar opvattingen met betrekking tot het onderwijs, zoals samen studeren en stimulerend onderwijs. De *leeroriëntaties* hebben betrekking op de motieven die leerlingen kunnen hanteren om te leren. Boven genoemde onderdelen worden met behulp van voorbeelditems nader toegelicht. Hierbij wordt eerst de formulering van het oorspronkelijke ILS-item ter vergelijking weergegeven.

Verwerkingsstijlen:

diepteverwerking:

"Ik probeer onderwerpen die in een cursus afzonderlijk behandeld worden samen te brengen tot een geheel"

"Ik probeer onderwerpen die bij een vak apart worden behandeld, samen te brengen tot een geheel".

stapsgewijze verwerking:

"Ik werk een leereenheid punt voor punt af en bestudeer elk afgerond stukje apart."

"Als ik een hoofdstuk uit een boek moet leren, werk ik dat hoofdstuk punt voor punt af en leer elk stukje apart."

concrete verwerking:

"Ik gebruik wat ik door een cursus leer bij mijn activiteiten buiten de studie."

"Ik gebruik wat ik op school leer ook bij dingen die ik buiten school doe."

Sturingsstijlen:

interne sturing:

"Ik voeg uit andere bronnen iets aan de studiestof toe."

"Ik vul de leerstof zelf aan door ook andere boeken of tijdschriften te lezen."

externe sturing:

"Ik leer alles precies zoals het in het cursusmateriaal staat."

"Ik leer alles precies zoals het in het boek staat."

stuurloosheid:

"Ik constateer dat ik het moeilijk vind om vast te stellen of ik de cursusstof voldoende beheers."

"Ik vind het moeilijk om na te gaan of ik voldoende geleerd heb."

Mentale leermodellen:

opbouw van kennis:

"Om te leren is het belangrijk dat ik in eigen woorden samenvat wat wordt bedoeld in de cursusstof."

"Om te leren is het belangrijk dat ik de leerstof in eigen woorden samenvat."

gebruik van kennis:

"Leren betekent voor mij kennis verwerven die ik kan gebruiken bij mijn dagelijks functioneren."

"Leren betekent voor mij dat ik dingen te weten kom die ik ook buiten school kan gebruiken."

opname van kennis:

"Onder leren versta ik proberen het studiemateriaal dat ik voor me krijg te onthouden."

"Onder leren versta ik dat ik moet proberen om de leerstof die ik krijg te onthouden."

samen studeren:

"Ik vind het belangrijk om van andere studenten adviezen te krijgen over hoe ik mijn studie het best aan kan pakken."

"Ik vind het belangrijk om van andere leerlingen tips te krijgen over hoe ik het beste kan leren."

stimulerend onderwijs:

"Het cursusteam moet duidelijk uiteenzetten wat belangrijk en wat minder belangrijk voor me is om te weten."

"De leraar moet duidelijk aangeven wat voor mij belangrijk en minder belangrijk is om te weten."

Leeroriëntaties:

beroepsgericht:

"Ik heb dit leerstofgebied gekozen omdat het werk dat ik ermee kan doen mij enorm interesseert."

"Ik zit op dit schooltype omdat ik het nodig heb voor mijn vervolgopleiding of latere beroep."

persoonlijk geïnteresseerd:

"Ik doe deze studie uit pure belangstelling voor de onderwerpen die worden behandeld."

"Ik zit op dit schooltype omdat ik belangstelling heb voor de vakken die ik hier krijg."

testgericht:

"Ik wil voor mezelf bewijzen dat ik een studie in het hoger onderwijs aankan."

"Ik wil voor mezelf bewijzen dat ik het niveau van dit schooltype aankan."

certificaatgericht:

"Mijn belangrijkste doel bij deze studie is om certificaten te behalen."

"Mijn belangrijkste doel op school is om hoge rapportcijfers te halen"

ambivalent:

"Ik twijfel eraan of dit leerstofgebied wel de juiste richting voor mij is."

"Ik twijfel eraan of dit schooltype wel het beste voor mij is."

De items zijn geformuleerd in termen van stellingen of opvattingen. Het in meerdere of mindere mate onderschrijven van een bepaald item wordt met behulp van een vijfpunts likertschaal gemeten.

Statistische analyses

Allereerst wordt nagegaan of de in de ILS-bk onderscheiden constructen bij leerlingen uit de onderbouw van het VO, intern consistent zijn terug te vinden. Om deze vraag te beantwoorden worden de Cronbach's alphas voor de verschillende ILS-bk constructen berekend. Vervolgens wordt nagegaan of de constructen onderscheidbaar van elkaar zijn en of de con-

Tabel 2. Cronbach's alphas leerstrategieën en smalle leerstijlen ILS.

Verwerking	Aantal items	Simons1*		Brugglas		Tweede leerjaar		Derde leerjaar		Leerjaar 1-3		Vermunt	
		(N=149)	Simons1* Kerst	Simons1* april/mei	Bergen Kerst	Bergen Kerst	Bergen* Kerst	Bockaerts Kerst	Bergen* Kerst	Simons2	N=41	KUB	OU
			(N=149)	(N=163)	(N=626)	(N=873)	(N=691)	(N=633)	(N=33)	(N=47)	(N=795)	(N=654)	
Stapsgewijs	11(10)	.74	.81	.73	.74	.76	.76	.76	.81	.72	.78	.79	
Analyseren	6(5)	.61	.72	.60	.60	.63	.65	.65	.69	.67	.79	.67	
Memoriseren	5	.59	.70	.57	.60	.58	.64	.64	.70	.56	.63	.79	
Concreet	5	.67	.71	.65	.72	.70	.73	.73	.74	.75	.71	.74	
Diepte	11(10)	.76	.85	.78	.78	.80	.84	.84	.85	.85	.85	.83	
Relateren	7(6)	.72	.83	.75	.72	.78	.80	.80	.83	.81	.83	.80	
Kritisch verw	4	.44	.63	.46	.54	.57	.67	.67	.54	.74	.72	.72	
Regulatie													
Stuurloos	6(5)	.59	.69	.62	.59	.67	.66	.66	.68	.75	.72	.68	
Externe sturing	11(10)	.65	.59	.62	.60	.54	.60	.60	.61	.44	.68	.78	
Ext. proces	6(5)	.58	.58	.55	.44	.37	.46	.46	.50	.18	.48	.67	
Ext. Resultaat	5	.52	.38	.46	.46	.38	.44	.44	.32	.24	.65	.71	
Interne sturing	11(10)	.74	.82	.77	.77	.79	.80	.80	.84	.82	.79	.81	
Zelf inhoud	4	.56	.72	.59	.61	.63	.73	.73	.74	.77	.73	.78	
Zelf proces	7(6)	.67	.76	.72	.69	.74	.72	.72	.76	.76	.73	.75	
Leermodellen													
Opname kennis	9(6)	.76	.72	.66	.78	.70	.70	.70	.78	.71	.77	.78	
Opbouw kennis	9(6)	.74	.79	.72	.83	.77	.77	.77	.85	.85	.78	.77	
Gebruik kennis	6(5)	.81	.83	.76	.76	.80	.80	.80	.83	.83	.70	.76	
Stim. onderwijs	8(5)		.81	.81	.65	.78	.78	.78	.88	.83	.88	.90	
Samen studeren	8(5)		.79	.79	.77	.73	.73	.73	.89	.81	.89	.93	
Leerorientaties													
Ambivalent	5		.77	.77	.65	.78	.78	.78	.79	.83	.82	.75	
Testgericht	5		.67	.67	.63	.72	.72	.72	.49	.71	.68	.86	
Certificaat	5		.77	.77	.71	.74	.74	.74	.66	.73	.80	.81	
Beroepsgericht	5		.72	.72	.63	.68	.68	.68	.62	.72	.71	.85	
Pers. interesse	5		.57	.57	.57	.59	.59	.59	.15	.69	.57	.74	

* In deze onderzoeken is de verkorte versie van de ILS-bk gebruikt. In kolom 1 staat het aantal items voor deze versie tussen haakjes.

structen op de door Vermunt gevonden wijze bij volwassenen samen te voegen zijn tot brede leerstijlen. Om deze vragen te kunnen beantwoorden wordt gebruik gemaakt van principale componenten analyses (PCA's). Tot slot wordt de relatie met schoolprestaties nagegaan door de Pearson product-moment correlaties met rapportcijfers te berekenen. Bij analyses op item-niveau worden alleen die leerlingen meegenomen die op de betreffende onderdelen een volledige dataset hebben. Bij analyses op schaalscores worden ook de leerlingen betrokken die op één van de schaalitems missende gegevens hebben.

RESULTATEN

Zijn Vermunt's ILS-schalen ook in de onderbouw van het VO aanwezig?

De door Vermunt veronderstelde leerstrategieën en smalle leerstijlen worden met behulp van Cronbach's alpha op interne consistentie (homogeniteit) onderzocht. De resultaten voor de verschillende leerjaren van de onderbouw van het voortgezet onderwijs zijn samengevat in Tabel 2 op de pagina hiernaast. Waarden onder de .60 zijn vet afgedrukt. Wanneer waarden in de tabellen ontbreken, betekent dit dat het specifieke onderdeel van de ILS in het betreffende onderzoek niet is afgenomen. De resultaten van Vermunt bij KUB-studenten en OU-studenten zijn als vergelijkingsmateriaal toegevoegd. Zoals blijkt uit Tabel 2 zijn de Cronbach's alphas voor de smalle leerstijlen (middenniveau), behalve twee waarden voor 'stuurloos', in het algemeen voldoende (.60 of hoger). Bij een aantal leerstrategieën (basale niveau), is de interne consistentie in de antwoorden duidelijk onvoldoende. Vooral problematisch zijn de beide strategieën van externe sturing van het leerproces 'externe sturing leerproces' en 'externe sturing leerresultaat', de verwerkingsstrategie 'kritisch verwerken', en het studiemotief 'persoonlijke interesse'.

Met andere woorden, de resultaten wijzen erop dat de ILS constructen op het midden-niveau, de zogenaamde smalle leerstijlen intern consistent terug te vinden zijn bij leerlingen van de eerste drie jaren van het VO. Deze leerlingen hebben echter problemen met de meer genuanceerde onderscheidingen die op het basale niveau in de specifieke leerstrategieën tot uitdrukking komen. Opgemerkt wordt dat voor de sturingsstrategie 'externe sturing van het leerproces' en het studiemotief 'persoonlijke interesse' ook door Vermunt lage interne consistenties in de beantwoording worden gerapporteerd bij KUB-studenten.

Zijn Vermunt's leerstijlschalen onderscheidbaar in de onderbouw van het VO?

Per onderdeel van de ILS-bk (verwerking, sturing, leermodellen) zijn er PCA's met varimax rotatie verricht, om na te gaan of de leerlingen de leerstrategieën en smalle leerstijlen op de veronderstelde wijze van elkaar onderscheiden. De leeroriëntaties zijn in eerste instantie buiten beschouwing gelaten. Gezien de geringe aantallen leerlingen in de onderzoeken Simons-1 en Simons-2, werden voor deze analyses alleen de data van het onderzoek van Bergen (drie leerjaren) en Boekaerts (leerjaar 2) gebruikt. De resultaten voor de drie leerjaren zijn sterk vergelijkbaar. Daarom bespreken we alleen de resultaten van het eerste leerjaar.

De resultaten van de PCA op de items die betrekking hebben op de verwerkingsstrategieën staan weergegeven in Tabel 3. Geconstateerd kan worden dat de drie verwerkingsstrategieën, 'stapsgewijze verwerking', 'concrete verwerking' en 'diepteverwerking', bij leerlingen in het eerste leerjaar niet zonder meer worden teruggevonden. De eerste factor is een combinatie van concrete en diepteverwerkingsstrategieën waarbij de schaal 'concrete verwerking' de hoogste lading heeft. De schaal 'relateren' laadt ook, zij het minder sterk, op de andere twee factoren. De tweede factor representeert in hoofdzaak de schaal 'memoriseren', maar wordt aangevuld met twee items van de schaal 'analyseren', een aantal items van de schaal 'relateren' en één item van de schaal 'kritisch verwerken'. De derde factor kan worden gekarakteriseerd als een combinatie van de schalen 'analyseren' en 'relateren', waarbij 'analyseren' de hoogste ladingen heeft. Ook een enkel item van de schalen 'kritisch verwerken' en 'memori-

seren' laadt op deze factor. Met andere woorden, leerlingen van het eerste leerjaar maken geen onderscheid tussen diepte- en concrete verwerking van de leerstof. Bovendien is er geen eenduidige 'stapsgewijze verwerking' te onderkennen. Enerzijds, zien we dat memoriserende en analyserende strategieën samenhangen, maar anderzijds wordt ook de diepteverwerkingsstrategie 'relateren' hieraan gekoppeld.

Tabel 3: PCA (3 factoren, varimax-rotatie) verwerkingsstrategieën leerjaar 1

	F1	F2	F3
Analyseren			
1		.55	
17			.62
23			.56
40		.40	.42
45			.60
53			
Memoriseren			
2		.73	
7		.67	
9		.36	
26		.57	
33			.46
Concreet			
3	.57		
14	.61		
22	.57		
48	.57		
52	.62		
Relateren			
6		.30	.31
10	.37		.40
13	.33	.43	
19	.41	.35	
25	.51		
34	.40		.43
35	.44		.35
Kritisch			
29	.39		.42
39	.51		
43	.49		
49		.37	
Eigenwaarde	6.28	2.03	1.17
Verkl. variantie 35.1%	23.3	7.5	4.3

De resultaten van de PCA op de sturingsitems zijn weergegeven in Tabel 4. De eerste factor representeert overwegend 'interne sturing', de tweede factor 'stuurloosheid', en de derde factor 'externe sturing'. Echter, een vijftal externe sturingsitems heeft dubbele ladingen: ze laden zowel op de eerste als op de derde factor. Inspectie van de items toonde aan dat het bij vier van de vijf items gaat om 'het opvolgen van aanwijzingen van de docent' of 'het oefenen met opdrachten die door de docent zijn verstrekt'. Item 32, bijvoorbeeld luidt: "Ik leer volgens de aanwijzingen die de leraar geeft". Item 47 "Ik gebruik de aanwijzingen van de leraar om precies te weten hoe ik iets moet aanpakken", item 11 "De aanwijzingen van de docent helpen me bij het leren", en item 44 "Als er bij wiskunde een be-

paalde oplossingsmethode is behandeld, oefen ik daar veel mee". Met andere woorden, gebruik maken van de hulp die de docent biedt hangt in de ogen van leerlingen van het eerste leerjaar VO samen met interne sturing van het leerproces. De docent is voor deze leerlingen het model om hun leerproces te sturen.

Tabel 4: PCA Reguleringsstrategieën (3 factoren, varimax rotatie) leerjaar 1

	F1	F2	F3
Extern proces			
4	.33		
5			.35
18			.57
32	.39		.52
38			.39
47	.48		.49
Extern resultaat			
11	.43		.33
12			.44
30			.41
44	.53		
55			.41
Stuurloos			
8		.57	
15		.63	
20		.57	
27		.65	
37		.54	
41		.57	
Intern inhoudelijk			
16	.49		
28	.44		
42	.58		
54	.61		
Intern proces			
21	.47		
24	.55		
31	.52		
36	.55		
46	.58		
50	.60		
51	.62		
Eigen waarde	4.63	2.29	1.92
Verkl. variantie 31.6%	16.6%	8.2%	6.9%

De resultaten van de PCA met betrekking tot het leermodel dat leerlingen hanteren zijn samengevat in Tabel 5. In de opvattingen van deze leerlingen met betrekking tot leren zien we een duidelijke overlap tussen de drie veronderstelde leermodellen. De eerste factor combineert items die 'gebruik van kennis' en 'opbouw van kennis' vertegenwoordigen, de tweede factor combineert items die 'opname van kennis' en 'opbouw van kennis' representeren, en de derde factor vertegenwoordigt 'opname van kennis' op zich. De resultaten tonen duidelijke overeenkomsten met hetgeen bij de verwerkingsactiviteiten werd gevonden. Daar zagen we immers, een gecombineerde factor van 'concrete' en 'diepteverwerking', en het opsplitsen van de 'stapsgewijze verwerking' in een op memoriseren en een met diepteverwerking (relateren) gecombineerde variant.

Tabel 5: Factorladingen items mentale leermodellen in een 3-factoren (varimax) oplossing voor leerjaar 1

	F1	F2	F3
Opname			
82			.69
86			
94			.64
100		.42	
106		.42	
107			.73
11		.40	
113			.57
Gebruik			
81	.71		
90	.71		
95			
102	.70		
108	.66		
114	.65		
Opbouw			
85	.60		
88	.47		
92			
96		.61	
98	.44		
104		.70	
116			
117		.61	
119		.69	
Eigen waarde	5.37	1.93	1.58
Verkl. variantie 38.6%	23.4%	8.4%	6.9%

Zijn de door Vermunt in het hoger onderwijs gevonden brede leerstijlen terug te vinden bij leerlingen van de onderbouw VO?

Volgens Vermunt is het leren van volwassenen te typeren in termen van brede leerstijlen, die hij definieert als een samenhangend geheel van metacognitieve sturingsstrategieën en cognitieve verwerkingsstrategieën, die zijn afgestemd op de persoonlijke leermodellen en leeroriëntaties. Er werden vier leerstijlen onderscheiden: betekenisgerichte, reproductiegerichte, toepassingsgerichte en ongerichte leerstijl. Om na te gaan of deze leerstijlen ook bij leerlingen in de onderbouw van het voortgezet onderwijs worden teruggevonden zijn in navolging van Vermunt PCA's uitgevoerd op de schaalcores van de ILS-bk. Voor deze analyses is gebruikgemaakt van de data van het longitudinale onderzoek van Bergen, omdat dit het enige onderzoek is waarbij data beschikbaar zijn voor alle aspecten van de ILS-bk.

De resultaten voor de brugklas zijn weergegeven in Tabel 6. De eerste factor representeert een combinatie van de betekenisgerichte en concrete leerstijl. Deze factor representeert in hoofdzaak 'diepteverwerking' en 'interne sturing'. Wel moet worden opgemerkt dat de als stapsgewijs gekenschetste strategie 'analyseren' door deze leerlingen wordt gehanteerd. Deze leerlingen zien leren als 'opbouw' en 'gebruik van kennis'. De tweede factor representeert leeroriëntaties en mentale leermodellen. In de visie van brugklassers dienen studieopvattingen en studiemotieven dus blijkbaar van de verwerkings- en sturingsstrategieën te worden onderscheiden. De derde factor representeert een reproductiegerichte leerstijl. Deze leerlingen

gebruiken tijdens het leren met name de strategieën 'memoriseren' en 'analyseren'. Hun leerprocessen en leerresultaten worden voornamelijk 'extern' gestuurd. Ze zien leren vooral als 'opname van kennis'. De vierde factor representeert de ongerichte leerstijl. Leerlingen met overwegend deze leerstijl hebben geen duidelijke voorkeur voor een bepaalde verwerkingsstrategie. Sturingsactiviteiten worden niet door hen zelf geïnitieerd en ze maken ook onvoldoende gebruik van sturing die uitgaat van de instructie. Ze worden gekenmerkt door 'stuurloosheid'. Verder geven deze leerlingen aan, leren vooral te zien als 'opname van kennis', ze hechten veel belang aan 'stimulerend onderwijs' en 'samen studeren', en worden gekenmerkt door een 'ambivalente' leeroriëntatie.

Tabel 6: PCA leerstijlen (varimax, 4 factoren) leerjaar 1

ILS-schaal	F1	F2	F3	F4
Verwerkingsstrategie				
Diepteverwerking				
Relateren & structureren	.77			
Kritisch verwerken	.67			
Stapsgewijze verwerking				
Memoriseren			.67	
Analyseren	.53		.58	
Concrete verwerking	.77			
Regulatiestrategie				
Interne sturing				
Leerproces & -resultaten	.70		.38	
Leerinhoud	.73			
Externe sturing				
Leerproces			.75	
Leerresultaten			.58	
Stuurloos leergedrag				.68
Mentale leermodellen				
Opbouw van kennis	.51	.52		
Opname van kennis		.47		
Gebruik van kennis	.40	.63	.46	.45
Stimulerend onderwijs		.39		.68
Samen studeren				.56
Leeroriëntaties				
Persoonlijk geïnteresseerd		.62		
Certificaatgericht		.71		
Testgericht			.68	
Beroepsgericht		.66		
Ambivalent		-.30		.70
Eigenwaarde	6.12	2.16	1.70	1.27
Verkl var (tot 56.2%)	30.6%	10.8%	8.5%	6.3%

De analyses voor het tweede leerjaar zijn samengevat in Tabel 7. De eerste en tweede factor vertonen belangrijke overlap. Het overgrote deel van de leermodellen en leeroriëntaties laden tegelijkertijd op deze beide factoren. Echter, de leermodellen laden dominant op de eerste factor, terwijl de leeroriëntaties de hoogste ladingen hebben op de tweede factor. Ook in het tweede leerjaar onderscheiden leerlingen dus leermodellen en leeroriëntaties van verwerkings- en sturingsstrategieën. De leermodellen en leeroriëntatieschalen onderscheiden echter niet

tussen de verschillende leerstijlen. Opvallend is de lading van 'stuurloosheid' op de eerste factor. De derde factor kan worden gekarakteriseerd als een combinatie van de betekenisgerichte en concrete leerstijl. Leerlingen met deze leerstijl worden gekenmerkt door diepteverwerkingsstrategieën als 'relateren' en 'kritisch verwerken'. Ze sturen hun leerproces voornamelijk intern, zowel qua inhoud als qua resultaat. Verbazingwekkend is echter dat ze tevens worden gekenmerkt door 'stuurloosheid', dat wil zeggen problemen zeggen te hebben bij het sturen van het leerproces. De vierde factor is niet eenduidig in termen van een leerstijl te karakteriseren. Deze leerlingen hanteren de diepteverwerkingsstrategie 'relateren', in combinatie met de op reproductiegerichte verwerkingsstrategieën, 'memoriseren' en 'analyseren'. Hun leerproces wordt geleid door een combinatie van 'interne sturing' en 'externe sturing'. Ze hebben geen duidelijke visie op leren of specifiek kenmerkende leeroriëntaties.

Tabel 7: PCA leerstijlen (varimax, 4 factoren) tweede leerjaar

ILS-schaal	F1	F2	F3	F4
Verwerkingsstrategie				
Diepteverwerking				
Relateren & structureren			.63	.50
Kritisch verwerken			.70	
Stapsgewijze verwerking				
Memoriseren				.80
Analyseren				.73
Concrete verwerking			.73	.31
Regulatiestrategie				
Interne sturing				
Leerproces & -resultaten			.56	.54
Leerinhoud			.75	
Externe sturing				
Leerproces				.60
Leerresultaten				.51
Stuurloos leergedrag	.40		.57	
Mentale leermodellen				
Opname van kennis	.79	.37		
Opbouw van kennis	.81	.31		
Gebruik van kennis	.80	.33		
Stimulerend onderwijs	.84			
Samen studeren	.84	.30		
Leeroriëntaties				
Persoonlijk geïnteresseerd	.56	.57		
Certificaatgericht	.36	.73		
Testgericht	.37	.70		
Beroepsgericht		.77		
Ambivalent	.44	.72		
Eigenwaarde	8.73	2.71	1.27	1.09
Verkl. variantie (tot 69%)	43.7%	13.5%	6.4%	5.5%

De resultaten voor het derde leerjaar zijn samengevat in Tabel 8. De eerste factor is opnieuw een combinatie van de betekenisgerichte en toepassingsgerichte leerstijl, gekenmerkt door de specifieke verwerkings- en sturingsactiviteiten en de leermodellen, zoals beschreven in de voorgaande twee leerjaren. Nu echter wordt deze leerstijl ook gekenmerkt door de leeroriën-

tatie 'persoonlijke interesse' en in iets minder sterke mate door 'testgerichtheid'. De tweede factor representeert een reproductiegerichte leerstijl en de derde factor de ongerichte leerstijl die gekenmerkt wordt door een 'ambivalente leeroriëntatie'. De vierde factor representeert de leeroriëntaties, en in mindere mate de mentale leermodellen. Opnieuw zijn de studiemotieven en de mentale leermodellen relatief onafhankelijk van de verwerkings- en sturingsstrategieën. Het feit dat een aantal leeroriëntaties ladingen vertoont op factoren die verwerkingsactiviteiten en sturingsactiviteiten representeren lijkt echter een eerste aanzet tot een mogelijke koppeling.

Tabel 8: Resultaten PCA leerstijlen (varimax, 4 factoren) derde leerjaar

ILS-schaal	F1	F2	F3	F4
Verwerkingsstrategie				
Diepteverwerking				
Relateren & structureren	.79			
Kritisch verwerken	.71			
Stapsgewijze verwerking				
Memoriseren		.75		
Analyseren	.43	.67		
Concrete verwerking	.77			
Regulatiestrategie				
Interne sturing				
Leerproces & -resultaten	.71	.45		
Leerinhoud	.78			
Externe sturing				
Leerproces		.61		.31
Leerresultaten		.52		
Stuurloos leergedrag			.77	
Mentale leermodellen				
Opname van kennis		.51	.53	.38
Opbouw van kennis	.53	.44		.32
Gebruik van kennis	.57			.51
Stimulerend onderwijs			.70	.34
Samen studeren			.67	
Leeroriëntaties				
Persoonlijk geïnteresseerd	.54			.54
Certificaatgericht		.34		.64
Testgericht	.37			.63
Beroepsgericht				.74
Ambivalent			.72	
Eigen waarde	6.96	2.26	1.80	1.24
Verkl. variantie 61.2%	34.7%	11.3%	9.0%	6.2%

Relaties met schoolprestaties

In Tabel 9 zijn de relaties tussen Vermunt's leerstrategieën en schoolprestaties (rapportcijfers Kerst, Pasen, en Eind schooljaar) voor het eerste en tweede leerjaar voortgezet onderwijs weergegeven. In het eerste leerjaar zien we dat de schaal 'stuurloosheid' negatief samenhangt met rapportcijfers. Dit geldt eveneens voor de opvatting van leren in termen van 'opname van kennis'. In het tweede leerjaar, bij een beduidend grotere steekproef, worden ook een aantal (positieve) verbanden tussen verwerkingsstrategieën en rapportcijfers gevonden. De sterkte

van de verbanden is echter gering: nagenoeg alle correlaties zijn lager dan .20. Bij deze steekproef zijn ook de relaties van 'stuurloosheid' met rapportcijfers zwak. Gezien de geringe verbanden tussen de verschillende leerstijlschalen en de schoolprestaties van de leerlingen zijn geen verdere analyses, zoals bijvoorbeeld regressie-analyses, verricht.

Tabel 9: Product-moment correlaties (*) ILS met rapportcijfers

	Brugklas voormeting (N=162)**					Simons & Kluvers Brugklas nameting (N=173)**					Boekaerts & Otten Tweede leerjaar (N=708)**				
	Eng	Gesch	Nederl	Wisk	Totaal	Engels	Gesch	Nederl	Wisk	Totaal	Frans	Gesch	Nederl	Wisk	Totaal
Verwerking															
Stapsgewijs											.12				
											.11				
											.14				
Analyseren											.10				
											.10				
											.24				
Memoriseren											.12				
											.09				
											.11				
Concreet															
Diepte														.11	
											.12				
Relateren														.10	
											.10			.10	
											.10				
Kritisch verw															
															.11
Regulatie															
Stuurloos	-.30	-.26	-.29		-.37	-.34	-.31	-.33	-.29	-.44	-.19	-.18	-.11	-.10	-.21
						-.32	-.37	-.39	-.28	-.43	-.13	-.18	-.18		-.18
												-.10	-.10		-.10
Externe sturing															
Ext. proces															
Ext. Resultaat															
Interne sturing															
Zelf inhoud												.12			
Zelf proces															-.09
Opvatting leren															
Opname kennis						-.25	-.19	-.22	-.22	-.26					
						-.25	-.23	-.23	-.19	-.26					
Opbouw kennis															
Gebruik kennis							.20								
															.19

* correlaties significant op .01 niveau. Brugklasonderzoek $r > .14$ of $r < -.14$ significant bij $P < .05$. Onderzoek tweede leerjaar $r > .07$ of $r < -.07$ significant bij $P < .05$.

** De ILS-resultaten op de voormeting brugklas zijn gecorreleerd met de rapportcijfers voor Kerst. De ILS-resultaten van de nameting brugklas zijn gecorreleerd met de achterenvolgens, de rapportcijfers voor Pasen en het einde van het schooljaar. De ILS-resultaten in het tweede leerjaar zijn, achterenvolgens, gecorreleerd met de rapportcijfers voor Kerst, Pasen en het einde van het schooljaar.

DISCUSSIE EN CONCLUSIES

De resultaten wijzen op de nodige problemen bij het gebruik van Vermunt's ILS in de onderbouw van het VO. De problemen zijn het grootst bij de leermodellen en leeroriëntaties. In tegenstelling tot Vermunt's resultaten bij universiteitsstudenten en volwassenen is er bij leerlingen in de onderbouw van het VO geen relatie tussen, enerzijds, modellen met betrekking tot leren en leeroriëntaties en, anderzijds, verwerkings- en sturingsstrategieën. Van brede leerstijlen zoals die gevonden zijn bij volwassenen, is in de onderbouw van het VO geen sprake. Dit betekent echter niet dat de ILS daarmee ook zonder meer onbruikbaar is geworden. Zoals ook Dochy en Kirschner (1993) al opmerkten, ligt de kracht van de ILS waarschijnlijk meer in de uitwerking van de verschillende onderdelen, dan in de specifieke koppeling van leerstrategieën, mentale leermodellen, en leeroriëntaties. Bovendien is het mogelijk, dat gezien de relatief jonge leeftijd van onze doelgroep, de koppeling nog niet tot stand is gekomen.

Vermunt's verwerkings- en sturingsstrategieën zijn wel bruikbaar voor de onderbouw van het VO, maar dienen onzes inziens te worden aangepast aan de belevingswereld van leerlingen in dit type onderwijs. De kennis die deze leerlingen over verwerkings- en sturingsstrategieën hebben, is nog weinig gedifferentieerd. De ILS-schalen die betrekking hebben op de concrete strategieën worden onvoldoende intern consistent door de leerlingen beantwoord en onvoldoende van elkaar onderscheiden. Vermunt's ILS is in de onderbouw van het voortgezet onderwijs vooral bruikbaar op het niveau van de smalle leerstijlen, het middenniveau. Naast deze algemene bevinding zijn nog een aantal meer specifieke resultaten van belang. In de onderbouw van het VO is er slechts sprake van drie brede leerstijlen; de toepassingsgerichte leerstijl is niet apart terug te vinden. De cognitieve strategie 'kritisch verwerken' is nog niet herkenbaar aanwezig. Bovendien is het onderscheid tussen interne en externe sturing niet eenduidig. Sturing door de docent hangt in de ogen van de leerlingen samen met interne sturing van het leerproces. Een laatste punt van aandacht is, dat de schaal 'stuurloosheid' een duidelijker verband (in negatieve zin) heeft met schoolprestaties dan de overige schalen.

Belangrijk om in gedachten te houden is, dat het specifieke karakter van het voortgezet onderwijs een rol kan spelen bij bovengenoemde problemen. Vrij algemeen wordt aangenomen dat leerlingen metacognitieve kennis en vaardigheden verwerven door middel van reflectie op het eigen leren en dat van anderen. Flavell (1971) introduceerde hiervoor de term metacognitieve ervaringen. Dit zijn ervaringen die zich voordoen wanneer leerlingen vastlopen, constateren dat hun gebruikelijke aanpak niet functioneert, bepaalde problemen niet kunnen oplossen, e.d. Nadenken over de oorzaken van het achterwege blijven van leren of juist over de oorzaken van onverwacht goede prestaties kunnen, aldus Flavell, leiden tot nieuwe metacognitieve kennis en op grond hiervan kunnen in combinatie met oefening (met feedback) nieuwe metacognitieve vaardigheden tot stand komen. In de onderbouw van het voortgezet onderwijs doen dit soort situaties zich mogelijk slechts in beperkte mate voor. Bepaalde leerstrategieën worden in het voortgezet onderwijs minder verwacht. Klatter (1996) bijvoorbeeld, rapporteert de volgende reacties op het item "Ik probeer te bedenken wat de onderwerpen uit verschillende lessen van een vak met elkaar te maken hebben".... De ene leerling constateert: "....we gaan meestal van het ene onderwerp naar het andere, dus dan hoort alles wel bij elkaar (....); dan heeft het wel wat met elkaar te maken. Daar ga ik nooit over nadenken." De andere antwoordt met de opmerking: "Nee, doe ik niet, want meestal heb je in het boek wel een bepaalde volgorde dat je snapt waar het over gaat". Zij concludeert "Uit de antwoorden van deze overigens slimme leerlingen blijkt dat zij, gegeven de voorstructurering van het onderwijs in de brugklas, op dit terrein weinig relaterende verwerkingsactiviteiten (hoeven) verrichten" (p. 314). Bovendien zijn de leerstrategieën die studenten gebruiken niet zonder meer op dezelfde wijze bij brugklassers vertegenwoordigd. Bijvoorbeeld, de mening van het leerboek of de docent ter discussie stellen, lijkt in dit type onderwijs niet zo aan de orde. Het ligt meer voor de hand om bij deze doelgroep kritisch verwerken op te vatten in de

zin van op begrip-gericht leren, bijvoorbeeld je afvragen "Wat staat hier nu eigenlijk precies", "Begrijp ik eigenlijk wel wat er staat", enz. Tot slot is Klatter's bevinding dat brugklassers vooral 'domeingericht', d.w.z. schoolvakgericht over hun leerproces nadenken, van belang. Zij stelt: "Als zij al hun leergedrag overdenken, blijft dit beperkt tot domeinspecifieke taken waarover in concrete termen wordt nagedacht" (p. 303).

Op grond van deze resultaten concluderen we het volgende met betrekking tot het ontwikkelen van een leerstijleninstrument voor het voortgezet onderwijs. Het niveau van de smalle leerstijlen sluit het meest aan bij de belevingswereld van deze doelgroep. Een aantal specifieke schalen, te weten 'kritisch verwerken' en 'interne' versus 'externe' sturing, dient grondig te worden herzien. Het feit dat het voornamelijk de schaal 'stuurloosheid' is die met leerprestaties samenhangt, maakt duidelijk dat het opsporen van 'door de leerling geconstateerde moeilijkheden of problemen' erg belangrijk is. We stellen dan ook voor om naast het bevragen van de frequentie van gedrag ook het al dan niet ervaren van problemen bij het ten uitvoer brengen van bepaald gedrag te inventariseren. De resultaten van Klatter ten slotte, wijzen erop dat bij leerlingen in de onderbouw van het VO het in 'algemene termen' over leerstijlen denken nog nauwelijks ontwikkeld is. Het is daarom belangrijk leerstijlen te concretiseren naar specifieke domeinen van schoolsituaties, zoals bijvoorbeeld schoolvakken. In een volgende fase van onderzoek zou kunnen worden nagegaan in welke mate leerlingen in staat zijn te reflecteren op deze domeinspecifieke leerstijlen. Het lijkt belangrijk om na te gaan of leerlingen die verschillen in domeinspecifieke leerstijlen, ook weet hebben van deze verschillen en zich bewust zijn van de condities waaronder ze een bepaalde stijl van leren hanteren.

LITERATUUR

- Amelvoort, J. van. (in voorbereiding). Nog zonder titel. Academisch proefschrift, Katholieke Universiteit Nijmegen. UNILO.
- ARO. (1994). Ruimte voor leren. Utrecht: Adviesraad voor het onderwijs.
- Boekaerts, M. (1993). Being concerned with well-being and with learning. *Educational Psychologist*, 28, 149-167.
- Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist*, 1(2), 100-112.
- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by policy makers, educators, teachers and students. *Learning and Instruction*, 7(2), 161-186.
- Boekaerts, M., & Simons, P.R.J. (1990). Motivatie en zelfregulatie als determinanten van onderwijseffecten. SVO onderzoeksaanvraag. Nijmegen: Katholieke Universiteit Nijmegen, Vakgroep Onderwijskunde.
- Boekaerts, M., & Simons, P.R.J. (1992). Motivatie en zelfregulatie als determinanten van onderwijseffecten. SVO vervolgaanvraag. Nijmegen/Leiden: Katholieke Universiteit Nijmegen, Vakgroep Onderwijskunde/Rijksuniversiteit Leiden, Vakgroep Onderwijsstudies.
- Boekaerts, M., & Simons, P.R.J. (1993). Leren en instructie: Psychologie van de leerling en het leerproces. Assen: Dekker & Van de Vegt.
- Busato, V.V., Prins, F.J., Hamaker, C., & Visser, K. (1995). Leerstijlonderzoek gerepliceerd: De samenhang tussen leerstijlen en intelligentie. *Tijschrift voor Onderwijsresearch*, 20, 332-340.
- Corno, L. (1986). The metacognitive control components of self-regulated learning. *Contemporary Educational Psychology*, 11, 333-346.
- Cross, D.R., & Paris, S.G. (1988). Development and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology*, 80, 131-142.
- Dochy, F.J.R.C., & Kirschner, P.A. (1993). Procesgerichte instructie moet aansluiten bij leermodellen, leerstrategieën en voorkennis van studenten. *Tijschrift voor Hoger Onderwijs*, 1, 57-62.
- Flavell, J.H. (1971). First discussant's comments: What is memory development the development of? *Human Development*, 14, 421-435.
- Hermans, D. (1977). Analyse en validering van een studieproblemen vragenlijst. Ongepubliceerd rapport, Katholieke Universiteit Nijmegen, IOWO.
- Klatter, E.B. (1996). Studievaardigheden in de brugklas. Een vergelijking tussen algemene en vakspecifieke verwerkingsstijlen van brugklasleerlingen. *Pedagogische Studiën*, 73, 303-316.

- Klein, W.C., Topman, R.M., & Van der Ploeg, H.M. (1993). Cognities, studiegewoonten en academische prestatie: De ontwikkeling van de Studie-Management en Academische Resultaten Test (SMART). *Nederlands Tijdschrift voor de Psychologie*, 49, 231-232.
- Klein, W.C., Topman, R.M., & Van der Ploeg, H.M. (1994). Cognition, study habits, test anxiety, and academic performance. *Psychological Reports*, 75, 1219-1226.
- Klerk, de, L.F.W., Lodewijks, J.G.C.L., & Simons, P.R.J. (1981). Autonomoos leren: Ontwerp van een onderzoeksproject over de aard, ontwikkeling en functies van zelfregulatiesystemen bij leerlingen in het voortgezet onderwijs (Intern rapport). Tilburg: Katholieke Universiteit Brabant, Vakgroep Onderwijspsychologie en Opleidingen.
- Kluvers, C., & Simons, P.R.J. (1992). Zelfregulatievaardigheden en COO. Een onderzoeksverslag. Nijmegen: Katholieke Universiteit Nijmegen, Vakgroep Onderwijskunde.
- Kluvers, C., Simons, P.R.J., & Tijmensen, L. (in voorbereiding). Training van zelfregulatievaardigheden. Een onderzoeksverslag. Nijmegen: Katholieke Universiteit Nijmegen, Vakgroep Onderwijskunde.
- Kolb, D.A. (1984). The process of experiential learning. In D.A. Kolb (Ed.), *Experiential learning: Experience as the source of learning and development* (pp. 20-38). Englewood Cliffs, N.J.: Prentice Hall.
- Koerts, E.M. (1995). De ondernemende leerling. Een onderzoek naar het leren studeren door HAVLOT-schematiseren in het kader van zelfstandig leren. Academisch proefschrift, Katholieke Universiteit Brabant.
- Marton, F., & Säljö, R. (1984). Approaches to learning. In F. Marton, D. Hounsell & N. Entwisle (Eds.), *The experience of Learning* (pp. 36-55). Edinburgh: Scottish Academic Press.
- Noij, L.G., Riteco-Slaats, A., & Lodewijks, J.G.L.C. (1995). Inventarisatie van leerstijl in het beroepsonderwijs: De ILS-MBO toegepast. Een onderzoeksverslag. Nijmegen: Katholieke Universiteit Nijmegen, Vakgroep Onderwijskunde.
- Otten, R., & Boekaerts, M. (1994). Metamotivatie en schoolsucces in het voortgezet onderwijs. Motivatie en handelingscontrole als determinanten van inzet, prestaties en vakkenpakketkeuze. Een onderzoeksverslag. Leiden: Rijksuniversiteit Leiden, Vakgroep Onderwijsstudies.
- Pask, G. (1988). Learning Strategies, teaching strategies, and conceptual or learning style. In R.R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 83-100). New York: Plenum Press.
- Prins, F.J., Busato, V.V., Hamaker, C., & Visser, K.H. (1996). Een bijdrage aan het (meta-)cognitieve deel van de Inventaris Leerstijlen. *Pedagogische Studiën*, 73, 108-122.
- Rogier, A.M.L. (1986). Invloed studeergedrag VWO op studieresultaten (Rapport nr. 27). Eindhoven: Technische Hogeschool, Bureau Studentenpsychologen.
- Roosendaal, L.A. (1994). Met stijl door het Hoger Onderwijs: Een meta-onderzoek naar leerstijlen van HBO- en WO-studenten. Tilburg: Katholieke Universiteit Brabant.
- Schouwenburg, H.C. (1994). Uitstelgedrag bij studenten. Academisch proefschrift, Rijksuniversiteit Groningen.
- Schouwenburg, H.C. (1996). Een onderzoek naar leerstijlen. *Tijdschrift voor Onderwijsresearch*, 21, 151-161.
- Schouwenburg, H.C., & Vugteveen, T.H. (1995). Voorspellende waarde van de Vragenlijst Studieproblemen voor het studiesucces van eerstejaars studenten. In H.C. Schouwenburg & J.T. Groenewoud (Red.), *Studievaardigheid en leerstijlen*. Proceedings van de 14e Landelijke Dag Studievaardigheden te Groningen. Groningen: Wolters-Noordhoff.
- Shuell, T.J. (1988). The role of the student in learning from instruction. *Contemporary Educational Psychology*, 11, 276-295.
- Simons, P.R.J., & Vermunt, J.D.H.M. (1986). Self-regulation of knowledge acquisition: A selection of Dutch research. In G. Beukhof & P.R.J. Simons (Eds.), *German and Dutch research on learning & instruction*. The Hague: S.V.O.
- Slaats, A., van der Sanden, J.M.M., & Lodewijks, J.G.L.C. (1994). Leerstijlen in het Middelbaar Beroepsonderwijs- Verslag van een interviewstudie. Paper gepresenteerd op de Onderwijs Research Dagen, Utrecht.
- Slaats, A., van der Sanden, J.M.M., & Lodewijks, J.G.L.C. (1994). Leerstijlen en persoonlijkheidskenmerken: In het Middelbaar Beroepsonderwijs. Paper gepresenteerd op de Onderwijs Research Dagen, Tilburg.
- Vermeten, Y., Vermunt, J.D.H.M., & Lodewijks, J.G.L.C. (1995). Changes in learning styles as a result of student oriented education. Paper gepresenteerd op de '6th European Conference for Research on Learning and Instruction', Nijmegen.
- Vermunt, J.D.H.M. (1992). Leerstijlen en sturen van leerprocessen in het hoger onderwijs. Naar procesgerichte instructie in zelfstandig denken. Amsterdam: Swets & Zeitlinger.
- Vermunt, J.D.H.M. (1995). Leerstijlen: Een overzicht en recente onderzoeksgegevens. In H.C. Schouwenburg & J.T. Groenewoud (Red.), *Studievaardigheid en leerstijlen* (pp. 51-72). Proceedings van de 14e Landelijke Dag Studievaardigheden te Groningen. Groningen: Wolters-Noordhoff.
- Vorst, H.C.M. (1993). Leren & Studeren. Handleiding en verantwoording bij vragenlijst en cursus studievaardigheid L&S. Nijmegen: Berkhout.

- Wang, M.C., Haertel, G.D., & Walberg, H.J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84, 30-43.
- Weersink, D. (1996). *Leerstijlen, studielastbeleving en studievoortgang in het HBO. Een onderzoek naar leerstijlen, studielastbeleving en studievoortgang bij HBO-studenten van de Hogeschool Nijmegen*. Doctoraalscriptie, Katholieke Universiteit Nijmegen, Vakgroep Onderwijskunde.
- Wetenschappelijke Raad voor het Regeringsbeleid (1986). *Basisvorming in het onderwijs*. 's-Gravenhage: Staatsuitgeverij (273 pp.).
- Zimmerman, B.J. (1989). Models of self-regulated learning and academic achievement. In B.J. Zimmerman & D.H. Schunk (Eds.), *Self-regulated learning and academic achievement. Theory, research and practice* (pp. 1-25). New York: Springer Verlag.
- Zimmerman, B.J. & Pons, M.M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, 80(3), 284-290.

Manuscript ontvangen 09-10-96

Definitief aanvaard 20-02-97

Constructie en evaluatie van een tekstbegripstoets voor de brugklas

Een instrument voor signaleren en typeren van leesproblemen

Hilde Hacquebord

Faculteit der Letteren, Rijksuniversiteit Groningen

ABSTRACT

This article describes the development and evaluation of a text comprehension test, as an educational tool for screening students in the first year of secondary education. It may be used by remedial teachers and school managers, not only for spotting weak readers, but also for hypothesizing about the types of reading problems. The test results yield information about the individual's reading style, which can be either compensating a lack of word knowledge, or a 'bottom up' reading style without bringing about meaning on the textual level. The typologie of readers is validated by video-observations and questionnaires.

INLEIDING

Remedial teachers en docenten in het voortgezet onderwijs weten uit ervaring dat niet alle leerlingen wat betreft leesvaardigheid goed voorbereid zijn op het voortgezet onderwijs. Met name de zogenaamde schoolse leesvaardigheid, het kunnen begrijpen van schoolboekteksten, is voor veel leerlingen en met name allochtonen, een probleem (Hajer en Meestringa, 1995). De laatste tijd komen er steeds meer empirische gegevens die deze ervaring ondersteunen. In de studies naar functionele geletterdheid worden percentages 'functionele analfabeten' genoemd die variëren tussen 7% (Wesdorp & Hoeksema, 1985) en 27% (Leseman, Sijssling & de Vries, 1992). In het internationaal vergelijkend IEA-onderzoek komt Nederland slecht uit de bus. Zo hebben de 9-jarige Nederlandse leerlingen een achterstand van een half jaar op (West-)Duitsland en van een vol jaar op Noorwegen, Zweden en Finland (De Glopper & Otter, 1993). Er is dus reden om zorgen te hebben over de leesvaardigheid van brugklasleerlingen. Veel scholen vinden het daarom belangrijk leesproblemen bij leerlingen tijdig te kunnen signaleren, zodat ze extra hulp kunnen bieden aan wie dat nodig heeft.

Leerlingen in het voortgezet onderwijs vertonen grote verschillen in leesvaardigheid. De verschillen betreffen niet alleen het algemene niveau, maar ook de aard van de leesvaardigheid. De problemen die brugklassers hebben met lezen zijn divers. In het algemeen hebben leerlingen in het reguliere voortgezet onderwijs geen problemen meer met de woordherkenning. Ernstige leesproblemen, waaronder die van dyslectische leerlingen, komen naar schatting voor bij slechts 1 à 2% van de leerlingen. Veel vaker komt het voor dat leerlingen, met name anderstaligen, in het voortgezet onderwijs kampen met leesproblemen als gevolg van een ontoereikende woordenschat. Uit eigen onderzoek is gebleken dat met name anderstaligen, maar ook sommige dyslectische leerlingen, uit zichzelf strategieën ontwikkelen ter compensatie voor hun moeilijkheden op het woordniveau. Anderen zijn zwakke lezers doordat ze

te weinig gericht zijn op de betekenis van een tekst in z'n geheel. Met name als het gaat om schoolboekteksten zijn veel leerlingen onvoldoende functioneel leesvaardig (Hacquebord, 1989, 1991a).

In dit artikel wordt een beschrijving gegeven van de Tekstbegriptoets voor de Brugklas (paragraaf 2) en vervolgens de verantwoording van het instrument met betrekking tot zowel de signaleringsfunctie (paragraaf 3 en 4) als de typeringsfunctie (paragraaf 5). In de hieropvolgende paragraaf wordt het onderwijskundige kader aangeduid waarin de toets wordt toegepast.

1. SIGNALEREN EN TYPEREN VAN LEESPROBLEMEN IN HET VOORTGEZET ONDERWIJS

Veel scholen voeren aan het begin van het jaar een algemeen screeningsonderzoek uit bij brugklassers. Een toets voor begrijpend lezen is hierbij zeer gewenst. Met name heeft men behoefte aan een instrument waarmee leesproblemen kunnen worden gesignaleerd en dat bovendien richting geeft aan de aanpak ervan. In verband met afname bij groepen waarin ook allochtone leerlingen zitten moet het instrument rekening houden met taalaspecten die bij tekstbegrip een grote rol spelen. De Tekstbegriptoets voor de Brugklas is ontwikkeld voor dergelijk screeningsonderzoek. Het instrument is oorspronkelijk gebruikt in onderzoek naar problemen van Nederlandse en Turkse leerlingen met schoolboekteksten (Hacquebord, 1989). Tijdens en na dit onderzoeksproject bleek dit toetsmateriaal veel gevraagd te worden door docenten. Voor onderwijsdoeleinden was de toets evenwel niet ontwikkeld en ook niet geschikt. Er ontbraken normen en betrouwbaarheidsgegevens. Ook was er onvoldoende zicht op de validiteit van het instrument, met name op de vraag of het instrument behalve voor evaluatie- en selectiedoeleinden ook kan worden ingezet voor een meer individueel gerichte diagnostiek.

Bij algemeen screeningsonderzoek gaat het erom op efficiënte en betrouwbare wijze leerlingen met eventuele leesproblemen te kunnen signaleren. Na deze screening is nader onderzoek noodzakelijk teneinde leerlingen ook adequate hulp te kunnen bieden al naargelang de aard van het probleem. Deze diagnostische en besluitvormende fase wordt op de meeste scholen uitgevoerd door gespecialiseerde docenten of door coördinatoren die ook belast zijn met organisatorische maatregelen die verbonden zijn aan de beslissing om bepaalde leerlingen extra hulp te bieden (Meijer, 1995). Scholen hebben hierbij zeer verschillende praktijken ontwikkeld, maar op vrijwel elke school bestaat een of andere vorm van zorgstructuur die meestal met name de brugklasleerlingen betreft. Bij de selectie en plaatsing van deze leerlingen kan de tekstbegriptoets worden gebruikt. Deze kan met name in de eerste twee fasen van het diagnostisch onderzoek worden ingezet.

1.1 Fasen in het diagnostisch onderzoek

Als bij de diagnostiek een hypothesetoetsend model wordt gehanteerd komt dit de betrouwbaarheid van het onderzoek ten goede (Pameijer, 1993). Bovendien maakt dit het beslissingsproces inzichtelijker, waardoor het overdragen van gegevens beter plaats kan vinden. De procedure waarin beslissingen over individuele leerlingen wordt genomen, verloopt bij voorkeur in een aantal fasen:

Fase I: signalering

In de eerste plaats gaat het om een betrouwbare *signalering* van leerlingen die eventuele extra hulp en begeleiding nodig hebben. Deze signalering vindt veelal plaats op basis van screeningsonderzoek. De Tekstbegriptoets kan met name worden ingezet in deze eerste fase, omdat hij geschikt is voor afname bij grote groepen en bovendien makkelijk, objectief en betrouwbaar is na te kijken.

Fase II: formuleren van een hypothese

De volgende fase van het diagnostisch onderzoek is *hypothesevormend*. Van leerlingen die gesignaleerd worden door een onvoldoende toetsuitslag, worden de toetsresultaten nader onderzocht. Het gaat er dan om een beslissing te kunnen nemen voor verdere selectie en eventuele individuele hulpverlening, hetzij om een eerste indruk te krijgen van het soort problemen waarmee deze leerlingen zitten. Zo zijn veel leesproblemen van allochtone leerlingen bijvoorbeeld in verband te brengen met taal- en met name woordenschattekorten. Andere leerlingen hebben bij tekstbegrip juist meer moeite met het achterhalen van de hoofdgedachte van een tekst. In deze fase van het onderzoek kunnen de toetsresultaten worden geraadpleegd, met name door te kijken naar de verschillende deelscores. Op basis van een zogenaamde profielanalyse van deze deelscores kan een hypothetische typering worden geformuleerd over de aard van de individuele problematiek (zie verder paragraaf 5).

Fase III: toetsen van de hypothese

In de derde, *hypothese-toetsende* fase, kunnen aanvullende observaties van individuele leerlingen worden verzameld. De uitslagen van een bepaalde toets zijn zelf nooit voldoende voor een betrouwbaar individueel beeld. Voor aanvullende observaties kan van verschillende instrumenten gebruik worden gemaakt. Zo zijn er specifieke deoltoetsen op bekende probleemgebieden als woordenschat (bv. Aarnoutse, 1993) en/of technisch lezen (voor het voortgezet onderwijs bijvoorbeeld De Klepel, Van den Bos e.a., 1994). De docent zal in deze fase met name moeten putten uit eigen ervaring en eigen observatievermogen. Met name willen wij hier wijzen op de mogelijkheid om informatie van de leerling zelf te krijgen, bijvoorbeeld op basis van een gesprek over diens leesgedrag en leesproblemen, eventueel in combinatie met een observatie van de manier waarop die leerling een bepaalde leesopdracht uitvoert. Ook is het mogelijk, bij grotere groepen leerlingen, dergelijke informatie schriftelijk te verkrijgen, bijvoorbeeld door vragenlijstjes aan de leerlingen voor te leggen. In het evaluatie-onderzoek dat hieronder is beschreven is van een schriftelijke vragenlijst gebruikgemaakt ter validering van de lezerstypen zoals die op basis van de profielanalyse tot stand zijn gekomen (zie paragraaf 5 en bijlage 3).

De Tekstbegriptoets voor de brugklas heeft in de eerste plaats een signaleringsfunctie (fase I) en levert daarnaast informatie op voor de hypothesevormende fase II.

2. DE TEKSTBEGRIPTOETS VOOR DE BRUGKLAS

Criteria

We hebben de volgende criteria gehanteerd bij het ontwikkelen van de Tekstbegriptoets voor de brugklas:

1. De toets is valide en betrouwbaar, voor zowel Nederlandse als allochtone leerlingen.
2. De toets is afgestemd op leerlingen in het voortgezet onderwijs, waarbij twee niveaus worden onderscheiden.
3. De toets kan extensief en efficiënt worden afgenomen door docenten en remedial teachers.
4. De toets kan zwakke lezers signaleren.
5. De toets heeft een diagnostische waarde: hij geeft aanknopingspunten voor verdere behandeling.

De toetsvorm

De toets bestaat uit twee versies van elk 5 teksten uit schoolboeken van ongeveer gelijke lengte (ca. 200 woorden). Versie A is bestemd voor VBO/Mavo en betreft tekstmoeilijkheid 1 en 2. Versie B is bestemd voor Mavo/Havo/Vwo en betreft tekstmoeilijkheid 2 en 3 (zie tabel 1).

De drie-keuzevragen bij de teksten betreffen tekstproposities op drie tekstlinguïstische niveaus. Deze keuze voor meerkeuzevragen bij tekstproposities wordt als volgt gemotiveerd:

1. Anderstalige leerlingen mogen niet benadeeld worden door ze op hun mogelijk geringe schrijf- of spreekvaardigheid aan te spreken. Ditzelfde geldt voor taalzwakke Nederlandse leerlingen.
2. Deze toetsvorm biedt de mogelijkheid verschillende aspecten van tekstbegrip op vergelijkbare wijze te meten.
3. De objectieve scoringswijze levert betrouwbare informatie op en maakt grootschalige afnames voor selectiedoeleinden mogelijk.
4. De afname en verwerking van de toets zijn bij deze toetsvorm relatief eenvoudig.

Aan de formulering van de vragen is veel aandacht besteed, want deze mag geen nieuwe begripsproblemen voor de leerlingen opleveren. De vragen zijn zo veel mogelijk gesteld in frequent gebruikte Nederlandse woorden. Als criterium is hierbij genomen het al dan niet voorkomen van een woord in het Basiswoordenboek Nederlands (De Kleijn en Nieuwborg, 1983).

Verschillende vraagtypen

Door de vragen betreffende hetzelfde tekstuele niveau bij elkaar te nemen, worden drie vraagtypen onderscheiden: micro-, meso- en macrovragen, die betrekking hebben op de onderscheiden begripsniveaus en de daarmee corresponderende tekstlinguïstische niveaus, ontleend aan Van Dijk en Kintsch (1983) en Meijer (1984). Operationeel gedefinieerd wordt onder het tekstuele micro-niveau verstaan proposities op het woord- en zinsniveau van een tekst waarvan we aannemen dat het congrueert met het proces van woordherkenning en de constructie van proposities. Op het meso-niveau, operationeel het alineaniveau van een tekst, vindt de constructie van lokale coherentie plaats en met macro-niveau wordt de globale tekstuele structuur verstaan waarvan de lezer een 'macrostructurering' vormt. Operationeel gedefinieerd betreft dit proposities m.b.t. het hoofdthema, de tekstsoort en de 'strekking' van de tekst (Galema (1989); voorbeelden in bijlage 1).

De proposities die in de eerste proefversie van de toets als ja/nee-stellingen zijn opgenomen, vormen een steekproef uit alle tekstuitspraken en relaties tussen tekstuitspraken die in principe in de tekst aanwezig zijn. Bij de selectie van de tekstproposities zijn niet de uitgebreide analyses van Meyer (1984) of Van Dijk en Kintsch (1983) gevolgd. Intuïtie voor de representativiteit van bepaalde proposities heeft voorop gestaan bij de keuze, naast praktische overwegingen, zoals de formuleerbaarheid van bepaalde uitspraken. De selectie van de items vond plaats op basis van validiteits- en betrouwbaarheidsoordelen van doctoraalstudenten Nederlands. De vaststelling of een bepaalde stelling behoort tot een van de drie typen, of het een micro-, een meso- of een macro-uitspraak is, is ook aan de beoordelaars voorgelegd. De uiteindelijke itemselectie heeft plaatsgevonden na psychometrische analyse op basis van de proefafnames met ja/nee-vragen. Vanwege een te gering discriminerend vermogen (als gevolg van een te grote raatkans bij de ja/nee-vragen) is besloten tot de constructie van drie-keuzevragen bij de geselecteerde proposities.

Tekstkeuze

In totaal zijn er acht teksten van circa 200 woorden. Deze teksten komen uit veel gebruikte aardrijkskundemethodes voor verschillende onderwijsniveaus, die representatief worden geacht voor de teksten waarmee leerlingen in de loop der jaren op school worden geconfronteerd. We hebben ons beperkt tot teksten uit aardrijkskunde/ wereldoriëntatieboeken omdat deze teksten variëren in onderwerp, van concreet (woningbouw) tot abstract (luchtverontreiniging); van dicht bij huis (bijvoorbeeld de wijk) tot ver weg (Amerika). Bovendien lijken deze teksten niet al te zeer gekenmerkt te worden door vakspecifiek taalgebruik.

Moeilijkheidsgraad

De teksten zijn representatief voor de moeilijkheid van teksten in de verschillende onderwijsniveaus. De bepaling van de moeilijkheidsgraad van teksten is gebaseerd op:

1. Docentbeoordelingen van teksten; docenten is gevraagd 40 tekstfragmenten paarsgewijs te ordenen op moeilijkheidsgraad, waarna een algemene rangordering van teksten kon worden vastgesteld, die vervolgens zijn verdeeld in 'makkelijk', 'middelmatic' en 'moeilijk'.
2. Linguïstische analyse van teksten; hierbij is gekeken naar de gemiddelde zinslengte, de frequentie waarmee de woorden voorkomen (voorkomend in het Basiswoordenboek Nederlands) en het percentage abstracta (zelfstandige naamwoorden die naar een abstract begrip verwijzen).
3. Psychometrische analyse van de toetsresultaten in een proefversie van de toets.

De drie moeilijkheidsgraden van teksten zijn als volgt gedefinieerd (zie tabel 1):

Tabel 1: Criteria van moeilijkheidsgraden voor teksten

criteria	Moeilijkheidsgraad 1	Moeilijkheidsgraad 2	Moeilijkheidsgraad 3
Docentbeoordeling	gemakkelijk	middelmatic moeilijk	moeilijk
Afkomstig uit boeken voor	- eind basisschool - brugklas VBO/MAVO	- tweede klas VBO - brugklas MAVO/HAVO	- tweede klas MAVO/HAVO - derde klas VBO/MAVO
Zinslengte	10 tot 12 woorden	12 tot 15 woorden	15 tot 18 woorden
Frequente woorden	85 tot 90 %	80 tot 85%	70 tot 80 %
Abstracta	0 tot 5 %	5 tot 10 %	10 tot 15 %

3. DE PSYCHOMETRISCHE EVALUATIE

De Tekstbegripstoets voor de brugklas is in een predefinitieve versie afgenomen bij 569 brugklasleerlingen verspreid over 13 scholen voor voortgezet onderwijs. De toets is in versie A (Vbo/Mavo) afgenomen bij 323 leerlingen (van wie 36% allochtoon) en in versie B (Mavo/Havo/Vwo) bij 246 leerlingen (van wie 34% allochtoon). Deze betrekkelijk kleine steekproeven achten wij representatief voor de landelijke situatie door de spreiding van mogelijk relevante schoolvariabelen, waarbij met het oog op het zuiverstellen van de toets met name het percentage allochtone leerlingen in acht is genomen. De scholen varieerden verder naar grootte, organisatietype, richting en geografische ligging. Van de zes scholen uit de Randstad en Utrecht waren er vier met een middelgroot percentage allochtone leerlingen (30-40%), in het Vbo was dit percentage nog groter. De andere scholen hadden kleinere percentages allochtone leerlingen (5-30%). Gemiddeld werden twee brugklassen per school getoetst door de docent die zich hiervoor beschikbaar had gesteld via de cursus Remedial Teaching van de Hogeschool van Utrecht. De docenten zijn tijdens een cursusbijeenkomst voorbereid op de toetsafname en hebben zelf de keuze voor de af te nemen toetsversie A of B bepaald.

De psychometrische evaluatie is uitgevoerd met behulp van de bij het éénparameter logistische model (OPLM) behorende software (Glas & Verhelst, 1993; Verhelst, Glas & Verstralen, 1995). Het éénparameter logistische model specificeert de relatie tussen de kans op een

goed antwoord en het discriminerend vermogen van een bepaalde vraag. Dit wordt op de volgende wijze uitgedrukt:

$$P_{vi} = \exp \alpha_i (\theta_v - \delta_i) / [1 + \exp \alpha_i (\theta_v - \delta_i)]$$

Waarbij P de kans op een goed antwoord is van de leerling v met tekstbegrip θ op vraag i met moeilijkheid d en discriminerend vermogen α .

Uit (1) blijkt dat de kans om een vraag goed te beantwoorden groter is naarmate het tekstbegrip toeneemt; hoe moeilijker de vraag, hoe lager de kans dat een leerling een goed antwoord geeft; en, de kans op een goed antwoord hangt slechts af van het tekstbegrip van het kind en niet van andere eigenschappen als bijvoorbeeld nationaliteit of culturele achtergrond.

Bij de evaluatie van de deoltoetsen is ten eerste een scoringsregel vastgesteld; per vraag is bepaald hoeveel punten een goed antwoord oplevert op basis van de vastgestelde moeilijkheid en discriminerend vermogen van het betreffende item (3.1). Ten tweede is de zuiverheid van de toets onderzocht en is onderzocht of de toetsen in staat zijn kinderen met een zwak tekstbegrip te onderscheiden van kinderen met een gemiddeld of sterk tekstbegrip (3.2). In onze context betekent dit dat de toetsscore alleen afhangt van de vaardigheid van het kind, en niet van zijn culturele/etnische achtergrond.

3.1 Het berekenen van de scores op de deoltoetsen

Uit (1) blijkt dat hoe groter het discriminerend vermogen van een vraag, hoe meer informatie het antwoord bevat over de mate van tekstbegrip van een leerling. Dit leidt tot de volgende scoringsregel die uit (1) valt af te leiden (Verhelst en Glas, 1995):

$$S_v = \sum_i \alpha_i x_{vi}$$

waarbij x het antwoord van kind i op vraag v voorstelt en S de score van het kind i is. Het discriminerend vermogen van vraag v wordt uitgedrukt door parameter α . Er vallen meer punten te verdienen met goede antwoorden op vragen met een hoog discriminerend vermogen dan op vragen met een laag discriminerend vermogen. Het aantal punten dat per vraag te verdienen valt, is dus gelijk aan het discriminerend vermogen van de betreffende vraag.

Het gevolg van deze scoringsprocedure betekent wel een ongemak voor de gebruikers, namelijk verschillen in maximale somscores tussen de (deel)toetsen. Teneinde deze scoringsregels voor docenten hanteerbaar te maken, zijn de toetsscores van de verschillende deoltoetsen en van de totale toets daarom op eenzelfde schaal gebracht door middel van een rekenkundige transformatie waardoor de puntentaallen per vraag, in hele getallen, sommen tot maximaal tot 100 (resp. 98 en 102 als gevolg van afronding). Hierdoor zijn dezelfde interpretatiecategorieën toepasbaar op de verschillende (deel)toetsen waardoor het gebruikersgemak wordt vergroot. De interpretatie van de scores volgens het Stroomschema 1a zijn van toepassing op beide versies en geldt voor de deelscores en voor de totaalscore.

Stroomschema 1a:

Fase I Signalering van (zeer) zwakke leerlingen: interpretatie van de scores

1. Bepaal de scores voor de Tekstbegrip-totaaltoets.
2. Selecteer de zeer zwakke en zwakke leerlingen als volgt:

Score	Interpretatie
< 40	zeer zwak
40 - 59	zwak
60 - 79	voldoende
80 - 89	goed
≥ 90	zeer goed

3.2 De toetszuiverheid en -betrouwbaarheid

Met behulp van het éénparameter logistische model is ten eerste de zuiverheid van de toets geëvalueerd. Een toets is zuiver als alle vragen van de toets zuiver zijn, in dit geval met betrekking tot cultureel/etnische achtergrond van de leerlingen. Van de 36 vragen die indicatief waren voor tekstbegrip op microniveau moesten er 4 verwijderd worden wegens onzuiverheid. De 36 vragen op mesoniveau waren alle zuiver en van de 36 macrovragen waren er 3 onzuiver. Er resteerden dus 32, 36 en 33 vragen om tekstbegrip op micro-, meso- en macro-niveau te meten. Voor elk kind wordt, behalve een gewogen score voor elk van deze deelttoetsen een totaalscore berekend, gedefinieerd als de som van de micro-, meso- en macroscores gedeeld door 3.

De interne consistentie van de gezuiverde toets is berekend met behulp van Cronbach's alpha (Carmines en Zeller, 1979). In tabel 3 staat Cronbach's alpha voor elk type tekstbegrip en elke steekproef weergegeven. Het blijkt dat de totale toets (101 vragen) hoge waarden heeft voor alfa in elk van de steekproeven. Dat betekent dat de totale toets uitstekend geschikt is om kinderen in te delen naar de mate van tekstbegrip. De toets kan dus zijn signaleringsfunctie betrouwbaar vervullen. De waarden van alpha voor de micro-, meso- en macrodeelttoetsen liggen rond de .60. Dit is geen hoge waarde, maar voldoende om kinderen met betrekking tot tekstbegrip op de verschillende niveaus in te delen met de labels '(zeer) zwak', 'gemiddeld' dan wel 'sterk'. In de procedurevoorschriften voor de profielanalyse (zie paragraaf 5) is rekening gehouden met deze matige betrouwbaarheidsmaten, zodat typering van lezerstypen alleen plaatsvindt wanneer sprake is van duidelijke verschillen in labeling op de verschillende deelttoetscores.

Tabel 2: Cronbach's alpha voor de zuivere deelttoetsen en totaaltoets bij allochtone en autochtone brugklasleerlingen op laag (vbo/mavo) en hoog (mavo)/havo/vwo) onderwijsniveau

groep/niveau	n	betrouwbaarheid			
		mic	mes	mac	tot
autocht laag	237	.62	.70	.60	.84
allocht laag	86	.51	.61	.59	.81
autocht hoog	183	.64	.65	.52	.84
allocht hoog	63	.64	.72	.61	.86

4. VERDELING VAN LEERLINGEN OVER SCORECATEGORIEËN PER (DEEL)TOETS

In de tabellen 4 en 5 worden voor beide toetsversies de scoreverdelingen gegeven van de 569 leerlingen van verschillende onderwijstypen op verschillende scholen in Nederland. In deze tabellen staat het percentage leerlingen dat in een bepaalde scorecategorie valt. Bij de versie Vbo/Mavo haalde dus 3.5% van de leerlingen een totaalscore van 40 of minder; 45% van de leerlingen haalde een micro-score van 60 - 79. Uit deze tabellen valt verder af te lezen dat bij elke (deel)toets en bij beide versies zo'n 30% van de leerlingen in de categorie 'zwak' of 'zeer zwak' valt. Bij een vergelijking tussen de verschillende deelttoetsen valt verder op dat bij versie A een relatief groot percentage leerlingen 'zeer zwak' scoort op de macrotoets, terwijl bij versie B juist de microdeelttoets een relatief groot deel 'zeer zwakken' oplevert.

Tabel 3: Geschatte landelijke percentages voor versie A

	totaal	micro	meso	macro
zeer zwak <40	3,5%	5%	7,5%	12,5%
zwak 40 - 59	28,5%	25%	22,5%	27,5%
voldoende 60 -79	50%	45%	45%	50%
goed 80 - 89	16%	20%	25%	9%
zeer goed \geq 90	1%	5%	5%	1%

Tabel 4: Geschatte landelijke percentages voor versie B

	totaal	micro	meso	macro
zeer zwak <40	4%	8%	5%	5%
zwak 40 - 59	26%	27%	25%	25%
voldoende 60 -79	50%	40%	45%	50%
goed 80 - 89	18%	17%	18%	16%
zeer goed \geq 90	2%	8%	7%	4%

Bovenstaande tabellen zijn opgenomen in de handleiding bij de toets, zodat gebruikers hun eigen resultaten kunnen vergelijken met het landelijke beeld.

5. ANALYSE VAN LEESGEDRAG EN INDELING IN LEZERSTYPEN

Zoals in paragraaf 2 werd beschreven is de toets geconstrueerd op basis van een analyse van de tekststructuur waarbij we een micro-, een meso- en een macrobegripsniveau van de tekst onderscheiden. Deze drie vormen van leesvaardigheid worden hier beschouwd als aspecten van tekstbegrip die in verschillende mate door leerlingen worden beheerst en/of toegepast als ze lezen. Op basis van de scores en scoreprofielen kunnen hypothesen opgesteld worden over de relatief sterke en zwakke aspecten van de leesvaardigheid van individuele leerlingen. Deze hypothesen kunnen worden getoetst op basis van verdere observaties van het leesgedrag, zoals die met name bij Remedial Teaching worden gedaan. De indeling in lezerstypen is bedoeld als richtinggevend moment in het proces van verdere diagnostiek en eventuele behandeling.

Stroomschema 1b:

Fase II: Typering van zwakke leerlingen; interpretatie van de deelscores

3. Bepaal de deelscores $Mi(cro)$, $Me(so)$ en $Ma(cro)$ van de (zeer) zwakke leerlingen.
4. Bepaal het lezerstype als volgt:

Scores Mi Me Ma

$Mi < 59$ en $Me < 59$ en $Ma < 59$

$Mi < 59$ en $Ma - Mi = > 12$ of $Me - Mi > 12$

$Mi > 60$ en $Mi - Ma > 12$ of $Mi - Me > 12$

Overigen:

totaalscore < 59 en verschil deelscores < 12

totaalscore < 40 en verschil deelscores < 12

interpretatie

probleemlezer

compenserende lezer

schoolse lezer

zwakke lezer

zeer zwakke lezer

Ter validatie van de lezerstypologie op basis van de toetsuitslag en vanuit het idee dat het scoreprofiel van een bepaalde leerling aanwijzingen geeft over het type lezer dat deze leerling is, hebben we video-observaties gedaan van 70 leerlingen uit de steekproef van 569. Deze 70 leerlingen vroegen we een (bewerkte) krantentekst (opgenomen in de methode Goed Nederlands, deel 1) hardop voor te lezen en een aantal vragen op de drie verschillende begripsniveaus (resp. het micro-, meso- en macroniveau) te beantwoorden. Vervolgens moesten ze een korte tekst voorlezen met een probleem dat gaandeweg de tekst wordt opgelost. We vroegen ze dit probleem hardopdenkend op te lossen. Op basis van de video-opnamen zijn deze leerlingen door twee onafhankelijke observatoren ingedeeld in lezerstypen op basis van de volgende beschrijving:

1. de probleemlezer

Deze lezers hebben moeilijkheden op allerlei niveaus: hun technische leesvaardigheid is waarschijnlijk zwak, waardoor het leestempo te laag is, ze hebben een geringe woordenschat, ze hebben weinig zelfvertrouwen in hun eigen leesvaardigheid en zijn (daardoor) niet erg gemotiveerd om te lezen. Ze denken niet na over een oplossing van hun problemen. Deze lezers komen gemakkelijk terecht in een negatieve spiraal: ze lezen niet vlot genoeg, ze denken dat ze het niet kunnen of ze hebben werkelijke problemen op woordniveau, daarom vermijden ze leessituaties waardoor ze relatief steeds slechter gaan lezen.

2. de compenserende lezer

Deze leerling laat een totaalbeeld zien dat net niet voldoende is. Dit type lezer heeft weliswaar moeilijkheden op het woordniveau, bijvoorbeeld door een niet toereikende woordenschat, maar legt zich daar niet bij neer. Deze lezer is erop gericht de betekenis van een tekst te achterhalen en kenmerkt zich door een probleemoplossende en actieve leeshouding. Zolang deze lezers gemotiveerd blijven om te lezen en gestimuleerd worden in hun actieve houding, kunnen ze hun 'handicap' op den duur overwinnen.

3. de schoolse lezer

Deze lezer heeft een bevangen en niet-actieve leeshouding. Begripsproblemen in de tekst worden niet herkend of worden niet opgelost door bijvoorbeeld verbanden in de tekst te leggen. Het leestempo is vlot, maar het lijkt of deze lezer over betekenisproblemen heen leest. Ondanks een goede technische leesvaardigheid en voldoende woordkennis schiet deze lezer tekort in het begrijpen waar de tekst nu eigenlijk over gaat. Deze lezer heeft niet geleerd dat een tekst als geheel een betekenis heeft.

De beschrijving van deze lezerstypen is ontwikkeld op basis van het al eerder aangehaalde eigen onderzoek (Hacquebord, 1989), waarin de Turkse leerlingen in vergelijking met de Nederlandse leerlingen vaker bleken te behoren tot het compenserende lezerstype op basis van de profielanalyse van de toetsresultaten. In vervolgonderzoek (Hacquebord, 1991a; Hacquebord, 1991b) is deze typologie op basis van profielanalyse verder geëxploreerd in relatie tot andere observatievormen en andere groepen leerlingen (dyslectici, zwakke lezers en taalzwakke leerlingen in het algemeen).

Er was voldoende overeenstemming tussen de indeling op grond van de video-observaties en de indeling op basis van de toetsuitslag wat betreft de probleemlezer ($\kappa = .88$, resp. $.92$ voor elke beoordelaar in relatie tot de toetsobservatie). De geobserveerde probleemlezer komt overeen met de categorieën '(zeer) zwak' in tabel 2 voor alle (sub)schalen. De indeling in compenserende of schoolse lezers leidde tot minder grote overeenkomsten zowel tussen de beoordelaars onderling ($\kappa = .61$) als tussen de video-beoordelaars en de toetsuitslag ($\kappa = .33$, resp. $.46$). Op grond van deze matig overeenstemmende observaties is besloten

alleen leerlingen met extremere scoreprofielen te karakteriseren in termen van 'schools' dan wel 'compenserend'. Dit betekent dat een verschil van minimaal 12 punten tussen de microscore en een van de andere deelscores vereist is voor de indeling in compenserende dan wel schoolse leesstijl (zie Stroomschema 1b). Bij de compenserende lezer is de microscore de laagste, bij de schoolse lezer is de microscore de hoogste. Bij kleine verschillen tussen deelscores is een typering niet betrouwbaar. Alleen bij relatief zwakke leerlingen is deze indeling in lezerstype vervolgens zinvol met het oog op remediëring.

Een tweede valideringspoging van de lezerstypologie op basis van de toetsuitslag is ondernomen bij de steekproef die beschreven is in paragraaf 3. Behalve de toetsdata zijn bovendien enquêtegegevens van de leerlingen over hun eigen leesgedrag verzameld. Deze leerlingenquête is geanalyseerd in relatie tot de toetsprofielen. Er waren 40 5-puntsschaaltjes bij uitspraken over het eigen leesgedrag, waarvan we er zestien hebben geselecteerd op basis van frequentieverdelingen en onderscheidend vermogen op basis van correlaties met de lezerstypen. Vervolgens zijn drie subschalen met kenmerkende uitspraken voor elk van de lezerstypen vastgesteld (zie bijlage 3). Deze drie subschalen van lezersuitspraken hebben een betrouwbaarheid van resp. .60 ('probleemitems'), .25 ('compenserende items') en .40 ('schoolse items') (Cronbach alpha). De groep van probleemlezers onderscheidt zich met name door de volgende uitspraken:

- "Ik verveel me als ik lees"
- "Ik ben niet zo'n vlotte lezer"
- "Ik lees tamelijk langzaam"
- "Er zijn in teksten vaak woorden die ik niet begrijp"
- "Als ik zo'n tekst goed wil lezen, lees ik die zachtjes hardop"
- "Ik ben vaak bang dat ik het niet goed begrijp"

Typerende uitspraken van de compenserende lezer zijn:

- "Ik vind het niet erg als ik iets niet meteen begrijp"
- "Ik lees vaak stukjes tekst nog eens over"
- "Een moeilijk woord sla ik gewoon over"
- "Voor het lezen denk ik na over wat ik er al van weet"
- "Er zijn in teksten vaak woorden die ik niet begrijp"
- "Als ik iets niet begrijp probeer ik te raden waarover het gaat"

De schoolse lezer kenmerkt zich door de volgende uitspraken:

- "Als ik lees wil ik precies weten wat er staat"
- "Ik lees niet langzaam"
- "Er zijn in teksten niet vaak woorden die ik niet begrijp"
- "Moeilijke woorden in een tekst brengen me niet in de war"
- "Een moeilijk woord sla ik niet over"

De gevonden lezerstypen zijn met paarsgewijze t-tests vergeleken op de drie itemsets en de vijftien items zijn vervolgens betrokken in een discriminantanalyse (in SPSS) die de classificatieresultaten opleverde die onderaan de volgende tabel zijn gegeven:

Tabel 5: Scores op itemsets, t-test en classificatie-resultaten voor lezerstypen op basis van toetsresultaten

	probleem- lezers (n=87) 1		compenserende lezers (n=55) 2		schoolse lezers (n=213) 3		t-test- resultaten
itemssets: (zie bijlage 3)	M	SD	M	SD	M	SD	
'probleemitems'	2.5	5.9	0.4	6.2	-4.4	6.3	1-2 .005 1-3 .000
'compenserende items'	4.3	4.4	4.7	4.9	2.5	4.3	2-1 .64 2-3 .004
'schoolse items'	-5.1	5.9	-1.9	5.1	.2	6.3	3-1 .000 3-2 .009
totaal correct geclassificeerde leerlingen: 57.3%	correct ingedeeld 57.6%		correct ingedeeld 59.3%		correct ingedeeld 56.7%		

Als valideringsinstrument van genoemde indeling in lezerstypen is de leerlingenquête matig betrouwbaar. 57.3% van de leerlingen wordt op basis van de enquête-uitslag bij hetzelfde lezerstype ingedeeld als op basis van de toetsuitslag. Er is overlap tussen kenmerken van de verschillende lezerstypen, hetgeen ook blijkt uit de scores (gemiddelden en standaarddeviaties, zie tabel 5) voor de verschillende itemssets. De probleemlezer onderscheidt zich het duidelijkst van de andere twee lezerstypen, de compenserende en schoolse lezer zijn moeilijker te onderscheiden op basis van de betreffende itemssets, waarschijnlijk omdat deze lezers deels overeenkomstige percepties hebben van wat 'goed lezen' inhoudt. Wanneer naast de Tekstbegrijptoets ook een woordenschattoets zou worden afgenomen, zal de compenserende lezer eerder te herkennen zijn aan een grote discrepantie tussen de woordkennis (laag) en tekstbegrip (relatief hoog), terwijl de schoolse lezer bij een relatief grote woordenschat zwak scoort op de tekstbegrijptoets.

De resultaten van de leerlingenquête is geen sterke validatie gebleken van de leerlingtypologie op basis van de toetsuitslagen. We moeten bedenken dat validatie op basis van zelf-schattingen van (zwakke) leerlingen uiteraard een hachelijke zaak is. In het kader van dit valideringsonderzoek van de gepresenteerde lezerstypologie kunnen we weinig zekerheid ontlenen aan deze leerlingenquête, die overigens wel duidelijk maakt dat er verschillen in percepties van leesgedrag bestaan tussen (groepen) lezers en dat deze een zekere overeenkomst hebben met de gemeten leesprestaties.

Het instrument kan de docent dienen als observatie-instrument in Fase III van het diagnostisch onderzoek (zie 1.1). Als instrument voor (individueel) gerichte diagnostiek is de leerlingenquête echter niet voldoende betrouwbaar. Hij kan wel als aanvullend observatie-instrument worden benut in de hypothesetoetsende fase III, maar met de nodige voorzichtigheid en in combinatie met andere instrumenten.

6. SAMENVATTING EN DISCUSSIE

In het voortgezet onderwijs wordt in toenemende mate gebruikgemaakt van screeningsonderzoek op het gebied van leesvaardigheid. Hierbij heeft men behoefte aan een toets die niet alleen zwakke lezers kan signaleren, maar die ook richting geeft aan verdere diagnostiek en remediëring van verschillende typen lezers. Enerzijds zijn er lezers die problemen hebben

met tekstbegrip als gevolg van een ontoereikende woordkennis (of woordherkenning, zoals bijvoorbeeld dyslectische leerlingen), anderzijds zijn er leerlingen die zwak zijn doordat ze te weinig gericht zijn op de betekenis van de tekst in z'n geheel. De aanwezigheid van allochtone leerlingen, die veelal problemen hebben op het gebied van leesbegrip als gevolg van een geringe woordkennis, maakt het bovendien noodzakelijk dat een tekstbegriptoets zuiver is ten aanzien van cultureel/etnische achtergrond.

In dit artikel wordt de constructie en evaluatie beschreven van een tekstbegriptoets voor de brugklas, die eenvoudig kan worden afgenomen door docenten teneinde zwakke lezers in de brugklas te signaleren. Daarnaast kan de docent de toetsuitslag nader analyseren om tot diagnostische hypothesen te komen, die eventueel verder kunnen worden getoetst op basis van aanvullende observaties. In het gerapporteerde onderzoek zijn observaties gedaan ter validering van de gepresenteerde lezerstypologie op basis van de toetsuitslag. Behalve video-observaties zijn uitspraken van de leerlingen over hun eigen leesgedrag geanalyseerd in relatie tot de toetsuitslag.

De tekstbegriptoets is een handzaam instrument, en is voldoende betrouwbaar en valide wanneer het gaat om de signaleringsfunctie. De toets kan ook zuiver genoemd worden ten aanzien van allochtone leerlingen. De betrouwbaarheid van de subschalen die ten grondslag liggen aan de lezerstypologie is - door het kleinere aantal items - kleiner, maar voldoende om leerlingen in globale categorieën (zeer) zwak, voldoende en (zeer) goed in te delen. De lezerstypologie zelf laat zich lastig valideren omdat leesgedrag moeilijk te observeren is. De overeenkomsten tussen de indelingen in lezerstype op grond van de toetsuitslag enerzijds en die op basis van video-observaties en enquêteresultaten anderzijds zijn voldoende aanvaardbaar om waarde te hechten aan de toets als instrument voor verdere typering van lezers. Uitspraken met betrekking tot deze typering zijn evenwel hypothetisch, maar kunnen wel richting geven aan verdere diagnostiek en hulpverlening.

Met dank aan Dr. Herbert Hoytink voor het uitvoeren en rapporteren van de OPLM-analyses en Dr. Anne Boomsma voor zijn adviezen op het gebied van statistische analyse. Beiden werken bij de Vakgroep Statistiek en Meettheorie van de Faculteit der Psychologische, Pedagogische en Sociologische Wetenschappen van de Rijksuniversiteit Groningen. Dank aan Drs. Sjoerd de Vos voor het uitvoeren van de overige statistische analyses, het meedenken en het commentaar.

LITERATUUR

- Aarnoutse, C. (1993). Woordenschattests bestemd voor het eerste leerjaar van het Vbo en van het Mavo, Havo en Vwo. Nijmegen: Berkhout.
- Carmines, E.G. & Zeller, R.A. (1979). Reliability and validity assessment. London: SAGE.
- Dijk, T.A. van & Kintsch, W. (1983) Strategies of discourse competence. New York: Academic Press.
- Galema, C. (1989). Tekstbegrip getoetst. Beschrijving van de constructie van toetsen voor tekstbegrip, woordkennis, woordenschat en grammatica voor de onderbouw van het voortgezet onderwijs. Publicatie van de Vakgroep Taalbeheersing nr 27. Rijksuniversiteit Groningen.
- Glas, C.A.W. & Verhelst, N.D (1993). Een overzicht van itemresponsmodellen. In: T.J.H.M. Eggen en P.F. Sanders (Eds.), Psychometrie in de praktijk (pp 179-238). Arnhem: Cito.
- Glopper, K. de & Otter, M. (1993). Nederlandse leesprestaties in internationaal perspectief. Amsterdam: SCO.
- Hajer, M. & Meestringa, T. (1995). Schooltaal als struikelblok, Muiderberg: Coutinho.
- Hacquebord, H.I. (1989). Tekstbegrip van Turkse en Nederlandse leerlingen in het voortgezet onderwijs. Dordrecht: Foris.
- Hacquebord, H.I. (1991a). Diagnose en behandeling van leesproblemen met begripend lezen in het voortgezet onderwijs. In: K.P. van den Bos & H. Nakken (red), Dyslexie '91 (pp 115-131). Amsterdam/Lisse: Swets & Zeitlinger.
- Hacquebord, H.I. (1991b). Signalering van leesproblemen d.m.v. tekstbegriptoetsing? In: M.M.H. Bax & W.

- Vuyk (red), Thema's in de Taalbeheersing (pp.213-225). Dordrecht: ICG Printing.
- Hacquebord, H.I. (1995). Tekstbegrijptoets voor de Brugklas. Groningen: Jacob Dijkstra.
- De Kleijn, P. & Nieuwborg, E. (1983). Basiswoordenboek Nederlands. Groningen: Wolters Noordhoff.
- Leseman, P.P.M., Sijsling, F.F. en Vries, E.M. de (1992). Zorgbreedte en instructieklimaat: aanknopingspunten voor de preventie van functioneel analfabetisme in het LBO en MAVO. In: Pedagogische Studiën, 69, 371-387.
- Meijer, W. (1995). Leesproblemen, diagnostiek en onderwijs. Over de veranderende rol van diagnostici bij de hulpverlening aan leerlingen met leesproblemen. In: Clemens, J & H. Hacquebord (red.), Diagnostiek van leesvaardigheid. (pp. 11-27). (Rain-reeks nr. 4). Delft: Eburon.
- Pameijer, N.K.(1993). Systematisering van het diagnostische proces: pleidooi voor het hypothesetoetsend model. Kind en adolescent 14, 2, 55-66.
- Verhelst, N.D. en Glas, C.A.W. (1995). The one-parameter logistic model. In: G.H. Fischer en I.W. Molenaar (Eds.) Rasch models Foundations, Recent Developments and Applications (pp 215-238).
- Verhelst, N.D., Glas, C.A.W. en Verstralen, H.H.F.M. (1995). One-parameter logistic model OPLM. Arnhem: CITO.
- Wesdorp, H. & Hoeksema, J.B. (1985). Voorstudie periodieke peiling van het onderwijsniveau. Deel 1: leesen schrijfpredaties van zesdeklassers. (SCO-rapport nr. 55). Amsterdam: SCO.

Manuscript ontvangen 24-04-96

Definitief aanvaard 11-02-97

BIJLAGE I

Een voorbeeldtekst en vragen op micro-, meso- en macroniveau

Tekst 1

ARM EN RIJK

- 1 De wereld is verdeeld in arme en rijke landen. Eénderde deel van de wereldbevolking woont in de rijke landen en tweederde deel in de arme landen. Rijke landen zijn bijvoorbeeld Nederland, Engeland, Duitsland en de Verenigde Staten. Arme
- 5 landen vindt men vooral in Afrika, Azië en Zuid-Amerika. Er zijn ook tussenvormen. Italië wordt wel tot de rijke landen gerekend. Maar er zijn ook heel armelijke streken in Italië. Nederland is in elk geval een rijk land. Dat wil zeggen dat hier bijna niemand echt gebrek lijdt, dat we goed te eten hebben,
- 10 dat er volop scholen staan en dat er overal dokters te vinden zijn. Heel anders is het in de arme landen. Daar heerst vaak voedselgebrek. Er zijn landen waar de mensen nooit genoeg eten krijgen. Elders krijgen ze wel genoeg, maar altijd hetzelfde.
- 15 Kinderen in Indonesië krijgen bijna altijd rijst te eten. Kinderen in de arme Afrikaanse landen krijgen vrijwel dagelijks een cassave (wortel) voorgeschoteld. Onvoldoende of eenzijdig eten tast de gezondheid aan. Dit is een van de oorzaken van de grote kindersterfte in de arme lan-
- 20 den. Volgens cijfers van 1968 stierven toen in Chili honderd van de duizend baby's voor hun eerste levensjaar. In Mexico waren het er 64 en op Ceylon 48. In Nederland waren het er maar 13. Dat scheelt nogal wat.

- 1** "Gebrek lijden" (regel 9) betekent hier:
- ziek zijn
 - goed te eten krijgen
 - niet genoeg eten krijgen
- 2** Italië behoort bij:
- de rijke landen
 - de arme landen
 - zowel de arme als rijke landen
- 3** Wat is juist?
- In Nederland krijgt iedereen genoeg te eten.
 - In Nederland eet bijna iedereen hetzelfde.
 - In Nederland eet bijna niemand echt goed.
- 4** Van de wereldbevolking
- heeft tweederde weinig kans op goed voedsel.
 - is éénderde ziek.
 - woont éénderde in een arm land.
- 5** "Elders" (regel 14) betekent hier:
- overal
 - ergens anders
 - nergens
- 6** Een andere titel van deze tekst kan zijn:
- "Ziekte in de arme landen"
 - "Voedselgebrek"
 - "De rijkdom in de wereld is ongelijk verdeeld"
- 7** Wat is juist?
- Altijd hetzelfde eten is goed voor de gezondheid.
 - Altijd hetzelfde eten is slecht voor de gezondheid.
 - Altijd hetzelfde eten heeft geen invloed op de gezondheid.
- 8** De Kindersterfte wordt kleiner als:
- de mensen minder gebrek lijden.
 - er volop scholen staan.
 - er overal dokters te vinden zijn.
- 9** Van de landen die in de tekst genoemd worden heeft Chili:
- de grootste kindersterfte
 - de kleinste kindersterfte
 - een gemiddelde kindersterfte
- 10** "Iemand iets voorschotelen" (regel 17) betekent hier:
- iemand iets laten zien
 - iemand iets in de school leggen
 - iemand iets te eten geven
- 11** Indonesië is een voorbeeld van:
- zowel een arm als rijk land
 - een arm land
 - een rijk land
- 12** Rijke landen hebben
- een grotere kindersterfte dan arme landen
 - een kleinere kindersterfte dan arme landen
 - geen kindersterfte

BIJLAGE II

Scorevoorschrift voor docenten met correctiemal en gewichten per goed antwoord

2 RICHTLIJNEN VOOR AFNAME EN VERWERKING VAN DE TOETS

Bij het gebruik van de toetsen zijn de volgende stappen te onderscheiden:

- 1 Kopiëren van het toetsmateriaal;
- 2 Afnemen van de tekstbegripstoets;
- 3 Verwerken van de resultaten aan de hand van normscores
- 4 Vaststellen van de lezersprofielen
- 5 Selecteren van leerlingen voor steunlessen of remedial teaching

2.1 Het kopiëren van het toetsmateriaal

Het toetsmateriaal mag door de school worden gereproduceerd, zodat elke leerling beschikt over:

- 1 Een toetsboekje bestaande uit vijf teksten met bijbehorende meerkeuze vragen;
- 2 Een antwoordformulier.

U kunt hiervoor gebruik maken van de kopieerbladen in deel II. U dient de bladen zodanig te kopiëren, dat op elke linkerbladzijde de tekst, en op de daarnaast liggende rechterbladzijde de vragen terecht komen.

Op het losse antwoordformulier vullen de leerlingen hun naam, klas en de antwoorden in. Het vetgedrukte blok onderaan het antwoordformulier is bestemd voor de scores van de leerlingen.

2.2 De afname

De meeste leerlingen moeten de toets kunnen maken in ongeveer 50 minuten. Vaak zijn leerlingen binnen die tijd klaar; geef de rest eventueel wat meer tijd, maar dring wel enigszins aan als leerlingen al te lang zitten te tobben en geen keuze kunnen maken.

De instructie bij het toetsboekje wordt voorafgaand aan de toets mondeling doorgenomen. Probeer te voorkomen dat ze zich meteen op de vragen storten! Het is belangrijk dat leerlingen de tijd nemen om eerst de tekst rustig door te lezen voordat ze aan de vragen beginnen. Bij het beantwoorden van de vragen mogen ze de tekst opnieuw raadplegen.

2.3 De verwerking van de resultaten

De toets wordt nagekeken met behulp van de correctiemallen op transparanten en een bijbeho-

rend antwoordformulier per leerling. Deze correctiemallen en het antwoordformulier vindt u in deel II.

Met behulp van de correctiemal berekent u op het antwoordformulier:

de deelscores voor de micro-, meso- en macro-deeltoetsen;

de totaalscore (de optelsom van de drie deeltoetsen gedeeld door 3).

De scores vormen de optelsom van punten, die u per goed beantwoorde vraag toekent. De correctiemal geeft aan hoeveel punten er voor een bepaalde vraag te verdienen zijn. De uitslagen worden ingevuld op het antwoordformulier.

Let op

Op basis van de subtotalen per **afzonderlijke tekst** kunt u geen interpretaties maken over de prestatie van uw leerling. Per tekst zijn de te behalen puntentotalen namelijk verschillend. Bij de ene tekst kunnen bijvoorbeeld meer micropunten worden verdiend, bij de andere meer macropunten. Ook zijn de subtotalen per tekst verschillend. De toetsversies dienen in hun geheel, dus elk met vijf teksten te worden afgenomen. Alleen op deze wijze is sprake van een betrouwbaar signaleringsinstrument en kunnen de normen gehanteerd worden.

2.4 Het hanteren van de normen

2.4.1. Bepaling van scores voor de individuele leerling

U bepaalt voor elke leerling het puntentotaal met behulp van de correctiemal.

*Micro-score: Het totaal aantal behaalde punten op de vragen op micro-niveau (zoals aangegeven op de correctiemal)

*Mesoscore: Het totaal aantal behaalde punten op de vragen op meso-niveau (zoals aangegeven op de correctiemal)

*Macroscore: Het totaal aantal behaalde punten op de vragen op macro-niveau (zoals aangegeven op de correctiemal)

*Totaalscore: Het totaal aantal behaalde punten op alle subtoetsen **gedeeld door 3**

De scores voor zowel de deeltoetsen als voor de totaaltoets zijn op dezelfde schaal gebracht, waardoor de interpretatie gemakkelijk wordt.

CORRECTIEAAL VERSIE A

	TEKST1	Mi	Me	Ma	TEKST2	Mi	Me	Ma	TEKST3	Mi	Me	Ma	TEKST4	Mi	Me	Ma	TEKST5	Mi	Me	Ma
vraag 1	a ■ b ■	4			a ■ b ■	2			■ b c 4				a ■ c 8				a ■ b ■			6
vraag 2	a ■ b ■	2			a ■ c		4		a ■ c			4	a ■ b ■	6			■ b c			4
vraag 3	■ b c 4				■ b c	3			a ■ c 10				■ b c	5			a ■ b ■			6
vraag 4	■ b c		6		a ■ c 4				a ■ b ■			6	a ■ b ■	6			■ b c			7
vraag 5	a ■ c 6				a ■ b ■		3		a ■ b ■			2	a ■ c	5			a ■ c			8
vraag 6	a ■ b ■		4		a ■ b ■		2		a ■ b ■			4	■ b c				■ b c			7
vraag 7	a ■ c	6			a ■ c			3	■ b c 4				■ b c	3			■ b c			4
vraag 8	■ b c		6		a ■ c 6				a ■ c			6	■ b c				a ■ c			6
vraag 9	■ b c		8		a ■ c 8				■ b c			6	a ■ c 4				a ■ c			8
vraag 10	a ■ b ■	4			a ■ b ■		3		■ b c			5	a ■ c	6			■ b c			2
vraag 11	a ■ c		2		■ b c		3		a ■ b ■			6	a ■ b ■				a ■ b c			
vraag 12	a ■ c		8		a ■ c		4		a ■ b ■			8	a ■ b ■	6			a ■ b c			

BIJLAGE III

Leerlingvragenlijst m.b.t. eigen leesgedrag

Bijlage 3 De leerlingenquête.

Tussen haakjes de scoringswijze; achter de items de typering van P= probleem-, S=Schoolse, C= compenserende lezer.

Hoe lees ik eigenlijk?

naam:

school:

Op school moet je veel lezen. Denk maar aan al die schoolboeken en het huiswerk dat je elke dag opkrijgt. Wij zijn benieuwd hoe je dat nu eigenlijk aanpakt. Daarom vragen we je de lijst met vragen hieronder te beantwoorden. Je antwoord kan niet goed of fout zijn, ieder leest op zijn eigen manier!

Stel je voor: je moet thuis een tekst lezen voor school, die niet te lang of te moeilijk is. Het gaat om een **zakelijke tekst** (dus geen verhaal) over een onderwerp dat een beetje saai is, maar wel belangrijk (bijv. **over het milieu**). Over die tekst krijg je later in de les vragen.

Als je zo'n tekst gaat lezen, hoe doe je dat?

Hieronder staan uitspraken over leesmanieren. In een leeg hokje zet je een **kruisje**, bijvoorbeeld:

0. Voor ik ga lezen, bekijk ik eerst hoe lang de tekst is 1 2 3 4 5

Als je van te voren meestal niet kijkt naar hoe lang een tekst is, zet je een kruisje onder 4 (zie vb. boven).

Als je van te voren zeker wel kijkt naar hoe lang een tekst is, zet je een kruisje onder 1 (vb hiernaast):

1 2 3 4 5

De cijfers betekenen:

1= klopt, mee eens (SCORE:1 punt)

2= soms wel, een beetje mee eens (SCORE:1punt)

3= ik weet het niet (SCORE:0punt)

4= meestal niet, niet helemaal mee eens (SCORE:-1punt)

5= klopt niet, niet mee eens(SCORE:-1punt)

- Ik verveel me bij het lezen van zo'n tekst. (P)
- Als ik lees wil ik precies weten wat er staat. (S, C)
- Ik vind het niet erg als ik iets niet meteen begrijp. (S,C)
- Ik lees tamelijk langzaam. (P, -S)
- Ik ben vooral in hoofdzaken van de tekst geïnteresseerd. (P,S)
- Ik lees vaak stukjes tekst nog een over. (C)
- Voor het lezen denk ik na over wat ik al weet van het onderwerp. (C)
- Als ik de tekst goed wil lezen, lees ik die zacht hardop. (P,-S)
- Ik ben vaak bang dat ik het niet goed begrijp wat ik lees. (P, -S)
- Ik weet niet wat ik moet doen als ik iets niet begrijp. (P, -S)
- Er zijn in teksten vaak woorden die ik niet begrijp. (C, P)
- Ik ben niet zo'n vlotte lezer. (P)
- Moeilijke woorden in een tekst brengen me in de war. (P, -S)
- Moeilijke woorden in een tekst zoek ik op in het woordenboek. (C, -S)
- Een moeilijk woord sla ik gewoon over.(C,-S)
- Als ik iets niet begrijp probeer ik te raden waarover het gaat. (C)

Multiniveau modellen voor de analyse van leerwinst vergeleken

Huub van den Bergh (*Utrechts instituut voor Linguïstiek*)
en Hans Kuhlemeier (*Cito*)

ABSTRACT

Differences between schools in effectiveness can be analyzed with five types of multilevel models. These models are: a univariate model for the analysis of variance, a univariate model for the analysis of covariance, a univariate model for gain scores, a multivariate model for the analysis of variance, and a multivariate model for gain scores.

The first model can be used to compare schools on the basis on their final achievement levels at a given moment. The other four models can be analyzed if pretest scores are available. In the model for the analysis of covariance the first model is extended with a covariate representing the pretest. In the univariate model for gain scores the difference between the posttest and pretest score is the dependent variable. In the multivariate model for the analysis of variance pretest and posttest scores are analyzed simultaneously. In the multivariate model for gain scores, the pretest score, and the difference score between posttest and pretest, are modelled simultaneously.

It is demonstrated that with the five multilevel models different hypotheses are tested, and therefore the models yield different results, although these differences can be converted to each other in some cases. It is also shown that the model for the analysis of covariance and both gain score models are likely to yield different results; in many cases the results of the commonly employed model for the analysis of covariance are hard to interpret in terms of differences in school or instructional effectiveness.

1. INLEIDING

Onderzoek naar instructie- en schooleffectiviteit is gebaseerd op de veronderstelling dat leerlingen in de ene klas of school meer leren dan in de andere (o.a. Levine & Lezotte, 1990; Willms, 1992). De analyse van deze verschillen tussen leerlingen, klassen en scholen in leerwinst stelt de onderzoeker voor het probleem van de keuze van een passend analysemodel. Dat dit een multiniveaumodel moet zijn, daarover bestaat heden ten dage grote overeenstemming (zie o.a.: Bryk & Raudenbush, 1992; Goldstein, 1995). Het onderwijs is immers hiërarchisch georganiseerd, met leerlingen genest binnen klassen, die op hun beurt weer genest zijn binnen scholen. Multiniveau-analyse houdt expliciet rekening met deze geneste structuur, levert meer accurate parameterschattingen op (Bryk, & Raudenbush, 1992) en biedt extra mogelijkheden tot interpretatie (Van den Bergh & Hoeksma, 1993). In dit artikel worden vijf multiniveaumodellen voor het analyseren van verschillen in eindniveau en leerwinst met elkaar vergeleken. De vijf modellen zijn:

1. Een univariaat variantie-analytisch model (UNVA) met het bereikte eindniveau als afhankelijke variabele. Dit model wordt vaak gebruikt als er slechts informatie voorhanden is over de prestaties van de leerlingen op één moment. Zijn de prestaties van de leerlingen

- op verschillende momenten gemeten (bij voorbeeld: aan het begin en einde van een cursusjaar), dan komen de vier andere modellen in aanmerking.
2. Een univariaat covariantie-analytisch model (UNCO) met het bereikte eindniveau als afhankelijke variabele en het beginniveau als covariaat.
 3. Een univariaat variantie-analytisch model voor de analyse van leerwinst (UNLW) met de verschilscore tussen eind- en beginniveau als afhankelijke variabele.
 4. Een multivariaat variantie-analytisch model voor de analyse van het gemiddelde begin- en eindniveau (MUVA), waarbij het begin- en eindniveau als afhankelijke variabelen geanalyseerd worden.
 5. Een multivariaat variantie-analytisch model voor het analyseren van het gemiddelde beginniveau en de gemiddelde leerwinst (MULW) waarbij de leerwinst geschat wordt als de afwijking van het beginniveau.

In het vervolg van dit artikel wordt ingegaan op de statistische en inhoudelijke verschillen tussen de vijf modellen. Aan de hand van een empirisch voorbeeld wordt geïllustreerd hoe de parameters geïnterpreteerd kunnen worden. Gedemonstreerd wordt dat met deze modellen verschillende hypothesen getoetst kunnen worden; er worden dus (deels) andere vragen beantwoord. Alleen via de nodige omwegen kunnen dezelfde hypothesen getoetst worden, en zijn de verschillende parameterschattingen tot elkaar te herleiden. Het doel van het artikel is onderzoekers in staat te stellen precies dat analysemodel te kiezen waarmee de gegeven onderzoeksvraag het meest efficiënt en adequaat beantwoord kan worden.

De gebruikte gegevens voor de voorbeelden zijn afkomstig uit een landelijke peiling van het niveau van het aardrijkskunde-onderwijs in het derde leerjaar van het Nederlandse voortgezet onderwijs (Kuhlemeier, Van den Bergh, Notté, Wagenaar, Verstralen & Cappers 1994). In deze peiling is de leerlingen zowel aan het begin als aan het einde van het derde leerjaar (onder meer) een zogeheten kerntoets voorgelegd. De itemscores blijken alle op een en dezelfde schaal uitgedrukt te kunnen worden. Hiermee kan de leerwinst in één schooljaar worden geanalyseerd. Daarnaast zijn (onder meer) de volgende achtergrondkenmerken van de leerlingen vastgesteld: sekse, doubleren in het derde leerjaar, en sociaal-economische en etnische herkomst. Deze entree- en achtergrondkenmerken zijn in het onderzoek opgenomen omdat geobserveerde prestatieverschillen tussen scholen niet zonder meer mogen worden toegeschreven aan verschillen in de effectiviteit van het gegeven onderwijs (Willms, 1992; Hoeben, 1994). De vooruitgang die in één jaar geboekt kan worden, is onder andere afhankelijk van de prestaties aan het begin van het schooljaar (zie bij voorbeeld: Bloom, 1976) en van achtergrondkenmerken zoals sekse, sociaal-economische achtergrond van de ouders en etniciteit (zie bij voorbeeld: Fraser, Walberg, Welch & Hattie, 1987).

Voor de overzichtelijkheid beperken we ons bij de voorbeelden tot het getalsmatig belangrijkste opleidingstype in het Nederlandse voortgezet onderwijs, namelijk het MAVO¹. De gepresenteerde analyses hebben betrekking op de prestaties van 5209 leerlingen in 241 derde klassen van 79 MAVO-opleidingen. De scores op de begin- en eindmeting zijn geschaald (zie: Kuhlemeier e.a., 1994), waarbij de scores op de beginmeting zijn uitgedrukt op een (arbitraire) schaal met een gemiddelde van 100 met een standaarddeviatie van 15.

In het vervolg worden eerst de vijf modellen nader beschreven, en de parameterschatting van een voorbeeld gepresenteerd. Daarna wordt ingegaan op de verschillen en overeenkomsten tussen de modellen in achtereenvolgens de fixed en random parameterschattingen. Tot slot worden consequenties voor verklarende variabelen kort aangestipt.

2. HET UNIVARIAAT VARIANTIE-ANALYTISCH MODEL (UNVA)

Het univariate variantie-analytische model voor het analyseren van verschillen tussen leerlingen, klassen en scholen in het bereikte eindniveau is het meest simpele van de vijf modellen. In de meest basale vorm ziet dit model er als volgt uit. Stel, Y_{ijk} is de score van leerling i in

klas j van school k . We kunnen deze score schrijven als:

$$Y_{ijk} = \beta_0 + X_{i0} + [e_{ijk} + u_{0jk} + v_{00k}] , \quad (1)$$

$$(i = 1, 2, \dots, I_{jk}; j = 1, 2, \dots, J_k; k = 1, 2, \dots, K) .$$

Het model bestaat uit twee delen: het fixed gedeelte en het random gedeelte (tussen vierkante haken). In het fixed deel van het model wordt slechts één 'verklarende variabele' onderscheiden (X_{i0}). Dit is een kolomvector éénen. Er wordt in het fixed deel dan ook maar één regressiecoëfficiënt geschat: β_0 . In een model, zonder verdere 'verklarende' variabelen, is deze te interpreteren als (een schatting van) het populatiegemiddelde. En, aan de hand van de standaardfout² kan getoetst worden of dit gemiddelde, of meer algemeen: deze regressiecoëfficiënt, afwijkt van nul.

In het random deel worden drie residuele scores onderscheiden: e_{ijk} , u_{0jk} en v_{00k} . De eerste (e_{ijk}) representeert de afwijking van leerling i van het gemiddelde van zijn klas, de tweede (u_{0jk}) de afwijking van klas j van het gemiddelde van school k , en de derde (v_{00k}) de afwijking van school k van het geschatte populatiegemiddelde. Gewoontegetroou nemen we aan dat deze residuele scores normaal verdeeld zijn rond nul (d.w.z.: $E[e_{ijk}] = E[u_{0jk}] = E[v_{00k}] = 0$) met een variantie van respectievelijk σ_e^2 , σ_u^2 en σ_v^2 . Bovendien nemen we aan dat de residuele scores van verschillende niveaus ongecorrleerd zijn ($r_{e,u} = r_{e,v} = r_{u,v} = 0$). De gemiddelde score op school k is derhalve $\beta_0 + v_{00k}$, en het gemiddelde van klas j van school k : $\beta_0 + v_{00k} + u_{0jk}$.

In onderzoek naar instructie- en schooleffectiviteit wordt aangenomen dat het er wat toe doet in welke klas en op welke school een leerling zit; op de ene school, in de ene klas leert een leerling meer dan op een andere school of in een andere klas. Vertaald naar het model in Vergelijking 1 betekent dit dat verwacht wordt dat zowel de variantie tussen klassen (σ_u^2) als de variantie tussen scholen (σ_v^2) afwijkt van nul. Aan de hand van de ratio van de variantieschattingen en de respectievelijke standaardfouten kan nagegaan worden of deze verwachtingen al dan niet verworpen moeten worden (zie: noot 2).

Het univariate variantie-analyse model kan eenvoudig uitgebreid worden met leerling-, klas- en/of schoolvariabelen. Met deze variabelen kan de variantie op het betreffende niveau of hogere niveaus verklaard worden.

De mogelijkheden van het univariate variantie-analytische model (UNVA) verduidelijken we aan de hand van een voorbeeld uit de landelijke niveauepeiling. Aan het einde van het derde leerjaar blijken duidelijke prestatieverschillen te bestaan tussen leerlingen, klassen en scholen. In Tabel 1 zijn de variantieschattingen op school-, klas- en leerlingniveau weergegeven.

Het intercept is geschat als 108.9. Omdat er in dit model geen verdere verklarende variabelen opgenomen zijn, is het intercept te interpreteren als de gemiddelde score aan het einde van het derde leerjaar. De tussen-scholen- en tussen-klassen-variantie is respectievelijk 24.6 en 7.3. Beide parameterschattingen wijken af van nul. We zouden dus kunnen concluderen dat het voor een leerling zowel uitmaakt op welke school hij zit, als in welke derde klas van die school hij zit. En, omdat de verschillen tussen scholen groter zijn dan die tussen klassen, kunnen we ook concluderen dat de school meer gewicht in de schaal legt dan de klas. De variantie tussen leerlingen, of niveau-1-variantie, bedraagt 271.4.

Het UNVA-model geeft informatie over de relatieve grootte van de geobserveerde prestatieverschillen tussen scholen, klassen en leerlingen op een bepaald moment in het onderwijs (d.w.z.: aan het einde van het derde leerjaar). Het geeft echter géén eenduidig antwoord

op de vraag naar de verschillen tussen scholen en klassen in de *effectiviteit* van het onderwijs. Of de verschillen tussen scholen en klassen veroorzaakt worden door verschillen ten gevolge van leerlingkenmerken, of door verschillen tussen klassen en scholen in bij voorbeeld instructiekenmerken, valt niet uit te maken. We kunnen hooguit constateren dat de gemiddelde prestaties van leerlingen aan het einde van het derde leerjaar van school tot school en van klas tot klas verschillen. Door het toevoegen van verklarende variabelen om de verschillen tussen scholen, klassen en leerlingen te verklaren, kunnen dus alleen verschillen in bereikt eindniveau, en geen verschillen in leerwinst verklaard worden.

Tabel 1: Verschillen tussen scholen, klassen en leerlingen in het model voor univariate variantie-analyse (UNVA; zie: Vergelijking 1; standaardfouten tussen haakjes)

	Schatting (se)
Intercept	Fixed parameter 108.93 (0.64)
Intercept (v_{00k})	Random parameters Schoolniveau 24.63 (5.18)
Intercept (u_{0jk})	Klasniveau 7.27 (2.21)
Intercept (e_{ijk})	Leerlingniveau 271.39 (5.44)

Om inzicht te krijgen in de leerwinst moeten de prestaties van de leerlingen op een eerder moment in de analyse betrokken worden. Door (bij voorbeeld) de prestaties aan het begin van het derde leerjaar in de analyse te betrekken, kunnen we het bereikte eindniveau corrigeren voor de prestaties aan het begin van het leerjaar en/of de leerwinst schatten. Voor de analyse van het gecorrigeerd eindniveau i.c. leerwinst komen verschillende multiniveau modellen in aanmerking. Op vier van deze modellen zullen wij hier ingaan: het univariate covariantie-analyse model (UNCO), het univariate leerwinstmodel (UNLW) en twee multivariate leerwinstmodellen (MUVA en MULW). Gedemonstreerd zal worden dat deze modellen andere schattingen opleveren en tot verschillende conclusies aanleiding kunnen geven.

3. HET UNIVARIATE COVARIANTIE-ANALYTISCH MODEL (UNCO)

Een veel gebruikt model voor de analyse van verschillen in leerwinst is het covariantie-analyse model (zie bij voorbeeld; Lockheed & Longford, 1991; Kreft & Aschbacher, 1994). In dit model wordt een beginmeting opgenomen als een verklarende variabele (of, om in de terminologie te blijven: als covariaat), waardoor het eindniveau gecorrigeerd kan worden voor de reeds bij aanvang aanwezige verschillen. *'By controlling for prior achievement, one is able to obtain estimates of learning gain over the intervening period. (...) Effectiveness is thus defined in terms of that part of the achievement not predicted by prior achievement'* (Hill & Rowe, 1996, p.9).

Bij covariantie-analyse wordt standaard aangenomen dat de regressielijnen van de eind- op de beginmeting parallel zijn. In effectiviteitsonderzoek is men veelal ook geïnteresseerd in variantie van de hellingshoek; de hellingshoek van de regressielijnen verschilt van klas tot

klas en/of van school tot school (Burstein, Kyung-Sung & Delandshire, 1989; Bryk & Raudenbush, 1992, p. 21, p. 60-127). Dat wil zeggen: leerlingen met dezelfde beginscore leren in de ene klas meer dan in de andere, terwijl leerlingen met een andere beginscore in de andere klas meer leren dan in de ene. Conform de traditie in schooleffectiviteitsonderzoek leggen wij geen restricties op aan de relatie tussen begin- en eindmeting; deze regressiecoëfficiënt mag variëren van klas tot klas en van school tot school.

Als we de score op de eindmeting aangeven met Y_{ijk} en de score op de beginmeting met BEG_{ijk} , dan kan het te analyseren model als volgt geschreven worden (waarbij de covariaat meestal gecentreerd wordt rond het algemeen gemiddelde: $BEG_{ijk} = BEG_{ijk} - BEG_{000}$):

$$\begin{aligned}
 Y_{ijk} = & \beta_0 * X_0 + \beta_1 * (BEG_{ijk} - BEG_{000}) + \\
 & [e_{ijk} + u_{00jk} + u_{10jk} * (BEG_{ijk} - BEG_{000}) + \\
 & v_{000k} + v_{100k} * (BEG_{ijk} - BEG_{000})]
 \end{aligned}
 \tag{2}$$

$$(i = 1, 2, \dots, I_{jk}; j = 1, 2, \dots, J_k; k = 1, 2, \dots, K).$$

In het fixed deel van het model worden twee parameters onderscheiden: het intercept (β_0), en een algemene regressiecoëfficiënt (β_1) voor de relatie tussen de begin- en eindmeting. (N.b. de 'verklarende variabele' X_0 is, zoals in alle modellen, een kolomvector met éénen.) Aan de hand van de geschatte standaardfouten kan bepaald worden of het intercept - na centering van de beginmeting -, en de regressie van de eind- op de beginmeting afwijken van nul.

In het random deel van model (tussen vierkante haken) worden vijf residuele scores onderscheiden: één op leerlingniveau (e_{ijk}), twee op klasniveau (u_{00jk} en u_{10jk}) en twee op schoolniveau (v_{000k} en v_{100k}). De afwijking van leerling i van het klasintercept wordt weergegeven met e_{ijk} , de afwijking van klas j van het schoolintercept met u_{00jk} , en de afwijking van het intercept van school k van het (geschatte) populatie-intercept met v_{000k} . De beide resterende residuele scores geven de verschillen tussen klassen (u_{10jk}) en tussen scholen (v_{100k}) aan in regressiecoëfficiënt van eind- op beginmeting (i.c. de variantie van het regressiecoëfficiënt β_1 in Vergelijking 2). Dus: v_{100k} representeert de afwijking van de regressie van de eind- op de beginmeting (β_1) voor school k . En, u_{10jk} geeft de afwijking van klas j van dit regressiegewicht van school k (d.w.z.: $\beta_1 + v_{100k}$) voor klas j .

Getoetst wordt of de interceptvariantie op klas- en schoolniveau afwijkt van 0, en of de variantie van het regressiegewicht afwijkt van 0, d.w.z. of de relatie tussen begin- en eindmeting invariant is over klassen respectievelijk scholen. Ook kan aan de hand van de standaardfouten nagegaan worden of de covariantie tussen de beide residuen van één niveau afwijkt van nul.

Vanzelfsprekend kan dit basismodel uitgebreid worden met diverse (andere) verklarende variabelen op school-, klas- en leerlingniveau. Met deze variabelen wordt dan getracht de variantie tussen leerlingen (σ_e^2), tussen klassen (σ_{u0}^2 voor de verschillen in intercept en σ_{u1}^2 voor de verschillen in hellingshoek) en tussen scholen (σ_{v0}^2 en σ_{v1}^2 voor respectievelijk de intercept- en de hellingshoekverschillen) te binden.

Toepassing van dit model op de gegevens uit de landelijke niveaupeiling levert de schattingen die in Tabel 2 zijn samengevat.

Het intercept van de eindmeting bedraagt 109.0. Aangezien de scores op de beginmeting gecentreerd zijn (rond het gemiddelde op deze meting), is het intercept hier de gemiddelde score op de eindmeting voor een leerling met een gemiddelde score op de beginmeting. Voor elk

punt hoger of lager op de beginmeting neemt de score op de eindmeting met een half punt toe of af. Het intercept in de hier uitgevoerde covariantie-analyse komt overeen met de schatting van de gemiddelde score van de leerlingen op de eindmeting (zie: Tabel 1). Daaruit kan de conclusie getrokken worden dat de gemiddelde leerling met een score van 100 op de beginmeting negen punten leerwinst geboekt heeft. Merk op dat dit niet gelijk hoeft te zijn aan de gemiddelde leerwinst, omdat de gemiddelde leerling op de beginmeting én de gemiddelde leerling op de eindmeting niet dezelfde leerling behoort te zijn.

Tabel 2: Parameterschattingen voor het model voor univariate covariantie-analyse (UNCO; zie: Vergelijking 2; standaardfouten tussen haakjes)

		Schatting (se)	
Intercept (B_0)		Fixed parameters	
Beginmeting (B_1) ^a		109.00 (0.56)	
		0.51 (0.02)	
		Random parameters	
	Intercept	Schoolniveau	Beginmeting
Intercept (v_{000k})	18.94 (3.88)		
Beginmeting (v_{100k})	0.00 (0.00)		0.00 (0.00)
		Klasniveau	
Intercept (u_{00jk})	2.39 (1.39)		
Beginmeting (u_{10jk})	0.17 (0.06)		0.012 (0.005)
		Leerlingniveau	
Intercept (e_{ijk})	215.60 (4.41)		

^a gecentreerd rond het gemiddelde.

Er blijken zowel verschillen tussen scholen als tussen klassen aangetoond te kunnen worden. Uit Tabel 2 blijkt dat op schoolniveau de interceptvariantie significant is; het eindniveau voor leerlingen met een gelijk beginniveau verschilt van school tot school. Er kan echter niet aangetoond worden dat de regressie van de eind- op de beginmeting varieert van school tot school; de variantie van dit regressiegewicht is op schoolniveau gelijk aan nul.

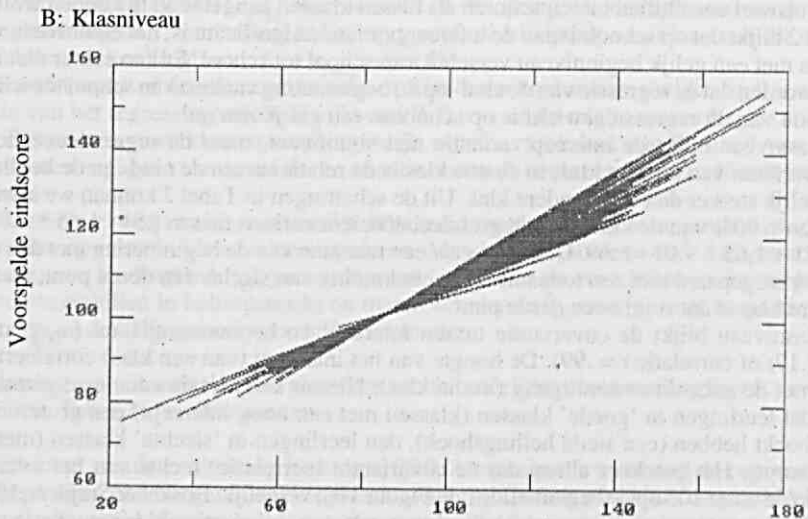
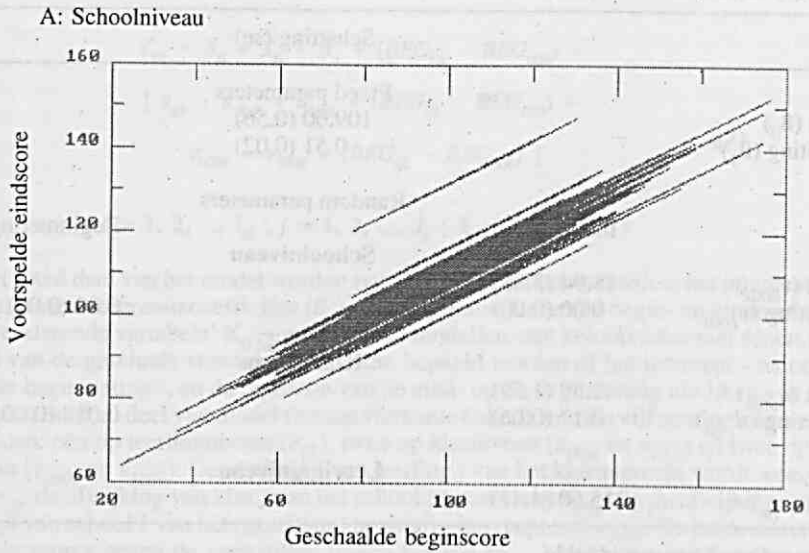
Op klasniveau blijkt de interceptvariantie niet significant, maar de regressiecoëfficiënt blijkt te variëren van klas tot klas; in de ene klas is de relatie tussen de eind- en de beginmeting duidelijk sterker dan in de andere klas. Uit de schattingen in Tabel 2 kunnen we afleiden dat voor zo'n 90% van de klassen de regressiecoëfficiënt varieert tussen $(.51 - 1.65 * \sqrt{.01} =)$.33 en $(.51 + 1.65 * \sqrt{.01} =)$.69. Derhalve gaat een toename van de beginmeting met één punt in de ene klas gepaard met een toename in de eindmeting van slechts één derde punt, maar in een andere klas is dat ruim twee derde punt.

Op klasniveau blijkt de covariantie tussen intercept en beginmeting ($\text{cov}(u_{00jk}, u_{10jk})$) positief (.17; of correlatie $r = .99$). De hoogte van het intercept (van een klas) correleert dus positief met de geboekte vooruitgang (in die klas). Hieruit kan *niet* de conclusie getrokken worden dat leerlingen in 'goede' klassen (klassen met een hoog intercept) een grotere leerwinst geboekt hebben (een steile hellingshoek), dan leerlingen in 'slechte' klassen (met een laag intercept). Het betekent alleen dat de covariantie (correlatie) rechts van het intercept positief is (de score 100 op de beginmeting (zie Figuur 1B); vergelijk: Bosker & Snijders, 1990). Hadden we het intercept op een ander schaalpunt gekozen, bij voorbeeld bij een beginscore

60, dan was de uitkomst geweest dat in klassen met een hoog intercept-residu de hellingshoek klein is (zie Figuur 1B). M.a.w. de correlatie tussen intercept en hellingshoek was negatief geweest. Centrering van een covariaat is van invloed op de covariantiecoëfficiënt.

In Figuur 1 is de geschatte relatie tussen de eind- en beginmeting visueel weergegeven.

Figuur 1: De relatie tussen begin- en eindmeting op school- (boven) en klasniveau (onder) volgens het model voor univariate covariantie-analyse (UNCO)



In Figuur 1A representeert elke lijn één school. Uit het gegeven dat alle lijnen evenwijdig lopen kunnen we concluderen dat de regressie van de eindmeting op de beginmeting gelijk is voor alle scholen. Dat de verschillende lijnen te onderscheiden zijn, en dat niet alle scholen door één lijn gerepresenteerd kunnen worden, is de weergave van de interceptvariantie. Deze geeft aan dat het eindniveau verschilt van school tot school.

In Figuur 1B representeert elke lijn één klas. Heel duidelijk is te zien dat de lijnen niet parallel lopen. De relatie tussen de begin- en eindmeting varieert dus van klas tot klas; het regressiegewicht varieert van klas tot klas³. Of, anders gezegd: er zijn klassen waarin goede leerlingen veel opsteken en er zijn klassen waarin slechte leerlingen veel opsteken. Aangezien de relatie tussen de scores op de begin- en eindmeting⁴ met enkel een lineaire term afdoende beschreven kan worden, is een klas die 'goed' is voor leerlingen met hoge scores op de beginmeting minder 'goed' voor leerlingen met lage scores op de beginmeting en vice versa. Eenvoudig rekenwerk leert dat de score op de eindmeting voor een leerling met een hoge score op de beginmeting (130) kan variëren van 119 tot 130 afhankelijk van de klas van een school waarin deze leerling zit. Voor een leerling met een lage beginscore (70) loopt de range in scores op de eindmeting tussen 81 en 96⁵.

4. HET UNIVARIATE LEERWINSTMODEL (UNLW)

Leerwinst wordt soms gemodelleerd door simpelweg de verschillscore tussen eind- en beginmeting als afhankelijke variabele in de analyse te betrekken (Fennessey & Salganik, 1983; Huff, 1989; Kuyper & Veenstra, 1995). Stel, BEG_{ijk} is de score van leerling i in klas j van school k op de beginmeting, en $EIND_{ijk}$ is de score van deze leerling op eindmeting, dan is de leerwinst gelijk aan: $EIND_{ijk} - BEG_{ijk}$. Het te analyseren multiniveaumodel kan dan als volgt geschreven worden:

$$(EIND_{ijk} - BEG_{ijk}) = \beta_0 * X_0 + [e_{ijk} + u_{0jk} + v_{00k}] , \quad (3)$$

$$(i = 1, 2, \dots, I_{jk}; j = 1, 2, \dots, J_k; k = 1, 2, \dots, K) .$$

Dit model voor verschillscores is in veel opzichten gelijk aan het eerder genoemde model voor variantie-analyse (UNVA). Het verschil tussen beide modellen is alleen gelegen in de responsvariabele: deze is hier het verschil tussen de eind- en de beginmeting, en in het UNVA-model de score op de eindmeting. Het intercept (β_0) in dit univariate model voor verschillscores, zonder verdere verklarende variabelen, representeert het gemiddelde verschil tussen de eind- en de beginmeting: de gemiddelde leerwinst. Dientengevolge wordt aan de hand van de fixed parameter een andere nulhypothese getoetst dan in het UNVA-model: het gemiddelde verschil tussen eind- en beginmeting is gelijk aan nul.

De varianties van de drie residuele scores, e_{ijk} , u_{0jk} en v_{00k} , aangeduid met respectievelijk σ_e^2 , σ_u^2 en σ_v^2 , representeren de verschillen op leerling-, klas- en schoolniveau in leerwinst. Op het tweede en derde niveau is de toetsing of de varianties afwijken van nul een toetsing of de leerwinst verschilt van klas tot klas respectievelijk van school tot school. De interpretatie van de derde variantiecomponent, σ_e^2 , is, zoals in alle modellen enigszins ambigu. Het betreft namelijk zowel de variantie tussen leerlingen (in leerwinst), als de errorvariantie die ontstaat door de onbetrouwbaarheid van de scores.

Toepassing van het model voor verschillscores op de gegevens van de landelijke niveaupeiling leidt tot de parameterschattingen die in Tabel 3 zijn samengevat.

Tabel 3: Parameterschattingen voor het univariate leerwinstmodel (UNLW; zie: Vergelijking 3; standaardfouten tussen haakjes)

	Schatting (se)
Intercept (B_0)	Fixed parameter 9.18 (0.55)
Intercept (v_{00k})	Random parameters Schoolniveau 18.13 (3.75)
Intercept (u_{0jk})	Klasniveau 1.10 (1.51)
Intercept (e_{ijk})	Leerlingniveau 269.10 (5.40)

Uit Tabel 3 blijkt dat de gemiddelde leerwinst 9.2 punten bedraagt. Ook blijkt de variatie in leerwinst tussen leerlingen groot (269.1). Daar staat tegenover dat er geen verschillen in leerwinst tussen klassen geconstateerd zijn. Op schoolniveau kunnen wel verschillen aangetoond worden. De standaarddeviatie van de schoolgemiddelden is 4.3. Hieruit kan afgeleid worden dat het verschil tussen de 5% hoogst en de 5% laagst scorende scholen 14.1 punten bedraagt; het verschil tussen deze beide extremen is dus - ongecorrigeerd voor instroomverschillen e.d. - ruim ($14.1 / 9.2 =$) anderhalf leerjaar.

5. HET MULTIVARIATE VARIANTIE-ANALYTISCH MODEL (MUVA)

De scores op de begin- en eindmeting kunnen ook opgevat worden als één vector van scores die met behulp van een multivariaat model geanalyseerd kunnen worden. In het multivariate variantie-analytische (MUVA) model onderscheiden we twee responsvariabelen: de score op de beginmeting en de score op de eindmeting. Het te analyseren model bestaat nu uit twee (zoals in dit voorbeeld, of uit meer) delen; voor elke responsvariabele één. Stel, Y_{1ijk} is de score van leerling i in klas j van school k op een beginmeting, en Y_{2ijk} is de score van deze leerling op een eindmeting. Het te analyseren model kan dan geschreven worden als:

$$\begin{aligned} Y_{1ijk} &= \beta_1 * X_{10} + [e_{1ijk} + u_{10jk} + v_{100k}] \\ Y_{2ijk} &= \beta_2 * X_{20} + [e_{2ijk} + u_{20jk} + v_{200k}] \end{aligned} \quad (4)$$

$$(i = 1, 2, \dots, I_{jk}; j = 1, 2, \dots, J_k; k = 1, 2, \dots, K).$$

Hierbij zijn X_{10} en X_{20} beide een kolomvector met éénen. De overige parameters in beide vergelijkingen hebben dezelfde interpretatie als die in het eerder besproken univariate model.

Vanzelfsprekend worden ook in het MUVA-model dezelfde assumpties gemaakt met betrekking tot de residuen.

Door introductie van twee dummy-variabelen kunnen de beide vergelijkingen ineen worden geschoven, en de parameters simultaan geschat worden. Stel, $D-BEG_{tijk}$ en $D-EIND_{tijk}$ zijn dummy-variabelen die 'aan staan' als een score de beginmeting ($D-BEG_{tijk} = 1$; $D-EIND_{tijk} = 0$) respectievelijk de eindmeting betreft ($D-BEG_{tijk} = 0$; $D-EIND_{tijk} = 1$), dan kan het MUVA-model als volgt geschreven worden:

$$Y_{tijk} = \beta_1 * D-BEG_{tijk} + \beta_2 * D-EIND_{tijk} + \\ [D-BEG_{tijk} * (e_{1ijk} + u_{10k} + v_{100k}) + \\ D-EIND_{tijk} * (e_{2ijk} + u_{20k} + v_{200k})] \quad (5)$$

$$(t = 1, 2 ; i = 1, 2, \dots, I_j ; j = 1, 2, \dots, J_k ; k = 1, 2, \dots, K).$$

In vergelijking 5 geven de parameters β_1 en β_2 de gemiddelden (intercepten) op de begin- en eindmeting aan. Daarnaast zijn op elk niveau (leerling-, klas- en schoolniveau) twee residuele scores onderscheiden; één voor de beginmeting en één voor de eindmeting (aan te duiden met respectievelijk: σ_{e1}^2 , σ_{e2}^2 , σ_{u1}^2 , σ_{u2}^2 , σ_{v1}^2 en σ_{v2}^2). Aan de covariantie tussen de beide residuele scores per niveau wordt geen restrictie opgelegd, zodat voor elk niveau een volledige (3X3) covariantiematrix geschat wordt, waarvan voor elk element nagegaan kan worden of dit afwijkt van nul.

Merk op dat het model in principe geen restricties oplegt aan de aan- of afwezigheid op één van beide meetmomenten; ook leerlingen die op één van de beide metingen afwezig waren kunnen in de analyse betrokken worden (i.t.t. het UNCO- of UNLW-model). Immers, het model legt niet de restricties op dat er van elke leerling (en elke klas of school) op beide momenten een record aanwezig is (vergelijk Goldstein & McDonald, 1988). Om redenen van vergelijkbaarheid hebben wij ons echter beperkt tot die leerlingen waarvoor op beide meetmomenten een score beschikbaar is.

De parameterschattingen voor het multivariate variantie-analytische model (MUVA) zijn samengevat in Tabel 4.

Uit Tabel 4 blijkt dat het gemiddelde op de beginmeting geschat wordt als 99.8, en het gemiddelde op de eindmeting als 108.9. Aan de hand van de (geschatte) standaardfouten kan getoetst worden dat deze beide gemiddelden afwijken van 0. Het gaat in veel gevallen echter niet om de vraag of de beide gemiddelden afwijken van nul (immers, dat is slechts een schalkingskwestie, en aangezien het meestal gegevens op intervalniveau betreft minder interessant), maar om de vraag of de gemiddelde leerwinst, $(108.90 - 99.78 =) 9.1$, significant is. Om te toetsen of de beide gemiddelden verschillen moet een aparte toetsing uitgevoerd worden⁶. In het onderhavige model is de toetsingsgrootte χ^2_1 gelijk aan 278.59, $p < .001$, zodat het verschil tussen begin- en eindmeting niet aan steekproeftevalligheden toegeschreven kan worden.

Op zowel de begin- als de eindmeting kunnen in het MUVA-model verschillen tussen scholen als tussen klassen aangetoond worden; de variantie van de begin- respectievelijk eindmeting wijkt op beide niveaus af van nul.

Op schoolniveau vinden we een redelijk positieve relatie tussen begin- en eindmeting ($r = .53$). Dit neemt niet weg dat de rangordering van scholen aan het einde van het derde leerjaar verschilt van die aan het begin van dit leerjaar. Hieruit kan geconcludeerd worden dat de derdeklassers op de ene school (gemiddeld) meer geleerd hebben dan op de andere school. Daarnaast blijkt dat de variantie tussen scholen aan het einde van het derde leerjaar (25.0)

groter is dan de variantie tussen scholen aan het begin van het derde leerjaar (9.6; een verschil dat bij toetsing significant blijkt: $\chi^2_1 = 9.24$; $p < .05$). De verschillen tussen scholen aan het einde van het derde leerjaar zijn dus groter dan de verschillen tussen scholen aan het begin van het derde leerjaar.

Tabel 4: Parameterschattingen voor het multivariate variantie-analytische model (MUVA; zie: Vergelijking 5; standaardfouten tussen haakjes)

		Schating (se)	
Beginmeting (β_1)	Fixed parameters		
	99.78 (0.44)		
Eindmeting (β_2)	108.90 (0.64)		
	Random parameters (correlaties boven de diagonaal)		
	Beginmeting		Eindmeting
	Schoolniveau		
Beginmeting (v_{100k})	9.61 (2.51)		.53
Eindmeting (v_{200k})	8.14 (2.92)		25.01 (5.23)
	Klasniveau		
Beginmeting (u_{10jk})	6.19 (1.76)		.92
Eindmeting (u_{20jk})	6.13 (1.64)		7.15 (2.20)
	Leerlingniveau		
Beginmeting (e_{1ijk})	208.73 (4.19)		.44
Eindmeting (e_{2ijk})	105.53 (3.69)		271.40 (5.44)

Op klasniveau is de correlatie tussen begin- en eindmeting zeer hoog ($r = .92$); klassen die op de beginmeting hoog presteren, doen dat ook op de eindmeting. Of, anders gezegd, de rangorde van klassen binnen scholen qua prestaties is aan het einde van het derde leerjaar vrijwel identiek aan de rangordering aan het begin van het derde leerjaar. Aangezien ook de tussen-klassen-variantie op beide meetmomenten (nagenoeg) onveranderd is, kan afgeleid worden dat er volgens het MUVA-model (vrijwel) geen leerwinstverschillen zijn tussen klassen.

Op leerlingniveau tenslotte is er ook sprake van een redelijk positief verband ($r = .44$); leerlingen met een hoge score op de beginmeting hebben ook overwegend een hoge score op de eindmeting. Wel blijkt dat de variantie tussen leerlingen toeneemt; aan het einde van het derde leerjaar zijn de verschillen tussen leerlingen groter dan aan het begin van dat leerjaar.

6. HET MULTIVARIATE LEERWINSTMODEL (MULW)

Nets als het MUVA-model kent het multivariate variantie-analytisch model voor het analyseren van leerwinst (MULW) twee afhankelijke variabelen, namelijk het begin- en het eindniveau. Door de introductie van een dummy-variabele (D-EIND_{ijk}), die alleen 'aan staat' (d.w.z.: D-EIND_{ijk} = 1) als het een score op de eindmeting betreft, kan de score op de eindmeting geschat worden als afwijking van de score op de beginmeting. Vandaar dat het multivariate leerwinstmodel (MULW) ook wel een model voor afwijkingsscores genoemd kan worden. Vanzelfsprekend moet, conform de uitgangspunten van instructie- en schooleffectiviteits-

onderzoek, aangenomen worden dat deze afwijking kan variëren van klas tot klas en van school tot school. In meer formele gedaante ziet dit model er als volgt uit:

$$Y_{hijk} = \beta_0 * X_0 + \beta_1 * D-EIND_{hijk} + \\ [e_{0ijk} + e_{1ijk} * D-EIND_{hijk} + \\ u_{00jk} + u_{10jk} * D-EIND_{hijk} + \\ v_{000k} + v_{100k} * D-EIND_{hijk}] \quad (6)$$

$$(h = 0, 1; i = 1, 2, \dots, I_j; j = 1, 2, \dots, J_k; k = 1, 2, \dots, K) .$$

Onder de gebruikelijke aannamen - een normale verdeling voor de residuele scores, en de residuele scores van verschillende niveaus zijn ongecorrleerd - kunnen de modelparameters als volgt geïnterpreteerd worden. We beginnen met de fixed parameters. Het intercept, β_0 , kan geïnterpreteerd worden als de gemiddelde score op de beginmeting. De tweede fixed parameter, β_1 , geeft het gemiddelde verschil tussen begin- en eindmeting aan. Dat wil zeggen: de gemiddelde leerwinst voor 'de' leerling in de derde klas.

Naast beide fixed effecten worden in het random deel van het model zes residuele scores gespecificeerd: drie voor verschillen op de beginmeting, en drie voor verschillen in leerwinst. De residuele scores op leerling-, klas- en schoolniveau worden aangegeven met respectievelijk e_{0ijk} , u_{00jk} , en v_{000k} (met varianties: σ^2_{e0} , σ^2_{u0} en σ^2_{v0}). De verschillen in leerwinst op de drie niveaus komen tot uiting in de resterende residuele scores: e_{1ijk} , u_{10jk} en v_{100k} (met varianties: σ^2_{e1} , σ^2_{u1} en σ^2_{v1}). De laatste residuele score, v_{100k} , geeft de afwijking van school k van de gemiddelde leerwinst. En, u_{10jk} , representeert de afwijking voor klas j van de gemiddelde leerwinst van school k , terwijl e_{1ijk} de afwijking van leerling i ten opzichte van de leerwinst in klas j van school k representeert. Aan de correlaties tussen de twee residuele scores op elk niveau worden geen restricties opgelegd, zodat op alle drie de niveaus een complete covariantie-matrix geschat wordt, en voor alle elementen nagegaan kan worden of deze afwijken van nul.

De parameterschattingen voor dit multivariate leerwinstmodel (MULW), of model voor afwijkingsscores, zijn samengevat in Tabel 5.

De gemiddelde score op de beginmeting (β_0) bedraagt 99,8, en de gemiddelde afwijking voor de score aan het einde van het derde leerjaar is 9,2; beide parameters wijken (gezien de standaardfouten) af van 0. Dus: de gemiddelde score aan het begin van het derde leerjaar is niet gelijk aan nul, én, meer relevant, de gemiddelde score aan het einde van het derde leerjaar is hoger dan de gemiddelde score aan het begin van dit leerjaar.

De variantie tussen scholen aan het begin van het derde leerjaar is 9,6. De variantie in leerwinst tussen scholen is 18,3; en beide parameters wijken af van 0. Dus: scholen verschillen qua gemiddelde score aan het begin van het derde leerjaar, en de gemiddelde leerwinst in dit leerjaar verschilt van school tot school. De covariantie tussen de beide residuele scores op schoolniveau is niet significant. De gemiddelde score van een school aan het begin van het derde leerjaar heeft niets van doen met de gemiddelde leerwinst in het derde leerjaar.

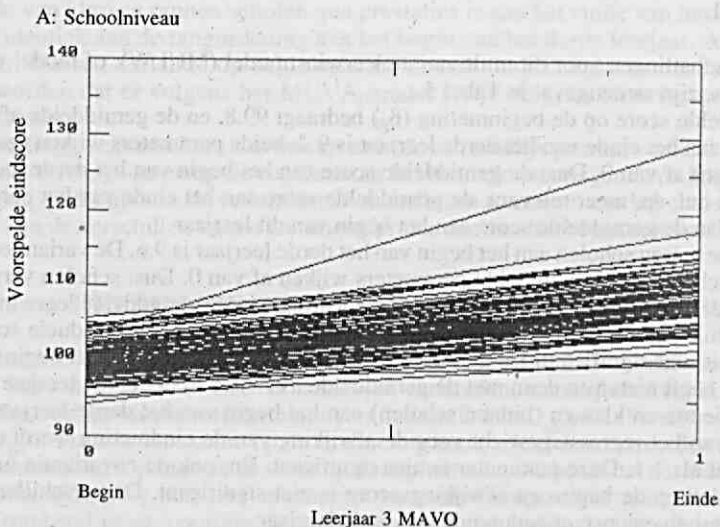
De variantie tussen klassen (binnen scholen) aan het begin van het derde leerjaar is 6,2. De variantie van het regressiegewicht voor de afwijking van de eindmeting wordt op klasniveau geschat als 1,1. Deze parameter is niet significant. En, ook de covariantie tussen de klasgemiddelen op de begin- en afwijkingsscore is niet significant. De verschillen tussen klassen zijn derhalve constant gedurende het derde leerjaar.

Tabel 5: Parameterschattingen voor het multivariate leerwinstmodel (MULW; zie: Vergelijking 6; standaardfouten tussen haakjes)

		Schatting (se)	
		Fixed parameters	
Intercept (β_0)		99.77	(0.45)
Leerwinst (β_1)		9.15	(0.55)
		Random parameters (correlaties boven de diagonaal)	
	Beginmeting	Schoolniveau	Leerwinst
Intercept (v_{000k})	9.61 (2.51)		
Leerwinst (v_{100k})	-1.47 (2.21)		-0.11
			18.34 (3.39)
		Klasniveau	
Intercept (u_{00jk})	6.19 (1.76)		-0.02
Leerwinst (u_{10jk})	-0.06 (1.21)		1.09 (1.51)
		Leerlingniveau	
Intercept (e_{0ijk})	208.73 (4.18)		-0.44
Leerwinst (e_{1ijk})	-103.20 (3.67)		269.07 (5.40)

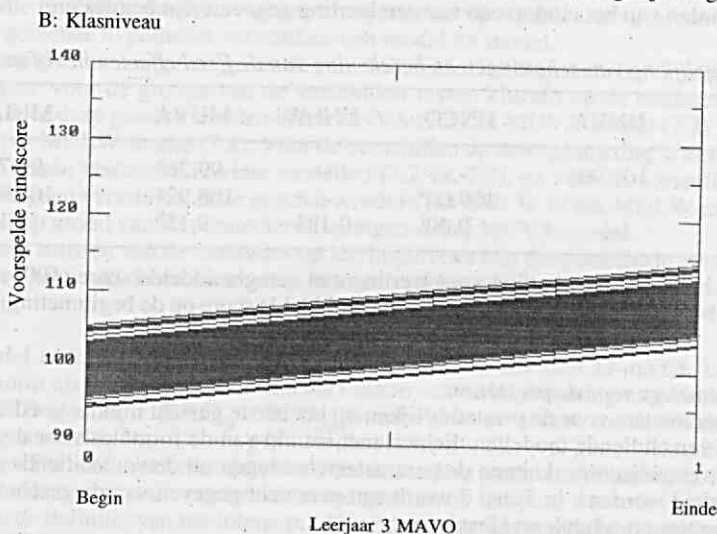
De beide variantieschattingen op leerlingniveau (beginmeting: 208.7; afwijkingsscore: 269.1) wijken af van nul. En, ook de covariantie tussen de begin- en de afwijkingsscore is significant. Dit impliceert dat leerlingen met een hoge beginscore overwegend een relatief lage afwijkingsscore hebben (en vice versa).

Figuur 2: De verschillen in leerwinst tussen scholen (onder) en klassen (volgende pag.) volgens het multivariate leerwinstmodel (MULW)



Ook deze geschatte verschillen tussen de eind- en beginmeting kunnen we visueel weergeven (zie: Figuur 2).

Uit Figuur 2A blijkt dat op schoolniveau de lijnen voor de verschillende scholen niet parallel lopen. Ergo, de relatie tussen de scores op de begin- en eindmeting is niet gelijk voor alle scholen; de leerwinst verschilt van school tot school (echter, vergelijk: Figuur 1A).



Uit Figuur 2B blijkt dat de verschillen tussen klassen aan het begin en het einde van het derde leerjaar even groot zijn; de lijnen die de verschillende klassen indiceren, lopen parallel. Er kan derhalve geconcludeerd worden dat de leerwinst niet verschilt van klas tot klas; de verschillen tussen klassen aan het begin van het derde leerjaar blijven intact gedurende dit leerjaar (echter, vergelijk: Figuur 1B).

7. DE MODELLEN VERGELEKEN

In deze paragraaf worden de fixed en random parameters van de vijf modellen met elkaar vergeleken. We beginnen met de fixed parameters voor de (gemiddelde) prestaties op de begin- en eindmeting, daarna komen de random parameters aan bod en ten slotte de parameters voor de fixed effecten van achtergrondkenmerken als sekse, zittenblijven, sociale en etnische achtergrond.

Fixed parameters voor de prestaties

Met de verschillende modellen worden andere hypothesen getoetst. De waarden van de fixed parameters verschillen dan ook van model tot model. Echter, na het nodige rekenwerk blijken de fixed effecten voor het gemiddelde prestatieniveau vergelijkbaar over de vijf modellen. In Tabel 6 wordt een overzicht gegeven van de geschatte fixed effecten, alsmede van de op grond van deze parameters berekende (geïnferreerde) waarden voor respectievelijk het (gemiddelde) eindniveau en de (gemiddelde) leerwinst.

Blijkens Tabel 6 maakt het voor het gemiddelde eindniveau (108.9) niet uit of dat nu geschat wordt in het univariate variantie-analytische model (UNVA), het multivariate variantie-analytische model (MUVA) of het multivariate leerwinst model (MULW). Ook is de gemiddelde leerwinst - iets meer dan negen scorepunten - nagenoeg gelijk in het univariate

leerwinstmodel (UNLW), in het MUVA- (108.90 - 99.78) en het MULW-model. In de verschillende modellen worden dus andere parameters geschat, en andere hypothesen getoetst.

Alleen bij het UNCO-model ligt aan de berekende waarde de aanname ten grondslag dat de gemiddelde leerling op de beginmeting ook de gemiddelde leerling op de eindmeting is. Daar staat tegenover dat in het UNCO-model, i.t.t. de andere modellen, wel een schatting gemaakt kan worden van het eindniveau van een leerling gegeven zijn beginscore.

Tabel 6: *Vergelijking van schattingen en berekening van de fixed effecten in vijf modellen*

	UNVA	UNCO	UNLW	MUVA	MULW
Beginmeting	108.90 ^g	--	--	99.76 ^g	99.77 ^g
Eindmeting	--	109.00 ^s	--	108.92 ^g	108.89 ^b
Leerwinst	--	9.00 ⁱ	9.18 ^g	9.15 ^b	9.12 ^g

^g: geschat; ^b: berekend; ^s: schatting voor leerling met een gemiddelde score (100) op de beginmeting; ⁱ: berekening voor leerling met een gemiddelde score op de beginmeting; --: geen informatie.

Random parameters voor de prestaties

De random parameters voor de prestaties lijken op het eerste gezicht minder goed vergelijkbaar over de verschillende modellen. Echter, met behulp van de formules voor de variantie van som- en verschilcores, kunnen de parameterschattingen uit de verschillende modellen tot elkaar herleid worden⁷. In Tabel 7 wordt een overzicht gegeven van de geschatte en berekende varianties op school- en klasniveau.

Tabel 7: *Vergelijking van schattingen en berekening van de random effecten in vijf modellen*

	UNVA	UNCO	UNLW	MUVA	MULW
Variantieschattingen op schoolniveau					
Beginmeting	--	--	--	9.61 ^g	9.61 ^g
Eindmeting	24.63 ^g	--	--	25.01 ^g	25.01 ^b
Leerwinst	--	18.17 ^s	18.13 ^g	18.34 ^b	18.34 ^g
Variantieschattingen op klasniveau					
Beginmeting	--	--	--	6.19 ^g	6.19 ^g
Eindmeting	7.27 ^g	--	--	7.15 ^g	7.16 ^b
Leerwinst	--	2.0 ^s	1.10 ^g	1.08 ^b	1.09 ^g
Variantieschattingen op leerlingniveau					
Beginmeting	--	--	--	208.73 ^g	207.36 ^g
Eindmeting	271.43 ^g	--	--	271.40 ^g	271.60 ^b
Leerwinst	--	251.70 ^s	269.10 ^g	269.07 ^b	269.07 ^g

^g: geschat; ^b: berekend; ^s: schatting voor leerling met een gemiddelde score (100) op de beginmeting; --: geen informatie

Na enig gepuzzel blijken de variantiecomponenten op school- en klas- en leerlingniveau zoals geschat in de diverse modellen redelijk overeenkomstig. Het maakt bij voorbeeld voor de variantie tussen scholen op de eindmeting niet uit of dit in het UNVA-model (24.6), of het

MUVA-model (25.0) geschat wordt, dan wel op grond van het MULW-model (25.0) berekend wordt (gezien de grootte van de standaardfouten zijn deze schattingen gelijkwaardig). Hetzelfde geldt voor de variantieschattingen voor de verschillen tussen scholen op de beginmeting (alleen MUVA en MULW), en de schattingen i.c. berekeningen van de verschillen in leerwinst tussen scholen (UNLW, MUVA en MULW). Wel wordt duidelijk uit Tabel 7 dat niet in alle modellen dezelfde variantiecomponenten op significantie getoetst (kunnen) worden; de getoetste hypothesen verschillen van model tot model.

Ook op klasniveau convergeren de verschillende schattingen i.c. berekeningen goed. Dat wil zeggen: voor de grootte van de verschillen tussen klassen op de eindmeting maakt het weinig uit of deze geschat worden in het UNVA (7.3), of MUVA-model (7.2), dan wel berekend in het MULW-model (7.2). Voor de verschillen op de beginmeting is er nauwelijks een verschil tussen beide multivariate modellen (7.2 vs. 7.2), en voor de variantie in leerwinst maakt het geen verschil of deze geschat wordt in het UNLW of het MULW-model, dan wel berekend op grond van de parameterschattingen in het MUVA-model.

Ook ten aanzien van de varianties op leerlingniveau kan deze conclusie gehandhaafd blijven. Voor geen van de drie varianties is er een substantieel verschil tussen de verschillende modellen.

Eén model is tot nog toe buiten beschouwing gelaten: het UNCO-model. In het UNCO-model wordt als enige het eindniveau als functie van de beginmeting uitgedrukt (middels de fixed coëfficiënt: $\beta_1 * [\text{BEG}_{ijk} - \text{BEG}_{000}]$), terwijl in de andere modellen de gemiddelde scores (gemiddelde beginscore, gemiddelde eindscore, dan wel gemiddelde leerwinst) centraal staan. Dit heeft ook consequenties voor de random parameterschattingen. In het UNCO-model wordt de variantie van het intercept geschat op dat punt waar de beginmeting nul is (conform de definitie van het intercept). Door de centrering van de beginmeting is dat in ons voorbeeld bij een score van 100 op de beginmeting. Indien blijkt dat de variantie van een covariaat op een niveau significant is, dan is de interceptvariantie, op dat niveau, afhankelijk van de plaats waarop het intercept gekozen is (zie: noot 3, en Reichardt, 1979, p. 155). Centrering of herschaling van de beginmeting heeft dus invloed op de variantieschattingen (zie: Figuur 1). In ons voorbeeld blijkt na centrering rond 100 de interceptvariantie niet significant. Echter, hadden we gekozen voor een centrering rond 102, dan was de interceptvariantie niet alleen groter ($\sigma^2_{u0} : 3.10$), maar gezien de standaardfout (se: 1.49) ook significant. Dus: niet alleen de grootte van de interceptvariantie, maar ook de toetsing of deze afwijkt van nul is afhankelijk van het schaalpunt waar deze gekozen wordt. De vraag of het eindniveau voor leerlingen met beginniveau X verschilt van klas tot klas (interceptvariantie!) hangt af van de waarde van X. In de huidige dataset kan de hypothese $H_0: (\sigma^2_{u0} | X = x) = 0$ verworpen worden voor $X \geq 102$. Dus, indien gecentreerd zou zijn rond 102, i.p.v. rond 100, dan was de interceptvariantie significant geweest; bij centrering rond 102 was geconcludeerd dat de verschillen tussen klassen voor leerlingen met een gelijk beginniveau significant is. En, traditioneel zou dan de conclusie getrokken zijn, dat de leerwinst verschilt van klas tot klas. Hopelijk is duidelijk dat deze conclusie voorbarig is. We kunnen namelijk hooguit concluderen dat het voor leerlingen met een gegeven beginscore (in ons voorbeeld groter dan 102) uitmaakt in welke klas zij zitten; voor leerlingen met een lagere beginscore kan niet aangetoond worden dat de interceptvariantie afwijkt van nul, en moet dus geconcludeerd worden dat het eindniveau voor deze leerlingen niet verschilt van klas tot klas.

Hoewel deze interpretatie recht-toe recht-aan is, willen wij er nog wel een kanttekening bij maken. Voor de variantie in hellingshoek kunnen verschillende alternatieve verklaringen naar voren gebracht worden. Zo zouden zij zowel tot stand gekomen kunnen zijn door bodem- en plafond-effecten, als door verschillen tussen klassen in de betrouwbaarheid van de beginmeting, als door (de geprefereerde) interactie tussen beginniveau en genoten onderwijs. Plafond-effecten zijn in het onderhavige voorbeeld een mogelijke verklaring, aangezien goede leerlingen (het gemiddelde plus 2 sd) er nauwelijks iets bij lijken te leren (zie noot 5).

Bodem-effecten lijken minder waarschijnlijk, maar (helemaal) uitgesloten zijn zij niet. Hetzelfde geldt voor verschillen in betrouwbaarheid van de beginmeting voor verschillende klassen; ook deze kunnen wij niet (geheel) uitsluiten. Daardoor kunnen de resultaten op basis van het UNCO-model (enigszins) biased zijn (vergelijk, Reichardt, 1979, p. 170).

Op schoolniveau geeft de variantie van het intercept aan in welke mate de scholen verschillen. Aangezien de regressie van eind- op beginmeting niet varieert van school tot school, is deze variatie onafhankelijk van het beginniveau. En, de verschillen in bereikt eindniveau tussen leerlingen op schoolniveau zijn onafhankelijk van hun beginniveau, maar wel afhankelijk van de school.

Samenvattend: in het UNCO-model wordt gedemonstreerd dat het eindniveau voor leerlingen met dezelfde beginscore lager dan 102 niet varieert klas tot klas, maar (mogelijk) wél voor leerlingen met een hogere score op de beginmeting. Waarbij opgemerkt zij dat bodem- en plafond-effecten, of verschillen tussen klassen in betrouwbaarheid van de beginmeting verantwoordelijk kunnen zijn voor de geobserveerde verschillen in hellingshoek.

Op schoolniveau blijkt het eindniveau voor leerlingen met een gelijk beginniveau te variëren van school tot school, en deze variatie in eindniveau tussen scholen is onafhankelijk van het beginniveau van een leerling. Hiermee levert het UNCO-model, waarin het eindniveau gemodelleerd wordt als (lineaire) functie van het beginniveau, een welkome aanvulling ten opzichte van de andere modellen waarin hetzij het gemiddelde eind- of beginniveau of de gemiddelde leerwinst gemodelleerd wordt. Slechts bij uitzondering (zie: Appendix) zullen het UNCO-model enerzijds en het UNLW-, MUVA- en het MULW-model anderzijds tot vergelijkbare uitkomsten leiden. Deze verschillen tussen MUVA- en MULW-modellen enerzijds en het UNCO-model anderzijds zijn een voorbeeld van wat wel Lord's paradox genoemd (Lord, 1967). Met dien verstande dat Lord zich beperkte tot de 'fixed' parameters (de verschillen in gemiddelde leerwinst). Echter, Lord's paradox strekt zich ook uit tot de random parameters.

Fixed parameters voor achtergrondkenmerken

De overeenkomsten en verschillen tussen de vijf modellen in te toetsen hypothesen komen ook tot uiting in de parameterschattingen voor de fixed effecten voor achtergrondkenmerken, zoals sekse, doubleren en sociale en etnische achtergrond. In het UNVA- en UNLW-model kunnen alleen regressies van de eindmeting respectievelijk leerwinst op de achtergrondkenmerken geschat worden. In de drie andere modellen kunnen twee effecten voor elk achtergrondkenmerk geschat worden. In het UNCO-model kan geschat worden of het eindniveau - conditioneel op beginprestaties en de overige achtergrondkenmerken - gelijk is voor verschillende typen leerlingen, en of de regressie van de eind- op de beginmeting varieert afhankelijk van het type leerling. In het MUVA-model kan vastgesteld worden of de verschillende typen leerlingen verschillen op de beginmeting en/of op de eindmeting. En, in het MULW-model tenslotte kan nagegaan worden of de verschillende typen leerlingen verschillen op de beginmeting en/of op de leerwinst. Hoewel in veel gevallen de parameterschattingen van het ene model afgeleid kunnen worden uit die van het andere model, zijn er toch subtiele verschillen die regelmatig over het hoofd gezien worden. Ter illustratie van de verschillende fixed effecten is Tabel 8 geconstrueerd. In deze tabel zijn de geschatte regressiegewichten voor enkele leerlingkenmerken weergegeven: sekse, doubleren in het basisonderwijs, doubleren in het eerste en/of tweede jaar van het voortgezet onderwijs, geboorteland ouders, thuistaal, en het gemiddelde opleidingsniveau van de ouders.

Sekse. Kijken we naar het effect van sekse, dan lijken de vijf modellen zo op het eerste gezicht verschillende uitkomsten op te leveren. Het UNVA-, MUVA- en MULW-model hebben gemeenschappelijk dat jongens aan het einde van het derde leerjaar gemiddelde 6.6 à 6.7 punten hoger scoren dan meisjes. Het UNCO-model wijkt hier duidelijk vanaf in de zin dat de voorsprong van jongens - conditioneel op de beginprestaties en de overige achtergrond-

kenmerken - tot 3.2 punten geslonken is (vergelijk: Hill & Rowe, 1996; Lockheed & Longford, 1991). Theoretisch geeft de fixed parameter in het UNCO-model aan hoe groot of hoe klein het gemiddelde prestatieverschil tussen jongens en meisjes zou zijn geweest in het hypothetische geval dat beide seksen een gelijke score zouden hebben behaald op de beginmeting (en op alle overige verklarende variabelen). Echter, daarbij geldt de aanname dat alle initiële verschillen inderdaad uitgepartialiseerd zijn (met behulp van de covariaat: de beginmeting), en dat is bij niet perfect betrouwbare metingen zelden het geval (zie ook: Lord, 1960). Dat hier niet alle initiële verschillen uitgepartialiseerd zijn, blijkt duidelijk uit de resultaten van het UNLW-, MUVA- en MULW-model; er is geen effect van sekse op de leerwinst in het UNLW- en MULW-model, en in het MUVA-model is het effect van sekse op de beginmeting nagenoeg gelijk aan dat op de eindmeting, en het netto resultaat is dus geen verschil in leerwinst tussen jongens en meisjes. Uit het MULW- en MUVA-model valt op te maken dat de verschillen tussen jongens en meisjes aan het einde van het derde leerjaar even groot zijn als aan het begin van dat leerjaar.

Tabel 8: De schattingen van verschillende fixed effecten per model

	Univariate var.analyse (UNVA)	Univariate covar.anal. (UNCO)	Univariate leerwinst (UNLW)	Multivariate var.analyse (MUVA)	Multivariate leerwinst (MULW)
Prestaties					
intercept	106.60 (.69)	107.50 (.69)	10.18 (.89)	--	94.56 (.75)
beginmeting	--	.50 (.03) ^a	--	94.56 (.75)	--
eindmeting	--	--	--	104.70 (.95)	10.13 (.89)
Sekse ($\rho : 0$; $\sigma^2 : 1$)					
intercept	6.61 (.46)	3.24 (.42)	-.14 (.47)	--	6.73 (.40)
beginmeting	--	.02 (.03)	--	6.73 (.40)	--
eindmeting	--	--	--	6.61 (.46)	-.13 (.46)
Doubleren bo (nee: 0; ja: 1)					
intercept	-1.66 (.75)	-1.56 (.68)	-1.37 (.76)	--	-.30 (.65)
beginmeting	--	.03 (.04)	--	-.30 (.64)	--
eindmeting	--	--	--	-1.62 (.74)	-1.33 (.76)
Doubleren vo (nee: 0; ja: 1)					
intercept	.02 (.52)	-.81 (.47)	-1.62 (.52)	--	1.65 (.45)
beginmeting	--	-.02 (.03)	--	1.65 (.45)	--
eindmeting	--	--	--	-1.62 (.51)	-1.60 (.52)
Geboorteland ouders (Nederland: 0; elders: 1)					
intercept	-2.73 (.76)	-.98 (.70)	.41 (.77)	--	-3.13 (.66)
beginmeting	--	--	--	-3.13 (.66)	--
eindmeting	--	.05 (.04)	--	-2.70 (.76)	.43 (.77)
Thuisstaal (Nederlands: 0; anders: 1)					
intercept	-6.07 (1.51)	-5.56 (1.41)	-4.19 (1.53)	--	-1.94 (1.31)
beginmeting	--	--	--	-1.94 (1.31)	--
eindmeting	--	-.15 (.08)	--	-6.10 (1.51)	-4.51 (1.53)
SES (laag (hooguit basisonderwijs): 1; hoog (universitair): 6)^b					
intercept	.61 (.21)	-.35 (.18)	-.09 (.21)	--	.70 (.18)
beginmeting	--	--	--	.69 (.18)	--
eindmeting	--	-.01 (.01)	--	.59 (.21)	-.09 (.21)

a: gecentreerd rond gemiddelde score op de beginmeting (d.w.z.: 100); b gecentreerd rond gemiddelde.

Merk op dat de schattingen van het effect van sekse in het UNVA- en UNLW-model eenvoudig te reconstrueren zijn op basis van de parameterschattingen in beide multivariate modellen. In het MUVA-model wordt net zoals in het UNVA-model het effect van sekse op de eindmeting direct geschat, en het verschil tussen beide parameterschattingen in het MULW-model geeft het verschil tussen jongens en meisjes op de eindmeting. Om de UNLW-schatting te reconstrueren kan het verschil genomen worden van de beide schattingen in het MUVA-model, maar het MULW-model geeft een directe schatting.

Doubleren. Nemen we bij voorbeeld het effect van doubleren in het voortgezet onderwijs dan blijken zittenblijvers volgens het UNVA- en het UNCO-model qua eindscore niet onder te doen voor niet-zittenblijvers. Volgens het UNLW-model is de leerwinst van zittenblijvers gemiddeld 1.6 punten lager dan die van niet-zittenblijvers. En, ook volgens het MUVA- en MULW-model is er een verschil tussen zittenblijvers en niet-zittenblijvers. Deze beide multivariate modellen leiden uiteindelijk tot dezelfde conclusie als het UNLW-model, maar bieden als extra informatie dat zittenblijvers aan het begin van het schooljaar 1.6 punten hoger scoren dan niet-zittenblijvers. Alleen op grond van het netto resultaat van deze beide effecten (MUVA: 1.65 - 1.62; MULW: 1.65 - 1.60) kan geen onderscheid gemaakt worden tussen beide groepen. Kennelijk raken zittenblijvers gedurende het schooljaar hun initiële voorsprong weer kwijt (vergelijk Kuhlemeier, Van den Bergh & Melse (1996) voor een vergelijkbare conclusie).

Thuistaal. Als laatste bespreken we hier het effect van thuistaal. Thuistaal heeft een effect op het eindniveau; leerlingen die thuis een andere taal dan het Nederlands spreken hebben aan het begin van het derde geen lagere score dan hun thuis Nederlands sprekende klasgenoten; noch in het MUVA-, noch in het MULW-model levert deze variabele een significante bijdrage aan de beschrijving van de gegevens. Aan het einde van het derde leerjaar hebben deze leerlingen wel degelijk een achterstand (UNVA en MUVA: 6.1). Hun leerwinst is duidelijk minder (UNLW: 4.2; MULW: 4.5). Uit het UNCO-model ten slotte blijkt dat een 'thuis-niet-Nederlands-sprekende-leerling' met de score 100 op de beginmeting gemiddeld 5.6 punten lager scoort op de eindmeting dan een klasgenoot die thuis wel Nederlands spreekt. Let wel, omdat de beginmeting random is op klasniveau kan dit effect verschillen van klas tot klas (vergelijk: noot 3).

8. CONCLUSIE

Aan de hand van een uitgewerkt empirisch voorbeeld is gedemonstreerd dat met vijf modellen die regelmatig gebruikt worden voor de analyse van 'leerwinst' deels andere hypothesen getoetst worden. Weliswaar zijn in de meeste gevallen de parameterschattingen van de vier modellen voor de analyse van gemiddelden (UNVA, UNLW, MUVA, en MULW) tot elkaar te herleiden, maar bij de interpretatie moet rekening gehouden worden met de specifiek getoetste hypothesen. Daarbij bieden de beide multivariate modellen (MUVA en MULW) meer mogelijkheden voor interpretatie dan de beide univariate modellen (UNVA en UNLW).

Het UNCO-model neemt een duidelijke uitzonderingspositie in. Hoewel dit model expliciete mogelijkheden biedt voor nuanceringen van de resultaten van modellen voor gemiddelden, zijn de parameterschattingen, met name indien de regressiecoëfficiënt voor de covariaat random is op een hoger niveau, niet zonder meer eenduidig te interpreteren. Het verdient dan ook aanbeveling om bij een keuze voor het UNCO-model ook één van beide multivariate modellen aan de gegevens op te leggen.

1. Opgemerkt zij echter dat de analyse in de andere opleidingstypen vergelijkbare resultaten oplevert.
2. De ratio van de absolute waarde van de parameterschatting en de bijbehorende standaardfout is groter dan 1.96: $|B| / \text{se}(B) > 1.96$; $p < .05$. Of voor random parameters: $\sigma^2 / \text{se}(\sigma^2) > 1.96$.
3. Dat wil zeggen: de variantie tussen klassen, gegeven de beginmeting, is heteroscedastisch. Voor elke score op de beginmeting kan de variantie tussen klassen geschat worden. In dit geval is dat: $\text{VAR}(\text{tussen klassen} | \text{BEGINSORE}) = 2.39 + 2 * 0.17 * \text{BEGINSORE} + 0.012 * \text{BEGINSORE}^2$. De variantie tussen klassen is derhalve een kwadratische functie van de beginscore. Dat wil zeggen: op het ene schaalpunt van de beginmeting is de variantie groter dan op het andere schaalpunt; voor sommige schaalpunten kan de variantie tussen klassen zelfs gelijk zijn aan nul (zie: Figuur 1B). In het onderhavige voorbeeld is de variantie tussen klassen minimaal voor een beginscore tussen 85 en 87.
4. Niet-lineaire relaties tussen begin- en eindmeting zijn niet significant.
5. In de klassen met de (5%) sterkste relatie tussen begin- en eindmeting (.69) is de score van een leerling met een hoge score op de eindmeting $109 + .69 * (130 - 100) = 130$, terwijl de score van een leerling met een lage score op de beginmeting $109 + .69 * (60 - 100) = 81$ is. Merk op dat door de centering van de beginmeting, het gemiddelde (100) van de beginscore van de leerling afgetrokken moet worden. In de (5%) klassen met de zwakste relatie tussen begin- en eindmeting is de score van een leerling met dezelfde hoge beginscore $109 + .33 * (130 - 100) = 119$, terwijl de score van een leerling met dezelfde lage beginscore $109 + .33 * (70 - 100) = 96$ is.
6. Hiervoor moet een zogenaamde contrast-toetsing uitgevoerd worden, waarbij gebruikgemaakt wordt van een te definiëren contrast-matrix. Met behulp van deze contrastmatrix C wordt de hypothese getoetst: $H_0: CB = 0$ tegen $CB \neq 0$, waarbij B een vector met geschatte regressiegewichten is. Voor het onderhavige voorbeeld geldt:

$$C = [1 \ -1 \ 0]$$

De toetsingsgrootheid χ^2 is dan $\hat{B}' C' (C \hat{\Sigma} C)^{-1} C \hat{B}$, waarbij met $\hat{\Sigma}$ de geschatte covariantie-matrix aangegeven wordt.

7. De variantie van somscores van toets X en Y is:

$$\text{VAR}(X + Y) = \text{VAR}(X) + 2 * \text{COVAR}(X, Y) + \text{VAR}(Y).$$

- De variantie van verschilcores van toets X en Y is:

$$\text{VAR}(X - Y) = \text{VAR}(X) - 2 * \text{COVAR}(X, Y) + \text{VAR}(Y).$$

9. LITERATUUR

- Bergh, H. van den & Hoeksma, J.B. (1993). Modelling development in education: A three level model. In J.H.L. Oud & R.A.W. van Blokland-Vogelsang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioural sciences*. Nijmegen: ITS.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw Hill.
- Bosker, R.J. & Snijders T.A.B. (1990). Statistische aspecten van multi-niveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-329.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.
- Burstein, L., Kyung-Aung, K. & Delandshire, G (1989). Multilevel investigation of systematically varying slopes: Issues, alternatives and consequences. In R.D. Bock (Ed.), *Multilevel analysis of educational data*. San Diego: Academic Press.
- Fennessey, J. & Salganik, L.H. (1983). Credible comparison of instructional impact: The rags procedure. *Educational Measurement: Issues and Practice*, 11, 13-17.

- Fraser, B.J., Walberg, H.J., Welch, W.W. & Hattie, J.A. (1987). Synthesis of educational productivity research. *International Journal of Educational Research*, 11, 145-252.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Hill, P.W. & Rowe, K.J. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement*, 7, 1-34.
- Goldstein, H. & McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, 53, 455-467.
- Hoeben, W.Th.J.G. (1994). Curriculum evaluation and productivity. *Studies in Educational Evaluation*, 20, 477-502.
- Huff, K. (1989). *A comparative analysis of achievement gains and related school characteristics*. Paper presented at the AERA on march 27-31 in San Francisco.
- Kreft, I.T.A. & Aschbacher, P.R. (1994). Measurement and evaluation in education: the value of multivariate techniques in evaluating an innovative high school reform program. *International Journal of Educational Research*, 21, 181-196.
- Kuhlemeier H., Melse, L. & Bergh, H. van den (1996). Comparison of two German language courses in Dutch secondary education. *Studies in Educational Evaluation*, 22, 181-205.
- Kuhlemeier, H., Bergh, H. van den, Notté, H., Wagenaar, H., Verstralen, H. & Cappers, R. (1994). *Balans van het aardrijkskunde-onderwijs in het derde leerjaar van het voortgezet onderwijs*. Arnhem: Cito.
- Kuyper, H. & Veenstra, D.R. (1995). De effectiviteit van NT2-onderwijs aan gevorderde tweede taalleerders in het voortgezet onderwijs. In: B. Creemers e.a. (Red.), *Onderwijsonderzoek in Nederland en Vlaanderen*. Groningen: GION.
- Levine, D.U. & Lezotte, L.W. (1990). *Unusually effective schools. A review and analysis of research and practice*. Madison: National Center for Effective Schools Research and Development.
- Lockheed, M.E. & Longford, N.T. (1991). School effects on mathematics achievement. In: S.W. Raudenbush & J.D. Willms (Eds.), *Schools classrooms, and pupils; International studies of schooling from a multilevel perspective*. San Diego: Academic Press Inc.
- Lord, F.M. (1960). Large sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association*, 55, 307-321.
- Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68, 304-305.
- Reichardt, C.S. (1979). The statistical analysis of data from nonequivalent group designs. In: T.D. Cook & D.T. Campbell, *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Willms, J.D. (1992). *Monitoring school performance: A guide for educators*. Washington, D.C.: Falmer.

APPENDIX

Het UNCO-model en de beide leerwinstmodellen leiden in één situatie wel tot vergelijkbare uitkomsten. Dat is het geval wanneer er a) een één-op-één relatie is tussen een begin- en eindmeting (d.w.z.: de regressiecoëfficiënt is gelijk aan 1.00), en b) deze regressiecoëfficiënt niet varieert tussen scholen of klassen ($\sigma_{01}^2 = 0$). Wanneer we voor het gemak uitgaan van een model met twee niveaus, waarbij Y_{ij} de score van leerling i in klas j op de eindmeting aangeeft, en X_{ij} de score van dezelfde leerling op de beginmeting, dan kan het te analyseren covariantie-analyse model in termen van gemiddelden geschreven worden als:

$$\bar{Y}_j = \beta_{0j} + \beta_{1j} * \bar{X}_j \quad (A.1)$$

In vergelijking (A.1) is β_{0j} het intercept, en β_{1j} de regressiecoëfficiënt van de eindmeting op de beginmeting. Wanneer we vergelijking (A.1) schrijven als multivariaat leerwinstmodel, en de index h gebruiken om de voorrespectievelijk eindmeting aan te geven, krijgen we (in termen van gemiddelden):

$$\bar{Y}_{hj} = \beta_{2j} + \beta_{3j} * Z_{hj} \quad (A.2)$$

waarbij Z_{hj} een dummy variabele indiceert welke 'aan' staat als de score een de eindmeting betreft ($Z_{hj} = 1$), en 'uit' staat als het de beginmeting betreft ($Z_{hj} = 0$). In feite worden in vergelijking (A.2) dus twee vergelijkingen in één gepresenteerd. Voor de beginmeting geldt:

$$\bar{Y}_{0j} = \beta_{2j} + \beta_{3j} * 0 \quad (A.2a)$$

en voor de eindmeting:

$${}_{ij}\bar{Y}_{1j} = \beta_{2j} + \beta_{3j} * 1, \quad (\text{A.2b})$$

In vergelijking (A.2b) representeert \bar{Y}_{1j} natuurlijk \bar{Y}_j uit vergelijking (A.1). En, in vergelijking (A.2a) representeert Y_{0j} natuurlijk \bar{X}_j uit vergelijking (A.1); β_{2j} is dus gelijk aan \bar{X}_j . Vergelijking (A.2b) kan dus herschreven worden als:

$$\bar{Y}_j = \bar{X}_j + \beta_{3j} * 1, \quad (\text{A.3})$$

wat hetzelfde is als:

$$\bar{Y}_j - \bar{X}_j = \beta_{3j}. \quad (\text{A.4})$$

Na enige manipulatie kunnen we vergelijking (A.1) schrijven als:

$$\bar{Y}_j - \bar{X}_j = \beta_{0j} + \bar{X}_j (\beta_{1j} - 1). \quad (\text{A.5})$$

Substitutie van vergelijking (A.4) in (A.5) levert als resultaat:

$$\beta_{3j} = \beta_{0j} + \bar{X}_j (\beta_{1j} - 1). \quad (\text{A.6})$$

Als β_{1j} ongelijk is aan 1 (én ten minste één van de \bar{X}_j 's ongelijk is aan 0), dan is de waarde van β_{3j} afhankelijk van \bar{X}_j . In dat geval zullen de resultaten van het multivariate leerwinstmodel en de covariantie-analyse dan ook moeten verschillen.

Aangezien de verschillende regressiegewichten (β_{0j} , β_{1j} en β_{3j}) geschreven kunnen worden als afwijking van schattingen van de populatieparameters (bij voorbeeld: $\beta_{0j} = \gamma_{00} + u_{0j}$), impliceert Vergelijking A.6 eveneens dat de variantieschattingen verschillen tussen de beide modellen indien β_{1j} ongelijk is aan één.

Alleen in het geval β_{1j} gelijk is aan 1, dan is de 'winst' in die klas gelijk aan het intercept van de covariantie-analyse van die klas. En, alleen in dat geval zullen beide modellen leiden tot eenzelfde oplossing.

Televisiekijken en lezen

Het effect van televisiekijken op de leesfrequentie en de leesprestaties in leerjaar 5 tot en met 8 van het basisonderwijs

Martha E. Otter, Rob Schoonen en Kees de Glopper
SCO-Kohnstamm Instituut van de Universiteit van Amsterdam

ABSTRACT

In a four-year longitudinal study, we investigated the effects of television viewing on leisure time reading and on reading achievement. This study was carried out in grades 3, 4, 5 and 6 of primary schools. Reading achievement was measured five times: at the beginning and end of grade 3 and at the end of grade 4, 5 and 6. Leisure time reading and television viewing were measured by means of diaries in grades 3, 4, 5 and 6 during eight, twelve, six and ten weeks, respectively. To investigate the relationships between television viewing, leisure time reading and reading achievement covariance structure analysis with LISREL is used (Jöreskog & Sörbom, 1988). The hierarchical structure of the data (students within classes) is taken into account by means of a multilevel procedure. In the early grade only, some very small effects of television viewing on leisure time reading were found. No effects whatsoever were found on the reading achievement of the children.

INLEIDING

Het leesgedrag van kinderen en volwassenen mag zich de laatste jaren verheugen in een ruime belangstelling. Begin 1996 bleek uit een rapport van het Sociaal Cultureel Planbureau dat er nog maar weinig gelezen wordt *om het lezen*. We kunnen tegenwoordig weliswaar niet meer om teksten heen: folders, brochures, handleidingen, ondertiteling etc., maar het lezen van boeken is over generaties ver teruggelopen (Knulst & Kraaykamp, 1996). Het 'lezen' lijkt meer en meer verdrongen te worden door het 'kijken' naar televisie of video. Deze ontwikkeling wordt over het algemeen als zorgelijk beschouwd.

Deze oprukkende kijkcultuur wordt behalve voor een vermindering van het lezen in de vrije tijd, ook verantwoordelijk gehouden voor de herhaaldelijk geconstateerde matige leesvaardigheid van leerlingen in het onderwijs (zie voor een overzicht Beentjes & Van der Voort, 1988, 1989). Beentjes en Van der Voort (1988, 1989) wijzen er echter op dat deze voornamelijk Amerikaanse onderzoeksresultaten niet zonder meer gegeneraliseerd mogen worden naar de Nederlandse situatie. Het blijft de vraag of de negatieve invloed van televisiekijken op de leesfrequentie en de leesprestaties ook in Nederland, opgaat. Zo kijken bijvoorbeeld Nederlandse jongeren beduidend minder naar de televisie dan Amerikaanse (Elley, 1994). Bovendien zijn de meeste Amerikaanse onderzoeken crosssectioneel van opzet waardoor een causale interpretatie van verbanden problematisch is (Beentjes & Van der Voort, 1988, 1989).

Sinds kort zijn er voor de Nederlandse situatie gegevens beschikbaar die enig licht werpen op de relatie tussen televisiekijken, leesfrequentie en leesvaardigheid. Koolstra (1993) heeft een tweejarige, longitudinale studie verricht naar de relaties tussen deze drie variabelen, waarbij leesfrequentie betrekking heeft op *buitenschools* lezen. In zijn onderzoek zijn twee

leeftijdsgroepen onderzocht (Cohort 1 en 2). Ten tijde van het eerste onderzoeksjaar zaten de kinderen in groep 4 (N=522) respectievelijk groep 6 (N=528). Met een tijdsinterval van steeds één jaar zijn drie metingen verricht. De hoofdvariabelen (telesiekijken, leesfrequentie en leesvaardigheid) zijn in zijn onderzoek als volgt geoperationaliseerd: een vragenlijst voor de frequentie van het telesiekijken, een vragenlijst voor de frequentie van het buitenschools lezen in boeken en ten slotte toetsen voor begrijpend lezen. Twee belangrijke onderzoeksvragen in de studie van Koolstra zijn: (1) 'Oefent de frequentie waarmee kinderen telesiekijken in een bepaald leerjaar X (i.c. leerjaar 4, 5, 6 of 7) een negatief effect uit op de leesfrequentie in het daaropvolgende leerjaar, X+1 (i.c. leerjaar 5, 6, 7 of 8)?' en (2) 'Oefent de frequentie waarmee kinderen telesiekijken in een bepaald jaar X een negatief effect uit op de leesprestaties in het daaropvolgend jaar X+1?'. Koolstra gaat dus uit van een uitgesteld effect van telesiekijken op de leesfrequentie en de leesprestaties.

Voor het effect van telesiekijken op de leesfrequentie is dit opmerkelijk, want men zou verwachten dat als men veel telesiekijkt, er minder tijd overblijft om (buitenschools) te lezen; de vrije tijd is nu eenmaal eindig en daarom zou men eerder direct (d.w.z. binnen hetzelfde onderzoeksjaar) een effect van telesiekijken op de leesfrequentie verwachten.

Dat veel telesiekijken (en daarom misschien weinig buitenschools lezen) niet meteen leidt tot slechte leesprestaties, is plausibel. Het idee is dat kinderen minder oefenen en minder ervaring opdoen in het lezen wat ertoe kan leiden dat hun leesprestaties niet optimaal zijn. Hoewel de relatie tussen het buitenschools of recreatief lezen en de leesprestaties op school discutabel is (vgl. Otter, 1993), moet men toch aannemen dat - als er al een effect op de leesprestaties optreedt - dat effect niet van de ene op de andere dag waarneembaar zal zijn. Koolstra's veronderstelling dat het om uitgestelde effecten op de leesprestaties gaat, is daarmee zeer plausibel.

Directe effecten komen in Koolstra's onderzoek niet aan de orde en - zoals hiervoor opgemerkt - is dat met name voor de relatie tussen kijkfrequentie en leesfrequentie opmerkelijk. Koolstra gaat ervan uit dat de kijkfrequentie via mediërende variabelen effect sorteert op de leesfrequentie. Als mediërende variabelen worden psychologische processen beschouwd als toenemende passiviteit of veranderende attitude of behoefte aan visuele informatie, die er uiteindelijk toe leiden dat lezen onaantrekkelijk of te inspannend wordt gevonden. Dergelijke psychologische kenmerken van personen veranderen ook niet van de ene op de andere dag (evenmin als de leesvaardigheid) en gegeven hun veronderstelde mediërende rol, onderzoek Koolstra uitsluitend hypothesen die uitgaan van uitgestelde effecten. Deze theoretische veronderstellingen van Koolstra laten overigens onverlet dat de vrije tijd eindig is en in principe maar een keer besteed kan worden. Men moet dus concluderen dat er voor zowel directe als uitgestelde effecten van telesiekijken op de leesfrequentie argumenten aan te voeren zijn.

Koolstra (1993) concludeert op basis van zijn data dat telesiekijken in het ene jaar (X) het lezen van boeken in het volgende jaar (X+1) doet verminderen. Dit geconstateerde effect van telesiekijken op lezen is nagenoeg onafhankelijk van leeftijd, geslacht, sociaal milieu en intelligentie van de kinderen. Evenmin maakt het uit naar welk type programma's er gekeken wordt: zowel het kijken naar amusement, drama, informatieve als het kijken naar kinderprogramma's heeft een negatieve invloed op het lezen van boeken, zo blijkt uit Koolstra's data.

Hoewel het geconstateerde effect van telesiekijken op de leesfrequentie statistisch significant is, moet men ook vaststellen dat het aangetoonde effect erg klein is: de kijkfrequentie verklaart slechts twee à tweeënehalf procent van de variantie in de leesfrequentie¹.

Voor het effect van het telesiekijken op de leesprestaties komt Koolstra tot soortgelijke conclusies. Ook hier constateert hij dat het telesiekijken een negatief effect heeft op de ontwikkeling van de leesprestaties. De negatieve invloed van telesiekijken is wederom nauwelijks afhankelijk van leeftijd, sociaal milieu, geslacht en intelligentie. Maar ook nu is het effect van telesiekijken klein te noemen: de kijkfrequentie verklaart minder dan één procent van de variantie in de leesprestaties² (zie ook noot 1).

De geringe effecten van telesiekijken op de leesfrequentie en leesprestaties zijn in zekere

zin teleurstellend. Een belangrijke, althans vaak geopperde, 'oorzaak' voor de teloorgang van het buitenschools lezen en de matige leesprestaties blijkt nauwelijks samen te hangen met het leesgedrag en de leesprestaties.

Hiervoor merkten we al op dat met name voor de verklaring van het leesgedrag een direct effect van televisiekijken evenzeer verwacht kan worden als een uitgesteld effect, dat wil zeggen dat de kijkfrequentie in jaar X een effect heeft op de leesfrequentie in datzelfde jaar X. Met betrekking tot het effect van televisiekijken op de leesprestatie hebben we eerder geconcludeerd dat een uitgesteld effect het meest *plausibel* is, onder meer omdat de leesvaardigheid een redelijk stabiele trek is die niet van de ene op de andere dag zal veranderen. Dit laat onverlet dat een direct effect van televisiekijken op de prestaties ook mogelijk is. Omdat de door Koolstra gerapporteerde effecten zeer klein zijn, mag men op voorhand niet uitsluiten dat de directe relatie tussen kijkfrequentie en leesprestatie hechter is dan men zou denken en dus een lage kijkfrequentie zich datzelfde jaar ook manifesteert in zwakke leesprestaties. Het lijkt daarom zinvol om een alternatief model met directe effecten van televisiekijken op het lezen op zijn juistheid te toetsen, zowel voor de leesfrequentie als voor de leesprestaties.

Daarbij is nog een kanttekening bij het onderzoek van Koolstra van belang. Koolstra heeft de lees- en kijkfrequentie gemeten met vragenlijsten die vragen bevatten waarvan de antwoordcategorieën vaak ambigu zijn en/of betrekking hebben op een lang tijdsbestek. Dergelijke kenmerken van vragen in enquêtes kunnen de validiteit van de antwoorden ernstig bedreigen volgens recente theorieën over het vraag- en antwoordproces (Cannel, Miller & Oksenberg, 1981; Graesser & Murachver, 1985; Tourangeau & Rasinski, 1988, Van der Pligt & De Vries, 1995). Negatieve effecten van de genoemde karakteristieken zijn ook empirisch aangetoond (Otter, Mellenbergh & De Glopper, 1995)³.

Het is denkbaar dat mogelijk geringe validiteit van de gebruikte meetinstrumenten het aantonen van (grotere) effecten heeft belemmerd. Wellicht leiden validere metingen van de lees- en kijkfrequentie tot duidelijker aantoonbare effecten van televisiekijken op het leesgedrag en de leesprestaties. Bij dit soort operationalisaties verdient het de voorkeur respondenten concrete vragen te stellen over perioden die nog eenvoudig uit het geheugen op te halen zijn. Tijdbudgetonderzoek of gebruik van adequate 'gisteren'-vragen in dagboeken zouden tot validere informatie kunnen leiden.

Samenvattend kan gesteld worden dat de vraag naar de mogelijke effecten van televisiekijken op de leesfrequentie en leesprestaties voor de Nederlandse context nog niet eenduidig en volledig beantwoord is. Nieuwe gegevens met andere operationalisaties kunnen verder bijdragen aan het inzicht in de relaties tussen de genoemde variabelen.

In een vierjarige longitudinale studie bij leerlingen in leerjaar 5 tot en met 8 van het basisonderwijs zijn met behulp van dagboeken en toetsen gegevens verzameld over hun lees- en kijkgewoonten en leesprestaties (Otter, 1993). Deze longitudinale data stellen ons in staat met andere, nieuwe gegevens de relatie tussen televisiekijken en lezen nader te onderzoeken. Hierbij gaat het in de eerste plaats om de volgende twee hoofdvragen (1a en 2a), die elk nog een aanvullende vraag met zich meebrengen (1b en 2b):

1. a) Heeft de frequentie waarmee kinderen televisiekijken een effect op de *frequentie* waarmee zij boeken lezen?
b) Is er hierbij sprake van een *direct* (i.e. binnen jaar X) dan wel een *uitgesteld* effect (van jaar X naar jaar X+1)?
2. a) Heeft de frequentie waarmee kinderen televisiekijken een effect op de *leesprestaties*?
b) Is er hierbij sprake van een *direct* (i.e. binnen jaar X) dan wel een *uitgesteld* effect (van jaar X naar jaar X+1)?

Om deze vragen te kunnen beantwoorden worden modellen gepostuleerd waarin uitgegaan wordt van uitgestelde dan wel directe effecten van televisiekijken op de leesfrequentie

respectievelijk de leesprestatie. Vergelijking van de passing van de twee typen modellen bij de data kan antwoord geven op de vraag of er sprake is van directe of uitgestelde effecten. Schatting en toetsing van regressiegewichten in deze modellen kan inzicht geven in de grootte en statistische significantie van eventuele effecten.

METHODE

Dit onderzoek is een uitbreiding van het onderzoek van Otter (1993, 1995). De steekproef van scholen (klassen) en leerlingen, het onderzoeksontwerp, de dataverzameling en het instrumentarium worden daarom hier slechts beknopt besproken, voor verdere informatie verwijzen we naar eerdergenoemde publicaties.

Scholen en leerlingen

Bij aanvang van het longitudinale onderzoek in september 1990 waren de respondenten 9-jarige leerlingen in het basisonderwijs (groep 5). Voor de steekproefopzet is Nederland in tien regio's ingedeeld. Een paar moeilijk bereikbare gebieden zijn vanwege hun ligging niet bij een van de regio's ingedeeld.

Vervolgens zijn per regio aselekt twintig scholen getrokken. Van deze twintig scholen zijn er steeds vier benaderd voor deelname aan het project. De benadering van de scholen was in strikte zin niet aselekt, omdat de proefleider van een regio (om praktische redenen) de scholen die het dichtst bij hem of haar in de buurt lagen als eerste benaderd heeft voor deelname. Er is overigens weinig reden om aan de representativiteit van de steekproef te twijfelen gezien de leesprestaties van de kinderen op de landelijk genormeerde leestoetsen (Otter & Schoonen, 1996a).

In het schooljaar 1990/1991 deden per regio vier scholen mee, 40 scholen met in totaal 42 leerkrachten en 973 leerlingen. Tijdens de vierjarige studie zijn leerlingen en scholen om verschillende redenen uitgevallen. Tabel 1 geeft het verloop van de leerlingaantallen gedurende het onderzoek. Uit de tabel blijkt dat de voornaamste reden voor uitval 'schooluitval' is geweest, in totaal 21%. Een verklaring hiervoor is dat de studie als een eenjarige studie was opgezet (zie De Gloppe, 1989; Otter 1993). Aan de scholen werd dus in eerste instantie medewerking gevraagd voor één jaar. Toen aan het einde van leerjaar 5 deelname werd gevraagd voor leerjaar 6 haakten vier scholen af door de komst van onervaren leerkrachten (in leerjaar 6) en schoolfusies.

Door het ontbreken van financiële middelen aan het einde van leerjaar 6 was niet duidelijk of het project kon worden voortgezet in leerjaar 7 en 8. Toen halverwege leerjaar 7 voortzetting van het project werd gewaarborgd, bleek dat vijf scholen verplichtingen waren aangegaan in andere onderzoeksprojecten. Om de school niet verder te belasten, zagen zij af van verdere deelname aan het onderhavige (arbeidsintensieve) project. Eén school haakte in leerjaar 8 af vanwege een langdurige ziekte van de leerkracht en de komst van een onervaren invaller⁴.

Incomplete datasets is een andere bron voor uitval, verantwoordelijk voor 14% van het totaal. Met name in leerjaar 5 lijkt het percentage incomplete datasets aan de hoge kant. De reden hiervoor is dat in leerjaar 5 de leerlingen twee maal zijn getoetst, zowel aan het begin als aan het einde van het leerjaar en dus kans op uitval groter is.

De derde bron van uitval vormen verhuizingen, doublures en verwijzingen naar het LOM, in totaal 12%.

Onderzoeksontwerp

In leerjaar 5 zijn in oktober 1990 en mei 1991 leestoetsen afgenomen (zie Tabel 2: T1 en T2). In leerjaar 6, 7 en 8 zijn leestoetsen in juni afgenomen (T3, T4 en T5). Gedurende het vierjarige longitudinale onderzoek hebben de leerlingen hun leesgedrag (leesboeken) en hun kijk-

gedrag (televisie en video) in dagboeken bijgehouden (zie Tabel 2: D01 t/m D36). In leerjaar 5 gedurende acht weken (D01 - D08), in leerjaar 6, 7 en 8 gedurende respectievelijk twaalf (D9 - D20), zes (D21- D26) en tien weken (D27 - D36).

Tabel 1: Overzicht van aantallen leerlingen per onderzoeksjaar en de uitval uitgesplitst naar drie categorieën. Steekproefgrootte bij aanvang: 973

Leerjaar	Bron van uitval	Uitval		In steekproef		aantal scholen
		N	%	N	%	
5 (1990/1991)	Incomplete datasets	87	9	886	91	40
6 (1991/1992)	Verhuisd/gedoubleerd/LOM	57	6	829	85	36
	Schooluitval	89	9	740	76	
	Incomplete dataset	9	1	731	75	
7 (1992/1993)	Verhuisd/gedoubleerd/LOM	31	3	700	72	31
	Schooluitval	102	10	598	61	
	Incomplete datasets	32	3	566	58	
8 (1993/1994)	Verhuisd	27	3	539	55	30
	Schooluitval	14	1	525	54	
	Incomplete datasets	10	1	515	53	
	Totaal	---	--	---	--	
		448	47	515	53	

Uit Tabel 2 blijkt dat in leerjaar 7 gedurende slechts zes weken het lees- en kijkgedrag met behulp van dagboeken is gemeten en dat deze dagboekmetingen uitsluitend in de tweede helft van het leerjaar zijn verricht. De reden hiervoor is de eerdergenoemde onzekerheid over de continuering van het project bij aanvang van het derde onderzoeksjaar.

Tabel 2: Design van de afname van de leestoetsen (T1 t/m T5) en metingen van het lees- en kijkgedrag (D01 t/m D36)

Onderzoeksjaar	groep	oktober-november	januari- juni	mei-juni
1990-1991	5	T1 D01 - D03	D04 - D08	T2
1991-1992	6	D09 - D12	D13 - D20	T3
1992-1993	7		D21 - D26	T4
1993-1994	8	D27 - D30	D31 - D36	T5

Dataverzameling

Proefleiders begeleiden de dataverzameling die de leerkrachten in hun klas uitvoerden. Alle proefleiders waren afkomstig uit het onderwijs en waren ervaren in het afnemen van toetsen en vragenlijsten.

Aan het begin van elk onderzoeksjaar hebben de proefleiders persoonlijk contact gezocht met de leerkrachten. In dit contact werd het draaiboek van het onderzoek besproken. Tevens werden de dagboeken van de leerlingen mondeling toegelicht. Voorafgaande aan deze bespreking hadden de leerlingen een aantal dagen de dagboeken bijgehouden, waardoor eventuele problemen besproken konden worden. De proefleiders hadden vóór en na afloop van

elke dagboekweek telefonisch contact met de leerkrachten. Hierdoor konden problemen snel gesignaleerd en opgelost worden.

Instrumentarium

Hieronder wordt het instrumentarium van het onderzoek toegelicht. Als eerste worden de leestoetsen besproken. Vervolgens komen de dagboeken aan bod waarmee de lees- en kijkfrequentie gemeten is.

Leestoetsen

De leestoetsen aan het begin en einde van leerjaar 5 waren identiek. De toets is ontwikkeld onder auspiciën van de International Association for Educational Achievement (IEA) in het kader van een internationaal onderzoek naar de opbrengsten van het onderwijs in begrijpend lezen (Elley, 1994; De Glopper en Otter, 1993). De toets bevatte vragen bij drie soorten teksten: verhalende teksten, zakelijke teksten en documenten.

De leesvaardigheid van de leerlingen aan het einde van leerjaar 6 is gemeten met een selectie uit de zakelijke en verhalende teksten van de IEA leesvaardigheidstoets en met een door het Cito ontwikkelde toets voor begrijpend lezen voor einde leerjaar 6 (toets E4).

De leesvaardigheid is aan het einde van leerjaar 7 gemeten met behulp van de hierboven al genoemde Cito-toets E4 en met Cito-toets E5.

Aan het einde van leerjaar 8 is de leesvaardigheid gemeten met Cito-toets E5 in combinatie met een selectie van teksten en vragen uit een toets die in het kader van het IEA-onderzoek ontwikkeld is voor leerlingen uit het tweede leerjaar van het voortgezet onderwijs (Elley, 1994; De Glopper & Otter, 1993). De selectie van teksten bestond uit drie verhalende en drie zakelijke teksten met in totaal 40 toetsvragen (zie Tabel 3).

Tabel 3: Beschrijving leestoetsen in leerjaar 5 tot en met 8: aantal teksten en vragen

Toetsmoment	Toets	Teksten	Vragen
Begin leerjaar 5	EA-toets 9-jarigen	21	66
Einde leerjaar 5	IEA-toets 9-jarigen	21	66
Einde leerjaar 6	IEA-toets 9-jarigen	7	33
	Cito-toets E4	6	25
Einde leerjaar 7	Cito-toets E4	6	25
	Cito-toets E5	6	25
Einde leerjaar 8	Cito-toets E5	6	25
	IEA-toets 14-jarigen	6	40

Om de toetsafnames te standaardiseren (over proefleiders en in de tijd) zijn alle afnames aan de hand van een draaiboek afgenomen.

Scoring. Alle vragen kenden een dichotome scoring in goed of fout, waarbij niet gemaakte vragen, ook aan het einde van de toets, als fout gerekend zijn.

Lees- en kijkfrequentie

Om redenen die in de inleiding al genoemd zijn, zijn de gegevens omtrent de lees- en kijkfrequentie van de kinderen verzameld met behulp van dagboekjes. Hieronder wordt in het kort uiteengezet hoe deze gegevens verzameld zijn. Tevens wordt er ingegaan op een verschil

tussen de dagboeken in leerjaar 5 en de overige leerjaren en op de scoringswijze van de dagboekgegevens. Ten slotte wordt aangegeven hoe er met ontbrekende waarnemingen is omgegaan.

Procedure. Alle leerlingen kregen voor elke onderzoekswEEK (zie Tabel 2) in leerjaar 5 tot en met 8 één dagboekje. Het boekje begon op dinsdag met vragen over hun lees- en kijkgedrag op maandag en eindigde de daaropvolgende week op maandag met vragen over hun gedrag op vrijdag, zaterdag en zondag. Voor elke dag was er één bladzijde in het dagboekje gereserveerd. De dagboekjes werden aan het begin van de ochtend ingevuld en het invullen nam zes à zeven minuten in beslag.

Verschillen tussen de dagboeken in leerjaar 5 en de overige leerjaren. In leerjaar 5 was elke schooldag onderverdeeld in drie dagdelen: (1) vóór schooltijd (2) ná schooltijd maar vóór het avondeten en (3) ná het avondeten. Na inspectie van de gegevens van de acht dagboekweken in 1990/1991 werd de conclusie getrokken dat met name lezen voor schooltijd zéér infrequent voorkwam. Daarom is besloten om het voorschoolse dagdeel in leerjaar 6 tot en met 8 achterwege te laten.

In alle onderzoeksjaren waren de weekeinddagen onderverdeeld in twee dagdelen: (1) vóór het avondeten en (2) na het avondeten.

Scoring. De dagboeken van de leerlingen zijn als volgt gescoord. Bij de vraag of kinderen gelezen, dan wel televisiegekeken hebben, is aan elk 'ja' antwoord de score 1 toegekend, aan elk 'nee' antwoord de score 0 (nul). Voor zowel het lezen van boeken als het televisiekijken zijn dag- en weekcores berekend. De maximale dagscore was voor schooldagen in leerjaar 5 drie en in leerjaar 6 tot en met 8 twee. De maximale dagscore in het weekeinde was in alle onderzoeksjaren gelijk, namelijk twee. De maximale weekscore bedroeg in leerjaar 5 negentien ($5 \cdot 3 + 2 \cdot 2$) en in leerjaar 6 tot en met 8 veertien ($7 \cdot 2$). De maximale jaarscore is steeds het aantal onderzoeksweken maal de maximale weekscore; in leerjaar 5 is dit 152 ($8 \cdot 19$), in leerjaar 6 168 ($12 \cdot 14$), in leerjaar 7 84 ($6 \cdot 14$) en in leerjaar 8 140 ($10 \cdot 14$).

Ontbrekende waarnemingen. De ontbrekende waarnemingen van de leerlingen zijn per gedragsmeting (lezen en televisiekijken) in alle onderzoeksjaren eerst op weekniveau verwerkt: indien méér dan de helft van de negentien (leerjaar 5) of veertien scores (leerjaar 6 tot en met 8) van een betreffende week ontbrak, is géén weekscore berekend, maar de week als 'ontbrekend' beschouwd. Als er minder dan de helft ontbrak, is de weekscore berekend door de ontbrekende scores te vervangen door het gemiddelde van het kind in die betreffende week. Tussen de 5 en 14% van de leerlingen hebben hiermee te maken gehad.

Bij de berekening van de jaarscores is dezelfde procedure toegepast, zij het dat nu al bij het ontbreken van meer dan 25% van de weekcores, de leerlingen uit het bestand verwijderd zijn (in totaal bijna tien procent van de leerlingen). In de andere gevallen zijn de ontbrekende weekcores vervangen door het gemiddelde van de resterende weekcores voor het betreffende kind.

Analyses

Hieronder worden eerst de analyses toegelicht die uitgevoerd zijn om de kwaliteit van het onderzoeksinstrumentarium (leestoetsen en dagboekgegevens) vast te kunnen stellen. Vervolgens worden de analyses besproken die betrekking hebben op het toetsen van de modellen voor de relatie tussen televisiekijken, buitenschools lezen en de leesprestaties.

Kwaliteit onderzoeksinstrumentarium

Van alle instrumenten (leestoetsen en dagboekmetingen) worden beschrijvende gegevens ge-

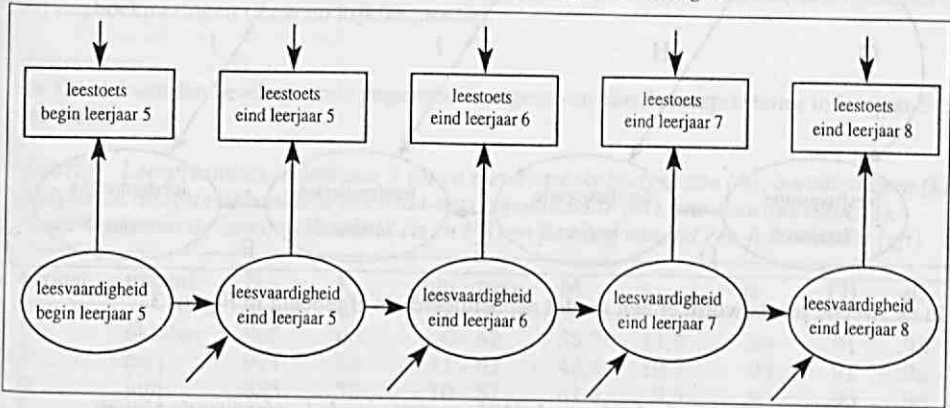
rapporteerd zoals gemiddelden en standaarddeviaties. Tevens wordt met behulp van de Kolmogorov-Smirnovtoets nagegaan of aangenomen mag worden dat de scores uit een normaal verdeelde populatie komen. Met het oog op deze aanname worden ook de verdelingen van de scores en de gepiektheids- en de scheefheidscoëfficiënten geïnspecteerd.

Per instrument worden vervolgens gegevens gerapporteerd over de interne consistentie volgens Cronbachs α en onder aanname van congenerieke metingen (Fleishman en Benson, 1987; verder: FB) en de passing ('goodness-of-fit' index: gfi) van een één-factormodel volgens Häggblunds instrumentele-variabelenmethode (1982) (zie Tabel 4 en 5).

Modelanalyses

In de eerste plaats zal voorafgaande aan de modeltoetsingen voor de beantwoording van de onderzoeksvragen, nagegaan worden of de leestoetsen een quasi-simplex vormen. Uit de literatuur is namelijk bekend dat in longitudinaal onderzoek, waar dezelfde variabelen gedurende een reeks van jaren bij dezelfde respondenten verzameld zijn, deze variabelen een quasi-simplexstructuur te zien geven (Jöreskog, 1970). In Figuur 1 is deze structuur voor de leestoetsen grafisch weergegeven.

Figuur 1: Het quasi-simplexmodel voor de leesvaardigheidsmetingen



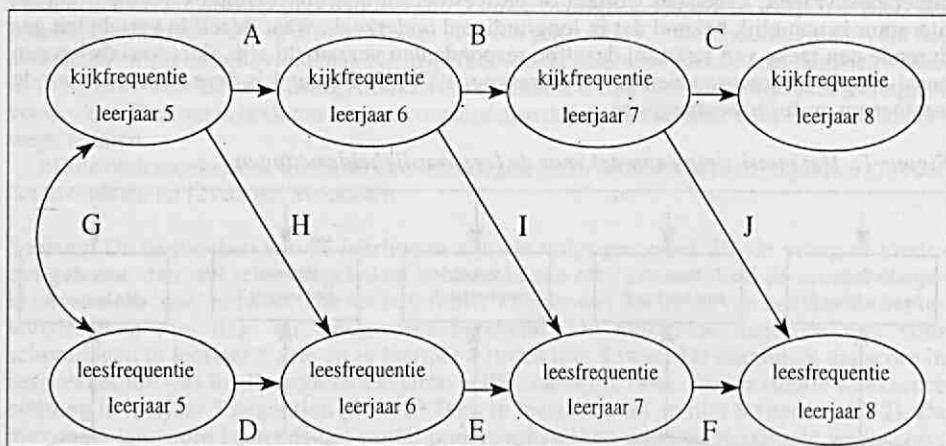
Uit Figuur 1 blijkt dat de leesscores worden opgevat als indicatoren van de latente variabele leesvaardigheid. Verder komt uit de figuur naar voren dat meetfouten zowel in de geobserveerde als in de latente variabelen zijn toegestaan. Voor de dagboekmetingen (lees- en kijkfrequentie) worden vergelijkbare modellen gepast. Indien uit de passing van de quasi-simplexmodellen zou blijken dat zij het databestand goed representeren, wordt bij de toetsing van de verdere modellen van deze structuur uitgegaan.

Het databestand van het onderhavige onderzoek is hiërarchisch van aard (leerlingen zijn onderdeel van klassen en scholen). Daarom is ervoor gekozen om de data te analyseren met covariantiestructuuranalyses (Jöreskog & Sörbom, 1988) waarbij met de hiërarchische structuur van de gegevens rekening wordt gehouden in een multiniveauprocedure van Muthén (Muthén, 1989, 1990). Deze benadering houdt in de onderhavige studie in dat, voorafgaand aan de modeltoetsingen met behulp van covariantiestructuuranalyses, de variantie op klasniveau volledig wordt uitgepartialiseerd.

De invloed van het televisiekijken op de leesfrequentie wordt in de eerste plaats onderzocht door een model te toetsen als afgebeeld in Figuur 2.

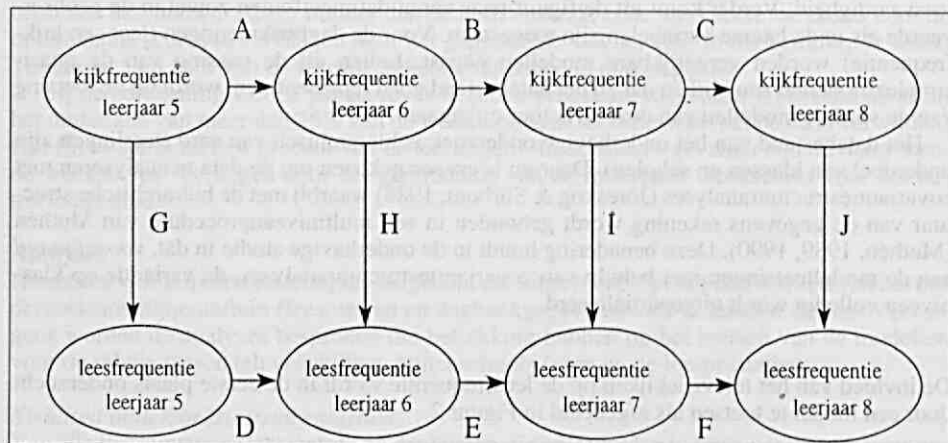
Volgens dit model heeft de frequentie waarmee kinderen televisiekijken in leerjaar X een effect op de frequentie waarmee kinderen thuis in boeken lezen in leerjaar X+1. Het (negatieve) effect van het televisiekijken op de leesfrequentie moet blijken uit significante (negatieve) regressies van leesfrequentie op de kijkfrequentie het jaar ervoor (relaties H, I, J). Het model staat ook indirecte effecten toe, bijvoorbeeld van het televisiekijken in jaar 5 op de leesfrequentie in bijvoorbeeld leerjaar 8. De indirecte effecten lopen dan via de kijkfrequentie en de leesfrequentie in leerjaar 6 of 7. In dat geval loopt de invloed van het televisiekijken bijvoorbeeld via de relaties A, B en J; H, E en F, of A, I, F.

Figuur 2: Model voor de invloed van de kijkfrequentie in leerjaar X op de leesfrequentie in leerjaar X+1: uitgestelde effecten



In de tweede plaats wordt er een model gepostuleerd als afgebeeld in Figuur 3.

Figuur 3: Model voor de invloed van de kijkfrequentie op de leesfrequentie binnen leerjaren: directe effecten



Volgens dit model heeft de frequentie waarmee kinderen televisiekijken in respectievelijk leerjaar 5, 6, 7 en 8 meteen in hetzelfde leerjaar een effect op de frequentie waarmee kinderen thuis in boeken lezen. Het (negatieve) effect van het televisiekijken op de leesfrequentie kan blijken uit de (negatieve) regressies van leesfrequentie op televisiekijken (relaties G, H, I of J). Het model staat ook indirecte effecten toe van het televisiekijken in jaar 5 op de leesfrequentie in bijvoorbeeld leerjaar 8. De indirecte effecten lopen via de kijkfrequentie en de leesfrequentie in leerjaar 6, 7 of 8. In dat geval loopt de invloed van het televisiekijken bijvoorbeeld via de relaties G, D, E en F of A, B, C en J.

De modellen die getoetst worden voor de invloed van het televisiekijken op de leesprestaties zijn gelijk aan die voor de invloed op leesfrequentie. In de modellen wordt alleen leesfrequentie vervangen door leesprestatie.

RESULTATEN

Kwaliteit van het onderzoeksinstrumentarium

Hieronder worden de resultaten gerapporteerd die betrekking hebben op de kwaliteit van het instrumentarium. Achtereenvolgens worden de resultaten weergegeven van de leestoetsen en de dagboekmetingen (lees- en kijkfrequentie).

Leestoetsen

In Tabel 4 worden beschrijvende gegevens weergegeven van de leesprestaties in leerjaar 5 tot en met 8.

Tabel 4: Leesprestaties in leerjaar 5 tot en met 8: steekproefgrootte (N), aantal vragen (k), laagste en hoogste behaalde score (min-max), gemiddelde (M), standaarddeviatie (s), schattingen van de betrouwbaarheid (α en FB) en passing van het één-factormodel (gfi)

Ljjaar	maand	N	k	min - max	M	s	α	FB	gfi
5	oktober	965	66	6 - 62	35.7	11.3	.90	.91	.95
5	mei	971	66	11 - 63	43.4	10.7	.90	.91	.96
6	juni	885	58	10 - 57	41.8	9.5	.90	.90	.96
7	juni	749	50	6 - 46	34.6	7.6	.89	.89	.99
8	juni	712	65	12 - 64	43.4	11.7	.92	.92	.97

Bij inspectie van de tabel blijkt dat de moeilijkheid van de opgaven aan het begin van groep 5 het grootst is geweest. De gemiddelde p-waarde van de toetsvragen bedraagt hier namelijk .54 (35.7/66). De leesvaardigheidstoets aan het einde van groep 6 had de laagste moeilijkheidsgraad. De gemiddelde p-waarde van de toetsvragen bedraagt in dat leerjaar .72. De gemiddelde p-waarden in de overige afnamen ontlopen elkaar niet veel. Deze zijn voor het einde van leerjaar 5, 7 en 8 respectievelijk .66, .69 en .67.

Verder is uit de tabel af te lezen dat in geen enkel leerjaar de kinderen de minimale of de maximale score behalen, hetgeen een aanwijzing is dat zowel bodem- als plafondeffecten uitgebleven zijn.

Met behulp van de Kolmogorov-Smirnov Goodness of Fit test is nagegaan of aangenomen mag worden dat de verschillende leesscores uit een normaal verdeelde populatie komen. Voor alle leestoetsen moet deze aanname verworpen worden ($p < .01$). Als we echter de ernst van de schendingen bekijken aan de hand van de histogrammen van de scoreverdelingen, de gepiektheids- en scheefheidscoëfficiënten, dan blijkt dat de schendingen gering zijn.

Uit Tabel 4 kan voorts worden opgemaakt dat de passing van het één-factormodel op alle

meetmomenten ruim voldoende tot zeer goed te noemen is: de 'goodness-of-fit' (gfi) is minimaal .95 en maximaal .99. De conclusie dat de leestoetsen in leerjaar 5 tot en met 8 min of meer één factor meten, mag getrokken worden.

Verder kan uit Tabel 4 worden afgelezen dat de betrouwbaarheid volgens Cronbachs α en volgens Fleishman en Bensons procedure goed genoemd mag worden; alle betrouwbaarheidsschattingen zijn minimaal .89.

Lees- en kijkfrequentie

Bij inspectie van Tabel 5 blijkt dat de kinderen in alle leerjaren minder frequent in boeken lezen dan naar de televisie kijken. In leerjaar 5 lezen zij gemiddeld op 18% (3.4/19) van de dagdelen en wordt er op 64% van alle dagdelen naar de televisie gekeken.

Tabel 5: Lees- en kijkfrequentiegegevens in leerjaar 5 tot en met 8: steekproefgrootte (N), aantal vragen (k), laagste en hoogste behaalde score (min-max), gemiddelde (M), standaarddeviatie (s), schattingen van de betrouwbaarheid (α en FB) en passing van het één-factormodel (gfi)

Onderdeel	Leerjaar	N	Min-Max	M	Sd	α	FB	gfi
leesfrequentie	5	900	0 - 19	3.4	3.4	.91	.91	.97
	6	903	0 - 14	1.8	1.9	.92	.92	.99
	7	775	0 - 14	1.3	1.8	.88	.88	1.00
	8	733	0 - 14	1.1	1.4	.81	.81	1.00
kijkfrequentie	5	901	0 - 19	12.2	3.9	.92	.92	.98
	6	907	0 - 14	8.8	3.4	.97	.97	1.00
	7	776	0 - 14	7.5	3.4	.94	.94	1.00
	8	735	0 - 14	8.1	3.4	.83	.83	.99

Uit Tabel 5 kan ook worden opgemaakt dat naarmate de kinderen ouder worden het percentage dagdelen waarop zij lezen, terugloopt: van gemiddeld 18% in leerjaar 5 naar gemiddeld 8% in leerjaar 8.

Verder kan uit Tabel 5 worden opgemaakt (kolom min-max) dat er in alle leerjaren kinderen zijn die extreem weinig (niet) of juist veel lezen of televisiekijken. Met behulp van de Kolmogorov-Smirnov Goodness of Fit test is nagegaan of aangenomen mag worden dat de verschillende lees- en kijkfrequentiescores uit een normaal verdeelde populatie komen. Met uitzondering van de gemiddelde kijkfrequentiescore in leerjaar 5 en 8, moet deze aanname voor alle frequentiescores verworpen worden ($p < .01$). Als we echter de ernst van de schendingen bekijken aan de hand van de histogrammen van de scoreverdelingen, de gepiektheids- en de scheefheidscoëfficiënten, dan blijkt dat de schendingen met betrekking tot de kijkfrequenties gering zijn. Ernstiger zijn de schendingen bij de leesfrequentiescores. Deze verdelingen zijn sterk asymmetrisch naar rechts uitlopend en aan de linkerkant sterk gepiekt. Dit betekent dat er veel kinderen zijn die aangeven zeer weinig te lezen en maar zeer weinig kinderen die aangeven veel te lezen. Daarbij wordt de scheefheid en de gepiektheid sterker naarmate de kinderen ouder worden. De groep kinderen die niet of weinig leest, neemt in de loop der jaren toe; het aantal kinderen dat veel leest, neemt daarentegen juist af.

Uit Tabel 5 blijkt dat de passing van het één-factormodel voor de weekcores van zowel lezen als televisiekijken zeer goed te noemen is: gfi is .97 of hoger.

Uit de tabel komt verder naar voren dat de betrouwbaarheid volgens Cronbachs α en Fleishman en Benson (1987) in leerjaar 5 tot en met 7 zeer goed genoemd mag worden. In leerjaar

8 is de betrouwbaarheid voor zowel lezen als televisiekijken nog steeds redelijk, doch duidelijk lager dan in de overige leerjaren. Uit de ons ter beschikking staande gegevens blijkt niet dat er in leerjaar 8 sprake is geweest van andere condities voor de dataverzameling (zie verder Otter & Schoonen, 1996a).

Modeltoetsingen

In de eerste plaats worden de resultaten gerapporteerd van de toetsingen betreffende de quasi-simplexstructuur van de longitudinaal verzamelde gegevens (leesprestaties, lees- en kijkfrequentie). Vervolgens worden twee modellen gepast voor de relatie tussen televisiekijken, buitenschools lezen en de leesprestaties.

Quasi-simplexstructuur van de longitudinale data

In Tabel 6 zijn de resultaten samengevat van de toetsing van de quasi-simplexmodellen. De parameters in het model zijn geschat volgens de 'maximum likelihood' (ML) methode⁵. In de tabel worden enkele toetsings- en passingsmaten weergegeven (χ^2 , df, gfi, 'adjusted goodness-of-fit' index (agfi) en de 'comparative-fit' index (cfi)).

Tabel 6: Toetsing ($\chi^2/df/p$) en passing (gfi, agfi, cfi) van het quasi-simplexmodel op de leesprestaties en kijk- en leesfrequentiescores. N is het aantal respondenten

	leesprestaties	kijkfrequentie	leesfrequentie
χ^2	3.18	.92	.87
df	3	1	1
p	.37	.34	.35
Gfi	.997	.999	.999
Agfi	.987	.992	.992
Cfi	1.000	1.000	1.000
N	485	551	550

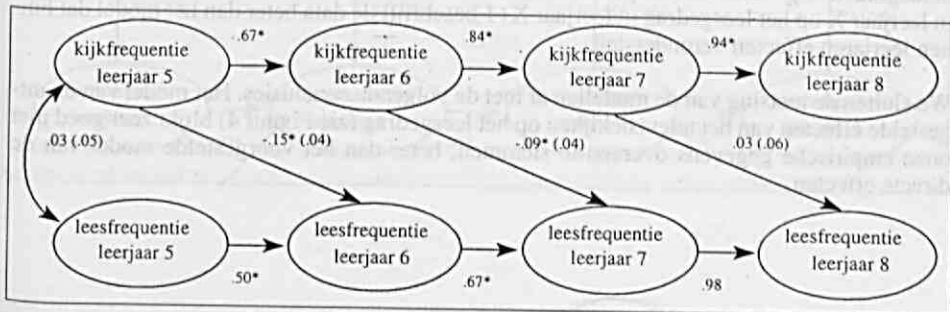
Uit Tabel 6 kan worden opgemaakt dat, ongeacht de gemeten variabele, het quasisimplexmodel de data goed beschrijft. Statistisch gezien kunnen de modellen niet verworpen worden ($p \geq .34$). Ook de beschrijvende maten duiden op een uitstekende passing. Zo ligt de χ^2/df verhouding rond de 1 en benaderen de passingsmaten gfi, agfi en cfi eveneens de 1.

Op grond van de resultaten in Tabel 6 is besloten bij de toetsing van de verdere modellen ervan uit te gaan dat de longitudinale gegevens een quasi-simplexstructuur hebben.

De effecten van televisiekijken op de leesfrequentie

UITGESTELDE EFFECTEN. Hieronder worden allereerst de resultaten weergegeven van het eerste model voor de invloed van het televisiekijken op de leesfrequentie (zie Figuur 4).

Figuur 4: Uitgestelde effecten van de kijkfrequentie op de leesfrequentie in termen van regressie en tussen haakjes de standaardfout (* = $p < 0.5$). Toetsing van het model: $\chi^2 = 15.13$, $df = 14$, $p = .37$ ($N = 550$), passing van het model: $Gfi = .993$, $Agfi = .983$ en $Cfi = .999$



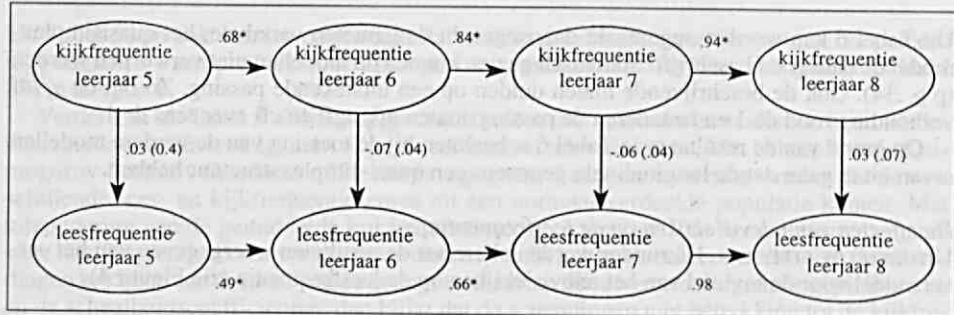
Uit de resultaten in Figuur 4 kan worden opgemaakt dat het model op statistische gronden niet verworpen wordt ($\chi^2=15.13$, $df=14$, $p=.37$). Daarbij geven de passingsmaten eveneens blijk van een goede passing ($gfi=.993$, $agfi=.983$, $cfi=.999$). Het model lijkt de empirische data goed te beschrijven.

Uit de regressiecoëfficiënten komt het volgende beeld naar voren. In de eerste plaats blijkt er in leerjaar 6 en 7 een significante negatieve regressie te zijn van de leesfrequentie op de kijkfrequentie het jaar daarvoor ($\beta = -.15$ met een standaardfout van $.04$, respectievelijk $\beta = -.09$ met een standaardfout van $.04$). Met andere woorden, de frequentie waarmee kinderen televisiekijken in leerjaar 5 en 6 heeft een negatieve invloed op hun leesfrequentie in respectievelijk leerjaar 6 en 7. De grootte van de regressiecoëfficiënten daalt echter van $-.15$ naar $-.09$ en is het jaar daarop niet meer significant ($.03$), hetgeen suggereert dat naarmate de kinderen ouder worden, de invloed van het televisiekijken op het leesgedrag afneemt en ten slotte verdwijnt. Hoewel de eerste twee regressiecoëfficiënten significant zijn, zijn ze over het algemeen laag. De kijkfrequentie in leerjaar 5 en 6 verklaart slechts ruim twee, respectievelijk minder dan één procent van de variantie in de leesfrequentie in leerjaar 6 en 7.

Uit Figuur 4 is verder op te maken dat naarmate de kinderen ouder worden hun lees- en kijkgedrag steeds stabiel wordt: 25% van de variantie in het leesgedrag in leerjaar 6 kan voorspeld worden uit het leesgedrag in leerjaar 5. Het leesgedrag in leerjaar 8 kan voor 96% voorspeld worden uit dit gedrag in leerjaar 7⁶. Voor televisiekijken geldt nagenoeg hetzelfde. Het kijkgedrag in leerjaar 6 kan voor 45% voorspeld worden uit het kijkgedrag in leerjaar 5, het gedrag in leerjaar 8 voor 88% uit het gedrag in leerjaar 7.

DIRECTE EFFECTEN. Om na te gaan of er mogelijk sprake is van grotere effecten binnen leerjaar is het tweede model voor directe effecten getoetst (zie Figuur 5).

Figuur 5: Directe effecten van de kijkfrequentie op de leesfrequentie in termen van regressie en tussen haakjes de standaardfout (= $p < .05$). Toetsing van het model: $\chi^2 = 28.44$, $df = 14$, $p = .01$ ($N = 550$), passing van het model: $Gfi = .987$, $Agfi = .967$ en $Cfi = .993$*



Uit de modelpassing blijkt dat het model op statistische gronden verworpen moet worden. Daarbij past het model ook slechter dan het model uit Figuur 4. Bij een gelijk aantal vrijheidsgraden is χ^2 ruim dertien punten hoger. Het model voor de invloed van televisiekijken in leerjaar X op het leesgedrag in leerjaar X+1 beschrijft de data beter dan het model dat binnen leerjaren effecten veronderstelt.

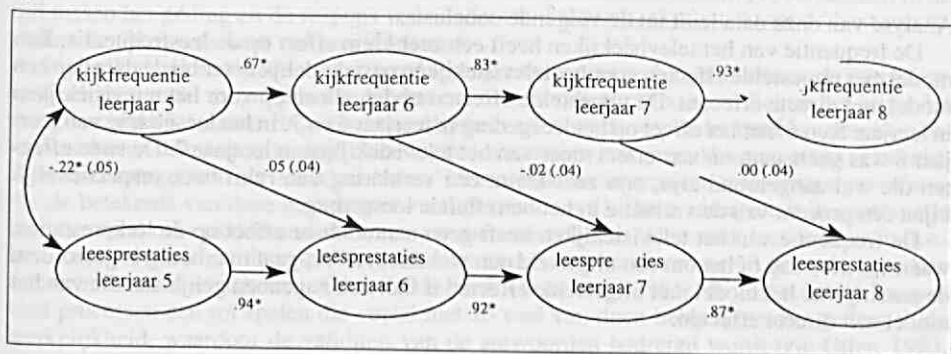
We sluiten de toetsing van de modellen af met de volgende conclusies. Het model van de uitgestelde effecten van het televisiekijken op het leesgedrag (zie Figuur 4) blijkt zeer goed met onze empirische gegevens overeen te stemmen, beter dan het voorgestelde model van de directe effecten.

Als we de uitgestelde effecten nader bekijken, blijkt dat telesiekijken in leerjaar 5 en 6 een (klein) negatief effect heeft op het leesgedrag in leerjaar 6 respectievelijk 7. Echter, de geringe invloed van het telesiekijken lijkt steeds verder af te nemen en in leerjaar 8 geheel verdwenen te zijn.

De effecten van telesiekijken op de leesprestaties

UITGESTELDE EFFECTEN. Hieronder worden de resultaten weergegeven van het model voor de invloed van het telesiekijken op de leesprestaties (zie Figuur 6).

Figuur 6: Uitgestelde effecten van de kijkfrequentie op de leesprestaties in termen van regressie en tussen haakjes de standaardfout (= p<.05). Toetsing van het model: $\chi^2 = 15.98$, $df = 14$, $p = .32$ ($N = 474$), passing van het model: $Gfi = .992$, $Agfi = .980$ en $Cfi = .999$*

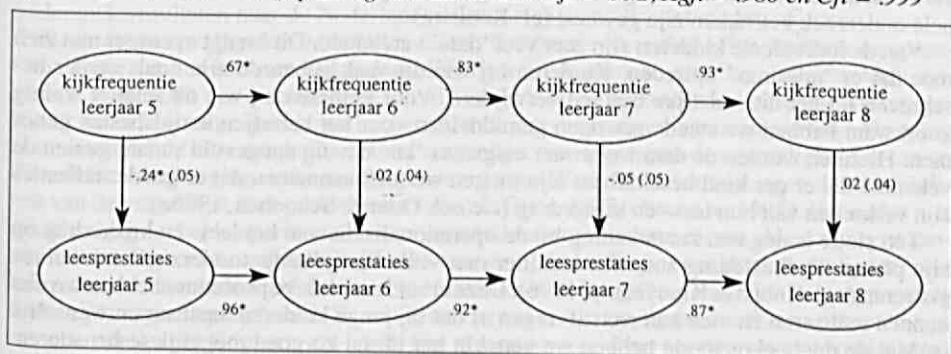


Uit de toetsing van het model blijkt dat het op statistische gronden niet verworpen kan worden ($\chi^2 = 15.98$, $df = 14$, $p = .32$). Ook de passingsmaten wijzen op een zeer goede passing van het model ($gfi = .992$, $agfi = .980$, $cfi = .999$).

De regressiecoëfficiënten laten zien dat er van een significante negatieve invloed van het telesiekijken op de leesprestaties géén sprake is. Wel is er sprake van een significante negatieve correlatie tussen het telesiekijken en de leesprestaties in het eerste leerjaar (-.22).

DIRECTE EFFECTEN. Om na te gaan of er wél sprake is van een effect binnen leerjaar is ook dit model getoetst (zie Figuur 7).

Figuur 7: Directe effecten van kijkfrequentie op de leesprestaties in termen van regressie en tussen haakjes de standaardfout (= p<.05). Toetsing van het model: $\chi^2 = 16.26$, $df = 14$, $p = .30$ ($N = 474$), passing van het model: $Gfi = .992$, $Agfi = .980$ en $Cfi = .999$*



Ook de passing van het model in Figuur 7 is goed te noemen. Het model lijkt even goed bij de data te passen als een model waarin uitgestelde effecten van de televisie op de leesontwikkeling worden verondersteld (zie Figuur 6): het verschil in χ^2 (met een gelijk aantal vrijheidsgraden) is namelijk te verwaarlozen.

De resultaten van de modellen zoals afgebeeld in Figuur 6 en 7 zijn overeenkomstig. Naar voren komt dat televisiekijken uitsluitend in leerjaar 5 negatief samenhangt (Figuur 6) of effect sorteert (Figuur 7) op de leesprestaties. In de overige jaren is er van een significant effect van televisiekijken op de leesprestaties geen sprake.

SAMENVATTING EN DISCUSSIE

Analyse van onze data leidt tot de volgende conclusies.

De frequentie van het televisiekijken heeft een zeer klein effect op de leesfrequentie. Een model met uitgestelde effecten voor het televisiekijken past duidelijk beter bij de data dan een model met directe effecten. De uitgestelde effecten treden alleen op voor het televisiekijken in leerjaar 5 en 6 met het effect op het leesgedrag in leerjaar 6 en 7. In het leesgedrag van leerjaar 8 was geen aantoonbaar effect meer van het televisiekijken in leerjaar 7. De twee effecten die wel aangetoond zijn, zijn zeer klein: een verklaring van ruim twee respectievelijk bijna één procent van de variantie in het betreffende leesgedrag.

De frequentie van het televisiekijken heeft geen aantoonbaar effect op de leesprestaties, waarmee de vraag of het om een uitgesteld dan wel direct effect gaat overbodig is geworden; de passing van het model met uitgestelde effecten is dan ook nagenoeg gelijk aan die van het model met directe effecten.

In het navolgende zullen we enkele kanttekeningen plaatsen bij onze uitkomsten en die uitkomsten vergelijken met het eerdergenoemde onderzoek van Koolstra (1993) dat ook de Nederlandse context betrof, maar dat tot enigszins andere conclusies lijkt te komen.

Zoals te verwachten bij longitudinaal onderzoek is er de nodige uitval aan 'respondenten' geweest tijdens het onderzoek. De redenen voor uitval stonden steeds los van het onderwerp van onderzoek, m.n. de noodzaak om scholen steeds opnieuw te moeten benaderen voor het vervolg heeft voor de grootste uitval gezorgd. We mogen ons inziens aannemen dat de uitval de representativiteit van de data en daarmee de generaliseerbaarheid van de conclusies niet ernstig geschaad heeft. Controle-analyses lieten zien dat de resterende groep nauwelijks in gemiddelde kijk- en leesfrequenties en leesprestaties afwijkt van de 'uitvallers' (zie ook noot 4). Het wordt dan onwaarschijnlijk dat door de uitvallers de relaties tussen de drie onderzochte variabelen zijn veranderd. Overigens zijn de analyses van de relaties tussen de drie variabelen (i.e. de modeltoetsingen) steeds uitgevoerd op de data van kinderen die bij het gehele onderzoek betrokken zijn geweest (cf. Koolstra).

Van de individuele kinderen zijn zeer veel 'data' verzameld. Dit brengt evenzeer met zich mee dat er 'missings' optreden. Kinderen bij wie dit vaak gebeurde zijn zoals eerder beschreven uit het uiteindelijke bestand verwijderd. Voor kinderen bij wie dit relatief weinig voorkwam hebben we steeds gewogen gemiddelden voor het betreffende tijdsbestek genomen. Hiermee worden de data weliswaar enigszins 'kunstmatig aangevuld', maar gezien de vele data die er per kind beschikbaar zijn mogen we toch aannemen dat er goede indicaties zijn verkregen van hun lees- en kijkgedrag (zie ook Otter & Schoonen, 1996a).

Ten slotte is nog een kanttekening bij de operationalisatie van het lees- en kijkgedrag op zijn plaats. Tijdbesteding zou men idealiter met volledig tijdbudgetonderzoek moeten registreren (vgl. Knulst & Kraaykamp, 1996). Deze arbeidsintensieve procedure hebben we niet kunnen realiseren en men kan zich afvragen of dat bij jonge kinderen haalbaar en wenselijk is. Met de dagboekmethode hebben we getracht het ideaal zo goed mogelijk te benaderen.

Over het algemeen weet men nog vrij goed wat men de dag ervoor (of na een weekend) twee dagen geleden gedaan heeft. Ongewenste effecten van herinneringsproblemen worden sterk gereduceerd en respondenten hoeven geen schattingen te maken over langere (en vaak ongedefinieerde) perioden, zoals dat vaak in vragenlijsten verlangd wordt. Bovendien hebben we empirische evidentie dat de leerlingen redelijk in staat waren de dagboeken adequaat in te vullen. In Otter (1993) zijn leesfrequentiegegevens verzameld bij zowel de leerlingen als hun ouders. Het verschil tussen de 'ouder'- en de leerlingdagboeken was dat de vragen in het 'ouderdagboek' altijd betrekking hadden op het leesgedrag van dezelfde dag (en dus niet zoals in de leerlingdagboeken op het leesgedrag van één, twee of drie dagen geleden). Uit de resultaten van dit onderzoek blijkt dat de 'ware' samenhang tussen de ouders en leerlingen .79 is (voor 'lezen in boeken'). Deze samenhang kan men toch als aanzienlijk beschouwen als men in acht neemt dat er sprake is van minstens drie foutenbronnen: (1) verschillen in de tijd tussen het gedrag en de respons voor ouders en kinderen, (2) fouten van de kant van de ouders en (3) van de kant van hun kinderen.

Bij de dagboeken zijn de vragen noodzakelijkerwijs eenvoudig gehouden waar eenmalige vragenlijsten meer mogelijkheden bieden om meer en dieper te vragen. Echter het blijkt dat de ogenschijnlijke nuanceringen in vragenlijsten vaak voor de respondent ambigu geacht moeten worden en daardoor niet altijd de beoogde informatiewinst geven die men ervan verwacht. Antwoordcategorieën als 'vaak' en 'soms' kunnen allerlei interpretatieproblemen geven omdat de betekenis van deze termen niet vastligt (zie noot 3). Deze ambiguïteit in combinatie met de veelal onbepaalde periode waarnaar men vraagt, b.v. 'Hoe vaak lees je thuis in boeken?' zorgt ervoor dat de respondent de antwoorden niet zonder meer uit het geheugen kan ophalen, maar moet schatten (Cannel, Miller & Oksenberg, 1981). Bij dit schatten kunnen veel processen een rol spelen die veelal niet zo veel van doen hebben met het gedrag in de werkelijkheid, waardoor de validiteit van de antwoorden bedreigd wordt (zie Otter, 1993, p. 19-30; 45-61). Hoewel hiermee niet gesteld mag worden dat de hier gebruikte dagboeken niet voor verbetering vatbaar zijn of perfect valide zouden zijn, geven we toch duidelijk de voorkeur aan de dagboekprocedure en hebben we voldoende indicaties dat de dagboeken (voldoende) valide gegevens opleveren. Zie ook Otter & Schoonen (1996a en b) voor de vergelijking van de kijk- en leesgegevens met andere bronnen.

In vergelijking met Koolstra (1993) komen we tot enigszins andere conclusies. Deze verschillende conclusies moeten in de eerste plaats beschouwd worden tegen de achtergrond van de verschillen in de onderzoekopzet en -uitvoering. Daarnaast moeten we ook de waarde van de verschillen in uitkomsten bekijken. We beperken ons hier tot een vergelijking met Koolstra's onderzoek omdat dit de enige studie in Nederland is die uitgebreid op de relaties tussen lees- en kijkgedrag en leesprestaties ingaat. Ook in buitenlands onderzoek is met wisselend succes de relatie tussen lees- en kijkgedrag en leesprestaties onderzocht, maar - zoals in de inleiding al aangegeven - het is niet zonder meer duidelijk dat deze m.n. Amerikaanse onderzoeksresultaten naar de Nederlandse situatie gegeneraliseerd mogen worden.

Een van de opvallendste verschillen tussen Koolstra's en onderhavige studie is de operationalisatie van het lees- en kijkgedrag. Koolstra heeft gekozen voor een (uitgebreide) eenmalige vragenlijst per meetmoment, terwijl in de hier gerapporteerde studie dagboeken gebruikt zijn waarmee gespreid over het betreffende jaar diverse metingen zijn verricht. De vraag is uiteraard of de uitvoerigheid van een vragenlijst opweegt tegen de eenmaligheid ervan en de validiteitsbedreigingen die eerder genoemd zijn. We hebben aanwijzingen dat de validiteit van de vragenlijst toch problematisch is (zie Otter, 1993; Otter e.a., 1995). Het is dus niet denkbeeldig dat beide operationalisaties verschillende constructen meten. Als we mogen aannemen dat dagboeken validere indicaties van de tijdbesteding geven dan vragenlijsten, dan zou men moeten concluderen dat de relatie tussen kijk- en leesgedrag enerzijds en leesprestaties anderzijds nagenoeg ontbreekt. Er zou dan evenwel een bevredigende verklaring moeten zijn voor de relaties die Koolstra wel vond. Een afdoende verklaring kunnen we niet zonder meer

geven, maar wel zij nog opgemerkt dat de modellen waarin Koolstra de genoemde relaties toetst uitgebreider zijn dan de door hem theoretisch verantwoorde modellen: diverse extra paden worden erin toegestaan die niet uit de te toetsen hypothesen afgeleid worden. Daarbij komt dat deze 'flexibele' modellen in vergelijking met de hier gepresenteerde modellen minder goed passen.

Bij het aantonen van relaties dient men zich goed te realiseren dat de significantie van relaties slechts gedeeltelijk bepaald wordt door de sterkte van de samenhang en voor een ander deel door de steekproefgrootte. Door het combineren van de twee cohorten van verschillende leeftijden krijgt Koolstra in zijn studie een zeer grote steekproef en daardoor meer onderscheidend vermogen. Kijken we alleen naar de resultaten van Koolstra's cohort 2, die in omvang en leeftijd vergelijkbaar is met onze steekproef, dan zien we dat er bij Koolstra evenmin aantoonbare effecten zijn voor televisiekijken op de leesprestaties (o.c., 89). In plaats van significantie als een 'arbitraire' scherprechter te gebruiken is het zinvol de gevonden effecten op een rijtje te zetten zoals in Tabel 7.

Uit Tabel 7 blijkt dat de regressiegewichten zoals gevonden door Koolstra en in onderhavige studie elkaar in absolute zin niet veel ontlopen. Hierbij moet aangetekend worden dat de kwaliteit van andere predictoren van de afhankelijke variabelen in het model ook een rol spelen. Zo zijn in ons model de longitudinale effecten in het quasimaximaal model veel sterker dan bij Koolstra; met andere woorden het leesgedrag en de leesprestaties op tijdstip T worden vooral bepaald door het gedrag respectievelijk de prestaties op tijdstip T-1. Bij Koolstra zijn die longitudinale effecten kleiner, zodat er meer ruimte blijft voor andere predictoren zoals televisiekijken.

Tabel 7: Vergelijking regressiegewichten van Koolstra (1993) en onderhavige studie (OSG) voor leerjaar 5 t/m 8

Leerjaar van → naar	tv → leesfrequentie		tv → leesprestatie	
	Koolstra (1993)	OSB (1996)	Koolstra (1993)	OSG (1996)
5 → 6	-.16*	-.15*	-.16*	.05
6 → 7	-.18*	-.09*	-.08	-.02
7 → 8	-.10*	-.03	-.00	.00

Dit alles in ogenschouw nemend lijkt de conclusie gerechtvaardigd dat televisiekijken in de midden- en bovenbouw van het basisonderwijs geen noemenswaardig effect uitoefent op de leesfrequentie en zeker nauwelijks effect op de leesprestaties. Televisiekijken en lezen zijn blijkbaar niet elkaars directe concurrenten: meer van het één leidt niet direct tot minder van het ander. Kennelijk zijn er nog andere factoren en andere vormen van vrijetijdsbesteding die voorkomen dat veranderingen of verschillen in kijkfrequentie direct merkbaar zijn in de leesfrequentie (of de leesprestaties). Andere verklaringmodellen voor het veronderstelde effect zoals de passiviteitshypothese (zie Koolstra, 1993) verliezen hun relevantie als er van televisiekijken geen effect uitgaat op lezen. Deze hypothesen die een dergelijk effect voorspellen of 'verklaren' moeten als gefalsifieerd beschouwd worden, als een dergelijk effect niet optreedt.

Dat televisiekijken het lezen nauwelijks beïnvloedt, kan mogelijk verklaard worden uit de zeer lage leesfrequentie van kinderen in combinatie met een veelheid aan andere vrijetijdsbestedingen. In een andere rapportage over dit onderzoek (Otter & Schoonen, 1996a) blijkt namelijk dat kinderen zeer infrequent lezen. De variantie in leesfrequentie speelt zich min of meer af in de marges van de vrijetijdsbesteding. Het al dan niet frequenter televisiekijken gaat niet per se ten koste van lezen. Kennelijk zit er 'tussen' deze twee vrijetijdsbestedingen nog

een 'buffer' van allerlei andere activiteiten die ervoor zorgt dat lezen en televisiekijken elkaar niet direct beïnvloeden⁷. Een factor die hierbij een rol kan spelen is het tijdstip waarop de activiteiten uitgevoerd worden. Men kan zich voorstellen dat met name het lezen zich beperkt tot tijdstippen waarop televisiekijken minder voor de hand ligt of minder bereikbaar is, zoals 's avonds in bed, tijdens reizen of wachten. Bovendien stellen televisiekijken en lezen andere condities aan de omgeving: televisiekijken moet men min of meer gezamenlijk doen, terwijl lezen een betrekkelijk solitaire bezigheid is. Meer inzicht in deze relaties, m.n. op individueel niveau, vergt gedetailleerde informatie over de tijdbesteding van respondenten zoals die met tijdbudgetonderzoek verkregen kan worden.

De implicaties voor onderwijs en opvoeding lijken dat van een strijd tegen het (vele) televisiekijken niet te veel verwacht mag worden, tenzij men bereid is tot het nemen van drastische maatregelen zoals de volledige uitbanning van de televisie. Het is in het hedendaags perspectief in ieder geval niet zo dat de (hoge) frequentie van televisiekijken de (enige) grote boosdoener is als het gaat om het geringe buitenschoolse lezen van veel kinderen en hun geringe leesprestaties.

Leesbevordering zal zich op het lezen zelf moeten richten en niet op de vermeende concurrent, waarbij het duidelijk moet zijn dat leesprestaties vooral door leesonderwijs bevorderd moeten worden en niet door de frequentie van het buitenschools lezen. Het buitenschools lezen zal het vooral moeten hebben van de bevordering van de intrinsieke motivatie van kinderen. Het lezen als interessante en boeiende vrijetijdsbesteding moet doel op zich zijn en leerlingen moeten wellicht geholpen worden deze vorm van lezen te ontdekken (vgl. Otter & Schoonen, 1996a). Tegen deze achtergrond zouden we meer moeten weten van de totstandkoming van die intrinsieke motivatie bij de (zeer kleine) groep van veellezers.

NOTEN

1. Omdat de kijkfrequentie in jaar X de leesfrequentie van het daaropvolgende jaar (X+1) beïnvloedt, die vervolgens de leesfrequentie in jaar (X+2) weer beïnvloedt, is er ook sprake van cumulatieve effecten. Als men rekening houdt met deze cumulatie is het effect van televisiekijken waarschijnlijk groter dan hier vermeld.
2. Bij Koolstra's psychologische verklaringsmodellen kan men ook enige kanttekeningen plaatsen. De mediërende variabelen 'leesattitude' en 'leesconcentratie' verklaren weinig variantie in de leesprestaties. Drie van de vier regressiegewichten zijn statistisch significant: in één geval gaat het om een percentage verklaarde variantie van ca. 3, in de overige twee om minder dan 1%. Bezwaarlijker is dat in de getoetste regressiemodellen allerlei andere paden verondersteld moeten worden die niet voortvloeien uit de te toetsen hypothesen; de bij deze paden behorende regressiegewichten zijn groter dan de theoretisch relevante regressiegewichten. Overigens passen ook deze uitgebreidere modellen maar matig bij de data, zodat enige reserve bij de regressiegewichten op zijn plaats is.
3. Voorbeelden van genoemde ambigue antwoordcategorieën zijn 'vaak' en 'soms'. De betekenis van 'vaak' en 'soms' ligt niet vast: Wat de ene respondent als 'vaak' betitelt, kan een andere respondent - met hetzelfde leesgedrag - als 'soms' kenschetsen (Payne, 1951; Speak, 1967; Belson, 1968; Belson, 1981; Groves, 1989). Bovendien hebben de meeste vragen in Koolstra's vragenlijsten betrekking op een ongedefinieerd tijdsbestek (b.v. 'Hoe vaak lees je thuis in boeken?'). Dit laatste betekent dat de antwoorden niet direct uit het geheugen ophaalbaar zijn, maar geschat worden (Cannel, Miller & Oksenberg, 1981). Bij het gokken of schatten van de antwoorden kunnen veel processen een rol spelen, ook processen die niet zo veel van doen hebben met het gedrag in de werkelijkheid (zie Otter, 1993, p. 19-30 en 45-61).
4. Met behulp van t-toetsen is nagegaan of de leerlingen van de scholen die zich al dan niet uit het onderzoek hebben teruggetrokken (tien versus dertig scholen) significant afwijkende leesprestaties hebben aan het begin en einde van leerjaar 5 van de resterende scholen. Dit bleek het geval. De leerlingen van de scholen die in de loop der jaren zijn uitgevallen, behaalden zowel op de voor- als op de nameting in leerjaar 5 statistisch significant lagere leesscores; uitgedrukt in Cohens effectmaat (d) voor verschillen in gemiddelden gaat het echter

om kleine tot middelmatige effecten. Deze uitval heeft de representativiteit van de steekproef niet ernstig aangetast. In leerjaar 6 en 7 blijken namelijk de leerlingprestaties op de leesvaardigheidstoetsen, ontwikkeld en genormeerd door het Cito, overeen te stemmen met de Cito-normgroep (voor leerjaar 6: zie Otter, 1993, p. 96; voor leerjaar 7 zie Otter & Schoonen, 1996a, bijlage 21). Bovendien verschilden de 'uitvallers' niet van de resterende steekproef als het gaat om de kijk- en leesfrequentie.

5. De ML-methode gaat ervan uit dat de variabelen normaal verdeeld zijn in de populatie. Uit de beschrijvende gegevens bleek dat deze aanname voor de leesfrequentie in leerjaar 5 tot en met 8 duidelijk verworpen moet worden. Daarom zijn de analyses ook uitgevoerd volgens de ongewogen kleinste-kwadratenmethode ('unweighted least squares': ULS), een methode die deze aanname niet doet. Als de resultaten met de ULS-methode afwijken van die met de ML-methode, zal dit gerapporteerd worden.

6. De schattingen volgens de ULS-methode zijn geheel overeenkomstig die volgens de ML-methode, uitgezonderd de regressiecoëfficiënt voor de latente variabele leesfrequentie in leerjaar 8 op die in leerjaar 7. Deze is volgens de ULS-schatting iets lager (.92), maar wél significant (zie ook Otter, 1995).

7. Historisch gezien heeft de komst van de televisie wel een enorm effect gehad op de tijdbesteding in het algemeen en het lezen in het bijzonder (Knulst & Kraaykamp, 1996).

LITERATUUR

- Beentjes, J.W.J. & Voort, T.H.A. van der (1988). Television's impact on children's reading skills: A review of research. *Reading Research Quarterly*, 23, 389-413.
- Beentjes, J.W.J. & Voort, T.H.A. van der (1989). Television and young people's reading behavior: A review of research. *European Journal of Communication*, 4, 51-77.
- Belson, W.A. (1968). Respondent understanding of survey questions. *International Review in Public Opinion*, 3, 1-13.
- Belson, W.A. (1981). *The design and understanding of survey questions*. London: Gower.
- Cannell, C.F., Miller, P.V., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological Methodology* (389-437). San Francisco: Jossey Bass.
- Elley, W.B. (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Elsevier Science Ltd.
- Glopper, K. de & Otter, M.E. (1993). *Nederlandse leesprestaties in internationaal perspectief*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam.
- Graesser, A.C., & Murachver, T. (1985). Symbolic procedures of question answering. In: A.C. Graesser & J.B. Black (Eds.), *The Psychology of Questions*. Hillsdale: Lawrence Erlbaum.
- Groves, R.M. (1989). *Survey errors and survey costs*. New York: Wiley.
- Häggglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika*, 47, 209-222.
- Jöreskog, K.G. (1970). Estimation and testing of simplex models. *The British Journal of Mathematical and Statistical Psychology*, 23, 2, 121-145.
- Jöreskog, K.G. & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS Inc.
- Jöreskog, K.G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the simplis command language*. New York: Hillsdale.
- Koolstra, C.M. (1993). *Television and children's reading. A three-year panel study*. Leiden: Centre for child and media studies.
- Knulst, W. & Kraaykamp G. (1996). *Leesgewoonten. Een halve eeuw onderzoek naar het lezen en zijn belagers*. Rijswijk: Sociaal en Cultureel Planbureau.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Presidential address to the psychometric society, July 1989. *Psychometrika*, 54, 557-585.
- Muthén, B. (1990). *Mean and covariance structure analysis of hierarchical data*. Los Angeles: UCLA Statistics Series, # 62.
- Otter, M.E. (1993). *Leesvaardigheid, leesonderwijs en buitenschools lezen: instrumentatie en effecten*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam.
- Otter, M.E. (1995). *Leesvaardigheid, leesonderwijs, buitenschools lezen en tv-kijken in groep 5 tot en met 8 van het basisonderwijs*. Amsterdam: SCO-Kohnstamm Instituut van de Universiteit van Amsterdam.
- Otter, M.E., Mellenbergh, G.J. & de Glopper, K. (1995). The relation between information-processing variables and test-retest stability for questionnaire items. *Journal of Educational Measurement*, 32, 2, 199-216.

- Otter, M.E. & Schoonen, R. (1996a). *Aap, noot, niets of het spook van de ontleding in het basisonderwijs*. Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam (rapport 427).
- Otter, M.E. & Schoonen, R. (1996b). *Waar blijft de tijd... Een onderzoek naar de tijdbesteding aan Nederlandse taal in leerjaar 5 tot en met 8 van het basisonderwijs*. Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam (rapport 444).
- Payne, S.L. (1951). *The art of asking questions*. Princeton: Princeton University Press.
- Pligt, J. van der & Vries, N.K. de (1995). *Opinies en attitudes. Meting, modellen en theorie*. Amsterdam/Meppel: Boom.
- Speak, M. (1967). Communication failure in questioning: errors, misinterpretations and personal frames of reference. *Occupational Psychology*, 14, 169-181.
- Tourangeau, R., & Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299-314.

Boekbesprekingen

Polder, K.J. (1996)

Implementatie van wetgeving in onderwijs: Beleidsevaluatie-onderzoek in basisonderwijs en volwasseneneducatie

Academisch proefschrift Universiteit van Amsterdam

Promotor A.M.L. van Wieringen

Handelseditie Leuven-Apeldoorn: Garant, ISBN 90-5350-488-5.

Integrale evaluatiestudies van de implementatie van het vernieuwingsconcept basisonderwijs zoals dat wettelijk verankerd is, zijn betrekkelijk schaars. Onder de veelzeggende titel 'Zo hard gelopen en toch nog zover te gaan' (1990) waagde de toenmalige staatssecretaris zelf een eerste poging. In 1994 heeft de Commissie Evaluatie Basisonderwijs de verrichtingen binnen deze onderwijssector beschouwd met de nadruk op effectiviteit van instructie en van de schoolorganisatie. In eerder onderzoek is dat vanuit een innovatietheoretische en organisatiekundige invalshoek gebeurd (Van Gennip, 1991). De dissertatie van Klari-Janne Polder (1996) vormt een welkome aanvulling. Zij bestudeerde welke beleids- en organisatorische factoren van invloed zijn op de invoering van onderwijskundige bepalingen zoals vervat in wetgeving. Primair doet zij dat vanuit een beleidsanalytische optiek.

Opbouw

De eerste twee hoofdstukken van dit kloeke proefschrift (428 pagina's) omvatten algemene beschouwingen over de gehanteerde hoofdbegrippen en het beleidsanalytisch kader. Voornamelijk is bij Nederlandse auteurs te rade gegaan; (internationale) literatuur over onderwijskundige innovatie- en implementatietheorieën blijft buiten beschouwing. Essentiële kenmerken van het beleidsproces passeren de revue. Gewezen wordt op de diverse niveaus waarop beleid wordt uitgevoerd (overheid, schoolexterne organisaties, school- en klasniveau) en die elk afzonderlijke 'beleidsprestaties' van 'beleidssubjecten' vereisen. De behandeling van een veelheid van analytische indelingen en begrippen omtrent beleid mondt uit in drie hoofdgroepen van beleidseffectiviteit beïnvloedende factoren, namelijk:

- a. kenmerken van de beleidstheorie (deze omvat beleidsdoelen en -instrumenten, waarbij middel-doel relaties worden verondersteld),
- b. kenmerken van de beleidsuitvoering, waaronder de capaciteit van de uitvoerende organisatie,
- c. de doorwerking van beleid in organisaties, o.a. via besluitvormingsprocessen.

Hoofdstuk 3 bevat een onderzoeksmethodologische verantwoording: uiteengezet wordt dat ook een niet strikt experimentele aanpak (met correlatieve analyses zoals regressie, LISREL) argumenten kan aandragen voor causale middel-doel relaties. De auteur typeert haar onderzoek als 'een output-evaluatie naar de beleidsprestaties op verschillende niveaus van de beleidsuitvoering' en als een 'doelbereikingsonderzoek': in welke mate worden de beleidsdoelen gehaald? Daartoe wordt de stand van zaken op basisscholen anno 1990 vergeleken met die in 1987.

Uiteraard ontbreekt een uitvoerige schets van de oogmerken en het invoeringsproces van de Wet op het basisonderwijs (WBO) niet. Dit hoofdstuk 4 is beschreven vanuit het perspectief van overheid en ondersteunende commissies. Afrondend worden de eerder vastgestelde kenmerken van het beleidsproces en -evaluatie hierop toegepast.

In hoofdstuk 5 worden de onderdelen van het conceptuele model voor de WBO-implementatie geoperationaliseerd. De beleidsimplementatie wordt op meerdere gebieden gemeten, zoals individualisering en differentiatie, zorgverbreding, onderwijsinhoudelijke integratie ko-lo en vakinhoudelijke ontwikkelingen. Voor de constructie van de 'schoolexterne factoren' is wederom een beroep gedaan op de beleidstheorie, terwijl voor de schoolinterne variabelen onderwijsorganisatorische inzichten dienstbaar geacht worden. De aandacht voor de organi-

satietheorie - naar opgave van de auteur de tweede theoretische pijler van deze studie - is echter nogal summier. Op pagina 105 komt de organisatietheorie even om de hoek kijken. Zonder omwegen is gekozen voor de 'open systeem benadering'. Meer in het bijzonder is in navolging van E. Marx onderscheid gemaakt tussen het organisatiekundig en het onderwijskundig subsysteem, waarbij het eerste als voorwaardelijk voor het tweede wordt gezien. Teruggerepen is op meetinstrumenten als de schoolleidersinterventieschaal van Van de Grift (1987) en naar onderzoek omtrent achterblijvende scholen. Bespreking van eerdere studies naar vernieuwing van het basisonderwijs leidt tot de conclusie dat de 'interne krachten van het systeem de primaire bron voor organisatieverandering vormen' (p.121) en niet de externe krachten vanuit het beleid. Een reeks hypothesen besluit dit hoofdstuk met als meest opmerkelijke dat er een verschil in implementatie zal zijn tussen scholen met een mannelijke en een vrouwelijke schoolleider (maar later niet bevestigd).

In hoofdstuk 6 volgt in extenso de beschrijving van de uitkomsten van de gehouden survey bij directeuren en leerkrachten van groep 2, 3 en 8 (respons rond vijftig procent). Daarbij worden de uitkomsten van 1990 vergeleken met gegevens van 1987. Ondanks voortgang op deelgebieden blijkt er nog een lange weg te gaan voordat de WBO-uitgangspunten ten volle ingevoerd zijn, bijvoorbeeld wat betreft individualisering en differentiatie en verbreding van het vormingsaanbod. De mate van onderwijskundige integratie van kleuteronderwijs en lager onderwijs blijkt geen vooruitgang meer te vertonen. De auteur spreekt van een verzadigingspunt. Vermeldenswaard is dat schoolleiders de inzet van het docententeam en de bereidheid tot veranderen als de belangrijkste bevorderende schoolinterne factor voor implementatie beschouwen. Datzelfde geldt voor de schoolbegeleidingsdienst als schoolexterne factor.

Interessant is hoofdstuk 7. Met de uiteindelijk geselecteerde variabelen (waarvan de lezer eerst nog de psychometrische verantwoording aantreft) wordt de regressie op de zes afzonderlijke afhankelijke variabelen (implementatiemaat) bepaald. De gestandaardiseerde gewichten zijn in de regel laag en de predictors wisselen gedeeltelijk per implementatiemaat. Als beste voorspeller komt uit de bus 'tevredenheid schoolleider met vorderingen schoolteam'. Vervolgens wordt het longitudinale model getoetst met LISREL 6, waarbij vooraf inhoudelijke verbanden verondersteld zijn. Twee latente variabelen indiceren de onderwijskundige vorderingen respectievelijk in 1987 (voormeting) en in 1990; alle overige (onafhankelijke) variabelen worden als direct gemeten opgevat. Opvallend is dat van de afhankelijke variabelen die de latente constructen vormen, de onderwijskundige integratie van kleuterschool en lagere school het meeste gewicht in de schaal legt. Naar voren komt - niet zo verrassend - dat de vorderingen anno 1987 het grootste effect hebben op de vorderingen in 1990 (totaal verklaarde variantie 57,4%). Daarnaast zijn er significante indirecte effecten: de gevorderdheid in 1987 beïnvloedt via andere variabelen de voortgang in 1990. Overige variabelen met een relatief groot effect zijn: 'tevredenheid met de vorderingen van het team' en 'profijt van de beleidsmaatregelen en de beleidsuitvoeringsorganisatie'. Schoolleidersinterventies blijken maar gedeeltelijk effect te hebben. Net zoals in ander onderzoek (Van Gennip) is de schoolgrootte, hier uitgedrukt als formatie-omvang, nauwelijks of niet van invloed op implementatie van vernieuwingen. Aangetekend moet worden dat bij de analyses van de uiteindelijke modellen zich relatief veel 'missings' voordoen. Jammer is dat de inhoudelijk/theoretische verklaringskracht gering is; de gevonden verbanden lijken toch iets tautologisch te hebben.

Hierna komen nog als toefigt (hoofdstuk 8) casestudies bij tien vernieuwende en bij tien achterblijvende basisscholen. Deze zijn geselecteerd op hun scores die de implementatie-gevorderdheid meten. Via een 'cross-site analysis' worden beide typen vergeleken op beleids- en organisatiefactoren. Belangrijk voor implementatie blijken: meer aandacht van de schoolleider (en andere teamleden) voor onderwijskundige managementtaken en de positieve wisselwerking tussen school en schoolbegeleidingsdienst. Ook de betrokkenheid bij de besluitvorming en consensus in het team over de beleidsdoelen doen ter zake.

Vergelijking van survey- en case-uitkomsten (hoofdstuk 9) leidt tot de slotsom dat schoolorganisatorische variabelen belangrijker zijn dan beleidsfactoren. De auteur stelt dat beleidsfactoren de implementatie wel bijsturen, maar niet vergaand bepalen. Meer in het bijzonder

wordt verondersteld dat het *ontwikkelingsniveau* van de school de achterliggende factor is van de voortgang van de school. Gezien het dereguleringsbeleid van de overheid en de nadruk op de eigen autonomie is Polder somber gestemd over de voortgang van scholen met geringe ontwikkelingscapaciteit.

Hoewel we inmiddels op pagina 240 aangekomen zijn, volgt nu een nieuw evaluatie-onderzoek rond de indertijd overstreden Kaderwet Volwasseneneducatie, een wet die slechts een kort leven beschoren is geweest. Ofschoon deze studie bescheidener van opzet is, passeren toch nog een survey en enkele cases de revue waarbij nu niet scholen maar voornamelijk gemeenten onderzoeksobject zijn. De lezer vraagt zich af waarom dit onderdeel ook nog aan het proefschrift is toegevoegd. De meerwaarde blijkt eerst bij de interpretatie van de uitkomsten. Polder's onderzoek naar deze innovatiepoging van het volwassenonderwijs laat zien dat hier juist beleidsvariabelen meer ter zake doen dan organisatievariabelen. Dit contrast met de implementatie van het vernieuwingsconcept basisonderwijs brengt Polder op pagina 371 tot een nieuwe these: *de ingrijpendheid van de beleidsmaatregel bepaalt of beleids- dan wel organisatiefactoren meer invloed hebben*. Interessant is deze stelling zeker, helemaal overtuigend echter niet. Ervoor pleit dat de Wet op het basisonderwijs in onderwijskundig opzicht scholen veel ruimte laat tot eigen invulling. De WBO dwingt niets af. Methodologisch gezien is het echter nog de vraag of de ingrijpendheid van de beleidsmaatregel inderdaad de cruciale verklaring vormt. Want behoudens wettelijk kader zijn er nog tal van wezenlijke verschillen tussen basis- en volwassenonderwijs, bijvoorbeeld qua complexiteit en schaal. Zo zijn in het ene geval directeuren en leraren onderzocht, terwijl in de andere studie gemeentebambtenaren als respondent fungeerden. De invulling van organisatie- en beleidsvariabelen varieerde bovendien per onderwijssector.

Kanttekeningen

Uiteraard zijn er ook kanttekeningen bij zo'n lijvig boekwerk te plaatsen. De meest wezenlijke is dat het geheel toch wel weinig c.q. eenzijdig theoriegestuurd is. Weliswaar wordt een beroep gedaan op beleidstheoretische inzichten maar deze omvatten voornamelijk begrippen en analytische indelingen. De tweede theoretische pijler waarop de auteur zich herhaaldelijk zegt te beroepen, de organisatie-theorie, komt nogal summier aan bod. Opvallend is dat er nauwelijks internationale literatuur is geraadpleegd.

Hoewel het boek helder is gestructureerd, had het proefschrift compacter kunnen zijn door een strakkere compositie en minder presentaties van cijfer. Voor een proefschrift vind ik het ontsierend (en soms een puzzel) reeksen variabelen aan te treffen onder hun cryptische SPSS-benaming in plaats van door een inhoudelijke aanduiding.

Al met al kunnen we blij zijn met de studie van Klari-Janne Polder. Er is gepoogd heel precies te analyseren hoe de voortgang te voorspellen is van implementatieprocessen. De empirie toont dat dit lastig is. Vooral nog is het laatste woord hierover nog niet gesproken. Polder interpreteert haar uitkomsten bij basisscholen met de mega-term 'ontwikkelingsvermogen'. Dat versterkt de noodzaak dit en aanverwante begrippen gedegen te conceptualiseren en te operationaliseren.

Literatuur

- Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit. Evaluatie van het basisonderwijs. Eindrapport*. Den Haag: Sdu.
- Gennip, J. van (1991). *Veranderingscapaciteiten van basisscholen. Onderzoek naar de functie van veranderingscapaciteiten binnen het innovatieproces basisonderwijs*. Nijmegen: ITS, Instituut voor Toegepaste Sociale wetenschappen.
- Grift, W. van de (1987). *De rol van de schoolleider bij onderwijsvernieuwingen*. 's-Gravenhage: VUGA.
- Ministerie van Onderwijs en Wetenschappen (1990). *Zo hard gelopen en toch nog zover te gaan: Evaluatie van basisonderwijs 1985-1988*. Den Haag: Sdu.

Joop Bokhove, Frank van der Schoot, Theo Eggen (1996)
Balans van het rekenonderwijs aan het einde van de basisschool 2.
Uitkomsten van de tweede peiling rekenen/wiskunde einde basisonderwijs
 Arnhem: CITO. Artikel nummer: 56651. Prijs: f 45,-

Het CITO heeft in 1996 rapport uitgebracht over de resultaten van de tweede peiling van het niveau van het rekenonderwijs op de basisschool. De tweede peiling vond voorjaar 1992 in groep 8 plaats. De opzet van de peiling is voor een groot deel overeenkomstig de eerste peiling uit 1987. Het peilingsonderzoek gebeurt in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap. De belangrijkste doelstelling van het peilingsonderzoek is periodiek gegevens te verzamelen over het onderwijsaanbod en de leerresultaten in het basis- en speciaal onderwijs. De gegevens moeten de mogelijkheid bieden om na te gaan in hoeverre de kerndoelen voor het basisonderwijs worden nagestreefd en bereikt. Peilingsonderzoek is een van de instrumenten van de centrale overheid voor de externe kwaliteitsbewaking van het onderwijs. Maar daarnaast zijn de resultaten van het peilingsonderzoek van belang voor onderwijsorganisaties, ontwerpers van methoden, schoolbegeleiders en leerkrachten, zo vermeldt het CITO in de inleiding van het rapport.

Achtergrond van de Periodieke peiling van het onderwijsniveau (PPON)

In de jaren tachtig groeide de overtuiging dat voor een zakelijke discussie over inhoud en niveau van het basisonderwijs empirische gegevens onontbeerlijk zijn. Het CITO kreeg van het Ministerie van O&W de opdracht periodieke peilingen van het onderwijsniveau voor te bereiden. De taak van het PPON-project is 'op basis van gegevensverzameling bij steekproeven van scholen en leerlingen per leer- en vormingsgebied, periodiek herhaald, een algemeen inzicht te verschaffen in het actuele leer- en vormingsaanbod, de onderwijsleerresultaten en de relaties daar tussen.' (Wijnstra, 1990).

Bij het eerste peilingsonderzoek voor het vak rekenen in 1987 zijn twee soorten instrumenten gebruikt: instrumenten om het leer- en vormingsgebied te inventariseren en instrumenten om rekentaalvaardigheden bij leerlingen te meten. Bij de inventarisatie van het leer- en vormingsaanbod is gebruikgemaakt van een schriftelijke vragenlijst voor de leerkrachten. Deze vragenlijst bevatte vragen over organisatorische kenmerken van het rekenonderwijs (lestijd op het lesrooster, aanwezige rekenmethode, differentiatievormen) en over de inhoud van het rekenonderwijs. De vragen over de inhoud werden gesteld aan de hand van een lijst van leerdoelen op het gebied rekenen/wiskunde. Voor het meten van rekentaalvaardigheden zijn 27 toetsen (zogenoemde vaardigheidsschalen) ontwikkeld. Deze opzet is in 1992 niet volledig herhaald.

De rekenpeiling van 1992

Het rapport bevat zeven hoofdstukken waarin achtereenvolgens aan de orde komen: (1) de uitgangspunten en opzet van het peilingsonderzoek; (2) de uitkomsten van het tweede peilingsonderzoek naar het onderwijsaanbod; (3) de uitkomsten van het tweede peilingsonderzoek naar de rekenprestaties in de domeinen basisoperaties, getallen en hoofdrekenen, cijferen, breuken, procenten, verhoudingen, meten, tijd, geld en meetkunde. De prestaties zijn in 1992 gemeten met 29 vaardigheidsschalen; (4) verschillen tussen diverse groepen leerlingen op de vaardigheidsschalen; (5) verschillen in de prestaties op de vaardigheidsschalen tussen de eerste en de tweede peiling.

De rekenprestaties worden besproken aan de hand van oordelen van deskundigen over a) minimale beheersing, b) voldoende mate van beheersing en c) gevorderde beheersing op de vaardigheidsschalen. Dit kan de interpretatie van de toetsresultaten vergemakkelijken.

Over het *onderwijsaanbod* wordt vermeld dat in 1992 meer scholen dan in 1987 een realistische rekenmethode gebruikten. Leerkrachten besteden in 1992 iets meer tijd aan

rekenen dan in 1987. De zakrekenmachine wordt vooral gebruikt door scholen met een realistische rekenmethode. Wat betreft de *rekenprestaties* lijken de resultaten erop te wijzen dat er ten opzichte van 1987 een vooruitgang is ten aanzien van de basisoperaties en het rekenen met procenten. Achteruitgang is er ten aanzien van de domeinen cijferen en meten (rekenen met maateenheden). De autochtone leerlingen hebben op vrijwel alle schalen hogere gemiddelde scores dan de allochtone leerlingen. Jongens scoren in alle domeinen, behalve cijferen, hoger dan meisjes.

Kanttekeningen bij de opzet van de rekenpeilingen

Nu over twee rekenpeilingen is gerapporteerd, kan worden vastgesteld dat het zwaartepunt van het onderzoek is komen te liggen bij de vraag wat leerlingen kunnen aan het einde van het basisonderwijs. De lezer die verwacht wijzer te worden over de inhoud en het niveau van het onderwijs en over veranderingen daarin, komt bedrogen uit. De informatie die over het rekenonderwijs gegeven wordt, beperkt zich tot de organisatorische randvoorwaarden van het rekenonderwijs. Evenals bij de eerste peiling is informatie verzameld over (a) welke methode er in groep 6, 7 en 8 wordt gebruikt, (b) hoeveel lestijd per week wordt besteed aan rekenen in deze groepen en (c) welke vorm van leerstofdifferentiatie men doorgaans toepast in de klas. Nieuw is dat men vaststelt (d) in welk deel van de rekenmethode de individuele leerlingen van groep 8 waren. Over het feitelijk aanbod van leerinhouden in de verschillende domeinen die met de toetsen zijn gemeten, wordt in tegenstelling tot de rapportage over de eerste peiling geen informatie gegeven. Er is slechts gevraagd hoever de leerling in leerjaar 8 met de methode is gekomen. De wijziging in de meting van het leerstofaanbod bij de tweede peiling is wellicht het gevolg van de bevinding dat de eerste meting van inhoud en niveau geen betrouwbare resultaten levert (zie Harskamp, Suhre en Willemsen, 1991). Door het ontbreken van gegevens over het feitelijke leerstofaanbod worden twee zaken onmogelijk. In de eerste plaats kan er geen vergelijking van veranderingen in het aanbod tussen de eerste en de tweede peiling worden gemaakt. In de tweede plaats zijn eventuele veranderingen in rekenprestaties tussen de eerste en tweede peiling niet te relateren aan het gegeven rekenonderwijs.

Kanttekening bij de resultaten van de tweede PPON-peiling

Uit de gegevens over het aanbod blijkt dat meer dan 80 procent van de leerlingen in groep 8 ten tijde van de toetsafname in het laatste rekenboek is of dat uit heeft. Deze gegevens staan in contrast tot de constatering van de Inspectie van het Onderwijs (1993) dat ruim 40 % van de leerkrachten uit de groepen 7 en 8 in dezelfde periode minder dan de 60% uit de rekenmethode heeft behandeld. De vraag is dan ook welke betekenis de informatie heeft dat de leerlingen in het laatste rekenboek werken. De onderzoekers merken zelf op (pagina 20) dat de rekenmethode in de hogere groepen minder stringent gevolgd wordt.

Het rapport gaat summier in op de methodologische aspecten van de constructie van de vaardigheidsschalen en het trekken van steekproeven. Aan de ene kant is dat voor de lezer prettig, maar aan de andere kant heeft de lezer hierdoor geen zicht op de vergelijkbaarheid van de vaardigheidsschalen en de steekproeven in beide peilingen. Bij de tweede peiling blijken de schalen uit de eerste peiling aangevuld te zijn met nieuwe toetsitems. We maken dit op uit de gepresenteerde items in de rapporten van beide peilingen en de verschillen in aantallen items tussen de peilingen. In de eerste peiling waren er per vaardigheidsschaal aanzienlijk minder items dan in de tweede peiling. Na schaalanalyse zijn ook in de tweede peiling bij de meeste schalen items verwijderd. Of dit items zijn die bij de eerste peiling wel zijn betrokken, blijft onduidelijk. Evenmin is duidelijk of de scores van beide peilingen op dezelfde of vergelijkbare items gebaseerd zijn en op welke wijze oude en nieuwe items aan elkaar zijn gerelateerd. Wellicht levert de

verankering van nieuwe aan oude items bij de meeste schalen geen interpretatieproblemen op voor de vergelijking van het prestatieniveau tussen de peilingen, maar het is toch wel prettig wanneer dat wordt toegelicht.

PPON hanteert een niet proportionele gestratificeerde *steekproefopzet*. Er zijn drie strata gekozen. Het eerste stratum bestaat uit scholen met kinderen waarvan de ouders overwegend een opleidingsniveau op MAVO-niveau of hoger hebben. Op de scholen uit het tweede stratum is het doorsnee opleidingsniveau van de ouders iets lager. De scholen uit het derde stratum worden vaker bezocht door kinderen van allochtone ouders met een laag opleidingsniveau. In vergelijking tot de eerste meting komen de aantallen in stratum 1 en 2 overeen. Het aantal scholen in stratum 3 is in tegenstelling tot de eerste peiling veel lager dan in strata 1 en 2 en selectief samengesteld naar regio. Dit roept vragen op ten aanzien van (a) de gevolgen van de selectiviteit voor de schatting van de toetsprestaties bij de peiling in 1992 en (b) de representativiteit van de deelnemende scholen ten opzichte van de deelnemende scholen vier jaar tevoren. Op de gevolgen van de selectiviteit wordt in het rapport nauwelijks ingegaan. Wel wordt bij vergelijking van de rekenprestaties tussen de eerste en de tweede peiling statistisch gecorrigeerd voor verschillen tussen de steekproeven in de kenmerken methodecategorie, stratum, formatiegewicht, geslacht en de mate waarin leerlingen vertraging hebben opgelopen (pagina 113). Kennelijk is men ervan overtuigd dat hiermee het probleem van de selectiviteit is opgelost. Ook al zou dat zo zijn, de kritische lezer blijft ondertussen zitten met de vraag hoe de twee groepen die bij beide peilingen op de diverse vaardigheidsschalen zijn vergeleken nu feitelijk zijn samengesteld. Meer informatie hierover zou prettig zijn evenals een toelichting op het waarom van de controle voor methodecategorie.

De uitkomsten van de vergelijking van de prestaties tussen reken/wiskundemethoden krijgen een voornamelijk plaats in het rapport. Meer voorbehoud zou daarbij op zijn plaats zijn. Er is namelijk veel kritiek mogelijk op de wijze van *analyseren* van methode-effecten. Ten eerste kan eraan getwijfeld worden of de keuze van een methode onafhankelijk is van de leerlingen die een school bezoeken (zie ook Harskamp en Suhre, 1986). Dat betekent dat bij het vaststellen van effecten grondig rekening gehouden moet worden met cognitieve leerlingkenmerken die medebepalend zijn voor de rekenprestaties en met relevante schoolkenmerken. Ten tweede is een multiniveau-analyse gewenst, waarin prestaties van leerlingen binnen een klas worden gemodelleerd, met controle voor verschillen in relevante leerling- en schoolkenmerken. Dat is niet gebeurd, terwijl al bij de eerste peiling werd erkend dat het gebruikte analysemodel niet correct was (Verhelst en Eggen, 1989). De tegenwerping toen was dat de ontwikkeling van passende multiniveau-software jaren vergt. Inmiddels is er wel het een en ander aan software beschikbaar. Men kan zich verder afvragen of het zin heeft om prestaties van leerlingen van verschillende methoden te vergelijken wanneer niet eens bekend is of een methode door de leerlingen als hoofdmethode is gebruikt en of de leerinhouden uit de methode voor het grootste deel zijn verwerkt. We zijn op dit punt al eerder ingegaan.

De vraag is algemener te stellen, namelijk of het zinvol is om de verschillen in effecten van rekenmethoden te onderzoeken als je niet weet in hoeverre de methoden zijn geïmplementeerd.

Het is prijzenswaardig dat voor de *interpretatie* van de toetsresultaten deskundigen zijn ingeschakeld. Het is echter niet-goed in te zien waarom per vaardigheidsschaal bij de ontwikkeling van een standaard voor minimale of voldoende beheersing slechts het oordeel van 50% van de deskundigen is gebruikt. Het betreft de 50% waarvan de oordelen het meest overeenkomen (zie pagina 16 van het rapport). Per vaardigheidsschaal kan steeds een anders samengestelde groep deskundigen tot deze 50% behoren. In welke mate de oordelen van de andere helft van de deskundigen over de gewenste standaard afwijken kan uit het rapport niet worden opgemaakt. Beter lijkt het ons om de oordelen over een standaard steeds te baseren op de gehele groep deskundigen.

Afsluiting

Het rapport presenteert op heldere wijze de onderzoeksresultaten van de tweede peiling voor rekenen/wiskunde. Er is tevens een bijlage met de toetsopgaven die voor veel leerlingen een plezierige uitdaging vormen. In het eerste hoofdstuk van het rapport wordt de lezer beloofd dat zowel de inhoud als het niveau van het rekenonderwijs worden gepeild. Na lezing van het rapport kun je slechts constateren dat de onderzoekers zich voornamelijk hebben beziggehouden met het bepalen van het prestatieniveau van leerlingen. Het CITO concludeert dat leerlingen op vrijwel alle toetsen in meerderheid onvoldoende scoren, maar laat de kans liggen om aan te geven waarover de discussie inzake het niveau van het rekenonderwijs en de verbetering ervan dient te gaan. Dit komt wellicht omdat het leerstofaanbod in de tweede peiling niet is onderzocht en niet is nagegaan welke factoren in het onderwijsaanbod van invloed zijn op de rekenprestaties. Hierdoor hebben de onderzoeksresultaten weinig praktische waarde. Het onderwijsbeleid wil niet alleen weten of de rekenprestaties van voldoende niveau zijn, maar ook welke initiatieven kunnen worden ontplooid om de kwaliteit van het reken/wiskunde onderwijs te verbeteren (zie Commissie Evaluatie Basisonderwijs, 1994). De kerndoelen geven aan welke leerinhouden scholen dienen aan te bieden. De scholen hebben vooral belang bij informatie over het aanbod dat moet worden gerealiseerd om tot voldoende prestaties met de leerlingen te komen. Een pluspunt van het rapport is dat het een gedetailleerd beeld geeft van de kerndoelen basisonderwijs en in hoeverre leerlingen de kerndoelen feitelijk bereiken. Het rapport verschijnt echter erg laat en de nieuwe peiling zal dit jaar waarschijnlijk op dezelfde voet doorgaan. Jammer, een discussie vooraf over het doel en de realisatie van de peilingen in de afgelopen jaren is gewenst.

E.G. Harskamp en C. Suhre zijn werkzaam bij het Gronings Instituut voor Onderwijs, Opvoeding en Ontwikkeling (GION) te Groningen.

Correspondentieadres: GION, Postbus 1286, 9701 BG Groningen.

Literatuur

- Harskamp, E.G., Suhre, C.J.M. (1986). *Vergelijking van rekenmethoden*. Groningen: RION.
- Harskamp, E.G., Suhre, C.J.M., Willemsen, T.F.W.P. (1991). *De kwaliteit van meetinstrumenten voor het aanbod van rekenen/wiskunde*. GION: Groningen.
- Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit: Evaluatie van het basisonderwijs*. Den Haag: SDU.
- Inspectie van het onderwijs (1993). *Feitelijk leer- en vormingsaanbod op de basisschool. Verslag ten behoeve van de Commissie Evaluatie Basisonderwijs*. De Meern: Inspectie van het onderwijs.
- Verhelst, N.D., Eggen T.J.M. (1989). *Psychometrische en statistische aspecten van peilingsonderzoek. PPO-n rapport nr. 4*. Arnhem: CITO.
- Wijnstra, J. M. (1990). Periodiek peilingsonderzoek: de opzet van het project PPO-n. *Pedagogische Studiën*, 67, 3, 139-150.

Inhoud

Van de redactie / **pag. 107**

Ontwikkelingen in onderwijskundig leiderschap in het basisonderwijs tussen 1989 en 1993 / **pag. 109**
W.J.C.M. van de Grift

Physics Problem Solving and the Transition from General Secondary Education to Higher Education / **pag. 123**
R. Taconis, M. Ferguson-Hessler en G. Verkerk

Vorbereidend lezen en rekenen in kleutergroepen: werkwijzen en resultaten / **pag. 145**
G.W.J.M. Driessen

Ontwikkeling van een adaptieve aanpakstrategie voor het schatten van hoeveelheden / **pag. 161**
L. Verschaffel, E. De Corte, C. Lamote en N. Dhert

Notities en Commentaren / **pag. 186**

Boekbesprekingen / **pag. 198**

Van de redactie

PUBLICEREN IN NEDERLANDSTALIGE WETENSCHAPPELIJKE TIJDSCHRIFTEN?

Onlangs heeft de redactie van het Pedagogisch Tijdschrift een manifest doen verschijnen waarin een oproep wordt gedaan tot behoud van het Nederlandstalig wetenschappelijk forum. Zij heeft zich daarbij ook gericht tot de voornaamste tijdschriften op het terrein van pedagogiek en onderwijskunde. De inhoud luidt als volgt:

Pedagogische wetenschappen en onderwijskunde zijn de laatste jaren geïnternationaliseerd. Het aantal buitenlandse publicaties, met name in het Engels, is sterk toegenomen. Onze participatie in internationale congressen en de contacten met buitenlandse collega's zijn duidelijk verbeterd. Kortom, de Nederlandse en Vlaamse pedagogiek en onderwijskunde zijn internationaal goed zichtbaar geworden. Een belangrijke stimulans hiervan is uitgegaan van de onderzoeksscholen, die minimaal (per vijf jaar) vijf publicaties als toelatingseis hanteren.

Deze positieve ontwikkeling heeft echter ook een schaduwzijde: de Nederlandse taal wordt momenteel sterk veronachtzaamd als medium van wetenschappelijke communicatie. Het onvoorziene neveneffect van de toelatingseis van de Nederlandse onderzoeksscholen en van vergelijkbare Vlaamse beoordelingsprocedures is, dat steeds minder in onze eigen taal wordt geschreven. De Nederlandstalige pedagogische en onderwijskundige tijdschriften, vanouds intermediair tussen wetenschappelijke vakgenoten en tussen wetenschap en praktijk, worden hiervan de dupe. En dat, terwijl de kwaliteit van deze tijdschriften veel beter is dan die van menig wel erkend buitenlands periodiek. Onze verantwoordelijkheid naar de betrokken maatschappelijke velden mag niet op deze manier verwaarloosd worden. Het lijkt wel alsof we over onze eigen problemen nog slechts in vreemde talen met elkaar kunnen communiceren. Dat kan niet de bedoeling zijn!

Er treden oneigenlijke verschuivingen op in wetenschappelijke belangstelling, enkel doordat bepaalde onderwerpen, hoewel ze dringend om onderzoek vragen, gemeden worden omdat ze minder makkelijk in buitenlandse tijdschriften worden opgenomen. Er ontstaan zelfs onbillikheden in de intercollegiale beoordeling, louter omdat het voor specialisaties met meer plaatsgebonden thema's minder voor de hand ligt internationaal te publiceren dan voor andere delen van het vak. Ook dat kan niet de bedoeling zijn!

Wij pleiten daarom voor een herwaardering van het Nederlands als wetenschappelijke taal, en van de Nederlandstalige tijdschriften als wetenschappelijk forum. Niet door de internationalisering terug te draaien, maar door van elke pedagoog en onderwijskundige óók degelijke bijdragen in onze eigen taal te blijven verwachten. De onderzoeksscholen moeten hun toelatingcriteria op dit punt bijstellen door, naast artikelen in de moderne vreemde talen (meervoud!), publicaties van hoog niveau in het Nederlands te eisen. Alleen zo kan het Nederlandstalig wetenschappelijk forum zijn functie naar behoren blijven vervullen.

De redactie voelt zich minder aangesproken door het pleidooi voor een herwaardering van het Nederlands als wetenschappelijke taal. De argumenten die aangevoerd zijn ter behoud van de Nederlandstalige wetenschappelijke tijdschriften worden daarentegen van harte onderschreven. Het verweer dat met Nederlandstalige vaktijdschriften kan worden volstaan gaat voorbij aan het gegeven dat de wetenschappelijk meest relevante en informatieve artikelen op het gebied van opvoedkunde en onderwijskunde veelal zeer toegankelijk van inhoud en

bewijsvoering zijn en derhalve ook voor grotere fora interessant. Te denken valt aan meta-analyses, reviews en analyses van veldexperimenten. Wanneer deze bij voorrang gepubliceerd worden in Nederlandstalige wetenschappelijke tijdschriften blijft niet alleen het forum van Nederlandse wetenschappers op de hoogte van de belangrijkste ontwikkelingen in hun discipline in Nederland maar ook het grotere forum van wetenschappelijk onderlegde professionals in de praktijk. Uiteraard bestaat daarna alsnog de mogelijkheid dezelfde artikelen voor een, vrijwel niet overlappend, internationaal forum te publiceren. De veronderstelling dat met een finale internationale versplintering van publicaties over tientallen en zelfs honderden internationale tijdschriften het functioneren van disciplines als pedagogiek en onderwijskunde in Nederland gediend is, berust op een misverstand.

De oproep aan de onderzoekscholen om voor toelating ook hoogwaardige artikelen in Nederlandstalige wetenschappelijke tijdschriften te laten wordt door de redactie dan ook van harte ondersteund.

De redactie

Dit artikel is oorspronkelijk verschenen in de Nederlandse Tijdschrift voor Onderwijsresearch, 1998, 23(1), 1-2. Het is hierin met de toestemming van de uitgeverij, de Nederlandse Vereniging van Onderwijsresearchers (NVO), afgedrukt. De Nederlandse Vereniging van Onderwijsresearchers (NVO) is een vereniging van wetenschappers die zich bezighouden met onderzoek naar het onderwijs. De Nederlandse Vereniging van Onderwijsresearchers (NVO) is een vereniging van wetenschappers die zich bezighouden met onderzoek naar het onderwijs. De Nederlandse Vereniging van Onderwijsresearchers (NVO) is een vereniging van wetenschappers die zich bezighouden met onderzoek naar het onderwijs.

De redactie van dit tijdschrift is verantwoordelijk voor de inhoudelijke kwaliteit van de artikelen die worden gepubliceerd. De redactie aanvaardt geen aansprakelijkheid voor schade van welke aard ook die voortvloeit uit het gebruik van de inhoud van dit tijdschrift. De redactie aanvaardt geen aansprakelijkheid voor schade van welke aard ook die voortvloeit uit het gebruik van de inhoud van dit tijdschrift.

De redactie aanvaardt geen aansprakelijkheid voor schade van welke aard ook die voortvloeit uit het gebruik van de inhoud van dit tijdschrift. De redactie aanvaardt geen aansprakelijkheid voor schade van welke aard ook die voortvloeit uit het gebruik van de inhoud van dit tijdschrift.

De redactie aanvaardt geen aansprakelijkheid voor schade van welke aard ook die voortvloeit uit het gebruik van de inhoud van dit tijdschrift. De redactie aanvaardt geen aansprakelijkheid voor schade van welke aard ook die voortvloeit uit het gebruik van de inhoud van dit tijdschrift.

Ontwikkelingen in onderwijskundig leiderschap in het basisonderwijs tussen 1989 en 1993

W.J.C.M. van de Grift

ABSTRACT

Educational leadership is the capacity or ability of a principal which finds expression in the initiation of school improvement, the creation of a learning oriented educational climate and the stimulation and supervision of teachers, so that the latter may execute their tasks as well and as effectively as possible. In 1989 and in 1993 teacher perceptions of educational leadership of principals in Dutch elementary education were evaluated by means of a Rasch scale. It turned out that the educational leadership of Dutch principals has been grown considerable. In both 1989 and 1993 the effect of educational leadership on pupil achievement was investigated. While in 1989 no significant relationship was found, the results of 1993 do however show a significant relationship between educational leadership and average pupil achievement over three succeeding years.

INLEIDING

In 1989 en in 1993 zijn in twee projecten met hetzelfde instrument gegevens verzameld over het onderwijskundig leiderschap van schoolleiders en de output van basisscholen (Van de Grift & Akkermans, 1991; Commissie Evaluatie Basisonderwijs, 1994ab). Secundaire analyse op de gegevensbestanden van beide projecten biedt de mogelijkheid de volgende vragen te beantwoorden:

1. Zijn onderwijsgeveenden in de periode tussen 1989 en 1993 hun schoolleider meer als onderwijskundig leider gaan ervaren?
2. Is er verband tussen het onderwijskundig leiderschap van schoolleiders en de output van basisscholen?

Voordat deze vragen worden beantwoord, wordt eerst ingegaan op de context waarbinnen Nederlandse schoolleiders hun werk doen. Daarna komen de kwaliteit van de steekproeven en de psychometrische eigenschappen van het gebruikte instrument aan de orde.

De context waarbinnen de schoolleider werkt

De Nederlandse schoolleider heeft zich de afgelopen veertig jaar ontwikkeld van het 'hoofden school', die op de kweekschool de hoofdakte had gehaald, tot directeur van een basisschool met een uitgebreid takenpakket op het gebied van het onderwijskundig leiderschap. Deze rol van de schoolleider is, onder invloed van een aantal belangrijke maatschappelijke, politieke en beleidsmatige factoren, nog steeds in ontwikkeling. Er treden veranderingen op in de schoolgrootte, er is meer eigen schoolbeleid noodzakelijk, de bekostigingswijze voor het personeel is in verandering en de taak op het gebied van de externe contacten is gecompliceerder geworden. Deze vier groepen veranderingen in de werksituatie van de schoolleider zijn van invloed op de rol van de schoolleider. Ze worden daarom nu behandeld.

De schoolgrootte is toegenomen in het basisonderwijs, omdat er hogere grenzen gesteld zijn aan het minimum aantal leerlingen en omdat onder invloed hiervan scholen gefuseerd

zijn. In het schooljaar 1992/1993 bedroeg het gemiddeld aantal leerlingen ongeveer 167, in 1993/1994 was dit 170. In het schooljaar 1994/1995 is dit gemiddeld aantal toegenomen met 13 leerlingen tot 183. Op dit moment ligt dit getal boven de 200 leerlingen. Het beleid dat gericht is op schaalvergroting in het basisonderwijs brengt niet alleen meer leerlingen per school, maar ook meer personeel per school met zich mee. Dit leidt tot ingewikkelder personeelsbeleid en complicaties op organisatorisch en huishoudelijk terrein.

Echter niet alleen de schoolgrootte is een element dat verandering brengt in de werksituatie van de schoolleider. Door de autonomievergroting en de afname van centrale regelgeving moeten schoolbestuur en schoolleider ook meer dan ooit eigen schoolbeleid gaan maken. Dit eigen schoolbeleid speelt niet alleen op onderwijsinhoudelijk vlak (kerndoelen, schoolgids, schoolplan), maar ook op het terrein van personele zaken en kwaliteitszorg.

De nieuwe bekostigingswijze voor het personeel, het zogenoemde formatiebudgetsysteem, maar ook de regeling voor arbeidsduurverkorting, de toename van het aantal parttimers en duobanen brengen een grotere complexiteit met zich mee in het voorbereiden en uitvoeren van het personeelsbeleid en dit heeft weer gevolgen voor het organisatorisch en huishoudelijk beleid in de school.

Ook de taak op het gebied van het onderhouden van interne en externe contacten is gecompliceerder en vooral uitgebreider geworden. Ging het vroeger vooral om contacten met ouders, de naburige kleuterschool, het voortgezet onderwijs en de begeleidingsdienst, nu vragen nieuwe beleidsontwikkelingen die het gevolg zijn van onder meer het 'Schevenings Beraad' en 'Weer samen naar school' om meer en vooral intensiever contacten met ouders, met andere basisscholen en met scholen voor speciaal onderwijs.

Vastgesteld kan worden, dat op het gebied van het management en op het gebied van het onderwijskundig leiderschap het aantal taken van schoolleiders duidelijk uitgebreid is. Dit leidt tot de vraag: Hoeveel tijd en gelegenheid hebben schoolleiders voor het uitoefenen van hun taken? Uit onderzoek van Stoel (1993) blijkt, dat schoolleiders gemiddeld gezien een weektaak hebben van ongeveer 50 uur. Uit evaluaties van de Inspectie (1994ab) onder bijna 400 scholen blijkt dat de gemiddelde schoolleider zo'n 43 uur per week op school aanwezig is. In de 'ambulante' uren moeten alle directietaken uitgevoerd worden. Slechts zo'n 12% van de schoolleiders had in het schooljaar 1991/1992 meer dan de helft van de werktijd beschikbaar voor het uitvoeren van de directietaken. In 1992/1993 groeide dit aantal schoolleiders met meer dan 20 uur formatie naar 14.7% en in 1993/1994 naar 15.6%. Van de tijd die beschikbaar is voor directietaken gaat ongeveer 30% naar administratieve taken, ongeveer 30% wordt besteed aan beleid maken al dan niet naar aanleiding van het beleid van het departement, eveneens zo'n 30% gaat naar managementtaken en zo'n 10% gaat naar externe contacten en werven van nieuwe leerlingen (zie ook Brandsma & Knuver, 1989; Knuver, 1993; Stoel, 1993). Schoolleiders blijken maar in zeer beperkte mate taken te delegeren naar andere leden van het schoolteam en de vervulling van de directietaken neemt vrij vaak de vorm aan van participatie in overlegsituaties (Stoel, 1993).

Deze context waarbinnen de schoolleider werkt bracht de Commissie Evaluatie Basisonderwijs tot de aanbeveling dat de gemiddelde basisschool zou moeten beschikken over minimaal 24 uur formatie voor de schoolleiding. Op dit moment is dat gemiddelde slechts ruim 15 uur. Hiervan is nog geen 5 uur beschikbaar voor het onderwijskundig leiderschap.

METHODE

Steekproeven

De bevindingen uit deze publicatie zijn gebaseerd op twee steekproeven: Een steekproef uit 1989 (Van de Grift & Akkermans, 1991) en een steekproef uit 1993 (Commissie Evaluatie Basisonderwijs, 1994a).

Er waren bij de uitvoering van het onderzoek dat in 1989 plaatsvond geen redenen voor

het trekken van een gestratificeerde steekproef. Daarom is volstaan met het trekken van een enkelvoudige steekproef. Ten behoeve van de representativiteit van de steekproef zijn aselekt 500 basisscholen uit de populatie getrokken. Deze scholen is om medewerking aan het onderzoek verzocht. Van 275 scholen werd deze medewerking ontvangen. Bij 250 scholen was de respons volledig. De representativiteit van deze respons is gecontroleerd op een aantal populatiekarakteristieken. Het bleek dat de verdeling van onderzochte kenmerken (denominatie, schoolgrootte en leerlingkenmerken) van het responderende deel van de steekproef niet of niet noemenswaard afweek van de populatie. De respons is voldoende groot om een nauwkeurigheid van ruim .06 te bereiken bij schattingen van proporties in de populatie. Geconcludeerd kan worden dat met behulp van de gegevens die bij deze responderende scholen verzameld zijn, voldoende nauwkeurige en voldoende representatieve uitspraken gedaan kunnen worden over de populatie. Van deze 250 scholen waren 73 scholen bereid om ten behoeve van het onderzoek hun leerlingen de door het Cito vervaardigde Entreetoets voor het laatste leerjaar te laten maken. De gegevens van deze scholen zijn benut voor het nagaan van het verband tussen onderwijskundig leiderschap en de voor milieu van herkomst gecorrigeerde leerlingprestaties uit 1989.

De door de inspectie in 1993 verzamelde gegevens waren bedoeld om evaluatieve uitspraken te kunnen doen over de proportie schoolleiders die voldoen aan de normen van de inspectie. Daarom is in dit project geopteerd voor de zeer grote steekproefnauwkeurigheid van .05. Om deze nauwkeurigheid (e) te bereiken is bij een populatie (N) van 8,402 scholen, onder gebruikmaking van formule $\{n=N/(1+N(e)^2)\}$, een steekproefgrootte (n) van 382 scholen nodig. Er waren ook in dit onderzoek geen redenen voor het trekken van een gestratificeerde steekproef. Daarom is ook hier volstaan met het trekken van één enkelvoudige aselekt steekproef. Om rekening te kunnen houden met onvermijdelijke uitval is aselekt een steekproef van 386 basisscholen getrokken. Gedurende de uitvoering van het project konden bij drie scholen geen gegevens verzameld worden. Hier hebben steeds bijzondere redenen en oorzaken, zoals overmacht, een rol gespeeld. Het aantal deelnemende scholen komt daarmee te liggen op 383 en blijft daarmee net boven het gewenste aantal van 382 scholen. De uitval is minder dan 1%. De deelnemende scholen zijn gecontroleerd op een aantal populatiekarakteristieken zoals: Denominatie, schoolgrootte en leerlingachtergronden. Hieruit bleek dat de verdelingen van kenmerken: denominatie, schoolgrootte en leerlingachtergronden, die in de steekproef van 383 scholen aangetroffen worden, niet of niet noemenswaard afwijken van de populatie. Geconcludeerd kan worden dat de gegevens die in deze steekproef verzameld zijn representatief geacht kunnen worden voor de populatie en dat de steekproef voldoende groot is om een nauwkeurigheid van .05 te bereiken bij schattingen van proporties in de populatie.

Om het verband tussen onderwijskundig leiderschap en de output van scholen uit de tweede steekproef te kunnen berekenen, zijn na toestemming van de betreffende scholen de gemiddelde Cito-eindtoetsresultaten over de jaren 1991, 1992 en 1993 bij het Cito opgevraagd. Slechts een enkele school onthield de toestemming om deze gegevens bij het Cito op te vragen. Aangezien niet alle scholen gedurende alle genoemde jaren deelnamen aan deze toets, zijn deze gegevens beperkt tot 174 basisscholen. Daarmee is de oorspronkelijke representatieve steekproef van 383 scholen meer dan gehalveerd. Het is daarom van belang om vast te stellen of de gegevens die bij deze groep scholen verzameld zijn voldoende nauwkeurig en voldoende representatief zijn voor een goed beeld van de situatie in de populatie. De nauwkeurigheid waarmee op basis van de steekproef van 174 'Citotoetssscholen' de proportie geschat kan worden van het aantal schoolleiders dat voldoet aan de norm van de inspectie, bedraagt 7.5%.

Om de representativiteit van de 'Citotoetssscholen' te beoordelen is een vergelijking gemaakt van de frequentieverdeling van enkele kenmerken van deze scholen met de scholen in de populatie. Ook hier bleken de verschillen beperkt te blijven tot enkele procenten als het gaat om denominatie en leerlingssamenstelling. Er is op dit punt dus geen reden om de representativiteit in twijfel te trekken. Wel bleken de 'Citotoetssscholen' gemiddeld meer leerlingen

te hebben, dan de gemiddelde basisschool in Nederland. Daarom is nader onderzocht wat de correlatie was tussen het onderwijskundig leiderschap van schoolleiders en de grootte van de school (i.e. het aantal leerlingen). De correlatie tussen onderwijskundig leiderschap en schoolgrootte bleek .07 te bedragen. Deze correlatie wijkt niet significant van 0 af. Er is dus geen lineair verband tussen de grootte van de school en de kwaliteit van onderwijskundig leiderschap. We kunnen dus vaststellen dat de resultaten op het punt van onderwijskundig leiderschap niet vertekend worden door het feit dat de Citotoetscholen gemiddeld iets groter zijn.

Samenvattend: De steekproef van 'Citotoetscholen' is voldoende groot en voldoende vergelijkbaar met de populatie om nauwkeurige en representatieve uitspraken te kunnen doen over het onderwijskundig leiderschap in deze populatie.

Instrument voor het evalueren van 'onderwijskundig leiderschap'

Onderwijskundig leiderschap is de vaardigheid van schoolleiders die tot uitdrukking komt in het initiëren van verbeteringen, het creëren van een op leren gericht schoolklimaat en het onderwijsgeevenden zodanig stimuleren en begeleiden dat zij hun taken zo effectief mogelijk uitvoeren. Deze definitie van onderwijskundig leiderschap bevat twee aspecten: een beschrijving van een bepaalde groep gedragingen van schoolleiders en een verwachting van wat deze gedragingen van schoolleiders bij onderwijsgeevenden teweegbrengen. Dit laatste moet uiteindelijk zijn resultaat vinden in de output van de school.

De vaardigheid 'onderwijskundig leiderschap' komt tot uitdrukking in vier groepen gedragingen van de schoolleider die onderling op elkaar afgestemd moeten zijn: het initiëren van verbeteringen, het creëren van een op leren gericht schoolklimaat en zowel het stimuleren als het begeleiden van onderwijsgeevenden. Het initiëren van verbeteringen betreft onder meer het actief meedoen aan de keuze van nieuwe onderwijsmethoden, het onderwijsgeevenden informeren en in contact brengen met nieuwe methoden en instructiematerialen en het prioriteit geven aan onderwijsmethoden die tot verbetering van de prestaties van de leerlingen kunnen leiden. Het creëren van een op leren gericht klimaat behelst onder meer het benadrukken van het belang van het regelmatig peilen van de leerlingprestaties en het aansporen van onderwijsgeevenden tot het verhogen van de prestaties van de leerlingen. Bij het stimuleren gaat het onder meer om: onderwijsgeevenden het gevoel geven dat hun werk belangrijk is en waardering tonen voor onderwijsgeevenden die hun werk goed doen. Begeleiden betreft het beschikbaar zijn als onderwijsgeevenden problemen willen bespreken; van teamvergaderingen waardevolle onderwijskundige bijeenkomsten maken; tijdig ingrijpen als er problemen optreden; belangstelling tonen voor wat er in de klassen gebeurt en onderwijsgeevenden adviseren, begeleiden en ondersteunen bij het verbeteren van de leerlingprestaties. De mate waarin deze gedragingen van schoolleiders door onderwijsgeevenden ervaren wordt, is indicatief voor de vaardigheid die we onderwijskundig leiderschap noemen.

Een aantal van deze gedragingen is voorwaardelijk voor ander gedrag: Van een schoolleider, die niet openstaat voor de problemen die onderwijsgeevenden met hun leerlingen ervaren, kan niet verwacht worden dat hij onderwijsgeevenden begeleidt en ondersteunt. Een schoolleider die geen op leren gericht klimaat creëert en die bij de keuze van nieuwe onderwijsmethoden niet let op de prestaties van de leerlingen zal weinig succes hebben in het aansporen van onderwijsgeevenden om regelmatig de leerlingprestaties te peilen. Sommige gedragingen zijn gemakkelijk door vrijwel alle schoolleiders te realiseren, andere vergen meer training en routine. Onderwijskundig leiderschap manifesteert zich in een dimensie van gedragingen die loopt van het openstaan voor onderwijsgeevenden tot en met het ondersteunen van onderwijsgeevenden in hun klaslokalen. Deze dimensie bevat net als een meetlat een aantal gradaties. Deze gradaties worden weergegeven aan de hand van de plaats die de items op de dimensie innemen. Om deze plaats te kunnen bepalen zijn de items gekalibreerd (Zie voor details Van de Grift & Akkermans, 1991). Dat wil zeggen: Voor elk item uit de schaal is de plaats op de dimensie (de σ -waarde) berekend. De σ -waarden van de dimensie zijn uitgedrukt in een logaritmische schaal. De schaal kan in theorie lopen van $-\infty$ tot $+\infty$. De gemiddelde σ -waarde

van alle items wordt vastgezet op 0. In tabel 1 worden de σ -waarden voor de items en hun bijbehorende standaardfout (s.e.) gepresenteerd. In tabel 1 kan afgelezen worden dat de itemparameters (de σ -waarden) goed gespreid zijn over de dimensie. Ze lopen van -2.76 tot en met 1.76.

Tabel 1: Schaal voor het evalueren van onderwijskundig leiderschap

items	σ	s.e.
01 leerkrachten kunnen bij de schoolleider terecht als ze problemen van leerlingen willen bespreken	-2.76	.25
02 de schoolleider doet actief mee aan de keuze van nieuwe onderwijsmethoden	-2.31	.23
03 de schoolleider geeft leerkrachten het gevoel dat hun werk belangrijk is	-1.18	.18
04 de schoolleider geeft ervoor dat teamvergaderingen waardevolle onderwijskundige bijeenkomsten zijn	-.27	.16
05 de schoolleider geeft hoge prioriteit aan onderwijsmethoden, die tot een verbetering van de leerresultaten van alle leerlingen kunnen leiden	-.25	.16
06 de schoolleider zorgt voor een op leren gericht schoolklimaat	-.10	.16
07 de schoolleider toont waardering voor leerkrachten die hun werk goed doen	.03	.16
08 de schoolleider bespreekt tijdig de problemen die kunnen optreden bij de uitvoering van het werk op school	.26	.16
09 de schoolleider informeert ons over nieuwe onderwijsmethoden en instructiematerialen	.31	.16
10 de schoolleider toont belangstelling voor wat er in mijn klas gebeurt	.31	.16
11 de schoolleider benadrukt het regelmatig peilen van leerlingprestaties	.70	.16
12 de schoolleider spoort leerkrachten aan tot aandacht voor het verhogen van de leerprestaties	1.13	.16
13 de schoolleider brengt ons in contact met nieuwe ideeën en benaderingen die de leerlingprestaties kunnen verbeteren	1.17	.16
14 de schoolleider geeft leerkrachten advies over de manier waarop de prestaties van zwakke leerlingen kunnen worden verbeterd	1.20	.16
15 de schoolleider begeleidt en ondersteunt leerkrachten bij het verbeteren van de leerprestaties	1.76	.17

antwoordmodaliteiten:

0 = nooit, bijna nooit, af en toe

1 = vaak, bijna altijd, altijd

schaalwaarden θ

0=-4.00 1=-3.36 2=-2.41 3=-1.76 4=-1.25 5=-0.82 6=-0.43 7=-0.08

8=0.26 9=0.60 10=0.95 11=1.32 12=1.75 13=2.29 14=3.11 15=4.00

Het Raschmodel biedt niet alleen de mogelijkheid om de plaats van de items op de dimensie aan te geven; ook de personen, in dit geval de schoolleiders, krijgen met behulp van de scores op een Raschschaal hun plaats op de dimensie. De vaardigheid van de schoolleiders wordt aangeduid met de Griekse letter θ . De schaal biedt de mogelijkheid om de schoolleiders in te delen in 16 vaardigheidsgroepen. Deze vaardigheidsgroepen zijn tezamen met hun bijbehorende ruwe score onderaan in tabel 1 weergegeven. Met het vaststellen van de θ - en de σ -waarden is de maatverdeling waarmee het onderwijskundig leiderschap gemeten kan worden, gereed. Om deze schaal als meetlat te kunnen gebruiken, is het van belang dat aan de volgende voorwaarden wordt voldaan.

De schaal moet voldoen aan:

1. de homogeniteitseisen, die in de klassieke testtheorie (Nunnally & Bernstein, 1994) gesteld

- worden (De schaal moet een KR-20 van ten minste .70 en een gemiddelde itemtotaalcorrelatie van ten minste .50 hebben.),
2. de assumpties die vanuit het Raschmodel aan een goede schaal gesteld kunnen worden (Dat wil zeggen de schaal moet unidimensioneel zijn en de items moeten parallelle itemkarakteristieke curves hebben.),
 3. diverse validiteitseisen (De conclusies die uit de scores op het instrument getrokken worden moeten in overeenstemming zijn met het meetdoel van het instrument.).
- Deze zaken worden in de volgende paragrafen behandeld.

Homogeniteit, unidimensionaliteit en paralleliteit van itemkarakteristieke curves

De schaal blijkt met een KR-20 van .82 en een gemiddelde itemtotaalcorrelatie van meer dan .50 te voldoen aan de eisen die in de klassieke testtheorie aan een homogene schaal gesteld worden.

Om de unidimensionaliteit te kunnen onderzoeken is nagegaan of de maatverdeling van de dimensie 'Onderwijskundig leiderschap' beïnvloed wordt door diverse kenmerken van de schoolleiding, van de school, en van de onderwijsgeveden en de klassen waarin de onderwijsgeveden werken. Voor elk van deze groepen van kenmerken zijn groepen respondenten onderscheiden. Voor elke groep zijn de σ -waarden berekend. Deze itemparameters zijn vervolgens met behulp van de Andersentoets (een χ^2 -toets) getoetst op gelijkheid. Alle χ^2 -en op één na bleken een overschrijdingskans te hebben die groter is dan .05. Hieruit kan geconcludeerd worden, dat de unidimensionaliteit van de schaal niet in gevaar gebracht wordt door de onderzochte kenmerken van de school en schoolleiding en evenmin door de kenmerken van de onderwijsgeveden (en hun klassen) die de beoordeling van het onderwijskundig leiderschap uitgevoerd hebben. Er is één uitzondering op het geheel. De schaal blijkt gevoelig te reageren wanneer de beoordeling van het onderwijskundig leiderschap gebeurt door leerkrachten die nog maar kort op school werken. Voor toekomstig gebruik van de schaal moet dan ook de eis gesteld worden dat onderwijsgeveden ten minste vier jaar werken op de betreffende school. (Zie voor details Van de Grift & Akkermans, 1991.)

Voorts is nagegaan of de items uit de schaal een gelijk discriminerend vermogen hebben. Dit gelijke discriminerend vermogen dat tot uitdrukking komt in het parallel zijn van de itemkarakteristieke curves, is nagegaan met behulp van de Martin-Löf test en de Andersentest. Beide toetsen zijn χ^2 -toetsen. De χ^2 -en van beide toetsen bleken hoge overschrijdingskansen te hebben. Daarmee is de paralleliteit van de itemkarakteristieke curves aangetoond.

Geconcludeerd kan worden dat de Raschschaal voldoet aan de assumpties van het Raschmodel. De items van de schaal hebben parallelle curves en de schaal is unidimensioneel. (Zie voor details Van de Grift & Akkermans, 1991.)

Validiteit

Een onderzoeksinstrument is valide als uit de scores op dat instrument conclusies getrokken kunnen worden die in overeenstemming zijn met het meetdoel van dat instrument. Het meetdoel van dit instrument betreft het onderwijskundig leiderschap van schoolleiders in het basisonderwijs. Onderwijskundig leiderschap is de vaardigheid van schoolleiders die tot uitdrukking komt in het initiëren van verbeteringen, het creëren van een op leren gericht schoolklimaat en het onderwijsgeveden zodanig stimuleren en begeleiden dat zij hun taken zo effectief mogelijk uitvoeren. Deze definitie van onderwijskundig leiderschap bevat twee componenten: een beschrijving van bepaalde groepen gedragingen van schoolleiders en een verwachting van wat deze gedragingen van schoolleiders bij onderwijsgeveden teweegbrengen. Van dit laatste veronderstellen we dat dit uiteindelijk zijn resultaat zou moeten vinden in de output van de school. Voor het validiteitsonderzoek betekent dit dat we gegevens moeten verzamelen om antwoord te kunnen geven op de volgende vragen:

1. Worden met de indicatoren van dit instrument (i.e. de ervaringen van onderwijsgeveden op een 15-tal aspecten) gegevens verzameld over het bedoelde gedrag van schoolleiders?

2. Gaat een hogere score op het onderzoeksinstrument gepaard met een effectievere taakuitvoering van onderwijsgeevenden?
3. Gaat een hogere score op het onderzoeksinstrument gepaard met een hogere output van scholen?

Wanneer onderzoek uitwijst dat we deze vragen positief kunnen beantwoorden, dan hebben we evidentie voor de validiteit van de schaal.

De eerste vraag wordt positief beantwoord als met dit instrument evaluatieresultaten worden bereikt, die niet afwijken van een andere procedure waarmee hetzelfde meetdoel beoogd wordt. Hierop wordt nu eerst ingegaan. In het project uit 1989 hebben onderwijsgeevenden het instrument over onderwijskundig leiderschap ingevuld ten behoeve van een aan een universiteit uitgevoerd onderzoek. In 1993 was er sprake van een andere situatie. Nu werd door de inspecteur aan de onderwijsgeevenden gevraagd om het instrument over de schoolleider in te vullen. Hierbij zou het risico kunnen ontstaan dat onderwijsgeevenden hun schoolleider ten overstaan van de inspectie 'een plezier' willen doen, of hem juist 'een streek' willen leveren. Mede om deze situatie onder controle te houden, hebben inspecteurs zich ook zelf een oordeel gevormd over het onderwijskundig leiderschap van de schoolleider. Daartoe voerden de inspecteurs gesprekken met onderwijsgeevenden en de schoolleider. Om tot hun oordeel te komen, werkten de inspecteurs met een daartoe ontworpen inspectie-instrument, een Likert-schaal, die uit 11 indicatoren bestaat. Deze schaal voldoet met een Cronbach's alpha van .79 aan de betrouwbaarheidseisen, die in de klassieke testtheorie worden gesteld. Daarmee beschikken we voor de steekproef uit 1993 over twee onafhankelijke procedures voor het vaststellen van de kwaliteit van het onderwijskundig leiderschap van schoolleiders. Beide procedures bevatten een normering voor de cesuur voldoende/onvoldoende.

Voor de Raschschaal die bij de onderwijsgeevenden is afgenomen, heeft de inspectie als norm gesteld dat schoolleiders ten minste aan 10 van de 15 items zouden moeten voldoen. Deze norm is ook gebruikt bij de evaluatie van het basisonderwijs (Zie Commissie Evaluatie Basisonderwijs, 1994a). Het bleek dat het inspectie-oordeel voor 8,6% van de schoolleiders afweek van de perceptie van de onderwijsgeevenden. Er waren ongeveer evenveel positieve als negatieve afwijkingen. Dit brengt met zich mee, dat het oordeel van de inspectie voor 91,4% overeenstemt met dat van de onderwijsgeevenden dat op de Raschschaal gebaseerd is. We kunnen daarom concluderen, dat de scores op de Raschschaal niet of nauwelijks tot andere vaststellingen leiden dan het oordeel van de inspectie. Dat en het feit dat beide instrumenten een correlatie van .73 met elkaar hebben, vormen belangrijke aanwijzingen voor de validiteit van de schaal.

De tweede vraag die centraal staat in het valideringsonderzoek is of een hogere score op de Raschschaal gepaard gaat met een effectievere taakuitvoering van onderwijsgeevenden. Daarom is bij de onderwijsgeevenden die het instrument over onderwijskundig leiderschap over hun schoolleider hebben ingevuld, nagegaan in welke mate zij de (reken)resultaten van hun leerlingen volgen door het nakijken van de (oefen)schriften en het op basis hiervan aan de leerlingen individueel, dan wel groepsgewijs, feedback geven. Deze Likertschaal, die uit drie indicatoren bestaat, voldoet met een Cronbach's alpha van .76 aan de betrouwbaarheidseisen die in de klassieke testtheorie gesteld worden. Het bleek dat de score van de onderwijsgeevenden op deze Likertschaal met een correlatie van .23 positief en significant samenhangt met de score van de schoolleider op de Raschschaal. Dit is een tweede aanwijzing voor de validiteit van de Raschschaal.

De derde vraag die in het valideringsonderzoek centraal stond is: Gaat een hogere score op het onderzoeksinstrument gepaard met een hogere output van de school? Als de scores op de Raschschaal geen voorspellende waarde hebben voor de output van basisscholen, dan is de waarde van de schaal van beperkte betekenis. In 1989 bleek dat er geen significant verband gevonden kon worden tussen het onderwijskundig leiderschap van de schoolleiders en de voor milieu gecorrigeerde prestaties van de leerlingen op de Entreetoets van het Cito. Voor de gegevens uit 1993 is opnieuw dit validiteitsaspect onderzocht. Het bestaan van een

significante samenhang tussen de score op de schaal en de hoogte van de output van de school is een aanwijzing voor de validiteit van de schoolleiderschapsschaal. Om dit te kunnen onderzoeken is een indicator nodig voor de output van basisscholen. Er zijn diverse maten die benut zouden kunnen worden om als indicator voor de output van basisscholen te dienen (zie ook Bosker, 1991):

1. de gemiddelde prestatiescore van leerlingen op een toets,
2. de gemiddelde prestatiescore van de leerlingen op een toets, nadat gecorrigeerd is voor leerlingkenmerken waarop scholen weinig of geen invloed kunnen hebben zoals intelligentie, sekse, nationaliteit en milieu van herkomst,
3. de gemiddelde prestatiescore van leerlingen op een toets, nadat met behulp van een voormeting gecorrigeerd is voor aanvangsverschillen tussen leerlingen,
4. de gemiddelde prestatiescore van leerlingen, die acht jaar basisonderwijs hebben gevolgd op een toets, nadat met behulp van een voormeting gecorrigeerd is voor aanvangsverschillen tussen leerlingen aan het begin van het basisonderwijs.

De vierde outputmaat is in theoretisch opzicht de beste maat. Dit geldt zeker als alle leerlingen die op de betreffende school acht jaar basisonderwijs gevolgd hebben (dus ook degenen, die wegens dubleren nog in groep 7 en in groep 6 zitten) deel uitmaken van de onderzochte leerlingengroep. Het enige dat hier nog aan toegevoegd zou moeten worden is een correctie voor het aantal leerlingen dat gedurende acht jaar door de betreffende basisschool naar het speciaal onderwijs is verwezen. De vierde outputmaat blijft echter in theorie de beste maat. In praktijk kan zelden aan al deze voorwaarden voor het samenstellen van deze outputmaat voldaan worden. De derde maat is een klassieke effectmaat, die vaak, zij het op kleine schaal, in onderwijsexperimenten gebruikt wordt. De eerste maat moet voor praktisch gebruik verworpen worden, omdat scholen nu eenmaal weinig invloed kunnen uitoefenen op de instroomkenmerken van leerlingen, terwijl deze instroomkenmerken toch een fors deel van de variantie in leerlingprestaties verklaren¹. In de praktijk van veel onderzoek wordt een variant van de tweede maat gebruikt. Dat wordt ook hier gedaan.

Voordat de vraag naar dit validiteitsaspect van het instrument beantwoord kan worden, moet duidelijk zijn dat de gekozen schooloutputindicator aan een drietal belangrijke voorwaarden voldoet:

1. scholen moeten verschillen op de outputindicator
2. de verschillen moeten stabiele verschillen tussen scholen zijn en
3. deze verschillen moeten vatbaar zijn voor de invloed van de school.

Immers, als er aan het eind van het basisonderwijs geen stabiele door scholen beïnvloedbare verschillen zijn tussen gemiddelde leerlingprestaties van scholen, dan heeft het weinig zin om onderzoek op te zetten naar de samenhang van deze verschillen met het onderwijskundig leiderschap van de directeuren. De verschillen tussen scholen in gemiddelde leerlingprestaties zullen dan verklaard kunnen worden met incidenten of toevallige factoren. De eerste vraag, die ten aanzien van de hier gekozen indicator gesteld moet worden, luidt dan ook: Zijn er over de jaren heen aan het eind van het basisonderwijs verschillen tussen de gemiddelde leerlingprestaties van scholen? (Zie Blok & Eiting, 1988; Van Batenburg, 1990; Bosker, 1991 en Blok & Hoeksma, 1993.) Daarna moet de vraag beantwoord worden of het noodzakelijk is om te corrigeren voor eventuele effecten van de schoolomgeving. We gaan eerst op de eerste vraag in; daarna op de tweede.

Het is uiteraard irrealistisch om te verwachten dat de output van een school jaar in jaar uit volkomen identiek is. Er zullen altijd ten minste kleine verschillen gevonden worden in de gemiddelde leerlingprestaties van dezelfde school over de verschillende leerjaren heen. Wanneer deze verschillen beneden een bepaalde grens blijven, dan kan de schooloutput stabiel genoemd worden. Deze grens wordt als volgt gedefinieerd: Als de verschillen tussen de verschillende schooljaren van dezelfde school kleiner zijn dan de verschillen die tussen scholen bestaan, dan kan de schooloutput stabiel genoemd worden. Technisch geformuleerd: De binnengroepsvariantie (de verschillen tussen de jaren binnen een school) moet kleiner zijn dan

de tussengroepsvariantie (de verschillen tussen de schoolgemiddelden over de jaren heen). De verhouding tussen deze beide variantiedelen wordt de intraclasscorrelatiecoëfficiënt genoemd. Om de verbanden tussen het onderwijskundig leiderschap van schoolleiders en de output van de scholen te kunnen berekenen, zijn na toestemming van de betreffende scholen de gemiddelde Cito-eindtoets resultaten over de jaren 1991, 1992 en 1993 bij het Cito opgevraagd. De benodigde berekeningen zijn uitgevoerd met behulp van het statistisch programma ML3, dat speciaal voor multiniveau-analyses ontwikkeld is door Prosser, Rasbash & Goldstein (1991). (Zie ook Goldstein, 1987.) De uitkomsten staan vermeld in tabel 2. Aan de hand van de resultaten in tabel 2 (bij het 0-model) kan geconstateerd worden, dat de intraclasscorrelatiecoëfficiënt .61 bedraagt. Hieruit blijkt dat de verschillen tussen scholen bijna twee keer zo groot zijn als de verschillen tussen de verschillende schooljaren van dezelfde scholen. Daarom kan geconcludeerd worden dat er sprake is van relatief stabiele verschillen in de gemiddelde Cito-eindtoetsresultaten van basisscholen.

De relatief kleine basisscholen in Nederland 'rekruteren' hun leerlingen uit een in omvang beperkte omgeving. Deze omgeving bestaat doorgaans uit slechts enkele wijken van een stad of dorp. Deze wijken hebben hun eigen specifieke sociaal-economische kenmerken. Op deze kenmerken van de wijken heeft de school nauwelijks of geen invloed, terwijl omgekeerd de wijk en haar bevolking wel een stabiele invloed kunnen hebben op de gemiddelde schoolprestaties. Daarom is onderzocht of er verband is tussen de gemiddelde leerlingprestaties van scholen en hun omgevingskenmerken. Om dit te kunnen berekenen, is de gebruikte regressievergelijking uitgebreid met een maat voor de schoolomgeving. De schoolomgeving is voor elk schooljaar afzonderlijk geoperationaliseerd als het gemiddelde leerlinggewicht² in het betreffende schooljaar. Wanneer dit gemiddelde leerlinggewicht voor de drie afzonderlijke schooljaren ingevoerd wordt in de regressievergelijking, dan blijkt de hoeveelheid nog te verklaren tussenscholenvariantie te zakken naar 56.2%. Ook de binnenscholenvariantie verandert een fractie, maar dat is te verwaarlozen. Uit dit laatste kan geconcludeerd worden dat de schoolomgeving relatief stabiel is. De resultaten zijn te vinden in tabel 2. Verder kan in tabel 2 geconstateerd worden, dat de regressiecoëfficiënt voor de schoolomgeving beduidend groter is dan 1.96 x de standaardfout. De schoolomgeving heeft dus een significant effect op de gemiddelde Cito-eindtoetsresultaten van basisscholen. Daarom moet voor dit effect gecorrigeerd worden voordat de variabele onderwijskundig leiderschap wordt ingevoerd in de regressie-analyse. Nu kan de vraag beantwoord worden of het onderwijskundig leiderschap van schoolleiders in het basisonderwijs verband houdt met een stabiele schooloutput waarop de school ook daadwerkelijk invloed kan hebben.

Tabel 2: Verklaring voor verschillen in de output van basisscholen

	te verklaren tussenscholenvariantie	te verklaren binnenscholenvariantie
0-model	61.3%	38.7%
+ schoolomgeving	56.2%	38.3%
+ onderwijskundig leiderschap	53.8%	38.3%

	regressiecoëfficiënt	standaardfout
intercept	554.8	2.17
B _{schoolomgeving}	-19.88	1.81
B _{onderwijskundig leiderschap}	3.03	1.30

In tabel 2 kan geconstateerd worden, dat de variabele onderwijskundig leiderschap inderdaad een weliswaar klein, maar positief significant effect heeft op de voor schoolomgeving gecorrigeerde stabiele output van basisscholen. Daarmee is ook dit aspect van de validiteit van het instrument aangetoond.

RESULTATEN

In tabel 3 wordt een overzicht geboden van de belangrijkste eigenschappen van de frequentieverdelingen die in 1989 en in 1993 gevonden werden.

Tabel 3: *Onderwijskundig leiderschap in 1989 en in 1993*

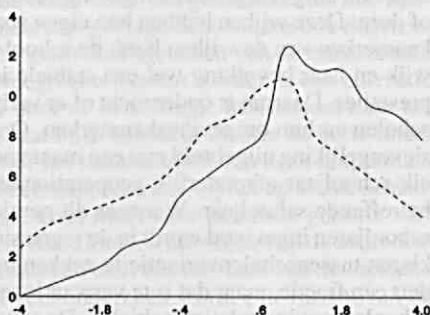
eigenschappen van de schaal:	1989	1993
gemiddelde (μ)	.34	1.62
standaarddeviatie (σ)	1.58	1.57

In het inspectie-onderzoek uit 1993 bleek dat de gemiddelde schoolleider een logitscore van 1.62 had. Hij voldeed daarmee aan bijna 12 van de 15 items voor onderwijskundig leiderschap. Daarmee voldeed de gemiddelde schoolleider overigens ruimschoots aan de door de inspectie gestelde norm. In 1989 bleek de gemiddelde schoolleider een logitscore van .34 te hebben. Op dat moment werd dus door de gemiddelde schoolleider aan 8 van deze 15 items voldaan (cf. Van de Grift & Akkermans, 1991). Deze verschillen komen tot uitdrukking in figuur 1. Op basis hiervan kan vastgesteld worden dat het onderwijskundig leiderschap, zoals de onderwijsgeevenden dat van hun schoolleider ervaren, de afgelopen jaren sterk tot ontwikkeling gekomen is.

Het bijzondere van een Raschschaal is dat niet alleen aangegeven kan worden aan hoeveel items een individu voldoet. Er kan binnen een zeker betrouwbaarheidsinterval ook aangegeven worden aan welke items dat individu voldoet. In 1989 werd het onderwijskundig leiderschap van de gemiddelde schoolleider getypeerd door het volgende gedrag: De onderwijsgeevenden kunnen bij de schoolleider terecht als ze problemen van leerlingen willen bespreken, de schoolleider doet actief mee aan de keuze van nieuwe onderwijsmethoden en geeft onderwijsgeevenden het gevoel dat hun werk belangrijk is. De schoolleider zorgt ervoor dat teamvergaderingen waardevolle onderwijskundige bijeenkomsten zijn en geeft hoge prioriteit aan onderwijsmethoden, die tot een verbetering van de leerresultaten van alle leerlingen kunnen leiden. De gemiddelde schoolleider zorgde in 1989 verder voor een op leren gericht schoolklimaat, hij toonde waardering voor onderwijsgeevenden die hun werk goed doen en hij besprak tijdig de problemen die kunnen optreden bij de uitvoering van het werk op school.

Tussen 1989 en 1993 is de gemiddelde schoolleider zijn gedragsrepertoire gaan uitbreiden met het informeren van onderwijsgeevenden over nieuwe onderwijsmethoden en instructiematerialen. Verder is de schoolleider belangstelling gaan tonen voor wat er in de klassen van de onderwijsgeevenden gebeurt, hij is het regelmatig peilen van leerlingprestaties gaan benadrukken en de schoolleider is onderwijsgeevenden gaan aansporen tot aandacht voor het verhogen van de leerprestaties.

Het volgende gedrag wordt in 1993 door de gemiddelde schoolleider nog niet vertoond: De onderwijsgeevenden in contact brengen met nieuwe ideeën en benaderingen die de leerlingprestaties kunnen verbeteren en adviseren over de manier waarop de prestaties van zwakke leerlingen kunnen worden verbeterd. Ook begeleidt en ondersteunt de gemiddelde schoolleider de onderwijsgeevenden nog niet bij het verbeteren van de leerprestaties.



Figuur 1: *Onderwijskundig leiderschap*
1989 - - -
1993 —

DISCUSSIE

De resultaten overziend, behoeft een drietal zaken nadere verklaring.

1. Zijn de bevindingen van een toename van het onderwijskundig leiderschap van Nederlandse schoolleiders in overeenstemming met ander onderzoek?
2. Als dat zo is: Hoe komt het dan dat het onderwijskundig leiderschap dat onderwijsgeevenden van hun schoolleider ervaren zo sterk is toegenomen?
3. En: Hoe komt het dat er in 1989 geen verband was tussen het onderwijskundig leiderschap van schoolleiders en de prestaties van de leerlingen, terwijl er in 1993 wel een (zwak) verband is?

We beginnen met een poging tot verklaring van het eerstgenoemde probleem. De observatie dat de onderwijsgeevenden tussen 1989 en 1993 hun schoolleider in toenemende mate als onderwijskundig leider zijn gaan ervaren, blijkt in overeenstemming te zijn met ander onderzoek dat is gericht op de zelfwaarneming door schoolleiders van hun begeleidingsvaardigheid. Uit een reeks van vier onderzoeken blijkt dat de begeleidingsvaardigheden van de gemiddelde schoolleider in het basisonderwijs in de loop van de tijd zijn toegenomen. We geven hier een korte samenvatting van deze onderzoeksresultaten. Een van de eerste³ Nederlandse instrumenten om het begeleidingswerk van schoolleiders te kunnen onderzoeken, is de schoolleiderinterventieschaal (Van de Grift, 1987). De schoolleiderinterventieschaal is een Raschschaal waarin 16 interventies als items opgenomen zijn. Het meetdoel van de schaal bestaat in het vaststellen van de begeleidingsvaardigheid van schoolleiders. Schoolleiders kunnen zelf in het instrument aangeven welke interventies zij ondernemen om een vernieuwingsproces op hun school te initiëren en te begeleiden. Dit instrument is tussen 1985 en 1996 vier keer afgenomen bij grote groepen schoolleiders. Dat biedt de mogelijkheid om een inzicht te verkrijgen in de ontwikkelingen op lange termijn.

In 1985 is de schoolleiderinterventielijst voor het eerst afgenomen. Aan deze afname hebben 108 schoolleiders uit het basisonderwijs meegewerkt. Deze schoolleiders werkten op hun scholen op dat moment aan uiteenlopende vernieuwingen. De gemiddelde schoolleider bleek toen 7 van de 16 interventies uit de schaal te realiseren (Van de Grift, 1987).

In 1988 is de schoolleiderinterventieschaal door Brandsma & Knuver (1989) bij 227 schoolleiders afgenomen als onderdeel van een onderzoek naar de effectiviteit van basisscholen. De gemiddelde schoolleider realiseerde toen 13 van de 16 interventies.

In 1990 onderzochten Stokking & Eberg (1991) in hun onderzoek naar regionale verschillen in het basisonderwijs met de schoolleiderinterventieschaal bij 245 scholen de begeleidingsvaardigheid van de schoolleiders. De gemiddelde schoolleider bleek 11 van de 16 interventies te realiseren bij vernieuwingen die gericht waren op individualisering, differentiatie en zorgverbreding.

Vijf jaar na het onderzoek van Brandsma en Knuver onderzocht Doolaard (1996) bij dezelfde schoolleiders uit het onderzoek van Brandsma & Knuver opnieuw de begeleidingsvaardigheid van deze schoolleiders. De gemiddelde schoolleider bleek nu aan 14 van de 16 interventies te voldoen.

Daarmee komen we aan het tweede verklaringsprobleem: Wij vermoeden dat de onderwijsgeevenden hun schoolleider in sterkere mate als onderwijskundig leider zijn gaan ervaren onder invloed van een aantal ontwikkelingen. Een eerste verklaring vormt de druk die op de schoolleiders is gaan ontstaan om leiding te gaan geven. Het oude model van het schoolhoofd met een aantal administratieve en beheersmatige taken voldeed al jaren niet meer. Dit model van de schoolleider dateert nog uit de jaren twintig en is zwaar verouderd. In artikel 27 van de lager-onderwijswet uit 1920 staat over de schoolleider het volgende:

'Aan het hoofd van elke school is een onderwijzer geplaatst, die den leeftijd van vijf en twintig jaren moet volbracht hebben en ten minste een diensttijd van drie jaren moet hebben vervuld als vast of tijdelijk onderwijzer aan eene lagere school of waarnemer volgens artikel 41'.

Deze onderwijzer die aan het hoofd van een school is geplaatst, heet in de jaren vijftig en zestig nog steeds 'hoofd ener school'. Hij had op de kweekschool de hoofdakte gehaald en hij was de onderwijzer van de hoogste klas, die 'ook nog' een aantal beheersmatige en administratieve taken moest uitvoeren.

De druk op schoolleiders om expliciet leiding te gaan geven, begon met het ontstaan van de basisschool in 1985. Het ontstaan van de basisschool uit de lagere school en de kleuterschool dwong de schoolleiders afstand te nemen van het oude 'hoofd der school'-model. Zij moesten of zij wilden of niet gaan leidinggeven aan de fusie van lagere school en kleuterschool. Verder moesten zij onder invloed van de nieuwe wet op het basisonderwijs met hun team een visie ontwikkelen en uitwerken in uitgangspunten en doelstellingen. Deze uitgangspunten, doelstellingen en de wijze waarop deze doelstellingen bereikt zouden moeten gaan worden, moesten op hun beurt weer in een schoolwerkplan worden vastgelegd. Over de hoeveelheid papier die hier het gevolg van was heeft vrijwel niemand nog een positief oordeel, maar het was wel het begin van leiding geven. Natuurlijk, de schoolleider was op dat moment nog steeds een *primus inter pares*. Niemand 'haalde het nog in zijn hoofd' om 'zijn kop boven het maaiveld uit te steken'. Maar het begin was geboren: Leiding geven begon te mogen. Al was het maar vanuit de rol van begeleider van de vernieuwingen. De schoolleider kon zich hierdoor gaan profileren als intern begeleider. Heel geleidelijk kreeg de schoolleider daardoor meer grip op het team: eerst alleen vanuit een initiërende rol, later ook vanuit een begeleidende rol. In de afgelopen jaren heeft dit proces een nieuwe wending gekregen. De vernieuwing van het basisonderwijs is steeds sterker in de ban geraakt van de effectiviteit van het onderwijs. Uit Amerikaanse en Britse onderzoeken was zeer nadrukkelijk naar voren gekomen dat een duidelijk en expliciet uitgeoefend leiderschap bevorderlijk was voor de resultaten van de leerlingen. Een aantal schoolleiders heeft hieruit wellicht de moed geput om 'de kop boven het maaiveld uit te steken'. Daardoor kunnen schoolleiders (meer) grip gekregen hebben op de prestaties van de leerlingen.

Een tweede met het voorafgaande samenhangende verklaring bestaat in het feit dat een groot aantal schoolleiders de afgelopen jaren is nageschoold. Onder invloed van de hierboven genoemde ontwikkelingen in het beleid omstond een algemeen gevoelde behoefte tot professionalisering van de schoolleiders op het gebied van onderwijskundig leiderschap. In de jaren zeventig werd met het ontstaan van de toen zogeheten pedagogische academie de specifieke schoolleidersopleiding, de zogenoemde derde leerkring, opgeheven. Er kwamen, aanvankelijk aarzelend, nieuwe impulsen op gang. Er ontstonden vakbladen, die speciaal op de schoolleider in het basisonderwijs gericht waren. De landelijke pedagogische centra gingen nascholingscursussen voor schoolleiders verzorgen. Voor deze nascholing zijn inmiddels specifieke instellingen van de grond gekomen. Van deze professionaliseringsactiviteiten kan een zeker effect vermoed worden op de begeleidingsvaardigheid van schoolleiders.

Last but not least is de afgelopen jaren een proces van fusie van basisscholen op gang gekomen. Dit fuseren vormt de derde verklaring; dit heeft ongetwijfeld een zekere 'survival of the fittest' onder de schoolleiders opgeleverd. Met dit proces van fuseren ging bovendien een schoolvergroting gepaard, waardoor nog weer een zwaarder appel op het leiderschap gedaan werd.

De laatste vraag die we in deze discussieparagraaf willen behandelen betreft de vraag waarom er in 1993 wel een verband was tussen onderwijskundig leiderschap en leerlingprestaties en een aantal jaren daarvoor niet. Allereerst moet dan opgemerkt worden, dat het slechts gaat om een zwak verband. Uiteraard moet men geen sterk verband verwachten: Op de eerste plaats valt er op schoolniveau, zoals ook uit dit onderzoek blijkt, slechts een gering percentage variantie in leerlingprestaties te verklaren. Op de tweede plaats spelen voor de op schoolniveau te verklaren variantie in leerlingprestaties ook andere schoolfactoren, dan het onderwijskundig leiderschap een belangrijke rol. We moeten hierbij denken aan de kwaliteit van het leerstofaanbod, de tijd en gelegenheid die de leerlingen krijgen om zich dit aanbod eigen te maken, de kwaliteit van het didactisch handelen van de onderwijsgevend en de mate waarin een

school in staat blijkt om adaptief onderwijs te leveren. Juist het besef dat het onderwijskundig leiderschap slechts kandidaat is voor de verklaring van een klein deel van de variatie in leerlingprestaties, maakt het van belang om de output van de school zo zuiver mogelijk te operationaliseren. In het onderzoek in 1989 was de stabiliteit van de schooloutput niet onder controle. Er werd toen een twee-niveau-analyse uitgevoerd waarin de individuele leerlingprestaties het eerste niveau vormden en de 'schoolgemiddelden' het tweede niveau. Wat hier niet onder controle is, is dat de schooloutput per jaar een zeker verschil kan vertonen. Scholen kunnen bij de ene leerlingengeneratie een output leveren waarin de gemiddelde leerling boven het verwachte niveau van de instroomkenmerken presteert, terwijl dat het volgende jaar bij de volgende leerlingengeneratie niet het geval is. Bij de analyses over de gegevens uit 1993 is voor deze instabiliteit gecontroleerd door gebruik te maken van de gegevens over een drietal jaren. Daardoor is voor de gegevens van 1993 een meer zuivere schooloutputindicator gebruikt. Wellicht dat dit een verklaring vormt voor het feit dat voor 1993 wel een verband tussen onderwijskundig leiderschap en schooloutput gevonden werd en voor de gegevens van 1989 niet.

Daarnaast is het ook van belang om een meer inhoudelijke overweging in de verklaring te betrekken. De gemiddelde schoolleider is in Nederland in 1993 gedrag gaan vertonen dat hij in 1989 nog niet vertoonde. Het gaat hierbij om de volgende zaken:

- * onderwijsgeveenden informeren over nieuwe onderwijsmethoden en instructiematerialen,
- * duidelijk belangstelling tonen voor wat er in de klassen van de onderwijsgeveenden gebeurt,
- * benadrukken van het regelmatig peilen van leerlingprestaties en
- * het aansporen van onderwijsgeveenden tot aandacht voor het verhogen van de leerprestaties.

Juist van dit gedrag mag men verwachten dat hiervan een positieve invloed uitgaat op de prestaties van de leerlingen.

LITERATUUR

- Batenburg, Th. A. van (1990). Variatie in schoolgemiddelden op de CITO-eindtoets basisonderwijs. In: *Tijdschrift voor Onderwijsresearch*, 15, 6, 362-369.
- Blok, H. & M.H. Eiting (1988). De grootte van schooleffecten: hoe verschillend presteren leerlingen van verschillende scholen. In: *Tijdschrift voor Onderwijsresearch*, 13, 1, 16-30.
- Blok, H. & J.B. Hoeksma (1993). De stabiliteit van het schooleffect in de tijd: een analyse op basis van vijf jaar Eindtoets Basisonderwijs van het Cito. In: *Tijdschrift voor Onderwijsresearch*, 18, 6, 331-342.
- Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. In: *Tijdschrift voor Onderwijsresearch*, 16, 4, 208-216.
- Brandsma, H.P. & J.W.M. Knuver (1989). *Basisschoolkenmerken als determinanten van het functioneren van leerlingen*. Groningen: RION.
- Commissie Evaluatie Basisonderwijs (1994a). *Inhoud en opbrengsten van het basisonderwijs*. De Meern: Inspectie van het Onderwijs.
- Commissie Evaluatie Basisonderwijs (1994b). *Zicht op kwaliteit*. De Meern: Inspectie van het Onderwijs.
- Doolaard, S. (1996). Onderwijsvernieuwingen krijgen meer kans: Schoolleiders begeleiden beter. *Basis-schoolmanagement*, vol. 9, nr. 4.
- Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Hans Huber Verlag.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Charles Griffin & Company Ltd.
- Grift, W. van de & W. Akkermans (1991). Schoolleiders en leerlingprestaties in het basisonderwijs. In: *Tijdschrift voor Onderwijsresearch*, 16, 4, 244-257.
- Grift, W. van de (1994). *Technisch rapport van het onderzoek onder 386 basisscholen ten behoeve van de evaluatie van het basisonderwijs*. De Meern: Inspectie van het Onderwijs.
- Gustafsson, J.-E. (1977). *The Rasch model for dichotomous items: theory, applications and a computer program*. Göteborg: Institute of education, University of Göteborg.
- Gustafsson, J.-E. (1979a). *Testing and obtaining fit of data to the Rasch model*. Göteborg: Institute of education, University of Göteborg.
- Gustafsson, J.-E. (1979b). *PML, a computer program for conditional estimation and testing in the Rasch model for dichotomous items*. Göteborg: Institute of education, University of Göteborg.

- Knuver, J.W.M. (1993). *De relatie tussen klas- en schoolkenmerken en het affectief functioneren van leerlingen*. Groningen: RION.
- Nunnally, J.C. & I.H. Bernstein (1994). *Psychometric theory*. New York: McGraw-Hill.
- Prosser, R., J. Rasbash & H. Goldstein (1991). *ML3, software for three-level analysis*. London: Institute of Education, University of London.
- Stoel, W.G.R. (1993). *De taakinhoud, de taakomvang en de taakbelasting van schoolleiders in het basisonderwijs*. Enschede: Universiteit Twente, OCTO.
- Stokking, K.M. & J. Eberg (1991). *Regionale verschillen in het basisonderwijs*. Utrecht: ISOR.
- Verhelst, N.D. (1993). Itemresponstheorie. In: Eggen, T.J.H.M. & P.F. Sanders (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Wright, B.D. & M.H. Stone (1979). *Best test design, Rasch measurement*. Chicago: Mesa Press.

NOTEN

1. Let wel: Er wordt hier steeds uitgegaan van de 'gemiddelde leerlingprestaties' per school. De prestaties van de afzonderlijke leerlingen spelen als het gaat om de schooloutput geen rol. De binnenschoolse variantie in leerlingprestaties (en dat is ongeveer 80%) en de factoren, die hierop van invloed zijn (zoals het milieu, de intelligentie en de motivatie van de leerlingen), vallen buiten dit bestek. Uitsluitend de tussenschoolse variantie, de verschillen tussen de scholen en de verklaringen hiervoor, zijn aan de orde als het gaat om de output van scholen.
2. Afhankelijk van het milieu van herkomst krijgt elke leerling een gewicht, dat een rol speelt bij de financiering van het onderwijs. Leerlingen van wie ten minste één van de ouders niet in Nederland geboren is en een laag beroeps- of opleidingsniveau heeft, krijgen een gewicht van 1.9; leerlingen van ouders uit de reizende en trekkende bevolking krijgen een gewicht van 1.7; leerlingen van ouders van schippers krijgen een gewicht van 1.4; leerlingen van wie ten minste één van de ouders een laag opleidings- en beroepsniveau heeft krijgen een gewicht van 1.25; en de overige leerlingen krijgen het gewicht 1.0.
3. Vrijwel gelijktijdig verscheen de Likertschaal van Van den Berg, Harskamp & Wolfgram (1985).

Manuscript ontvangen: 18-12-'95

Definitief aanvaard: 01-07-'97

Physics Problem Solving and the Transition from General Secondary Education to Higher Education

Ruurd Taconis*, Monica Ferguson-Hessler

*Department of Physics Education Faculty of Applied Physics,
Eindhoven University of Technology, The Netherlands
& Graduate school of Teaching and Learning, University of Amsterdam,
The Netherlands*

Gerrit Verkerk

*Department of Physics Education Faculty of Applied Physics,
Eindhoven University of Technology, The Netherlands*

* Also affiliated to IVLOS, University Utrecht, The Netherlands

ABSTRACT

This study is an attempt to tackle the problem of the decreasing level of performance of students entering the technical and scientific sectors of higher education from general secondary education. Teachers in higher education have indicated a lack of problem-solving skills as a probable cause. A set of general and problem-type specific problem-solving activities in the domain of physics at secondary level is identified. Skills in carrying out these activities are found to be correlated to success in the first year of higher education. General secondary education, however, is shown to give little attention to these skills. As a result of this, many students may experience great difficulties in taking the step from secondary to higher education. The results of the study provide a basis for tackling these problems. In this way, the results may support the current restructuring of Dutch general secondary education. Although the study was designed for problem solving in the domain of physics in Dutch education, a similar design could also be applied to other domains and educational systems.

This research was commissioned by the Institute of Educational Research in the Netherlands (SVO) at the request of the Dutch Ministry of Education and Science.

INTRODUCTION

In many countries, problems arise in the transition from secondary to higher education. This is also the case in the Netherlands. Appendix A provides the non-Dutch reader with the necessary information on the Dutch educational system. This 'transition problem' was one of the motives for the restructuring of Dutch general secondary education currently taking place (Ministerie van Onderwijs en Wetenschappen, 1991; 1992; Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs, 1994).

The 'transition problem' is particularly serious in the technical and scientific sector of both higher vocational education and university. In these sectors, a majority of the students experiences difficulty in finishing the propaedeutic program within one year, as is required by the curriculum. The causes of this transition problem have not yet been fully identified. Teaching

staffs have indicated a lack of problem-solving skills as a probable cause (Van Dyck, 1990).

This study concentrates on these skills and on the conditions for their acquisition. Many other factors influence study performance at the beginning of higher education, for instance motivation, effort, intelligence, and study techniques. These important factors are not taken into account in this study, as we concentrate on the mastery of problem-solving activities and its relation to domain knowledge.

AIM

Our study aims at the identification of problem-solving skills relevant to success in higher technical and scientific education, but getting little attention in physics as taught in general secondary education. The identification and detailed description of these skills provides a concrete basis for measures that could facilitate the transition from secondary to higher education. Examples are (a) the adaptation of curricula in higher education and for physics in general secondary education, (b) improvements of instructional methods, and (c) discussions on the policy concerning the admission of students to higher education.

THEORETICAL BACKGROUND

Factors influencing success in higher education

Various factors influencing success in higher education have been identified (Pascarella, 1985). Among these are personal aptitudes (for instance intelligence), social skills, affective factors, and cognitive abilities. Some of these desired dispositions are part of the personal capacities of a student, whereas others can be acquired during education. These will be called 'qualifications'. We define *qualifications for study in higher education* in the following way: "*knowledge, skills and attitudes that are needed for the accomplishment of study tasks in higher education, and that have been acquired, or could be acquired, in secondary education*". Study tasks in scientific and technical disciplines in higher education involve problem solving. Some of the skills involved in problem solving may belong to the curriculum of secondary education; that is to say they are qualifications. The question remains which problem-solving skills are qualifications and which are not.

PROBLEM SOLVING

Study tasks in higher education often contain problems that have to be solved by the application of knowledge acquired previously. We speak of a problem if a subject is presented with an initial situation and a goal that has to be reached, but is not capable of reaching the goal directly, that is to say without carrying out some form of manipulation of the information given, using his or her domain knowledge (De Jong, 1986). What constitutes a problem depends on the particular domain and on the knowledge of the subject. As a consequence a 'situation - goal combination' that forms a problem for one student, may not be a problem for another student.

Problem-solving activities and skills

We have adopted the view of cognitive psychology that problem solving is a form of information processing (Newell & Simon, 1972). This means that problem solving is described as involving the execution of several cognitive and metacognitive activities. Cognitive activities operate on external information and on knowledge in the memory of the problem solver. Metacognitive activities are defined as activities that regulate these cognitive activities.

Cognitive problem-solving activities are characterized by the kind of information and knowledge they process and produce. For example, 'analyzing' operates on the information provided in the problem statement and produces a mental problem-representation, whereas 'planning' operates on this representation and produces a plan for the solution.

We use the term *problem-solving skill* to denote the personal mastery of a student of a cognitive activity - or *problem-solving activity* - which is required in order to pass from a given situation to the goal specified. In this study, we aim at defining and describing problem-solving activities that are fundamental or elementary in a psychological sense. Hence, problems require a combination of cognitive and metacognitive skills.

The measurement of problem-solving skills requires the identification and description of the problem-solving activities. Problem-solving activities, and the associated skills, differ in their range of applicability and their domain of relevance. We distinguish two main types of problem-solving activities which are situated at the ends of a continuum: general activities, and problem-type specific activities. *General problem-solving activities (and skills)* are relevant to the solution of almost all types of problems in a wide variety of domains. Examples are 'analyzing' and 'planning'. *Problem-type specific problem-solving activities (and skills)* are useful and applicable only to a limited class of problems (e.g. problems concerning statics; a sub-domain of physics). These activities require specific knowledge of a certain law or principle, its application, and the sort of situations where it can be applied in order to reach the goal. An example is 'choosing a suitable integration path for the calculation of a potential difference in an electrostatic field'. Most problem-solving activities are situated in the continuum between these two extreme types, for instance: 'structuring of information given by means of a diagram' or 'systematic checking on the forces acting on a mechanical system'.

A general problem-solving activity, such as analyzing, can encompass several sub-activities that are still general in the sense of being applicable in a variety of domains. Analyzing a particular problem usually comprises a conglomerate of activities such as 'collecting information', 'interpreting the situation and the goal', and 'structuring the information'. On the other hand, in the application to a particular problem, for example a mechanics problem, analyzing and its sub-activities may take on a problem-type specific form. 'Structuring the information' then comprises the problem-type specific activity 'constructing a free-body diagram'.

It has been the subject of a fundamental discussion by researchers (Alexander & Judy, 1988) whether subjects use general or problem-type specific skills in solving problems. Moreover, experts and novices appear to differ in the degree to which they apply domain-knowledge dependent problem-solving activities. Novices master not as many problem-type specific skills. In practice, it may be difficult to tell whether a particular subject uses general skills within the context of a specific problem, or whether he or she uses situation-specific skills. Leaving alone this fundamental discussion, we will describe skills and activities on both the general and the problem-type specific level.

In this study, which is devoted to problem-solving activities in physics, the activities examined are defined on the basis of work by De Jong (1986), Ferguson-Hessler (1989), and Mettes and Pilot (1980). Taking this work as a starting point, five general problem-solving activities and skills are distinguished: 'analyzing', 'planning', 'elaborating', 'calculating', and 'checking'. These are shown in the first column of Figure 1. The second column of Figure 1 shows a selection of the sub-activities (and skills) used in our research. The third column shows examples of problem-type specific activities (or skills).

Some of the activities in Figure 1 need further explanation. Planning, in this domain, usually focuses on the identification of a so-called *central relation* (abbreviation: CR), that is to say a law or principle of physics, that can play a central role in the solution of the particular problem. The selection of such a central relation may be all the planning necessary to solve

Figure 1: Overview of the cognitive problem-solving activities considered in this article.

General problem solving activities	General sub-activities	Problem-type specific (examples)
Analyzing	Collecting information	* reading a diagram of forces * reading an elec. circuit
	Sketching the situation	
	Analyzing the situation	* analysis of forces
	Interpreting implicit information	
	Analyzing the goal	
	Elaborating one's own knowledge	
	Drawing a diagram	* of forces * of elec. circuit * of elec./magn. field
Planning	Selecting one or more Central Relations (CR)	
	Evaluating the applicability of the CR's	
	Selecting a standard solution method	* decomposition of vectors into horizontal and vertical components to analyse the motion of a thrown object
	Identifying sub-problems	* identifying the sub-problem of - determining the charge density - determining the moment of inertia
	Constructing a solution route	* stating $F_{\text{tot}} = 0$ and $M_{\text{tot}} = 0$ as a mathematical set of equations for static problems
Elaborating	Making strategical choices about the application of the formalism	* Choose: a system, the coordinate-axes, the centre of rotation, a contour for integration, or an earthpoint (ground) in an electrical circuit
	Partitioning the problem into analogous parts	
	Transforming the problem	
	Eliminating answers (Multiple-choice questions)	
	Interpreting information (on basis of the solution plan)	
Calculating		* algebraic operations * calculations with numbers * vector operations * solving a set of linear equations
Checking	Checking (partial) results	* by recalculating * by formal arguments * by reality

a particular problem. The solution of other problems, however, requires several central relations to be combined or to be used in sequence, that is to say the identification of sub-problems.

Planning may also involve the recognition of the applicability of a standard solution method, as in the Systematic Problem Approach (SPA) introduced by Mettes and Pilot (1980). A standard solution method is defined as a problem-solving algorithm or heuristic which has been taught explicitly.

Elaboration may be needed to decide how to proceed. In physics, this typically involves '*making strategical choices about the application of the formalism*'. Such choices are arbitrary from the perspective of the scientific formalism. From a strategic point of view, however, there are often only a few options that lead to practical calculations. In mechanics, for example, one may have to choose a coordinate-system strategically in order to find components for the forces that allow manageable calculations.

The role of the knowledge base in problem solving

Domain knowledge plays a critical role in problem solving in a semantically rich domain such as physics (Chi, Feltovich & Glaser, 1981; Alexander & Judy, 1988; Perkins & Salomon, 1989). De Jong and Ferguson-Hessler (1993) distinguish four types of knowledge necessary for effective problem solving in physics. These types are based on the function of the knowledge within the problem-solving process. Of these types, three are clearly domain-specific and the fourth has some general components as well. *Declarative knowledge* consists of facts and principles of the specific domain. *Procedural knowledge* comprises the actions that can be carried out, and how this is done. *Situational knowledge* is defined as knowledge allowing the recognition of the crucial features of the problem which govern the applicability of the declarative and procedural knowledge (Taconis, 1995). *Strategic knowledge* is knowledge of problem-solving strategies, either general or domain specific. The total of the knowledge that is relevant for problem solving in a certain domain is denoted as the *knowledge base*. A knowledge base adequate for problem solving contains elements of all four types defined above (Ferguson-Hessler, 1989).

Problem-solving activities differ in the kind of knowledge required for successful performance. Problem-type specific activities require domain knowledge specific to the problem type (i.e. knowledge about the solution methods for problems about objects thrown horizontally). On the other hand, both general and problem-type specific activities may involve general scientific concepts. Such concepts are defined to be applicable in a wide range of situations, for instance the physical concepts of 'equilibrium' and 'system'. These concepts are general from a formal point of view, but from a psychological point of view this is not necessarily true for the concept a particular subject holds. A subject may be able to use such concepts in a limited class of problems, but not recognize their relevance to other types of problems due to a lack of transfer of his knowledge.

In addition to the content of the knowledge base, its structure is crucial for effective problem solving. Chi, Feltovich and Glaser (1981) showed that the expert's knowledge-base is built up of so-called '*problem schemata*'. A problem schema represents knowledge applicable to a basic type of problem together with its typical problem-solving approach. Knowledge of all the types discussed above are integrated within it. Ferguson-Hessler and De Jong (1987; de Jong & Ferguson-Hessler, 1986) have described such problem schemata for the field of electricity and magnetism.

The measurement of problem-solving skills

As concluded above, the measurement of problem-solving skills requires the identification and description of problem-solving activities. However, the specification of the activities or performed to solve a certain problem enables us to determine whether a student did carry out this activity or not, but does not offer a complete measure of his or her skill. As a consequence

of the dependence of problem-solving skills on the knowledge base, two other aspects must be taken into account: (a) the presence in memory of the knowledge required, and (b) the accessibility of this knowledge. The last aspect depends on the circumstances under which the activity is performed. In order to quantify these aspects, we have defined the 'situated level of difficulty' of an activity. Key factors determining this measure are: the complexity and the unfamiliarity of both the situation and the activity. In addition, the 'level of performance' (e.g. swiftness, accurateness) can be defined. Only by combining measures of these different aspects of problem-solving activities it is possible to measure the problem-solving skills of a student.

DESIGN

In order to identify problem-solving skills relevant to success in higher education but getting little attention in secondary education, we have chosen a step by step approach and formulated four research questions, which are investigated in four partial studies. The studies are restricted to the topics 'mechanics' and 'electricity and magnetism' which are known to produce serious difficulties to first year students in higher scientific and technical education (Hulshof, 1990).

Research questions

1. Which problem-solving skills are *formally* required in the examination syllabi for physics in higher general secondary education (HAVO) and pre-university education (VWO)?
2. Which problem-solving skills are required to successfully complete *actual* problems of these final examinations?
3. Which problem-solving skills are required to successfully complete exams from the first year of higher vocational education (HBO) and university (WO)?
4. What is the relation between success in the first year of higher education (HBO and WO) and the level of problem-solving skills at the moment of entering higher education?

Overview of the studies

The first study concerns the analysis of the examination syllabi. Since skills depend on the knowledge base, both the required skills and the required knowledge are considered. The consistency of the educational objectives concerning skills with those concerning knowledge is evaluated.

In the second study the problem-solving skills actually tested in the final examination are identified by analyzing a sample of examination tasks from the national part of the final examination (see appendix A). The school-bound part may differ from school to school and no analysis was performed. We have no indications that the school-bound parts systematically require more problem-solving skills. The results of this study are compared with the results of the analysis of the examination syllabi since the examination tasks might represent tacit objectives concerning problem solving not described in the syllabi.

In the third study, typical problems which cause difficulties for students in first year examinations of various forms of higher education are analyzed. Both technical and other disciplines that demand physics in the first year (e.g. medical, agricultural) are included. The skills identified are compared with the skills demanded in the (national part of the) final examination of secondary education. This yields a list of problem-solving skills relevant to success in higher education, but getting little attention in actual general secondary education.

In the fourth study, the skills identified in study 3 are implemented in a problem-solving test, which essentially requires only the knowledge covered by general secondary education curricula. This test is administered to students starting their study in a discipline of the technical or scientific sector of higher education. The scores on the problem-solving test are

correlated with two indices of 'success' in higher education: (a) the mean grade for the pro-paedeutic examinations, and (b) the fraction of examinations passed.

Research techniques used in the studies 2 and 3

Before describing the four studies, we treat in some detail the method used to identify problem-solving skills required by examination tasks in study 2 and 3. This analysis comprises the identification of both the relevant problem-solving activities and the conditions under which the corresponding skills have to be demonstrated (Gardner, 1985; Venselaar, Van der Linden & Pilot, 1986). The identification of the relevant problem-solving activities requires the reconstruction of an adequate and complete problem-solving process in terms of elementary problem-solving activities. The identification of the conditions under which the skills have to be demonstrated requires the determination of the 'situated level of difficulty' for each of the activities demanded. Consequently our method of analysis consists of two phases. For each of these we have designed a separate procedure.

Phase 1: Reconstruction of the problem-solving process

As a theoretical basis for the reconstruction of the problem-solving process, we used a list of problem-solving activities for physics problem-solving based on teaching experiences and on experiences collected in earlier research on problem solving in physics (e.g. de Jong, 1986). Figure 1 gives a selection of the most important activities.

Since we analyze examination tasks, we focus on the problem-solving activities which are to be tested. Together, these activities form the '*Intended Solution Process*' (abbreviation: ISP) of the problem. The ISP is reconstructed for each problem studied, and is used as a description of ideal task performance. In general, an ISP has to meet three requirements. It has to be:

1. *successful* in that it produces a correct answer if executed correctly,
2. in *accordance* with the skills and knowledge which are intended to be tested by the problem,
3. *inclusive* in the sense that it essentially includes problem-solving activities in our catalogue (Figure 1) as necessary or useful.

Due to the first two requirements, the ISP will be uniquely defined for most examination problems. Moreover, the requirement of accordance prevents the inclusion within the ISP of skills that are in no way part of the curriculum. The requirement of inclusiveness excludes questions that are not a problem according to our definition, and for which no ISP can be found. An ISP constructed in agreement with the requirements specified above, can be said to represent the actual educational goals concerning problem-solving activities tested in examinations.

The construction of a valid ISP involves four steps:

1. collection of the problems, their standard solution, the answers, and information on the curriculum which preceded the examination,
2. solution of the problem and construction of a provisional ISP,
3. verification of the validity of the provisional ISP, and
4. adaption if necessary.

Step 1 can involve the selection of a subset of the problems, for example concentrating on problems typically causing difficulties for relatively many students.

Step 2 is performed on sub-tasks of the problems (i.e. question 1a, 1b, 1c, etc.) which are the units of analysis. In this two researchers cooperate. The researchers aim at an ISP in harmony with the general intention of the problem within the examination. This intention is derived from the additional material gathered in step 1. For identifying each of the problem-solving activities, a comprehensive guideline for coding has been developed (Taconis,

Stevens & Ferguson-Hessler, 1992). In identifying the problem-solving activities, two researchers cooperated and discussed until consensus was reached on the identification of the problem-solving skills. Each of the activities defined in the catalogue (see Figure 1), can show various degrees of relevance to the completion of the ISP. A particular activity can be either 'useless', 'useful but not essential', or 'essential' for the completion of the ISP. Both 'essential' and 'useful' activities are included in the ISP.

In step 3 it is checked whether the ISP is really in accordance with the intentions of the teacher by consulting him or her. Teachers confirmed the validity of the ISP constructed by the researchers in the vast majority of the cases. Step 3 leads to adaptation of the ISP if necessary until an ISP is found that represents the actual educational goals for problem-solving at the level examined.

Phase 2: Analysis of the ISP

In the second phase, the ISP for each of the problems is analyzed to measure the required skills. Two researchers perform this analysis together. Since the degree of relevance of the various problem-solving activities has already been determined, the analysis of the ISP concentrates on the 'situated level of difficulty' of the activities previously identified as 'useful' or 'essential'. The required level of performance (e.g. speed) is not taken into account in our analysis.

The 'situated level of difficulty' of a problem-solving activity for a particular problem is operationalized in terms of three aspects: (a) complexity of the activity, (b) lack of familiarity of the activity with respect to the preceding education and (c) support within the task (e.g. by previous sub-tasks of the same problem). In other words, we define three measures, *complexity*, *unfamiliarity* and *support*, which together give information of the 'situated level of difficulty' of the problem.

The measure 'complexity' is defined as an ordinal variable which can take either of five values ranging from 1 (simple) to 5 (extremely complex). Depending on the activity it refers to, it reflects: (a) the number of the assumptions involved in analyzing the problem, (b) the number of sub-problems to be recognized in planning the solution, or (c) the number of steps involved in calculating the solution of the problem.

The measure unfamiliarity is defined as an ordinal variable which can take either of three values ranging from 1 (familiar) to 3 (new). These values reflect whether the activity under consideration has been attended to or practised within the preceding education. Moreover, unfamiliarity depends on the degree to which the test problem resembles the exercises and examples used during the preceding education. A comprehensive guideline for coding (Taconis, Stevens & Ferguson-Hessler, 1992) has been developed containing detailed criteria for assigning the values for complexity, unfamiliarity, and support, for each of the problem-solving activities. Validity of the guideline was confirmed by a panel comprising a researcher on problem solving, an expert in physics education and experienced teachers from both secondary and higher education.

THE FOUR STUDIES

Study 1: Problem solving in the syllabi for Dutch general secondary education

Aim of this study is to determine the extent to which problem solving is explicitly implemented in the examination syllabi. Our description is limited to the key findings. A more detailed description can be found in Taconis, Stevens and Ferguson-Hessler (1992).

Data collection and method of analysis

Two physics examination syllabi were analyzed: The (old) CMLN syllabus (Het eindexamenbesluit MAVO-HAVO-VWO, 1984), and the (present) WEN syllabus (Eindexamenprogramma's

natuurkunde, 1989). Both describe examination objectives for general secondary education (HAVO) as well as for pre-university education (VWO). Each of the syllabi comprises two sections: knowledge description and skill description. The knowledge description sections were analyzed to identify the types of the required knowledge elements and the way these are related within the section. In this, the model of a problem schema as presented by Ferguson-Hessler (1989) was used. The result indicates to what degree problem schemata are described. It was combined with the results of the identification of demanded problem-solving skills from the skill description section.

Results and conclusions: Answer to research question 1

The skill description sections of both the CMLN and the WEN syllabi include some problem-solving skills (e.g. finding a central relation), but leave out other skills, which are just as relevant at the level of secondary education. For example: an essential skill like checking a (partial) solution, is not mentioned in the WEN syllabi.

Looking at the various types of knowledge elements described in the knowledge sections, we found that these sections all focus on declarative knowledge, and do not include relations between knowledge elements. Procedural knowledge forms a small fraction of the description, and strategic knowledge is hardly treated at all. In the WEN syllabi, situational knowledge is indicated, but this is mainly restricted to physical situations in which the students should be able to apply their knowledge: the so-called 'concepts of context' (Dutch: contextbegrippen) (Eindexamenprogramma's natuurkunde, 1989). Although the WEN syllabi do contain situational knowledge, its relevance for problem solving is not made explicit.

We conclude that problem solving does not play an important role in the syllabi. Problem-solving skills are only partially described. The knowledge description sections do not reflect the objective of the acquisition of a knowledge base adequate for problem solving (problem schemata). Cross-references between the skill description sections and the knowledge description sections are absent.

Study 2: Problem-solving skills in the final examinations of general secondary education

This study is aimed at an explicit description of the problem-solving skills required by or helpful to passing the final examinations of general secondary education. This can be regarded as a reconstruction of the educational goals concerning problem-solving skills which are actually tested in the final examinations.

Data collection and method of analysis

Examination tasks were collected from the national part of the examinations (1988, 1989 and 1990) for both HAVO and VWO. These all concern the CMLN syllabi.

Nearly all examination tasks required one or more of the problem-solving activities identified in Figure 1. For each of these the ISP, the Intended Solution Process (see the section on research techniques) was constructed. Experienced teachers were consulted in order to validate the ISP's constructed.

The examination problems were analyzed to determine which problem-solving activities were required for the solution, distinguishing general and problem-type specific activities. In each case it was decided whether the activity was useless, useful, or essential to the solution.

For HAVO, all problems on the topic 'electricity and magnetism' are of the same type: the 'network-type' which focuses on calculations of elements of an electrical circuit (Taconis, Stevens & Ferguson-Hessler, 1992). Since we have taken the position that generality of a problem-solving skill requires that it is successfully used in a *variety* of problems, the demands these problems pose to the students cannot be uniquely identified as either the general skills or problem-type specific skills. Therefore, only the more efficient problem-type specific

analysis was performed. For all problems analyzed, the 'situated level of difficulty' was determined as described under 'Phase 2' in the section on research techniques.

Results

Table 1 (left part) shows the results of the analysis of general problem-solving activities. The entries depict percentages of problems for which the skill could be evaluated ('valid percentages').

Table 1: General problem-solving activities in examination problems. The entries depict 'valid percentages'. An asterisk indicates a significance at a 10% level of the difference when comparing HAVO to HBO, or VWO to University. The abbreviation 'Mech.' indicates the topic 'Mechanics' while 'E&M' indicates the topic 'Electricity, Magnetism and Electrical circuits'.

		Final examination problems			Stumbling-block problems from propaedeutic examinations		
		HAVO		VWO	HBO	University	
		Mech. (n = 27)	Mech. (n = 23)	E&M (n = 18)	Mech. (n = 19)	Mech. (n = 32)	E&M (n = 16)
ANALYZING							
Total	useless	46%	32%	28%	16*	26%	24%
	useful	21	32	24	17*	19	25
	essential	33	36	48	67*	55	51
Collecting information	useless	-	-	-	-	-	-
	useful						
	essential						
Make sketch	useless	52	30	33	5*	15*	0
	useful	37	57	61	58*	41*	69
	essential	11	13	6	37*	44*	31
Analyzing the situation *	useless	0	0	6	0*	3	0
	useful	48	30	33	5*	16	12
	essential	52	70	61	95*	81	88
Analyzing the goal	useless	66	52	50	37*	41*	68
	useful	19	35	11	16*	9*	13
	essential	15	13	39	47*	50*	19
Elaborating on one's knowledge	useless	100	100	100	100		100
	useful	0	0	0	0	occa-	0
	essential	0	0	0	0	sional	0
Make diagram	useless	81	53	50	37*	7	19
	useful	4	30	17	5*	19	31
	essential	15	17	33	58*	44	50

Table 1 (continued)

		Final examination problems			Stumbling-block problems from propaedeutic examinations		
		HAVO	VWO		HBO	University	
		Mech. (n = 27)	Mech. (n = 23)	E&M (n = 18)	Mech. (n = 19)	Mech. (n = 32)	E&M (n = 16)
PLANNING							
Planning by Central Relation	useless	33	4	11	23	23	0
	useful	15	13	0	0	3	0
	essential	52	83	89	77	74	100
Evaluating the applicability of CR's	useless	89	74	94	65*	81	88
	useful	11	9	0	0*	3	6
	essential	0	17	6	35*	16	6
Selecting a standard solu- tion method	useless	100	100	100			100
	useful	0	0	0	often	rare	0
	essential	0	0	0			0
Identifying partial problems	useless	19	30	7	6	3	6
	useful	4	0	0	0	0	13
	essential	77	70	83	94	97	81
Constructing a solution route	useless	43	66	55	17*	41	56
	useful	23	4	0	18*	7	0
	essential	34	30	45	65*	52	44
ELABORATING							
Making strategical choices	useless	59	65	83	41	31*	31
	useful	0	9	6	0	10*	13
	essential	41	26	11	59	59*	56
Partitioning the problem	useless	-	-	100	-	-	68
	useful			0			13
	essential			0			19
Transforming the problem	useless	-	-	-	-	-	-
	useful						
	essential						
Eliminating answers	useless	100	100	100		100	100
	useful	0	0	0	occasional	0	0
	essential	0	0	0		0	0
Interpreting information	useless	22	44	22	18	32	6
	useful	0	4	11	6	3	13
	essential	78	52	67	76	65	81
CALCULATING							
Calculating	useless	12	9	33	0	6	0
	useful	7	0	0	12	3	0
	essential	81	91	67	88	91	100

For a fairly large fraction of the problems from general secondary education, most activities under 'analyzing', 'planning' and 'elaborating' are neither necessary nor useful. This is particularly true for higher general secondary education (HAVO).

Table 2 (left part) shows the distribution of the values found for the aspects of the 'situated level of difficulty' averaged over the problems and over the essential or useful problem-solving activities. The entries depict percentages of problems for which the skill could be evaluated ('valid percentages').

Table 2: The percentages of the aspects of the 'situated level of difficulty' averaged across the problems. The entries depict percentages of problems for which the skill could be assessed ('valid percentages'). An asterisk indicates a significance at a 10% level of the difference when comparing HAVO to HBO or VWO to University. The abbreviation 'Mech.' indicates the topic 'Mechanics' while 'E&M' indicates the topic 'Electricity, Magnetism and Electrical circuits'. For HAVO and HBO, the analysis of the topic 'Electricity, Magnetism and Electrical circuits' was only performed on the problem-type specific level. Such problems from higher general secondary education ($n = 10$) and for higher vocational education ($n = 10$) are not mentioned here.

		HAVO	VWO		HBO	WO	
		Mech. ($n = 27$)	Mech ($n = 23$)	E&M ($n = 18$)	Mech. ($n = 19$)	Mech. ($n = 32$)	E&M ($n = 16$)
Support	yes	10%	11%	8%	7%	13%	9%
	no	90	89	92	93	87	91
Unfamiliarity	standard	70	68	71	54	48*	53
	partly known	29	28	27	33	39*	41
	unknown	2	4	3	13	13*	6
Complexity	simple	34	22	17	7*	19*	15
	slightly c.	46	57	50	32*	28*	33
	complex	13	17	29	30*	20*	26
	very c.	7	4	3	18*	18*	21
	extremely c.	0	0	1	14*	15*	5

The problem-solving activities relevant to solving the problems as intended, are mostly 'standard' and hardly ever 'unknown from preceding education and exercises'. In most cases the activities are only 'slightly complex' but often 'simple' and rarely 'complex' (median: 'slightly complex'). Support provided within the problems themselves is present in only a few cases.

In addition to the analysis of the above CMLN examination, a similar analysis was performed on the 1995 WEN examination. Its aim was to compare the examination tasks corresponding to these syllabi.

The HAVO examination includes 10 problems on mechanics and 6 problems on electricity and magnetism. The VWO examination comprises 6 mechanics problems and 5 problems on electricity and magnetism. These numbers are comparable to the numbers found for the CMLN examinations (Table 1).

In comparing the results of the analysis with those of the CMLN exam-problems, the non-parametric Kolmogorov-Smirnov test (Hays, 1988) was used to test the identity of frequency

distributions. No indications for a significant ($p < .1$) difference with respect to the 'situated level of difficulty' were found. This is also the case for the vast majority of the various problem-solving skills.

The WEN examination tasks for HAVO significantly ($p < .1$) more often require the use of a central relation for planning and significantly ($p < .1$) less often require 'the identification of partial problems' and 'the making of strategic choices'. It is noted that strategic choices are also required significantly ($p < .1$) less often in the other WEN examination tasks.

Conclusions: Answer to research question 2

It is concluded that the tasks from the national part of the physics examination for the topics under investigation do not require comprehensive problem-solving skills. This is in accordance with the conclusion of study 1. The problems mainly require the identification of one single central relation that can be found without deep understanding of the problem by means of the knowledge taught. 'Calculating' is prominent in the solution of these problems. Hence the examination problems do not require the students to have their domain knowledge organized according to problem schemata.

Study 3: Problem-solving skills in the propaedeutic examinations in the technical and scientific sectors of higher education

This study aims at an explicit description of the problem-solving skills required or helpful in passing first year examinations from technical and scientific disciplines of higher vocational education and university.

Data collection and method of analysis

A large number of problems was collected from examinations of various institutes and disciplines of higher vocational education and universities. The institutes were selected at random and spread throughout the country. The disciplines were selected for the following reasons: (a) they require entering students to have taken physics in secondary education, (b) physics knowledge is elaborated during tertiary education, and (c) the first year curricula contain at least one of the topics under investigation. In addition, the disciplines were spread over the technical sector (9 studies from HBO, 7 studies from WO), the para-medical sector (3 studies from HBO, 1 study from WO), and the agricultural sector (1 study from HBO, 1 study from WO).

In order to identify the skills not mastered by the students, thus hindering them in passing these examinations, the analysis concentrates on so-called 'stumbling-block problems'. A stumbling-block problem is defined by three criteria: (a) less than 50 percent of the first year students are able to solve the problem adequately, (b) the problem is relevant to the further study of that discipline, and (c) the problem represents the core of a particular topic or discipline.

For higher vocational education, 19 stumbling-block problems on mechanics and 10 stumbling-block problems on electricity, magnetism and electrical circuits were identified. For university education, 32 stumbling-block problems on mechanics and 16 stumbling-block problems on electricity, magnetism and electrical circuits were identified.

The problems were analyzed in the same way as the problems from the secondary final examinations by the procedure described above. In order to allow the valid reconstruction of the ISP for each of the problems, additional information was gathered from the teachers. It concerned: (a) the curriculum, (b) the textbooks used, (c) teaching practice, and (d) standard solutions. The teachers were consulted in most cases to validate the ISP's and to check whether the ISP makes implicit educational goals explicit.

Results

General level

All stumbling-block problems required problem-solving skills and are 'problems' in the sense of our definition. Table 1 (right part) shows that 'calculating' is a central skill needed to solve

the stumbling-block problems. Stumbling-block problems often require a substantial analysis of the problem. The problems often involve a set of central relations rather than one single central relation. This requires additional planning of a solution route. Making strategical choices about the application of the formalism is required for a majority of the stumbling-block problems.

Table 2 (right part) shows that the relevant problem-solving activities are standard in about half of the cases and rarely unknown from preceding education and exercises. The activities are mostly only 'slightly complex' but rarely 'simple' and in more than 50% of the cases at least 'complex' (median: 'complex').

Problem-type specific level

At the problem-type specific level four problem types are identified that are typical for the stumbling-blocks in the first year of higher education: STROT- (STatics or ROTation) problems, Network-problems, Contour-problems, and Integration-problems (for a more detailed description see: Taconis, Stevens & Ferguson-Hessler 1992).

STROT-problems occur most frequently (approximately 40% of the stumbling-block problems). They dominate stumbling-block problems from higher vocational education but also occur frequently as stumbling-block problems from university. STROT-problems may concern either a dynamic or a static situation and may involve linear as well as rotational motion or degrees of freedom. Figure 2 presents a STROT-problem together with the associated problem-type specific activities.

Conclusions: Answer to research question 3

All stumbling-block problems typically require problem-solving skills. This supports our initial assumption that a lack of these skills is one of the causes of poor performance in higher education, in the technical and scientific sectors in particular, since stumbling-block problems represent examination problems particularly related to low success in higher education.

The vast majority of the stumbling-block problems belongs to four problem types. Hence the examination problems require the students to have their domain knowledge organized according to the corresponding problem schemata. Problems of these types occur more often in the curricula of some studies than in others, but the types of problems causing the first year students to stumble are the same in all cases.

Comparison of stumbling-block problems with problems from general secondary education

This result can be compared with the results of study 2 on the CMLN examination-tasks (national part). This comparison identifies skills that seem to need more attention in either secondary or in the first year of higher education in order to improve the students' results in the first year examinations. Again the Kolmogorov-Smirnov test (Hays, 1988) was used to test which differences are statistically significant. In the cases where the comparison turned out to produce significant differences ($p < .1$), the entries in Table 1 and Table 2 are marked with an asterisk.

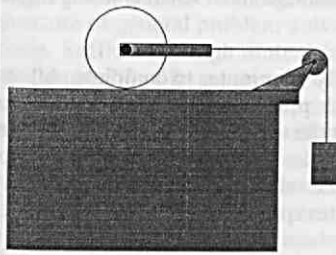
Table 1 shows that the skills 'analyzing' and 'planning' - constructing a solution route in particular - are required more often in HBO stumbling-block problems than in HAVO examinations. For university, Table 1 shows that 'analyzing' and 'making strategical choices about the application of the formalism', are required more often in stumbling-block problems from university than they are in VWO examinations.

An additional comparison concerning the 'situated level of difficulty' (Table 2) shows that stumbling-block problems on mechanics are significantly more complex than the problems from the final examinations. Those of university are also significantly less familiar with respect to the preceding education.

On a problem-type specific level, it is concluded that the general secondary education examination-tasks are not of the problem types to which the stumbling-block problems most

frequently belong. First year students appear to lack important problem-type specific skills, e.g. skills associated with the STROT-problem type like 'formulating a set of equations'.

Figure 2: An example of a problem of the STROT-problem type with the associated problem-solving activities. The degree of relevance of these skills may vary with the actual problem.

 <p>A hollow ball moves over a table without slipping.</p> <p><u>Calculate the acceleration.</u></p>	
Collecting relevant information	Reading a technical diagram
Analyzing	Drawing a diagram of forces in two dimensions Making geometrical calculations and interpretations
Planning	Formulating a set of two or more linear equations
Elaborating (Making strategical choices about the application of the formalism)	Choosing strategical sub-systems Choosing a strategical centre of rotation Choosing a strategical set of coordinate axes
Calculating	Solving a set of two or more linear equations in algebraic symbols Decomposition and addition of forces in two dimensions
Checking	Mostly recalculating the answer in a second (independent) way

Study 4: The relation between success in the first year of the technical sector of higher education, and students' problem-solving skills

We assumed that one of the causes of poor performance in the first year of higher education is the lack of problem-solving skills of the students. The results of study 3 suggest that this is due to the skills identified in the stumbling-block problems. The above comparison showed that there is little attention for these in secondary education, a finding supported in study 1. We hypothesize that mastery of these skills on entrance in higher education, correlates positively with success in the first year. In the fourth study, this hypothesis is tested empirically.

Design

In order to test our hypothesis, a problem-solving test was developed, which could be administered to students at the start of higher education. This test measures the skills which were found to be relevant to the solution of the stumbling-block problems, but is based on knowledge acquired in secondary education. The results of the test were related to various measures of success in the first year of higher education. It is well known that students' grades in the final physics examination of secondary education are strongly correlated to success in the first year (Pascarella, 1985; Nijgh, 1987; Janssen & De Neve, 1988). In study 4 we searched for additional relations between the problem-solving skills students have when entering higher education and their success in the first year.

The test used consists of 4 problems and takes approximately 90 minutes to complete. All test problems belong to the same types as the stumbling-block problems, and they require approximately the same skills. An essential difference is that the test requires knowledge from secondary education only. This could be achieved, for example, by translating a STROT-problem into a problem concerning only linear motion, a topic which is covered in secondary education. Appendix B shows this example. However, the test problem still requires many of the skills typically used in solving STROT-problems.

The problem-solving test was administered to 191 students in one of the first weeks of their first academic year. These students were from three universities and two institutes for higher vocational education. The subjects' grades from the secondary education examination, and grades achieved in the first half year of higher education were collected.

Success in the first year of higher education was operationalized in two types of indices: the fraction of all examinations passed, called 'pass', and the average grade on the attempted examinations, called 'grade'. In order to eliminate the bias due to differences in average grades between groups of students from different institutes, and between different examinations from the same institute, the measure 'grade' was calculated on the basis of z-scores for each examination.

Method of analysis of the test-results

The analysis of the test results concerns the measurement of the skills demonstrated by the subjects in written solutions to test problems (Gardner, 1985; Mettes & Gerritsma, 1986). For this we used the following method.

First the ISP (the Intended Solution Process) for the test-problem is constructed. The ISP is then used as a mask in scoring the occurrence in the solution of each of the problem-solving activities defined (Venselaar, Van der Linden & Pilot, 1986). This implies that only those skills are measured that concern the activities described in the ISP. The results of these measurements are called 'ISP-scores'. A set of 4 ISP-scores is used for each of the problems. These ISP-scores reflect the problem-solving skills: 'analyzing', 'planning', 'elaborating' ('making strategical choices about the application of the formalism' in particular), and 'calculating'.

The ISP-scores are defined to take one of three ordinal values: 'faulty and/or inadequate', 'partly incorrect or partly adequate' and 'correct and adequate'. Whenever an activity that is essential to the solution of the problem is not performed, 'faulty' is assigned. Again a comprehensive guideline for coding was used (Taconis, Stevens & Ferguson-Hessler, 1992). An example of the guideline as applied to a specific problem is presented in Appendix B. The analysis and scoring was performed by two researchers, working independently. The inter-rater agreement was estimated by means of Cohen's kappa. A value of 0.7 was found which is acceptable according to Heuvelmans & Sanders (1993). However, our categories are actually ordinal rather than nominal - the situation kappa is designed for. Cases of disagreement virtually always involved neighbouring classes minimizing the magnitude of the disagreement between the judges. Hence, kappa is a pessimistic estimate of quality of the ordinal me

asurement which can therefore be considered satisfactory (Taconis, Stevens & Ferguson-Hessler, 1992).

In principal, the ISP-scores per problem for the various skills may hold contributions from general skills as well as from problem-type specific skills. However, ISP-scores per problem are taken primarily as indices of problem-type specific skills in order to make unambiguous decisions.

In order to arrive at variables reflecting general skills, averages have been taken across all test problems for each of the four ISP-scores. Provided that the scores for the various problems show sufficiently high mutual correlation coefficients, these averages are a reliable measure of general problem-solving skills. This procedure is analogous to the formation of a scale. Sufficiently high mutual correlation coefficients result in a corresponding high value of Cronbach's α , a measure of the homogeneity of the scale. When α is at least .7, the mean score of the items comprising the scale may be taken to represent a common concept; a general problem-solving skill in our case.

The analysis is made more complex by the fact that the lack of a particular skill may obstruct the problem-solving process. This prevents the subject from getting to the point where it is necessary to demonstrate a certain other skill. Therefore, this last skill cannot be measured. For example: A missing or inadequate analysis can block the problem-solving process and thus hinder the measurement of e.g. skills in solving a set of equations. This produces missing values for the problem-solving skill that requires the (previous) execution of other problem-solving activities.

An index of the 'total problem-solving skill' was constructed by taking the average of the ISP-scores per problem for all test-problems and problem-solving skills.

Results

As expected (Nijgh, 1987; Janssen & De Neve, 1988), strongly significant ($p < .01$) Pearson correlation coefficients ($r = .46$ up to $r = .64$) were found between the indices for success in the first year of higher education and the physics grade for the final examination of secondary education. Pearson correlation coefficients were also calculated between the measures for general problem-solving skills and both indices for success in higher education. The result is shown in Table 3.

A series of analyses of covariance (Hays, 1988) was carried out with the indices for success in higher education as dependent variables and the secondary education physics grade as a covariate. The various general problem-solving skills were treated as factors. The effect of the covariate was calculated before determining the effect of the factor. Values of F ranging from $F = 3.4$ up to $F = 4.0$ were found for 'grade' as an index of success in higher education, and the 'total problem-solving skill' as a factor. These values are all significant at a level of $p < .07$. Higher values of F were found for some of the general problem-solving skills: 'analyzing' (HBO), and 'planning' (HBO and university) and 'making strategic choices' (HBO). For these skills, values of F up to 7 ($p < .05$) were found.

The implication of this result is that a significant positive correlation between the general problem-solving skills and success in higher education, tends to exist also within groups of students that achieved the same grade for their physics final examination in secondary education. Using the method of 'splitting the table' (in Dutch: 'tabelspitsing'; Swanborn, 1981 p. 326) this can be shown explicitly (Taconis, Stevens & Ferguson-Hessler, 1992). Though a fairly high correlation exists between the grade achieved for the physics final examination and success in higher education, the problem-solving skills as measured in the test are found likely to add an additional factor that is associated with success.

To find out which problem-type specific skills possibly contribute to this result, the various scores per test-problem were correlated with the indices of success in higher education. This

Table 3: The correlations of success in higher technical education and the scores for the various general problem-solving skills. The values are all significant at a level of $p < .1$. One asterisk indicates that $p < .05$, two asterisks indicate that $p < .01$.

	HAVO/HBO		VWO/UNIVERSITY	
	mean grade achieved ($n = 91$)	fraction of examinations passed ($n = 91$)	mean grade achieved ($n = 91$)	fraction of examinations passed ($n = 100$)
Analyzing	.33*	.56*	.41**	.37**
Planning	.32*	.19*	.37**	.37**
Elaborating (e.g. making strategical choices)	.36*	.29*	ns	ns
Calculating	.19	ns	.40*	.42*
Total	.47*	.30*	.46**	.43**

yielded particularly high values for Spearman's rank correlation coefficients for skills associated with the test-problems derived from the STROT-problem type. For students from higher vocational education, a significant ($p < .01$) value of $r_s = .36$ for Spearman's rank correlation coefficient is found for 'formulating a set of equations'. For the skill 'strategically choosing sub-systems' we find $r_s = .42$ ($p < .01$). For university students additional significant ($p < .01$) values were found for 'algebraic calculations using mathematical symbols' ($r_s = .39$).

Conclusions: Answer to research question 4

We have shown that first year students who do better on the entrance problem-solving test are likely to do better in the first examinations in scientific and technical disciplines. This also is the case within groups of students that achieved the same physics grade for the secondary education final examination. This supports our hypothesis and makes it likely that attention to these skills in secondary education would improve results in the first year¹.

'Analyzing' and 'planning' and the problem-type specific skills associated with the STatics-or-ROTation-problem type emerge as relevant. Algebraic skills appear to be relevant for university students in particular. Other problem-solving skills appear to be less relevant in our study, but it cannot be concluded from our data that these are irrelevant.

OVERALL CONCLUSIONS

We have developed a method for the identification and description of problem-solving skills which form relevant prerequisites for success in higher education and get little attention in secondary education. We have described these skills explicitly.

A possible interpretation of our results is that a causal impact exists of the presence of particular problem-solving skills at the start of higher education, on the achievement of students in the technical sectors of higher vocational education and university. The particular skills indicated as relevant are: the general problem-solving skills (analyzing, planning, making strategical choices, doing algebraic calculations) and problem-type specific problem-solving skills (e.g. choosing a free body for force calculations).

We conclude that:

1. Secondary education does not systematically attend to problem-solving skills (see study 1 and 2). Neither the examination syllabi, nor the problems in the national part of the final examinations require problem-solving skills.
- 2a. Disciplines of higher education requiring physics as a prerequisite, demand problem-solving skills in their propaedeutic examinations (study 3).
- 2b. Many students lack these skills, even after they have taken courses in higher education (study 3).
- 2c. Most stumbling-block problems belong to either of four problem types that occur in a wide variety of studies in higher education. This stresses the relevance of problem-type specific skills and the need for domain knowledge adequate for these skill i.e. problem schemata associated to these problem types.
3. The mastery of the skills indicated on entry in higher education correlates with success in technical and scientific sectors of higher education. This tends to be the case even within groups of students that achieved the same physics grade for the secondary education final examination (study 4). A causal relationship can be a possible interpretation of this.

Discussion and relevance to education

Our results on the transition problem are relevant for both secondary and higher education.

According to the Ministry of Education, a lack of communication contributes to problems in the transition from secondary to higher education. Therefore the formation of 'pedagogical communication networks' (e.g. Galesloot, van Asselt & Gierman, 1996) is promoted in which teachers of both types of education can study and discuss each other's curricula. A description of relevant skills and problem types as performed in this paper can contribute to the agenda of such a discussion. Meanwhile, it can stimulate discussions within both secondary and higher education on the educational objectives and instructional strategies concerning problem solving.

Our study poses two questions to secondary education. The first is whether problem-solving skills ought to be given attention in general secondary education. The second is whether these skills can effectively be taught at this level. If both these questions are found to have a positive answer, is it possible to consider problem-solving skills as qualifications in the sense of our definition. If this is the case, the problem-solving skills identified should be incorporated into the new examination syllabi that are to replace the current (WEN) examination syllabi.

Recently, the final draft of the new syllabi was published (Gravenberch, 1996). No major differences are found between the knowledge description sections of the new syllabi and that of the WEN syllabi (Gravenberch, 1996). On the whole, the new syllabi put more emphasis on skills. The skill listed in the skill description sections of the new syllabi (the so called 'domain A') are: language skills, mathematical skills, skills in information processing and learning, skills in using computers, skills in designing, and skills in planning and doing experiments. It is striking that problem-solving skills are not listed explicitly, even though some of the skills included do have some overlap with problem-solving skills. In his comments on the syllabi, Gravenberch (1996) occasionally mentions problem-solving skills focusing on the relation between problem-solving processes en learning and on the relevance of acquiring a systematic approach in solving problems (SPA).

From this, we conclude that the new syllabi do not put significantly more emphasis on the problem-solving skills we indicated as relevant to success in higher education or the kind of knowledge base these require (problem schemata).

In future, it may be necessary to measure the effects of the various skills in the new syllabi on success in higher education. The research approach presented in this study can be used as a basis for such an evaluation of the effects of the restructuring of secondary education on the transition problem.

On the other hand, our results and explicit description of problem-solving skills can stimulate awareness of the demands to be met by first year students in higher education. Our study suggests that a lack of problem-solving skills hinders success in higher education. This stresses the importance of developing effective pedagogical methods for teaching problem-solving skills in both secondary and higher education. In search of effective instructional strategies, it is important to note that a lack of problem-solving skills can stem from an inadequate knowledge base as is stressed by our finding of the relevance of problem-type specific skills. Exclusively stressing the need for a step by step approach in problem solving will - in our view - not be sufficient. Additional attention to forming an adequate knowledge base (problem schemata) is necessary.

At the moment further research is being conducted along this line, aiming at the design of instructional strategies for the acquisition of the problem-solving skills described here (Taconis, 1995; Taconis & van Hout-Wolters, submitted).

REFERENCES

- Alexander, P.A., & Judy, J.E. (1988). The interaction of domain specific and strategic knowledge in academic performance. *Review of Educational Research*, 58, 375-404.
- Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Dyck, M. van (1990). *Studievaardigheden in het HBO en WO: wat moet? wat kan? [Study skills in higher education: What is necessary? What is possible?]*. Paper presented at the 'Landelijke Dag Studievaardigheden', Feb. 1990. Eindhoven, The Netherlands: Eindhoven University of Technology.
- Eindexamen programma's natuurkunde [Physics examination syllabi] (1989). *Uitleg*, 21, 3-48. Zoetermeer, The Netherlands: Ministry of Education and Science.
- Ferguson-Hessler, M.G.M. (1989). *Over kennis en kunde in de fysica [On knowledge and expertise in physics]*. Unpublished Doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.
- Ferguson-Hessler, M.G.M., & Jong, T. de (1987). On the quality of knowledge in the field of electricity and magnetism. *American Journal of Physics*, 55, 492-497.
- Galesloot, L., Asselt, R. van, & Gierman, I. (1996). Betere aansluiting via onderwijsnetwerken [Better transition through pedagogical communication networks]. *Tijdschrift voor Hoger Onderwijs en Management*, 3, 34 - 39.
- Gardner, M.K. (1985). Cognitive Psychological Approaches to Instructional Task Analysis. *Review of Research in Education*, 12, 157-195.
- Gravenberch, F. (1996). *Natuurkunde: Voorlichtingsbrochure havo/vwo [Physics: the new examination syllabi for secondary education]*. Enschede, The Netherlands: SLO.
- Hays, W.L. (1988). *Statistics*. New York, Holt Rinehart & Winston.
- Het eindexamenbesluit MAVO-HAVO-VWO, (1984). In *Informatieboek voor natuurkundeleraren* (pp. 127-449). Amsterdam, The Netherlands: d'Witte Leli.
- Heuvelmans, A.P.J.M., & Sanders, P.F. (1993). Beoordelaarsbetrouwbaarheid [Agreement between judges]. In T.J.H. Eggen and P.F. Sanders (Eds.), *Psychometrie in de praktijk [Psychometric practice]* (pp. 443-469). Arnhem, The Netherlands: CITO.
- Hulshof, M.J.F. (1990). *Welke kwalificaties vraagt het wetenschappelijk onderwijs, literatuurstudie voor de ARVO naar gevraagde kennis en vaardigheden in de propedeuze [What qualifications does university require?]*. Nijmegen, The Netherlands: IOWO.
- Janssen, P.J., & Neve, H. de (1988). *Studeren en doceren aan het hoger onderwijs [Studying and teaching in higher education]* (pp. 138-145). Leuven, Belgium: ACCO.
- Jong, T. de (1986). *Kennis en het oplossen van vakinhoudelijke problemen [Knowledge based problem solving]*. Unpublished Doctoral dissertation, Eindhoven University of technology, Eindhoven, The Netherlands.
- Jong, T. de, & Ferguson-Hessler, M.G.M. (1986). Cognitive structures of good and poor novices problem solvers in physics. *Journal of Educational Psychology*, 78(4), 279-288.
- Jong, T. de, & Ferguson-Hessler, M.G.M. (1996). Types and Qualities of Knowledge. *Educational psychologist*, 31(2), 105-114.
- Mettes, C.T.C.W. & Gerritsma, J. (1986). *Probleemoplossen [Problem solving]*. Utrecht: Spectrum.
- Mettes, C.T.C.W. & Pilot, A. (1980). *Over het leren oplossen van natuurwetenschappelijke problemen [On*

- learning to solve scientific problems*]. Unpublished Doctoral dissertation, University of Twente, Enschede, The Netherlands.
- Ministerie van Onderwijs en Wetenschappen (1991). *Nota: Profiel van de tweede fase van het voortgezet onderwijs [An outline for upper secondary education]*.
- Ministerie van Onderwijs en Wetenschappen (1992). *Vervolgnota: Profiel van de tweede fase van het voortgezet onderwijs [An outline for upper secondary education]*.
- Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.
- Nijgh, G.J. (1987). *De voorspellende waarde van VWO-eindexamencijfers voor de resultaten bij de natuurkundestudie [The predictive value of pre-university final examination grades, for success in studying physics in university]*. Internal report, Eindhoven. The Netherlands, Eindhoven University of Technology.
- Pascarella, E.T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research*. New York, NY: Agathon Press.
- Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 16, 16-25.
- Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs (1994). *Tweede fase: scharnier tussen basisvorming en hoger onderwijs - een uitwerking op hoofdlijnen [Upper secondary education: link between basic education and tertiary education - main features]*. Den Haag, The Netherlands: Ministry of education and science.
- Swanborn, P.G. (1981). *Methoden van sociaal-wetenschappelijk onderzoek [Methods in the social sciences]*. Meppel, The Netherlands: Boom.
- Taconis, R. (1995). *Understanding-based problem-solving: Towards a qualifications-oriented physics education*. Unpublished Doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.
- Taconis, R., Stevens, M., & Ferguson-Hessler, M.G.M. (1992). *Probleemoplossingsvaardigheden in het natuurkundeonderwijs van de bovenbouw van het voortgezet onderwijs [Problem-solving skills in physics teaching in the final years of secondary education]*. Eindhoven University of technology, Eindhoven, The Netherlands.
- Taconis, R. & Hout-Wolters, B.H.A.M van (submitted). Systematic Comparison of Solved Problems as a Cooperative Learning Task.
- Venselaar, K., Linden, H.J.L.J. van de, & Pilot, A. (1986). *Mechanica in het eerste studiejaar [Mechanics in the first year of university]* (OC-report No. 56). Enschede, The Netherlands: University Twente.

FOOTNOTE

1. This does not imply that educational improvements within higher education could not lead to better results. This however is not within the scope of this study.

Manuscript ontvangen: 03-02-'95

Definitief aanvaard: 01-07-'97

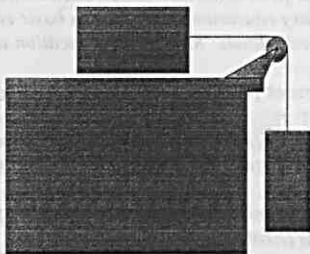
APPENDIX A: ON THE DUTCH EDUCATIONAL SYSTEM

In the Dutch educational system there are two types of general secondary education preparing for higher education: pre-university education (VWO) and higher general secondary education (HAVO) which prepares for higher vocational education (HBO). In both, students select their courses within certain limits. For each course chosen, the student has to take a final examination. These final examinations consist of two parts: A school-bound part and a national part. Grades ranging from 1 for utterly poor to 10 for excellent are calculated as the average of these two parts of the examination. The mark 6 is just sufficient to pass the examination.

There is no separate entrance examination to higher education. The institutes of higher education, however, can require that students applying for admission have taken specific courses

in general secondary education, two at the most. They cannot, however, require that applying students have achieved certain grades in them. For most of the disciplines in the technical and scientific sectors, physics and mathematics are required. Currently, a restructuring of secondary education is being implemented aiming at a more direct connection between secondary education and higher education (Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs, 1994).

APPENDIX B: EXAMPLE FROM THE PROBLEM-SOLVING TEST



Two wooden blocks are connected by a thin string. The string runs over a pulley which can rotate without friction. The tension in the string is F_s . One of the blocks hangs from the table. The mass of the hanging block is m_2 . The other block has a mass m_1 and can slide over the table surface in which case it experiences a frictional force F_f .

Show the validity of the equations below.

(1) $m_2g - F_f = (m_1 + m_2)a$

(2) $F_s = m_2(g - a)$

GUIDELINE FOR CODING (EXCERPT)

Problem-solving skills		Criteria	ISP-score
Analyzing	Force diagram	Perfect diagram	3
		No diagram	0
		Mistake in diagram	subtract 1
		Stated that $a = 0$	subtract 1
		Given equations ignored	subtract 1
		Argumentation about whether or not the blocks start moving	add 1
		Perfect solution but no explicit analysis	missing
Planning	Stating a set of equations	Correctly derived	3
		No use of $F = ma$	1
		Makes no sense	0
		Equations only derived for particular case ($a = 0$)	1
		Starts with given equations	missing
Elabora-ting	Choose free body strategically	Two free bodies	3
		Only m_1 OR m_2	2
		Only m_1+m_2	1
		No free body chosen	0
Etcetera

Vorbereidend lezen en rekenen in kleutergroepen: werkwijzen en resultaten*

Geert W.J.M. Driessen

Instituut voor Toegepaste Sociale Wetenschappen (ITS), Universiteit van Nijmegen

ABSTRACT

This article describes the results of a study into the relationship between teacher and pupil characteristics in forms 1 and 2 (nursery classes; age 4-6) of primary education. The study discerns three groups of variables: pupils' socio-ethnic background; pre-reading and pre-arithmetic test results; teaching method (intensity of steering/programmatic activities). The main question is whether the pupils' test results vary as a function of teaching method after controlling for socio-ethnic background on the individual as well as the class level. The sample consists of 446 teachers and 5490 pupils. The results of multilevel analyses show that, although there is ample variance in the variables, teaching method is of no relevance in explaining test results. The most important factor is the pupils' socio-ethnic background. Belonging to the Dutch or migrant working-class category has a negative effect on the pre-reading and pre-arithmetic test results. The teaching method effects (i.e., the lack of effects) do not differ for classes with or without migrant pupils. This does not alter the fact, however, that pupils in classes with relatively many working-class pupils perform worse, apart from their individual socio-ethnic background.

1. INLEIDING

1.1. Evaluatie van het kleuteronderwijs

In 1994 publiceerde de Commissie Evaluatie Basisonderwijs (CEB, 1994) een overzichtsstudie over het niveau en de kwaliteit van het onderwijs aan jonge kinderen, dat wil zeggen kinderen van 4 tot 7 à 8 jaar. Speciale aandacht kreeg de problematiek van de integratie van kleuter- en lager onderwijs. De commissie concludeert dat scholen slechts in beperkte mate doelen en inhouden formuleren voor het onderwijs aan jonge kinderen. Veelal gebeurt dat in algemene termen en kunnen ze niet zonder meer worden vertaald in concrete leer- en ontwikkelingsdoelen. Scholen slagen er ook onvoldoende in om door middel van het leer- en vormingsaanbod een ononderbroken ontwikkelingsproces te realiseren bij deze kinderen. Van een flexibele overgang van groep 2 naar groep 3 is vaak geen sprake, omdat pas aan het begin van groep 3 - klassikaal - wordt gestart met het leren lezen en rekenen. De pedagogisch-didactische aanpak in de kleutergroepen (groep 1 en 2) is onvoldoende doelgericht en voorgestructureerd. Onder 'pedagogisch-didactische aanpak' verstaat de commissie het sturend dan wel voorwaardenscheppend handelen van de leerkracht, de wijze waarop de vorderingen en het leer- en ontwikkelingsverloop van de leerlingen worden geëvalueerd en geregistreerd en de wijze waarop de leerkracht binnen de groep de materialen en werkvormen varieert en afstemt op de individuele behoeften van de leerlingen. Met betrekking tot de opbrengsten van het onderwijs in de kleutergroepen, in termen van kennis, vaardigheden en attitudes die van belang zijn voor een voorspoedig verlopende schoolloopbaan, constateert de commissie dat die nauwelijks systematisch zijn nagegaan. Met name is er onvoldoende bekend over de voorbereiding van jonge kinderen op het leren lezen en rekenen.

1.2. 'Leren' in het kleuteronderwijs

Het eindoordeel van de commissie is niet erg positief. Als een van de oorzaken voor deze situatie verwijst zij naar het vasthouden aan de 'verworvenheden' van het voormalige kleuteronderwijs als argument om niet te hoeven veranderen. Een belangrijk punt daarbij betreft de plaats van het 'leren' van kinderen in kleutergroepen. Cruciale vragen luiden: Moeten kinderen in kleutergroepen überhaupt wel leren? Zo ja, bij wie moet dan het initiatief liggen? En hoe moet daar vervolgens vorm aan worden gegeven?

Rond deze discussie zijn inmiddels tal van antagonismen ontstaan: voorwaardenscheppend versus sturend, ontwikkelingsgericht versus programmagericht, sturend versus volgend, leren versus ontwikkeling, spelen versus leren, kindgericht versus leerstofgericht, incidenteel versus intentioneel leren (vgl. ARBO, 1990; CEB, 1994; Janssen-Vos, 1993; Van Kuyk, 1991; Van Oers, 1993). Daarnaast is er sprake van een zekere - en tot verwarring leidende - 'schoolvorming' rond deze visies, bijvoorbeeld Ervaringsgericht onderwijs (EGO; Laevers, 1990), Ontwikkelingsgericht onderwijs (Janssen-Vos, 1993), Basisontwikkeling (Janssen-Vos, 1990; 1994), een geïntegreerde benadering (Van Kuyk, 1991), en een benadering die meer gericht is op specifieke groepen van leerlingen, het zogenaamde Leerlinggericht onderwijs (Mooij, 1994).

De posities op genoemde dimensies duiden voor een groot deel op gradaties van doelgerichtheid in pedagogisch-didactisch handelen. De belangrijkste achtergrond daarbij is het scholingsconcept dat men voor ogen heeft. Van Parreren (1985) typeert het onderwijs aan kleuters als 'ontwikkelingsgericht' en het onderwijs aan oudere leerlingen als 'programma-tisch/intentioneel gericht'. Ook de ARBO (1990) volgt dit onderscheid. De CEB (1994) spreekt in dit verband over 'voorwaardenscheppend' en 'sturend'. Het scholingsconcept dat veel leerkrachten in groep 1 en 2 voor ogen staat, wordt gedomineerd door ontwikkelingsgerichte trekken: veel waarde wordt gehecht aan de belangstelling van het kind, het zelfontdekkend leren, het scheppen van een rijke omgeving met uitnodigend materiaal (Janssen-Vos, 1989). Leerkrachten tonen een zekere beduchtheid voor prestatiegerichtheid en de introductie van methoden en toetsen (Van Gennip & Ruijs, 1992; Van Kuyk, 1992). Er zijn echter ook tegenbewegingen aan te wijzen. In programma's die gericht zijn op het vergroten van kansen van - met name allochtone - achterstandsgroepen (bv. Opstap), zijn de elementen van programma-tisch/intentioneel leren duidelijk herkenbaar. De veronderstelling is dat allochtone kleuters meer baat bij een dergelijke, sturende aanpak hebben en dat de leerkrachten hun pedagogisch-didactisch handelen daar dan ook in het bijzonder op zouden moeten richten (Van Kuyk, 1993; Van der Leij, 1993; vgl. ook Commissie (Voor)schoolse Educatie, 1994; Pels, 1995). Ook in de theorievorming over schooleffectiviteit en effectieve instructie staat het gericht handelen door de leerkracht centraal: bewaking van de effectieve leertijd, doelgerichte instructie en feedback, frequente evaluatie, differentiatie (Creemers, 1994).

1.3. De praktijk in de kleutergroepen

Wanneer we de sterke punten van effectieve instructie leggen naast de dominante praktijk in kleutergroepen, dan moet geconstateerd worden dat deze kwaliteitskenmerken hoogstens impliciet aan bod komen. Het kleuteronderwijs kent een grote mate van terloopsheid en ongedwongenheid. Voorwaarden scheppen wordt veelal belangrijker geacht dan sturen. Voorzover er sprake is van sturing staat het geven van opdrachten centraler dan het verzorgen van instructie en feedback. De feedback is weinig gericht en bij de observaties staat de registratie van de gemaakte werkjes, en niet het proces en de resultaten, centraal (Reezigt, Guldemond & Ros, 1993).

Eén element van de effectieve-instructietheorie betreft differentiatie, het rekening houden met verschillen tussen leerlingen. Leerkrachten differentiëren voornamelijk door verschillende eisen aan leerlingen te stellen. Kinderen die goed meekunnen, mogen zelfstandiger werken en krijgen moeilijkere taken. Minder snelle kinderen krijgen extra hulp en gemakkelijker werkjes. Deze differentiatiepraktijken kunnen echter leiden tot vroegtijdige acceptatie

van (blijvende) verschillen tussen leerlingen (vgl. Van der Hoeven-Van Doornum, 1990). Er zijn aanwijzingen dat dit niet alleen opgaat voor allochtone arbeiderskinderen, maar in nog sterkere mate voor allochtone leerlingen. Hierbij aansluitend kan worden opgemerkt dat onderzoek naar leerlingen in achterstandssituaties laat zien dat er markante prestatieverschillen zijn naar sociaal-etnisch herkomstmilieu. Bovendien blijken deze verschillen zich al bij de entree van de kinderen in het onderwijs te manifesteren en lijkt het erop, dat ze gedurende de onderwijsloopbaan alleen maar groter worden (Meijnen & Riemersma, 1992). Niet alleen speelt de sociaal-etnische achtergrond op individueel niveau een rol, uit recent onderzoek komen aanwijzingen dat dat ook voor het groeps/klasniveau geldt. Wanneer er binnen een klas een hoge concentratie is van achterstandsleerlingen, kan dat namelijk nadelige gevolgen hebben voor de individuele prestaties.

Uit het voorgaande zal duidelijk zijn geworden dat er geen consensus bestaat over de inrichting van het onderwijs aan jonge kinderen; in elk geval is er een spanningsveld tussen ontwikkelingsgericht en programmagestueerd onderwijs. Ook de CEB heeft het - in wat andere bewoordingen - over de subtiele balans tussen voorwaarden scheppen en sturen. Interessant is dat de commissie beide benaderingen doelgericht acht; een en ander is afhankelijk van de te realiseren leer- en ontwikkelingsdoelen. Anders gezegd: sturen waar het moet en voorwaarden scheppen waar het kan. De vraag is wat bij het voorbereidend lezen en rekenen de meest adequate aanpak is.

2. PROBLEEMSTELLING

Het bovenstaande samenvattend kunnen we concluderen dat er slechts weinig bekend is over de opbrengsten van het kleuteronderwijs en tot welke oorzaken deze kunnen worden herleid¹. We doelen wat dit laatste betreft niet alleen op kenmerken van het onderwijs, i.c. de pedagogisch-didactische aanpak van leerkrachten, maar ook op leerlingkenmerken zoals het herkomstmilieu. Teneinde meer inzicht te verkrijgen in dit type aspecten en de relaties daartussen, is in 1995 een onderzoek uitgevoerd met de volgende *vraagstellingen*:

1. Welke prestaties leveren kleuters met verschillende sociaal-etnische achtergronden op het gebied van het voorbereidend lezen en rekenen?
2. In welke mate hanteren leerkrachten in de kleutergroepen specifieke pedagogisch-didactische werkwijzen ter voorbereiding op het leren lezen en rekenen? Zijn er daarbij verschillen tussen klassen mét en klassen zonder allochtone leerlingen?
3. In hoeverre spelen de leerlingachtergronden en leerkrachtwerkwijzen een rol bij de verklaring van prestatieverschillen op het gebied van het voorbereidend lezen en rekenen? Wat is daarbij de rol van de sociaal-etnische samenstelling van de klas waarin de leerling zit? Verschillen eventuele effecten van de werkwijze naar de aanwezigheid van allochtone kinderen in de klas?

Modelmatig is de achterliggende verwachting dat de prestaties van de leerlingen in eerste instantie worden beïnvloed door hun sociaal-etnische achtergrond. Het afkomstig zijn uit een lager sociaal-economisch milieu en het toebehoren tot een allochtone bevolkingsgroep hangen negatief samen met de prestaties. Wat de pedagogisch-didactische werkwijze betreft, verwachten we dat er verschillen zijn tussen leerkrachten naar de mate waarin ze sturend handelen. We gaan er hierbij van uit dat de sturing sterker zal zijn in klassen mét allochtone kleuters dan in klassen zonder allochtone kleuters. We verwachten ook dat de prestaties van de kinderen zullen variëren met de pedagogisch-didactische aanpak van hun leerkrachten: meer sturing zal samengaan met hogere prestaties. Daarbij zullen er ook gevolgen optreden van de sociaal-etnische samenstelling van de klas op de individuele prestaties. Naarmate er meer kleuters uit achterstandsgroepen in de klas zitten, zullen de kleuters slechter presteren.

Bovendien verwachten we dat het effect van de werkwijze op de prestaties zal variëren voor klassen mét en klassen zónder allochtone kleuters.

In het vervolg van dit artikel beschrijven we eerst de onderzoeksopzet waarvoor is gekozen en de wijze waarop de centrale variabelen zijn geoperationaliseerd. Daarna gaan we over tot de feitelijke beantwoording van de vraagstellingen². We sluiten het artikel af met enkele discussiepunten; we zullen daarbij tevens terugkomen op enkele recentelijk verschenen verwante studies.

3. METHODE

3.1. Steekproef

Ter beantwoording van de onderzoeksvragen maken we gebruik van gegevens die zijn verzameld in het kader van het cohort-onderzoek Primair Onderwijs 'PRIMA' (vgl. Jungbluth & Meijnen, 1994). Dit onderzoek is in het schooljaar 1994/95 gestart met een eerste dataverzamelingsronde op circa 800 scholen voor basis- en speciaal onderwijs en met in totaal ongeveer 70000 leerlingen. Binnen PRIMA zijn met behulp van verschillende onderzoeks-instrumenten gegevens verzameld bij leerlingen en hun ouders, bij scholen, directies en groepsleerkrachten. Voor een uitgebreide verantwoording van de opzet en dataverzameling verwijzen we naar Jungbluth, Van Langen, Peetsma & Vierke (1996) en Van Langen, Vierke & Robijns (1996).

De *algemene* opzet van PRIMA is zodanig dat eind 1994, begin 1995 bij de leerlingen uit de groepen 2, 4, 6 en 8 toetsen zijn afgenomen en dat tegelijkertijd de leerkrachten die in het jaar daaraan voorafgaand aan de betreffende leerlingen les gaven een vragenlijst hebben ingevuld over hun pedagogisch-didactische werkwijze. Voor deze opzet is gekozen omdat de leerkrachten van het voorafgaande jaar langer ervaring hebben met de betreffende leerlingen dan de leerkrachten bij wie ze ten tijde van de toetsafname in de klas zaten (nl. een vol versus een half jaar) en dat bovendien eventuele effecten van de werkwijze dan ook waarschijnlijk in sterkere mate kunnen worden toegeschreven aan de vroegere leerkrachten. Voor het onderhavige artikel richten we ons op de gegevens die betrekking hebben op de kleutergroepen, dat wil zeggen groep 1 en 2 van de basisschool³.

Bij het samenstellen van de analyse-steekproef hebben we een selectie genomen uit het totale PRIMA-bestand. Criteria die we daarbij hanteerden zijn: alleen leerlingen waarvan gegevens van al de genoemde onderzoeksinstrumenten beschikbaar zijn, alleen klassen met gegevens over minimaal drie leerlingen, en alleen leerkrachten waarvan ten minste vijf indicatoren van de pedagogisch-didactische aanpak bekend zijn. Het uiteindelijke bestand telt 5490 leerlingen, afkomstig uit 446 klassen van 323 scholen. Om te controleren of er eventueel sprake was van selectieve uitval zijn de gemiddelden en standaarddeviaties van de meest relevante variabelen berekend, en ook de correlaties tussen de leerkracht-, respectievelijk leerling-variabelen, voor zowel het uiteindelijke analysebestand als de uitgevallen respondenten in de deelbestanden. Hieruit bleek dat het analysebestand een vrij goede afspiegeling vormt van de oorspronkelijke deelbestanden.

3.2. Operationalisatie centrale variabelen

Gezinsachtergronden

Het onderzoek is wat de leerlingen betreft gestart in groep 2. Aangezien er eigenlijk niet sprake is van een nul-meting, en er dus niet gecontroleerd kan worden voor entree-capaciteiten, hanteren we als substituuat de (individuele) *sociaal-etnische achtergrond*. Ter indicatie daarvoor is gebruik gemaakt van de weegfactor zoals die binnen het Onderwijsvoorrangsbeleid (OVV) functioneert. De drie grootste categorieën die in dit verband worden onderscheiden zijn: (a) allochtone kinderen van ouders met een laag opleidings- of beroepsniveau krijgen

weefactor 1.9; (b) autochtone kinderen van ouders met een laag opleidings- en beroepsniveau krijgen weefactor 1.25; (c) de overige kinderen krijgen weefactor 1.0, dit zijn de niet-achterstandsleerlingen.

Leerkrachtkenmerken

Aan de leerkrachten van groep 1, zoals gezegd de leerkrachten bij wie de leerlingen een jaar voordat ze werden getoetst in de klas zaten, is een vragenlijst voorgelegd, die is gericht op verschillende onderwerpen. Voor het onderhavige onderzoek is ter indicatie van de *pedagogisch-didactische werkwijze* een selectie genomen uit drie thema's: (a) ontwikkelingsaspecten; (b) speciale aandachtspunten; (c) aansluiting groep 1/2 - groep 3. De vragen zijn gebaseerd op een uitgebreide literatuurstudie, waarbij is gezocht naar factoren waarvan in zowel kwantitatief als kwalitatief empirisch onderzoek is gebleken dat ze verschillen in onderwijsprestaties konden verklaren (vgl. Overmaat & Ledoux, 1995). De vragen zijn daarmee dus niet geformuleerd vanuit één bepaald theoretisch oogpunt, maar vooral geselecteerd op empirische relevantie. De vragen bestrijken zowel de aanpak in groep 1 en 2, als ook de overgang van groep 2 naar groep 3. Omdat de kleuters in het merendeel van de gevallen in een combinatieklas zitten en gedurende de kleuterperiode onderwijs krijgen van dezelfde leerkracht, was de verwachting dat deze laatste doorgaans zonder problemen alle vragen zou kunnen beantwoorden. In totaal zijn 59 vragen uit de vragenlijst geselecteerd. Het betrof 17 aspecten van de werkwijze van de leerkracht; na factor- en betrouwbaarheidsanalyse bleven er uiteindelijk nog 13 over. In Tabel 1 geven we een beknopte karakteristiek van de resulterende aspecten.

Tabel 1. Aspecten pedagogisch-didactische werkwijze, (schaal)karakteristieken

kenmerk	n items	α
<i>Ontwikkelingsaspecten:</i>		
aanwezigheid systematisch programma ontwikkelingsaspecten	9	.91
aanwezigheid kant-en-klare methode ontwikkelingsaspecten	9	.78
wijze van monitoring ontwikkelingsaspecten	6	.69
gebruik screeningslijsten/toetsen ontwikkelingsaspecten	9	.93
mate van oefening en instructie begrippen en denkvaardigheden	6	.88
toetsing begrippen en denkvaardigheden t.b.v. overgang groep 2 → 3	6	.92
planning voorwaardenscheppende en sturende activiteiten	1	*
<i>Speciale aandachtspunten:</i>		
nadruk in curriculum op training/voorbereidend lezen/rekenen	7	.70
<i>Aansluiting groep 1/2 - groep 3:</i>		
overleg groep 1/2 - groep 3 over leerstof lezen en rekenen	2	.97
systematische voorbereiding aanvankelijk lezen	1	*
gebruik voorlopers leesmethode	1	*
systematische voorbereiding aanvankelijk rekenen	1	*
gebruik voorlopers rekenmethode	1	*

* N.v.t.

Toetsresultaten

Als *cognitieve maten* zijn de scores van de kinderen op twee onderdelen van het Cito-leerlingvolgsysteem genomen, namelijk Ordenen en Begrippen. De twee toetsen zijn groeps-

gewijs afgenomen door externe testleiders; hierbij waren doorgaans tevens de betrokken groepsleerkrachten aanwezig. De afname heeft plaatsgevonden in de eerste maanden van 1995. Een uitgebreide verantwoording van de keuze voor deze instrumenten is reeds gegeven in Driessen, Van Langen en Oudenhoven (1994); voor de verantwoording van de toetsen zelf verwijzen we naar Van Kuyk (1992) en Verhoeven (1993). We beperken ons hier tot het geven van enige karakteristieken.

Volgens Van Kuyk (1992) is ordenen een verzamelbegrip voor een aantal vaardigheden die de algemene cognitieve ontwikkeling van jonge kinderen bevorderen; daarnaast ontwikkelen ze ook meer specifiek het getalbegrip en het leren rekenen. De toets Ordenen is opgebouwd rond drie principes: classificeren, seriëren, en vergelijken en tellen. De toets als geheel telt 42 items; de betrouwbaarheid (KR-20) bedraagt .88. Met de Begrippentoets (Verhoeven, 1993) wordt vastgesteld in hoeverre leerlingen begrippen beheersen die voortdurend worden gebruikt bij het leren lezen en rekenen. Voorbeelden zijn: veel, weinig, eerste, laatste. In feite wordt met de toets bepaald of de leerlingen aan het aanvankelijk lees- en rekenonderwijs kunnen beginnen. De toets bestaat uit twee delen: begrippen in concrete situaties uit de leefwereld van kinderen, en begrippen die verband houden met afspraken over leesrichting en de verdeling van woorden en zinnen op de bladspiegel. De toets telt in totaal 60 opgaven; de betrouwbaarheid (KR-20) bedraagt .93. De uiteindelijke toetsscores zijn bepaald als het percentage goed gemaakte items. Er wordt van uitgegaan dat de resultaten op Begrippen en Ordenen een indicatie geven van het niveau dat de kleuters hebben met betrekking tot de voorbereidende lees- en rekenvaardigheid.

4. RESULTATEN

4.1. Beschrijving van de variabelen

Gezinsachtergronden

De verdeling van de steekproef van 5490 leerlingen naar sociaal-etnische achtergrond is als volgt: OVB-weefactor 1.9: 19.2%; 1.25: 33.8%; 1.0: 47.0%. De OVB-weefactor is een compacte variabele die zowel informatie over het sociaal-economisch milieu als het herkomstland combineert⁴. Bovendien - zo blijkt uit uitgevoerde analyses - hangt die ook nog eens lineair samen met de toetsresultaten. De sociaal-etnische achtergrond wordt als controlevariabele in de analyses opgenomen. Deze variabele kent een gemiddelde van 1.72 met een standaarddeviatie van .76.

Leerkrachtkenmerken

Tabel 2 geeft een overzicht van de intensiteit van de verschillende aspecten van de pedagogisch-didactische werkwijze van de leerkrachten ($n = 446$). De oorspronkelijke variabelen zijn hiertoe omgezet in z-scores en vervolgens gedichotomiseerd volgens de regel: $z \geq .5$ wordt 1, en andere waarden worden 0. Boven een z-score van .5 bevindt zich ongeveer 30% van de leerkrachten. De dichotomieën worden zo als indicatie van de mate van sturing opgevat, dus zwak versus sterk sturend. Omdat we ook willen nagaan of er een samenhang is met de aanwezigheid van allochtone leerlingen in de klas, hebben we in Tabel 2 de werkwijze-scores ook uitgesplitst naar 'wel/niet allochtonen in de klas'.

We beperken ons hier tot een korte toelichting op de tabel. De gemiddelden die genoemd staan bij de afzonderlijke werkwijze-aspecten betreffen steeds het percentage leerkrachten dat die werkwijze 'intensief' toepast. Bij wijze van voorbeeld het volgende: wat het gebruik betreft van programma's met betrekking tot ontwikkelingsaspecten, zien we dat 26% van de leerkrachten dat intensief doet. Hoewel het verschil statistisch niet significant is, blijkt dat het percentage leerkrachten in klassen zonder allochtone kinderen dat zich intensief met dergelijke programma's bezighoudt, hoger ligt dan het percentage leerkrachten in klassen met

allochtone leerlingen. Als we de tabel als geheel overzien, dan springen er enkele opvallende punten uit. Het percentage leerkrachten dat op het eind van groep 2 hun leerlingen toetst in verband met hun overgang naar groep 3 is groot (64%). Het percentage leerkrachten dat relatief veel nadruk legt op voorbereidend lezen, rekenen en schrijven daarentegen is gering (19%). Vermeldenswaardig is verder dat er zich in statistische zin nauwelijks verschillen voordoen tussen leerkrachten in klassen mét en klassen zónder allochtone kinderen. Slechts op twee aspecten zou van een op 5%-niveau significant (alhoewel nauwelijks relevant: $Eta=.11$) verschil gesproken kunnen worden, namelijk wat betreft de planning van voorwaarden-scheppende en sturende activiteiten, en het gebruik van voorlopers voor rekenen. In beide gevallen is het percentage leerkrachten dat dat intensief doet in klassen zónder allochtonen groter.

Tabel 2. Pedagogisch-didactische werkwijze, naar aanwezigheid allochtone leerlingen in de klas (gemiddelden)

	allochtone lln.		totaal	Eta	p
	nee	ja			
programma ontwikkelingsaspecten	30	23	26	.08	.08
methode ontwikkelingsaspecten	21	26	24	.06	.22
monitoring ontwikkeling	18	23	21	.07	.16
screeningslijsten/toetsen	39	35	37	.04	.42
oefening/instructie begrippen/denkvaardigheden	34	32	33	.02	.69
toetsing begrippen en denkvaardigheden	60	67	64	.08	.11
planning voorwaardensch. en sturende activiteiten	53	42	47	.11	.02
nadruk training/vorbereidend lezen/rekenen	20	19	19	.01	.80
overleg groep 1/2-3 leerstof lezen/rekenen	28	23	25	.07	.17
vorbereiding aanvankelijk lezen	44	44	44	.00	.97
gebruik voorlopers leesmethode	23	24	24	.01	.80
vorbereiding aanvankelijk rekenen	36	33	34	.02	.62
gebruik voorlopers rekenmethode	39	29	33	.10	.04
werkwijze totaal	35	33	34	.05	.27

We hebben geprobeerd de onderscheiden werkwijze-aspecten samen te voegen tot één of meer betrouwbare scores. Daartoe is op de 13 pedagogisch-didactische aspecten factor-analyse uitgevoerd. De correlatiematrix liet echter zien dat de kenmerken vrijwel ongecorrleerd zijn. In situaties als deze is factor-scoring niet zinvol en zal een *telling* van het aantal aspecten dat intensief wordt toegepast een index zijn voor het 'niveau' van het functioneren/handelen van de leerkrachten, i.c. hun werkwijze. Daartoe hebben we het percentage intensieve werkwijze-aspecten berekend. In de tabel staan de betreffende gemiddelden bij 'werkwijze totaal'. Het gemiddelde bedraagt 34%. Inhoudelijk betekent dit dus dat ongeveer een derde van het aantal werkwijzen een sterk-sturend pedagogisch-didactisch karakter heeft. Evenmin als bij de afzonderlijke aspecten blijkt ook hier dat er geen verschil bestaat tussen klassen mét en klassen zónder allochtone kinderen.

Etnische samenstelling klas

Om een indicatie te krijgen van de specificerende werking van de aanwezigheid van allochtone kinderen in de klas zal dit kenmerk als aparte variabele in de analyses worden meegenomen. Dat gebeurt in de vorm van de dichotomie wel/geen allochtonen in de klas⁵. In dit onderzoek geeft 57% van de leerkrachten les aan klassen met allochtone leerlingen.

Toetsresultaten

De gemiddelde score (het percentage goed) op Ordenen voor de totale steekproef bedraagt 64.7 met een standaarddeviatie van 18.5. Analyse van de subtoetsen maakt duidelijk dat met name het onderdeel 'seriëren' aan de moeilijke kant is geweest. Het gemiddelde op Begrippen is 73.3 met een standaarddeviatie van 17.1. Hier drukt het onderdeel 'leesbegrippen' het algemeen gemiddelde. Als we de gemiddelde scores uitsplitsen naar sociaal-etnische achtergrond, dan krijgen we het volgende beeld. Op Ordenen zijn de gemiddelden voor achtereenvolgens de 1.0-, 1.25- en 1.9-leerlingen: 70.8, 63.2 en 52.4 ($Eta=.37$); voor Begrippen zijn de scores: 78.5, 73.8 en 59.5 ($Eta=.41$). Evident is hiermee dat er al bij aanvang van het onderwijs zeer grote verschillen naar sociaal-etnische achtergrond bestaan. Met name de allochtone arbeiderskinderen starten al met een forse achterstand.

4.2. De verklaring van verschillen in toetsresultaten

Hierna gaan we in op de centrale vraag van dit artikel, namelijk welke relaties er bestaan tussen de kenmerken op leerling- en leerkracht/klasniveau. In verband met de verschillende niveaus waarop de gegevens beschikbaar zijn, wordt gebruikgemaakt van multilevel-analyse (vgl. Bosker & Snijders, 1990). Voor de hierna te bespreken analyses hebben we het programma MLn gehanteerd (Rasbash & Woodhouse, 1996).

De analyses kunnen worden opgevat in termen van modeltoetsingen. Samenvattend kunnen we deze modellen als volgt weergeven:

Model 0. In een multilevel-analyse wordt de totale variatie in de toetsscores van de leerlingen gesplitst in een deel leerlingniveau-variantie en een deel leerkrachtniveau-variantie. In model 0 zijn op geen van beide niveaus voorspellers opgenomen. Dit model laat zien hoeveel van de variatie in toetsscores aan elk van de twee niveaus gebonden is. De leerlingvariantie kan vervolgens vanuit leerlingkenmerken worden voorspeld, de leerkrachtvariantie uitsluitend vanuit leerkrachtkenmerken, respectievelijk geaggregeerde leerlingkenmerken.

Model 1. Nu wordt de sociaal-etnische achtergrond op het leerlingniveau ingevoerd als voorspeller van de toetsresultaten. Bekeken wordt de mate waarin de toetsscores gecorrigeerd worden voor leerling-achtergrond.

Model 2. Op groepsniveau worden de sociaal-etnische samenstelling van de klas en de interactie tussen de individuele sociaal-etnische achtergronden en de classesamenstelling ingevoerd. Hieruit wordt duidelijk wat de zelfstandige werking van de classesamenstelling is.

Met de invoering van deze twee factoren en hun interactie zijn de toetsscores dan op beide niveaus volledig gecorrigeerd voor de invloed van de achterstandsfactor. Dit model gebruiken we vervolgens als basis om de extra werking van de werkwijze-aspecten te evalueren.⁶

Model 3. Elk van de werkwijze-aspecten (totaal en afzonderlijk) wordt ingevoerd.

Model 4. De variabele 'allochtonen in de klas' wordt aan de predictoren in model 2 toegevoegd. Hiermee wordt het totale effect van wel/niet allochtone kinderen in de klas op de prestaties vastgesteld.

Model 5. Het product van werkwijze en allochtonen in de klas wordt toegevoegd om een indruk te krijgen van het lineaire interactie-effect. Teneinde een goede schatting te verkrijgen worden ook de betreffende hoofdeffecten tegelijk met de productterm in de voorspelling opgenomen. Dit geeft dan een indicatie van het interactie-effect van allochtone kinderen op het volledig gecorrigeerde effect van werkwijze. Het geeft antwoord op de vraag of het gecorrigeerde effect van werkwijze verschilt voor klassen met en klassen zonder allochtone kinderen.

De analyses resulteren in een maat die aangeeft hoeveel procent variantie er wordt verklaard, te weten R^2 . Deze R^2 -waarden zijn (afgezien van Model 0 uiteraard) berekend ten opzichte van de variantie in een voorgaand model, en wel als volgt:

- Model 0 ten opzichte van de totale variantie, dat wil zeggen: hoe is de totale variantie verdeeld over leerling- en leerkrachtniveau?

- Model 1 en 2 ten opzichte van Model 0, dat wil zeggen: welk deel van de variantie op elk van de niveaus wordt verklaard door de ingevoerde kenmerken?⁷

Teneinde te onderzoeken of er verschillen zijn tussen de twee cognitieve maten voeren we de analyses apart uit voor Ordenen en voor Begrippen. In Tabel 3 geven we de resultaten van de analyses met Ordenen en in Tabel 4 met Begrippen. Het Groot Gemiddelde is voor elk model van Ordenen vrijwel gelijk aan 65, en voor Begrippen 73. In de twee tabellen staan de samenhangen en - tussen () - standaardfouten vermeld. Effecten met een p-waarde van <.01 zijn met * gemarkeerd; (zeer zwakke) effecten met een p-waarde van <.05 met *.

Tabel 3. Resultaten multilevel-analyses m.b.t. Ordenen (ongestandaardiseerde regressie-coëfficiënten en () standaardfouten; Grand Mean: 65)

	model					
	0	1	2	3	4	5
REGRESSIE-COËFFICIËNTEN						
<i>Leerlingniveau:</i>						
soc-ethnische achtergrond		-8.1 (3)**	-7.6 (4)**			
<i>Leerkrachtniveau:</i>						
soc-ethn. klassensamenstelling			-2.3 (1.0)*			
interactie			-1.0 (.8)			
allochtone lln. in de klas					-2.9 (1.2)*	
werkwijze totaal				3.3 (2.4)		-4.7 (4.8)
programma ontwikkelingsasp.				1.6 (1.0)		.2 (2.0)
methode ontwikkelingsasp.				1.5 (1.1)		1.5 (2.1)
monitoring ontwikkeling				1.1 (1.1)		-2.9 (2.3)
screeningslijsten/toetsen				2.5 (.9)**		1.7 (1.9)
oef./instr. begrippen/denkv.				1.2 (1.0)		-3.7 (1.9)
toetsing begrippen/denkv.				.2 (.9)		-1.3 (1.9)
planning voorw./sturende act.				.8 (.9)		-1.6 (1.8)
nadruk training/voorb. lezen/rek.				-1.5 (1.2)		-2.8 (2.3)
overleg groep 1/2-3 lezen/rek.				-1.7 (1.1)		-1.5 (2.1)
voorbereiding aanvankelijk lezen				.1 (.9)		-.1 (1.8)
gebruik voorlopers leesmethode				-.5 (1.1)		.4 (2.1)
voorbereiding aanvankelijk rekenen				.3 (1.0)		.2 (1.9)
gebruik voorlopers rekenmethode				1.3 (1.0)		-2.3 (1.9)
VARIANTIECOMPONENTEN						
verklaarde variantie leerlingniveau	70%	7%				
verklaarde variantie leerkrachtniveau	30%	32%	2%			
deviantie	46487	45965	45955			

* p <.05; ** p <.01

Tabel 4. Resultaten multilevel-analyses m.b.t. Begrippen (ongestandaardiseerde regressie-coëfficiënten en () standaardfouten; Grand Mean: 73)

	model					
	0	1	2	3	4	5
REGRESSIE-COËFFICIËNTEN						
<i>Leerlingniveau:</i>						
soc-etnische achtergrond		-7.8 (.3)**	-7.2 (.3)**			
<i>Leerkrachtniveau:</i>						
soc-etn. klassensamenstelling			-4 (.9)			
interactie			-4.1 (.7)**			
allochtone ll.n. in de klas					-1.8 (1.1)	
werkwijze totaal				-4 (2.4)		-7 (4.8)
programma ontwikkelingsasp.				-1.4 (1.0)		1.1 (2.0)
methode ontwikkelingsasp.				.0 (1.1)		5.3 (2.1)*
monitoring ontwikkeling				1.0 (1.1)		-4.4 (2.3)
screeningslijsten/toetsen				2.9 (.9)**		.7 (1.8)
oef./instr. begrippen/denk.v.				2.9 (1.0)		1.7 (1.9)
toetsing begrippen/denk.v.				1.4 (.9)		-3.0 (1.9)
planning voorw./sturende act.				-1.1 (.9)		.0 (1.8)
nadruk training/voorb. lezen/rek.				1.0 (1.1)		-3.1 (2.3)
overleg groep 1/2-3 lezen/rek.				.1 (1.0)		-2.8 (2.1)
voorbereiding aanvankelijk lezen				-1.1 (.9)		2.6 (1.8)
gebruik voorlopers leesmethode				.1 (1.0)		-3.0 (2.1)
voorbereiding aanvankelijk rekenen				-1.4 (.9)		2.8 (1.9)
gebruik voorlopers rekenmethode				-1.1 (1.0)		.9 (1.9)
VARIANTIECOMPONENTEN						
verklaarde variantie leerlingniveau	66%	8%				
verklaarde variantie leerkrachtniveau	34%	29%	3%			
deviantie	45489	44916	44874			

* p <.05; ** p <.01

Uit Tabel 3 (Model 0) kunnen we wat Ordenen betreft aflézen dat 70% van de toetsvariantie aan het leerlingniveau is gebonden en 30% aan het leerkrachtniveau. De samenhang tussen de toetsresultaten en de sociaal-etnische achtergrond van de kinderen (Model 1) is -8.1 met een standaardfout van .3, hetgeen dus significant is. Inhoudelijk betekent dit dat een verschil van één categorie van de sociaal-etnische achtergrond gemiddeld genomen samengaat met 8.1 punten verandering op de toetscores. Door de sociaal-etnische achtergrond wordt 7% van de leerlingniveau-variantie verklaard en 32% van de leerkrachtniveau-variantie. Het resterende deel leerlingniveau-variantie blijft in de volgende modellen (2 - 5) hetzelfde. De variantie op leerkrachtniveau wordt nu verder verklaard vanuit leerkracht/klaskenmerken. De sociaal-etnische samenstelling van de klas (Model 2) heeft nog een relatief zwak zelfstandig effect op de toetsresultaten (namelijk -2.3; standaardfout 1.0). Dit betekent dat klassen met relatief veel achterstandskinderen (autochtoon/allochtoon) een extra lage toetscore halen bovenop het effect van instroom zoals dat al door de individuele sociaal-etnische achtergrond is

afgesplitst. Samen met de interactie verklaart de classesamenstelling 2% van de variantie op leerkrachtniveau.

Dan de toetsing van de pedagogisch-didactische werkwijze (Model 3). Het aspect met de sterkste samenhang met de toetsscores - het gebruik van screeningslijsten/toetsen - verklaart slechts 1.9% van de resterende leerkrachtvariantie. Dit werkwijze-effect bedraagt 2.5 met een standaardfout van .9, en is daarmee significant op 1%-niveau. Dit houdt in dat in klassen waarin deze werkwijze intensief wordt toegepast de gemiddelde score op Ordenen 2.5 punt hoger ligt dan in klassen waarin dit niet het geval is. De resultaten onder Model 4 laten zien dat er een significant effect is van de aanwezigheid van allochtone kinderen in de klas van 2.9. De gegevens in de laatste kolom van de tabel (Model 5) maken duidelijk dat de in Model 3 gevonden samenhangen niet verschillen tussen klassen mét en klassen zónder allochtone kinderen; het interactie-effect is namelijk in geen enkel geval significant op 1%-niveau, en evenmin op 5%-niveau.

Vergelijking van de analyseresultaten van de toets Ordenen met die van Begrippen maakt duidelijk dat de verschillen tussen beide betrekkelijk gering zijn. Het 0-model laat zien dat er bij Begrippen ten opzichte van Ordenen wat minder variantie wordt gebonden op leerling-niveau, en wat meer op leerkracht/klasniveau. Onder Model 2 blijkt dat het effect van de classesamenstelling bij Begrippen niet significant is, terwijl dat van de interactie dat nu wel is. Dit laatste houdt in dat de sociaal-etnische achtergrond in klassen met veel achterstandskinderen een kleinere rol speelt als voorspeller van de toetsprestaties dan in klassen met weinig achterstandskinderen. Evenals bij Ordenen blijkt bij Begrippen dat er niet meer dan één van de aspecten van de pedagogisch-didactische werkwijze significant (op 1%-niveau) samenhangt met de toetsprestaties; dat is ook hier het gebruik van screeningslijsten/toetsen. Dit aspect verklaart 2.5% van de leerkrachtvariantie. Het effect van allochtone kinderen in de klas is bij Begrippen in tegenstelling tot bij Ordenen niet significant. Onder Model 5 ten slotte, blijkt dat er ook nog een gering interactie-effect is (op 5%-niveau) van de aanwezigheid van kant en klare methodes gericht op ontwikkelingsaspecten. Dit betekent dat het intensief hanteren van deze methodes in klassen met allochtone kinderen een positief effect heeft op de Begrippenscores van ongeveer 2.6 punten (t.o.v. klassen met allochtone kinderen waarin deze methodes niet intensief worden gehanteerd). In klassen zonder allochtone kinderen heeft het intensief hanteren een negatief effect (t.o.v. klassen zonder allochtone leerlingen waar die methodes niet worden gehanteerd).

5. CONCLUSIES

Samenvattend luidt de conclusie dat voor de verklaring van verschillen in toetsresultaten met betrekking tot het voorbereidend lezen en rekenen de sociaal-etnische achtergronden van de leerlingen de belangrijkste factor vormt. Autochtone, maar vooral allochtone arbeiderskinderen (en dan met name Marokkaanse en Turkse kleuters) presteren slechter dan niet-arbeiderskinderen. De belangrijkste - en voor het onderhavige onderzoek tegelijkertijd teleurstellende - conclusie is dat de pedagogisch-didactische werkwijze van de leerkrachten er op een enkele uitzondering na verder niet toe doet. De analyses maken ook duidelijk dat het effect van de werkwijze - of eigenlijk: het ontbreken daarvan - niet verschilt voor klassen mét en klassen zónder allochtone kinderen. Wel blijkt dat kleuters in klassen met relatief veel autochtone en allochtone arbeiderskinderen het extra slecht doen op de toets Ordenen, los van hun individuele sociaal-etnische herkomst. Op het bovenstaande is er één uitzondering: van alle onderscheiden aspecten van de pedagogisch-didactische werkwijze is er namelijk één met een significante samenhang met de toetsprestaties. Het betreft de intensiteit waarmee de leerkrachten bij het in de gaten houden van de ontwikkeling gebruikmaakten van screeningslijsten en toetsen. Kleuters in klassen waar de leerkrachten vaker gebruikmaken van dergelijke instrumenten, halen hogere scores op zowel Ordenen, als ook op Begrippen. Mogelijk kan dit

positieve effect verklaard worden uit een soort toetservaring: de betreffende kleuters zijn beter bekend met het maken van toetsen.

Vanwaar de - wat de pedagogisch-didactische werkwijze betreft - teleurstellende resultaten? Het is gebleken dat er qua werkwijze flink wat variatie is onder de groepsleerkrachten wat betreft de mate van sturing in hun handelen. Desondanks komt uit de analyses naar voren dat deze werkwijze niet samenhangt met de door ons bestudeerde kenmerken van leerlingen en klassen. Een punt dat opkomt bij het zoeken naar een verklaring hiervoor is het tijdsaspect. Het zou kunnen zijn dat de kinderen nog te kort (één à anderhalf schooljaar) te maken hebben gehad met de werkwijze van de leerkracht om op basis daarvan een effect te verwachten op de testresultaten. Daar zou aan kunnen worden toegevoegd dat ook de intensiteit (i.c. individuele leertijd) nog niet zo sterk is geweest. Met name in het eerste leerjaar is de individuele leertijd waarschijnlijk relatief gering. Mede omdat in de (veelal gecombineerde) kleutergroepen de klasgrootte niet zelden boven de 35 kinderen ligt, zijn er signalen dat het 'echte' voorbereidende leren vooral in groep 2 gestalte krijgt. Toch kan dit niet het hele verhaal zijn. Immers, in de leerkrachtenvragenlijst is rekening gehouden met de verschillende fasen in groep 1 en 2 en de mate van sturing daarbij. En bij de analyses bleek slechts één aspect significant samen te hangen met de prestaties. Hoewel dit op zich verder niets aan deze verklaring toevoegt, zou nog kunnen worden opgemerkt dat van de compenserende werking van onderwijs in zijn algemeenheid niet al te veel verwacht mag worden. Evident is namelijk dat achterstanden van bepaalde groepen van kinderen bij hun entree in het onderwijs al erg groot zijn. Van leerkrachten kan nauwelijks verwacht worden dat zij binnen een relatief kort tijdsbestek deze achterstanden kunnen wegwerken. Wel geeft dit aan dat daar zo mogelijk al vroeg(er) iets aan gedaan moet worden (vgl. Commissie (Voor)schoolse Educatie, 1994). Het feit dat de momenteel in gebruik zijnde stimuleringsprogramma's niet alle en in alle opzichten tot de verwachte resultaten leiden, mag geen reden zijn om niet verder te blijven zoeken naar wegen om tot een reductie van onderwijsachterstanden te komen.

Zoals we in de inleiding al opmerkten, zijn er onlangs enkele publicaties verschenen waarin verslag wordt gedaan van aanverwant onderzoek. Afsluitend gaan we hierna op enkele daarvan in, omdat zij wellicht aanwijzingen geven voor de verklaring van het ontbreken van effecten van de aanpak op de prestaties.

Stevens & Harskamp (1996) rapporteren dat slechts 3% van de leerkrachten de beheersing van specifieke leervoorwaarden (rekenen en lezen) als belangrijkste doelen voor groep 2 noemt. Opvallend hierbij is dat ondanks deze lage prioriteit 56% van de leerkrachten (m.n. de 'kindsturende') toch activiteiten aanbiedt die gericht zijn op deze voorwaarden. Het lijkt er met andere woorden op dat er een zekere discrepantie bestaat tussen datgene wat de leerkrachten zeggen na te streven en wat ze concreet aanbieden, hun pedagogisch-didactische aanpak. Stevens & Harskamp verklaren dit door erop te wijzen dat leerkrachten zich door de eisen van buitenaf gedwongen voelen aan de leervoorwaarden te werken, maar dit eigenlijk liever niet doen. Naar aanleiding van deze bevinding zou ook voor ons onderzoek de vraag gesteld kunnen worden in hoeverre effecten te verwachten zijn van de pedagogisch-didactische aanpak van leerkrachten die daar wellicht niet (helemaal) achter staan.

Van Kuyk & Verhoeven (1996) vergeleken een ontwikkelingsgerichte interventie in de kleutergroepen met een traditionele aanpak, en gingen bovendien na of een programmagerichte stimulering in aanvulling op de ontwikkelingsgerichte benadering een positief effect had op de ontwikkeling van achterstandskinderen. De ontwikkelingsgerichte interventie bleek effectief te zijn en de combinatie van ontwikkelings- en programmagerichte elementen bleek een extra effect te bewerkstelligen. Van Kuyk & Verhoeven suggereren dat de effecten sterker zouden zijn geweest wanneer de programma's intensiever waren geweest en langer hadden geduurd (2 à 3 jaar). Alhoewel in ons onderzoek niet sprake is van een interventieprogramma, maar van een meerdere of mindere mate van sturing, sluit dit toch aan bij onze eerder gemaakte opmerkingen met betrekking tot de tijdsduur en -intensiteit.

In dit verband kan ook gewezen worden op de bevindingen van Schonewille & Van der Leij (1995). Zij hebben berekend dat kinderen in groep 1 gemiddeld genomen niet meer dan 1 minuut individuele aandacht krijgen op een les van 40 minuten. Verrassend daarbij is dat allochtone leerlingen niet meer, maar juist minder aandacht krijgen dan hun autochtone klasgenoten. Dit zou mede het in ons onderzoek gevonden negatieve effect van sociaal-etnische herkomst kunnen verklaren.

Of overigens bij een langere tijdsduur ook sterkere effecten van pedagogisch-didactische werkwijze kunnen worden verwacht, valt op basis van onderzoek van Van der Velden (1996) te betwijfelen. Hij onderzocht - op een voor een belangrijk deel soortgelijke wijze als wij deden - onder meer de effecten van de pedagogisch-didactische aanpak op de taal- en rekenprestaties in de groepen 6 en 8. Tot zijn verrassing bleken er zich geen effecten voor te doen. Op basis van een aanvullende literatuurstudie schrijft hij dit toe aan het feit dat de doorgaans als 'effectief' aangemerkte kenmerken in de praktijk eigenlijk niet of nauwelijks effectief blijken te zijn, met andere woorden dat de empirische evidentie vooralsnog zwak is en in ieder geval niet ondubbelzinnig is aangetoond. Mogelijk dat ook de voor ons onderzoek gedestilleerde kenmerken toch nog een nadere verificatie behoeven. Bovendien zou moeten worden nagegaan hoe het met de interne samenhang van de maat voor leerkrachtfunctioneren gesteld is; in ons onderzoek bleken de verschillende kenmerken immers nauwelijks gecorreleerd te zijn.

Ten slotte het onderzoek van Sontag & Meijnen (1995). Ook deze onderzoekers troffen geen effecten aan van verschillen in gedrag van leerkrachten, van organisatieverschillen of inhouden van curricula. Noch op een woordenschatstest afgenomen in groep 1, noch op een test voor performale intelligentie bleken er verschillen te zijn op klasniveau die verschillen tussen leerlingen zouden kunnen verklaren. Wél bleek één kenmerk op het niveau van de school er toe te doen: de beoordeling van de teamconsensus door de leerkracht had een positief effect op de leerlingprestaties. Deze bevinding sluit in bepaalde opzichten aan bij onze naar aanleiding van het onderzoek van Stevens & Harskamp (1996) getrokken conclusie, dat van leerkrachten die niet echt gemotiveerd zijn tot het aanleren van specifieke leervoorwaarden niet gauw resultaten in die richting verwacht kunnen worden.

Tot slot. Voor ons onderzoek hebben we gegevens geanalyseerd die via schriftelijke vragenlijsten bij groepsleerkrachten zijn verzameld. Hoewel het aantal vragen over de pedagogisch-didactische werkwijze met 59 als niet-gering kan worden gekwalificeerd, is het de vraag of met deze methode alle relevante informatie boven tafel kan worden gehaald. Zeker geldt dit omdat in ander onderzoek wordt gesuggereerd dat er verschillen zijn tussen datgene wat de leerkrachten zeggen te doen en wat ze daadwerkelijk doen. Ter verhoging van de validiteit van dergelijke informatie valt het daarom aanvullend aan te bevelen in onderzoek meer gebruik te maken van technieken als observaties en diepte-interviews. Verder zou in toekomstig onderzoek nog specifiekere onderscheid dienen te worden gemaakt tussen de aanpak in groep 1 en die in groep 2. In dergelijk onderzoek zou bovendien een accurate indicatie moeten worden verkregen van de tijd die de leerkracht aan de verschillende onderdelen van het curriculum en aan de individuele kleuters besteedt. Er is aanleiding daarbij de relatie tussen de doelen die moeten worden nagestreefd, de motivatie die de leerkracht daartoe heeft en de concrete uitwerking daarvan in de onderwijspraktijk aan een kritische analyse te onderwerpen.

NOTEN

1. Daar is overigens na voltooiing van het onderhavige onderzoek verandering in gekomen door de publicatie van enkele themanummers van *Pedagogische Studiën* (72, 3 & 4) en het *Tijdschrift voor Onderwijsresearch* (21, 1) over dit onderwerp. Om de 'chronologische ontwikkeling' geen geweld aan te doen komen we daar pas later in dit artikel op terug.

2. Het onderzoek dat ten grondslag ligt aan dit artikel is uitgevoerd in opdracht van SVO. Daarbij diende gebruik te worden gemaakt van gegevens die waren verzameld in het kader van het cohort-Primair Onderwijs. Voor meer gedetailleerde informatie verwijzen we naar Driessen & Claassen (1996).
3. Opgemerkt kan worden dat deze (algemene) opzet voor groep 1 en 2 eigenlijk minder relevant was. Het is namelijk gebleken dat driekwart van de leerkrachten voor een combinatieklas groep 1/2 staat, en dat het merendeel van de leerlingen dus ten minste twee jaar bij dezelfde leerkracht zit.
4. Ter verdere informatie: De verdeling naar herkomstland is: Nederland: 76%; Antillen, Suriname en Molukken: 3%; Turkije: 8%; Marokko: 6%; Mediterrane landen: 1%; Oost-Azië: 1%; overig allochtoon: 5%. Qua opleidingsniveau zijn vier niveaus onderscheiden: binnen 12% van de gezinnen hebben de ouders maximaal lager onderwijs gevolgd, 34% heeft maximaal lbo, 36% heeft maximaal mbo, en 19% heeft hbo of wo gevolgd. Voor de totale steekproef ligt het gemiddelde opleidingsniveau met een score van 2.6 tussen lbo en mbo in. Daarbij zijn er echter grote verschillen naar OVB-weegfactor. Het ouderlijk opleidingsniveau van 1.0-leerlingen ligt op een gemiddelde van 3.3, terwijl dat van 1.25-leerlingen 2.2 en dat van 1.9-leerlingen niet meer dan 1.7 is ($E_{\alpha}=.71$).
5. Opgemerkt moet worden dat dit een vrij grove indeling is. Omdat de betreffende vragen die hierover in de leerkrachtenvragenlijst zijn gesteld niet altijd correct werden beantwoord, konden we hier helaas niet nader differentiëren naar mate van concentratie van allochtone leerlingen.
6. Bij de analyses is gebruikgemaakt van een random intercept model; het intercept in de predictie van toetscore op leerlingniveau is dus als random variabele gedefinieerd en de regressiecoëfficiënt van de sociaal-etnische achtergrond met toetsscore als fixed. Dit intercept betreft dus een gecorrigeerde toetsscore berekend vanuit een algemene correctiefactor voor sociaal-etnische achtergrond. Als er feitelijk toch systematische verschillen zouden bestaan tussen groepen in de sterkte van de samenhang achtergrond-toetsscore, dan zou dit als consequentie hebben dat groepen met een relatief lage regressiecoëfficiënt worden ondergecorrigeerd en groepen met een hoge coëfficiënt overgecorrigeerd. We hebben - mede vanuit praktische overwegingen - voor deze aanpak gekozen, omdat uit analyses is gebleken dat er nagenoeg geen systematische tussen-groepsvariatie in regressiecoëfficiënten bestaat, met name als de klassensamenstelling nog als predictor van de regressiecoëfficiënt wordt toegevoegd. Voor Ordenen verklaart variatie in de gecorrigeerd coëfficiënt op klassenniveau nog slechts 1% van de systematische variantie, en voor Begrippen niet meer dan 4%. Op basis daarvan hebben we besloten deze gecorrigeerde regressiecoëfficiënt als fixed te beschouwen. Dit houdt wel in dat in de predictie van de toetsscore de interactie tussen de sociaal-etnische achtergrond van de leerling en de sociaal-etnische klassensamenstelling steeds als extra predictor moet worden opgenomen. Hoewel de interactie bij Ordenen niet significant is, en dus in principe weggelaten had kunnen worden, hebben we omwille van de uniformiteit van de analyses ook daar deze interactie als extra correctiefactor toegevoegd.
7. Omdat voor Ordenen en Begrippen samen 72 multilevel-analyses zijn uitgevoerd, is alleen voor model 0, 1 en 2 een berekening gemaakt van de variantieverklaring. De devianties van de modellen 3, 4 en 5 worden niet gepresenteerd, evenmin als de parameter-schatting van de constante. Deze laatste is voor elk model onder Ordenen ongeveer 65 en onder Begrippen 73. De reden voor deze beslissing ligt enerzijds in de presenteerbaarheid van alle gegevens en anderzijds in het feit dat deze gegevens geen extra informatie opleveren. De p-waarde voor toetsing van een bepaald werkwijze- of interactie-effect, bepaald vanuit de z-waarde van dat effect, is namelijk gelijk aan de p-waarde van het bijbehorende deviantie-verschil.

LITERATUUR

- ARBO (1990). *Spelen en jongleren. Advies over de vernieuwing van het onderwijs aan jonge kinderen*. Zeist: ARBO.
- Bosker, R., & Snijders, T. (1990). Statistische aspecten van multi-niveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-329.
- Commissie Evaluatie Basisonderwijs (1994). *Onderwijs aan jonge kinderen. Deelrapport 3*. Den Haag: SDU.
- Commissie (Voor)schoolse Educatie (1994). *(Allochtone) kleuters meer aandacht*. Rijswijk: Ministerie van WVC.
- Creemers, B. (1994). The history, value and purpose of school effectiveness studies. In D. Reynolds et al., *Advances in school effectiveness research and practice* (pp. 9-23). Oxford: Pergamon.

- Driessen, G., & Claassen, A. (1996). *Vorbereidend lezen, rekenen en schrijven. Toetsprestaties en leerling- en leerkrachtenmerken in de kleutergroepen van het basisonderwijs*. Beek: Tandem Felix.
- Driessen, G., Langen, A. van, & Oudenhoven, D. (1994). *De toetsen voor het cohort Primair Onderwijs. Verantwoording*. Nijmegen: ITS.
- Elt, M. den, Kuyk, J. van, & Meijnen, G. (1996). Culture and the kindergarten curriculum in the Netherlands. *Early Child Development and Care*, 123, 15-30.
- Hoeven-Van Doornum, A. van der (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen. Een studie naar leerkrachtverwachtingen, leerkrachtheffektiviteit en schooleffectiviteit*. Nijmegen: ITS.
- Gennip, J. van, & Ruijs, A. (1992). *Aan de basis. Onderzoek naar de pedagogisch-didactische kwaliteit van het onderwijs aan jonge kinderen*. Nijmegen: ITS.
- Grinten, M. van der, Jacobs, E., & Stufkens, N. (1996). *De onderbouw in beweging. Inventarisatie van kansrijke praktijksituaties in de onderbouw van Rotterdamse OVB-scholen*. Utrecht: SAC.
- Harskamp, E. (1993). Spelen op school. *Pedagogische Studiën*, 70, 337-351.
- Janssen-Vos, F. (1989). *Ontwikkelingsgericht onderwijs in de onderbouw*. Amsterdam: APS.
- Janssen-Vos, F. (1990). *Basisonwikkeling. Kleuters in de basisschool*. Assen/Maastricht: Van Gorcum.
- Janssen-Vos, F. (1993). Goed onderwijs voor kleuters. *Ontwikkelingsgericht versus programmagericht onderwijs. Vernieuwing. Tijdschrift voor Onderwijs en Opvoeding*, 52, (3), 8-12.
- Janssen-Vos, F. (1994). Basisonwikkeling. Ontwikkelingsgericht werken met jonge kinderen. In D. van Loggum & B. Bekkers (Red.), *Het jonge kind tussen onderwijs en welzijn* (pp. 69-88). Amsterdam: Averroës.
- Jungbluth, P. (Red.) (1994). *Onderzoeksinstrumenten voor het PRIMA-cohort*. Nijmegen: ITS.
- Jungbluth, P., Langen, A. van, Peetsma, T., & Vierke, H. (1996). *Leerlingegevens basisonderwijs en speciaal onderwijs. Technische rapportage PRIMA-cohortonderzoek 1994/95*. Amsterdam/Nijmegen: SCO/ITS.
- Langen, A. van, Vierke, H., & Robijns, M. (1996). *Veldwerkverslag basisonderwijs en speciaal onderwijs. Technische rapportage PRIMA-cohortonderzoek 1994/95*. Nijmegen/Amsterdam: ITS/SCO.
- Jungbluth, P., & Meijnen, W. (1994). *Cohortstudie Primair Onderwijs (PRIMA). Masterplan: Kernpunten van opzet en uitvoering*. Nijmegen/Amsterdam: ITS/SCO-Kohnstamm.
- Kuyk, J. van (1991). Een schets van het onderwijs aan jonge kinderen (4-8 jaar). *Pedagogische Studiën*, 68, 378-389.
- Kuyk, J. van (1992). *Ordenen. Handleiding*. Arnhem: Cito.
- Kuyk, J. van (1993). Visie op onderwijs aan jonge kinderen en hulp aan kinderen met achterstanden. In B. van Oers & F. Janssen-Vos (Red.), *Visies op onderwijs aan jonge kinderen* (pp. 15-35). Assen: Van Gorcum.
- Kuyk, J. van, & Verhoeven, L. (1996). Stimulering van geletertheid bij kleuters. *Tijdschrift voor Onderwijsresearch*, 21, 33-53.
- Laevers, F. (1990). *De betrokkenheid van de lerende als criterium voor effectief onderwijs*. Leuven: Katholieke Universiteit Leuven.
- Leij, A. van der (1993). Uitgangspunten voor een geconcentreerde aanpak. In A. van der Leij & E. Kappers (Red.), *Zorgverbreding. Bijdragen uit speciaal onderwijs aan basisonderwijs* (pp. 43-67). Nijkerk: Intro.
- Leij, A. van der, Meijnen, G., & Leseman, P. (1995). De leer- en opvoedingssituatie van jonge autochtone en allochtone leerlingen in het perspectief van hun schoolcarrière. *Pedagogische Studiën*, 72, 162-171.
- Leseman, P., Sijlsing, F., Jap-A-Joe, S., & Şahin, S. (1995). Gezinsdeterminanten van de cognitieve ontwikkeling van vierjarige Nederlandse, Surinaamse en Turkse kleuters. *Pedagogische Studiën*, 72, 186-205.
- Meijnen, G.W., & Riemersma, F. (1992). *Schoolcarrières: een klassenkwestie? De schoolloopbanen van leerlingen, bezien in relatie tot de invloeden die maatschappelijke en binnenschoolse determinanten hierop hebben. Een literatuurstudie*. Amsterdam/Lisse: Swets & Zeitlinger.
- Mooij, T. (1994). Begaafd in groep 1/2. Omgaan met (hoog)begaafde kinderen 1. *De wereld van het jonge kind*, 22, 102-106.
- Mulder, L. (1996). *Meer voorrang, minder achterstand? Het Onderwijsvoorrangsbeleid getoetst*. Nijmegen: ITS.
- Oers, B. van (1993). Onderwijsvisies en de vernieuwing van het onderwijs aan jonge kinderen: aanzet tot een discussiekader. In B. van Oers & F. Janssen-Vos (Red.), *Visies op onderwijs aan jonge kinderen* (pp. 1-14). Assen: Van Gorcum.
- Overmaat, M., & Ledoux, G. (1995). *De constructie van de instrumenten school- en klaskenmerken voor het cohort-onderzoek primair onderwijs. Verslag van twee voorstudies*. Amsterdam: SCO.
- Parreren, C. van (1985). Van kleuter tot schoolkind: continuïteit in het leren? *Pedagogische Studiën*, 62, 174-183.
- Pels, T. (1995). Kleuters verdienen méér dan individu-gerichte aandacht. *Jeugd en samenleving*, 25, 139-147.
- Rasbash, J., & Woodhouse, G. (1996). *MLn command reference. Version 1.0a*. London: University of London.
- Reezigt, G. (1996). Onderzoek naar vormgeving en effecten van onderwijs aan jonge kinderen in Nederland. *Tijdschrift voor Onderwijsresearch*, 21, 7-32.
- Reezigt, G., Guldmond, H., & Ros, A. (1993). *Effectief onderwijs aan jonge kinderen*. Groningen: RION.

- Reezigt, G., & Ros, A. (1994). De kwaliteit van het onderwijs aan jonge kinderen. *School & Begeleiding*, 41, (11), 17-19.
- Schonewille, B., & Leij, A. van der (1995). De rol van de leerkracht voor allochtone en autochtone leerlingen in het eerste jaar van het basisonderwijs. *Pedagogische Studiën*, 72, 242-257.
- Sontag, L., & Meijnen, G. (1995). De invloed van school- en klaskenmerken op de cognitieve en linguïstische vaardigheden van kleuters. *Pedagogische Studiën*, 72, 258-272.
- Stevens, L., & Harskamp, E. (1996). Actuele opvattingen over onderwijs aan jonge kinderen. Een onderzoek onder leerkrachten van groep 2 in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 21, 81-94.
- Velden, L. van der (1996). *Context, visie, aanpak en effectiviteit. De bestrijding van achterstanden van Nederlandse leerlingen in het basisonderwijs*. Groningen: GION.
- Verhoeven, L. (1993). *Begrippentoets. Handleiding*. Arnhem: Cito.

*Met dank aan J. Doesborgh en J. Hutjes voor hun methodologische adviezen en hulp bij de MLN-analyses.
Adres: Postbus 9048, 6500 KJ Nijmegen.

Manuscript ontvangen: 17-07-'96

Definitief aanvaard: 01-07-'97

Ontwikkeling van een adaptieve aanpakstrategie voor het schatten van hoeveelheden

Lieven Verschaffel, Erik De Corte, Christian Lamote en Nele Dhert
Centrum voor Instructiepsychologie en -Technologie, K.U. Leuven

SAMENVATTING

In deze studie wordt de ontwikkeling van een handige strategie voor het schatten van aantallen onderzocht vanuit het theoretisch perspectief van 'strategic change' (Lemaire & Siegler, 1995). Daartoe werd gebruikgemaakt van een eenvoudige schattaak waarbij drie groepen subjecten (20 universiteitsstudenten, 20 zesdeklassers en 10 tweedeklassers) het aantal blokjes - geplaatst in een 10 X 10 rooster - zo goed mogelijk moeten proberen te schatten. Het schatten van een dergelijke hoeveelheid blokjes kan in principe op twee verschillende manieren gebeuren: (1) via het herhaald optellen van schattingen van groepen van blokjes waarin de totale hoeveelheid is onderverdeeld (= optelprocedure) of (2) door het geschatte aantal lege vakjes af te trekken van het (geschatte) totaal aantal vakjes in het rooster (= aftrekprocedure). Een rationele taakanalyse maakt duidelijk dat de meest efficiënte globale schatstrategie bestaat uit het adaptief gebruik van beide procedures, rekening houdend met de verhouding tussen het aantal gekleurde blokjes en lege vakjes. Op grond van deze rationele taakanalyse luidt de eerste hypothese dat het gebruik van deze adaptieve strategie zal toenemen met de leeftijd en volgens de tweede hypothese is er een positief verband tussen de mate waarin deze adaptieve strategie gebruikt wordt enerzijds en de schatprestaties anderzijds. Convergerende empirische evidentie afkomstig van verschillende soorten van data (nl. patronen van reactietijden en schatuitkomsten voor de 100 experimentele items en zelfrapporteringsgegevens) biedt steun voor beide hypothesen. Vanuit methodologisch perspectief toont deze studie vooral de waarde aan van Beem's (1995a) segmentatiemodel voor de fijnmazige analyse van adaptieve keuzen tussen verschillende procedures en voor de bestudering van de relatie tussen dergelijk adaptief proceduregebruik bij cognitieve taken en prestaties op deze taken.

ABSTRACT

This study investigates the development of a clever strategy for estimation of numerosity from the theoretical perspective of 'strategic change' (Lemaire & Siegler, 1995). A simple estimation task was used in which participants of three different age groups (20 university students, 20 sixth-graders and 10 second-graders) had to estimate 100 numerosities of colored blocks presented in a 10 X 10 rectangular grid. Generally speaking, this task allows two distinct estimation procedures: either repeatedly adding estimations of groups of blocks (= addition procedure) or subtracting the estimated number of empty squares from the (estimated) total number of squares in the grid (= subtraction procedure). A rational task analysis indicates that the most efficient overall strategy for this estimation task consists of the adaptive use of both procedures, depending on the ratio of the blocks to the empty squares. Starting from this

rational task analysis, it was first hypothesized that there would be a developmental difference in the adaptive use of the two procedures, and according to the second hypothesis this adaptive use would result in better estimation accuracy. Converging evidence from different kinds of data (i.e., patterns of response times and of correct and incorrect estimations on the 100 experimental items, and retrospective reports) supported both hypotheses. From a methodological point of view, the study shows the potential of Beem's (1995a) 'segmentation analysis' for unravelling subjects' adaptive choices between different procedures in cognitive tasks and for examining the relationship between these adaptive choices and performance.

INLEIDING

Volgens Lemaire en Siegler (1995) kan een prestatieverbetering op een cognitieve leertaak minstens vier verschillende soorten strategische veranderingen ('strategic changes') reflecteren: (1) de verdwijning van oude en de opkomst van nieuwe strategieën, (2) een verschuiving in de frequentie waarmee bestaande strategieën gebruikt worden, (3) een toename van de vlotheid en de accuraatheid waarmee deze strategieën gehanteerd worden en (4) een toename van het adaptief karakter van de strategiekeuzen.

Op basis van een beknopt overzicht van de beschikbare literatuur rond 'strategic change' komen Lemaire en Siegler (1995) tot de conclusie dat '(...) usually, when the contribution of strategic change to learning is recognized at all, it is viewed solely in terms of acquisition of new, more effective approaches' (p. 94). Op basis van eigen en ander recent onderzoek bij verschillende soorten van leertaken uit diverse inhoudsgebieden komen zij echter tot de conclusie dat de drie overige vormen van 'strategic change' minstens even essentieel zijn. Zo blijkt uit hun eigen studie in verband met het leren van 'de tafels van vermenigvuldiging' dat de prestatieverbeteringen van een groep tweedeklassers in de loop van het schooljaar juist niet voortvloeiden uit het gebruik van nieuwe strategieën - immers, van meetaf aan maakten de leerlingen zowel gebruik van 'fact-retrieval strategies' als van 'backup strategies' -, maar eerder toe te schrijven waren aan veranderingen in de drie andere aspecten van 'strategic change'. Vooral veranderingen met betrekking tot het laatste aspect - nl. de optimale afwisseling tussen het gebruik van de 'fact-retrieval strategy' en de 'backup strategy' in functie van de moeilijkheid van de taak - bleken cruciaal te zijn voor de verbetering van de leerprestaties gedurende het schooljaar.

Ook andere onderzoekers hebben geconstateerd dat kinderen - soms reeds op erg jonge leeftijd - verschillende strategieën op adaptieve wijze gebruiken bij allerlei soorten van leertaken, zoals optellen, aftrekken, vermenigvuldigen, klokkezen, lezen en spellen van woorden (Geary & Brown, 1991; Lefevre, Greenham & Waheed, 1993; Siegler, 1988; Siegler & Jenkins, 1989) en dat dergelijk adaptief strategiegebruik een sterke samenhang vertoont met goede leerprestaties (Geary & Burlingham-Dubree, 1989).

In onderhavige studie wordt de ontwikkeling van adaptief strategiegebruik en de samenhang daarvan met leerprestaties bestudeerd bij een ander soort van cognitieve c.q. wiskundige taken, nl. het schatten van hoeveelheden. Sowder (1992) onderscheidt drie soorten reële (probleem)situaties waarin er sprake is van schatten: (1) het schatten van de uitkomst van een berekening (ofwel in de plaats van ofwel ter controle van een preciese berekening), (2) het schatten van meetkundige grootheden (bijv. lengtes, oppervlaktes, volumes,...) en (3) het schatten van hoeveelheden (bijv. het aantal mensen dat aanwezig is op een receptie, de plaats die iemand inneemt in een rij wachtenden...). Het verschil tussen de tweede en de derde vorm is dat bij het schatten van afmetingen gewerkt wordt met continue grootheden, terwijl het bij schatten van aantallen om discrete aantallen gaat (Sowder, 1992). De taak waarmee in onderhavige studie gewerkt is, heeft zoals gezegd betrekking op de derde soort van schattaken, nl. het schatten van hoeveelheden.

Hoewel algemeen aanvaard wordt dat schatten een vorm van rekenen is die bij uitstek beroep

doet op het bewust nadenken over en het flexibel inzetten van oplossingswijzen rekening houdend met de kenmerken van de taak en de context waarin die wordt gepresenteerd (Sowder, 1992; Treffers & De Moor, 1990), zijn schattaken nog nauwelijks gebruikt in onderzoek naar de ontwikkeling van adaptief strategiegebruik. Dit geldt inzonderheid voor de derde soort van schatten, nl. het schatten van hoeveelheden (Sowder, 1992).

Eén belangrijke uitzondering hierop vormt het onderzoek van Newman en Berger (1984). Zij bestudeerden de ontwikkeling van de vaardigheid in en de strategieën voor het schatten van ranggetallen. Kinderen uit de kleuterschool, het eerste en het derde leerjaar speelden een soort van 'darts' spel op een microcomputer. Ze moesten het ranggetal schatten dat met een pijltje werd aangeduid op een verticaal gepresenteerde, lege getallenlijn waarvan enkel de twee eindpunten (nl. 1 en 23) aangegeven waren. Bij iedere beurt werd de kinderen gevraagd om een schatting te maken totdat ze erin slaagden het ranggetal te geven dat het pijltje aanwees. Achteraf werd hen ook gevraagd om mondeling verslag uit te brengen over de manier waarop ze tewerkgegaan waren bij het uitvoeren van deze schattaak. Door middel van een vergelijking van de schatprestaties bij kleine (2 tot 8), middelmatige (9 tot 15) en grote ranggetallen (16 tot 22), stelden de onderzoekers eerst en vooral een evolutie vast in de schatprestaties: de derdeklassers slaagden er niet enkel in om de kleine en de grote ranggetallen erg accuraat te schatten, maar maakten ook relatief nauwkeurige schattingen van die uit de middelste regio; de jongere kinderen daarentegen slaagden er enkel in de kleinste getallen goed te schatten. Ten tweede werd ook een duidelijke ontwikkeling in (gerapporteerd) strategiegebruik vastgesteld: terwijl de meeste kinderen uit de kleuterklas bij alle beurten gebruikmaakten van één vaste, uniforme strategie, die neerkomt op simpelweg voorwaarts tellen, verklaarde de meerderheid van de eersteklassers en bijna alle derdeklassers een flexibele strategie toe te passen waarbij de ene keer voorwaarts en de andere keer achterwaarts geteld werd, afhankelijk van de grootte van het te zoeken ranggetal. Verder bleek met de leeftijd de neiging toe te nemen om eigen referentiepunten te creëren en te gebruiken voor beurten waarbij het te zoeken ranggetal zich niet in de nabijheid van één van de beide eindpunten bevond. Kortom, uit de patronen van de prestaties en uit de bijhorende zelfrapporteringen van de drie groepen kon afgeleid worden dat met toename van de leeftijd ook de neiging groeit om adaptief tewerk te gaan bij het schatten van ranggetallen, en dat deze toenemende adaptiviteit resulteert in betere schatprestaties.

In onderhavige studie staat niet het schatten van ordinale maar van kardinale hoeveelheden centraal. Meer bepaald gaat het om een taak waarbij het aantal vierkante blokjes in een 10 X 10 rooster moet worden geschat. Naar analogie met wat in de studie van Newman en Berger (1984) gevonden werd, kunnen bij het schatten van dergelijke aantallen in theorie twee verschillende aanpakstrategieën onderscheiden worden: (1) een stereotiepe strategie waarbij dergelijke aantallen steeds via eenzelfde vaste procedure (noot 1) geschat worden, nl. via het herhaald optellen van schattingen van groepen van blokjes waarin de totale hoeveelheid is onderverdeeld (= optelprocedure), of (2) een adaptieve strategie waarbij men - afhankelijk van het gegeven aantal blokjes - de ene keer de hierboven beschreven optelprocedure gebruikt en een andere keer het geschatte aantal *lege* vakjes aftrekt van het totaal aantal vakjes in het rooster (= aftrekprocedure).

Het doel van deze studie was tweeledig. Ten eerste wilden we een bijdrage leveren tot de methodologie van het onderzoek naar adaptief strategiegebruik bij cognitieve c.q. wiskundige taken door gebruik te maken van een veelzijdiger en geavanceerder manier van data-analyse dan in de studie van Newman en Berger (1984) en in de meeste andere vermelde studies op het terrein van 'adaptive strategy change' gehanteerd werd. Meer bepaald pasten wij Beem's (1995a, 1995b) 'program for fitting two-phased segmented curve models with an unknown change point' toe, - een programma dat ons toeliet om de mate van adaptief strategiegebruik af te leiden uit de reactietijden en de schatuitkomsten van subjecten én om het verband tussen adaptief strategiegebruik en de schatprestaties te bepalen. Ten tweede wilden we weten of de evolutie in de richting van toenemende adaptiviteit qua strategiegebruik én de samenhang

daarvan met de schatprestaties, zoals die gevonden is in de studie van Newman en Berger (1984) en in de andere vermelde studies, ook terug te vinden was bij taken waarin cardinale hoeveelheden geschat moeten worden.

RATIONELE TAAKANALYSE EN ONDERZOEKSVRAGEN

Zoals vermeld, bestond de taak erin om een gegeven aantal vierkante blokjes in een 10 X 10 rooster te schatten. Elk van de 100 vakjes van dit rooster was dus ofwel leeg ofwel gevuld met een gekleurd vierkant blokje, hetgeen betekent dat het totaal aantal van blokjes in het rooster steeds een getal tussen 1 en 100 was (zie Figuur 1a en 1b voor twee voorbeelden van schattaken).

Op basis van een rationele taakanalyse kan men ruwweg twee algemene strategieën om deze schattaak aan te pakken onderscheiden: een optelstrategie (of ADD-strategie, naar de Engelse term 'addition strategy') en een minimum- (of MIN-) strategie. Gebruikers van de ADD-strategie voeren bij elke schattaak dezelfde optelprocedure uit: zij verdelen de blokjes in een aantal groepen en tellen de schattingen van deze groepen vervolgens bij elkaar op, *ongeacht het aantal blokjes dat aangeboden wordt*. Wie de MIN-strategie toepast, maakt de ene keer gebruik van een optelprocedure en de andere keer van een aftrekprocedure. Bij deze laatste procedure wordt eerst een schatting gemaakt van het aantal lege vierkantjes in het rooster, waarna deze schatting afgetrokken wordt van het totaal aantal vakjes waaruit het rooster bestaat. Of er voor een optel- dan wel voor een aftrekprocedure gekozen wordt, hangt af van de verhouding tussen het aantal vierkante blokjes en lege vakjes. Wanneer het rooster relatief weinig blokjes telt (zoals in Figuur 1a), wordt gekozen voor de optelprocedure, terwijl bij items met relatief veel blokjes (zoals in Figuur 1b) op de aftrekprocedure overgestapt wordt.

Deze rationele taakanalyse maakt duidelijk dat voor de schattaak uit onderhavige studie de adaptieve MIN-strategie globaal genomen efficiënter is dan de stereotiepe ADD-strategie. Immers, zodra het rooster (veel) meer blokjes dan lege vakjes bevat, leidt de MIN-strategie in principe tot (veel) snellere en (veel) betere schattingen.

De bruikbaarheid en de efficiëntie van de MIN-strategie is evenwel afhankelijk van de mate waarin men bepaalde kennis-elementen en vaardigheden die voor het uitvoeren van deze strategie (en met name van de daarin vervatte aftrekprocedure) vereist zijn, onder de knie heeft. Allereerst moet men vertrouwd zijn met het rechthoekmodel van de vermenigvuldiging als rekenkundige operatie; men moet zich met andere woorden onmiddellijk realiseren dat de grootte van het rooster te achterhalen is door het aantal rijen te vermenigvuldigen met het aantal kolommen. Ten tweede moet het product van deze vermenigvuldiging (in dit geval: $10 \times 10 = 100$) geautomatiseerd zijn, zodat dit tafelpduct onmiddellijk en trefzeker uit het langetermijngeheugen kan worden opgediept. Ten derde moet men aftrekkingen van het type $a - b = .$ (in dit geval $100 - x = .$) vlot mentaal kunnen uitrekenen. Ten slotte moet men weten vanaf welk punt er het best overgestapt wordt van de optel- naar de aftrekprocedure. Een rationele taakanalyse suggereert dat de preciese locatie van dit 'keerpunt' (mede) bepaald wordt door de grootte van het rooster: bij een 20 X 20 rooster zal dit keerpunt bijvoorbeeld allicht verder liggen dan bij een 10 X 10 rooster. Doch het is aannemelijk dat de preciese locatie van dit keerpunt niet louter bepaald wordt door dit objectief taakkenmerk, maar ook door de subjectieve inschatting door de oplosser van de cognitieve energie ('mental load') die nodig is om de aftrekprocedure met succes te kunnen uitvoeren in vergelijking met de energie nodig voor de uitvoering van de optelprocedure.

Uitgaand van deze rationele taakanalyse en rekening houdend met de resultaten van het onderzoek van Newman en Berger (1984) en de andere vermelde studies op het terrein van 'strategy choice', werden in onderhavige studie de volgende twee hypothesen geformuleerd.

Ten eerste: het gebruik van de ADD-strategie zal afnemen en het gebruik van de MIN-strategie zal toenemen met de leeftijd. Hiervoor zijn verschillende argumenten te geven. Ten

eerste vereist de toepassing van de MIN-strategie de beheersing van verschillende wiskundige concepten (m.n. vermenigvuldigen, aftrekken) en vaardigheden (zoals het vermenigvuldigen en de procedure om het even welk getal af te trekken van 100), die normaliter pas in de loop van de aanvangsjaren van de basisschool tot ontwikkeling komen (Verschaffel & De Corte, 1996). Ten tweede is er het metacognitieve aspect (De Corte, Greer & Verschaffel, 1996): zoals in de rationele taakanalyse aangegeven, veronderstelt het efficiënt gebruik van de MIN-strategie het inzicht dat de aftrekprocedure vanaf een bepaald punt meer aangewezen is dan de optelprocedure. Uit onderzoek is bekend dat dergelijke metacognitieve kennis over de efficiëntie van verschillende werkwijzen geleidelijk en slechts moeizaam tot ontwikkeling komt (Schoenfeld, 1992). Om deze eerste hypothese te toetsen, vergeleken we het gebruik van de MIN-strategie bij verschillende leeftijdsgroepen. Meer bepaald gingen we na of het gebruik van deze handige strategie steeg naarmate de leeftijd van de subjecten toenam.

Volgens de tweede hypothese bestaat er een positief verband tussen de mate van gebruik van de MIN-strategie en de schatprestaties. Deze hypothese werd getoetst door in de verschillende leeftijdsgroepen na te gaan of subjecten die meer (systematisch) gebruikmaakten van de MIN-strategie inderdaad ook betere schatprestaties leverden.

Met het oog op een adequate toetsing van beide hypothesen dienden we echter een manier te vinden om te achterhalen of - en zo ja, in welke mate - er gebruikgemaakt werd van één van beide aanpakstrategieën. Daarvoor hebben we hoofdzakelijk beroep gedaan op de fijnmazige analyse van de reactietijden en de schatuitkomsten met behulp van Beem's (1995a, 1995b) 'program for fitting two-phased segmented curve models with an unknown change point', dat verderop toegelicht wordt. Zoals gezegd, laat deze analysetechniek toe om het gebruik van de MIN-strategie precieser te bestuderen dan in voorgaande studies rond 'strategy choice' het geval was. Doch naast deze analyse van de reactietijden en schatuitkomsten hebben we ook verscheidene andere gegevens verzameld (zoals zelfrapporteringsgegevens) en andere analyses verricht (waaronder de analyse van systematische schatfouten en van zgn. 'voltreffers').

METHODE

Subjecten

Aan het onderzoek namen 50 subjecten deel: 20 universiteitsstudenten (tussen 20 en 22 jaar), 20 zesdeklassers (van 11 of 12 jaar) en 10 tweedeklassers (tussen 7 en 8 jaar). Het geringer aantal tweedeklassers was te wijten aan het praktisch probleem dat het onderzoek onmogelijk tijdens de lessen kon plaatsvinden.

In het Vlaamse wiskundecurriculum wordt weinig of geen aandacht besteed aan het tellen of schatten van grote hoeveelheden (Verschaffel & De Corte, 1966). We gaan er dus van uit dat de subjecten de MIN-strategie niet systematisch onderwezen werd.

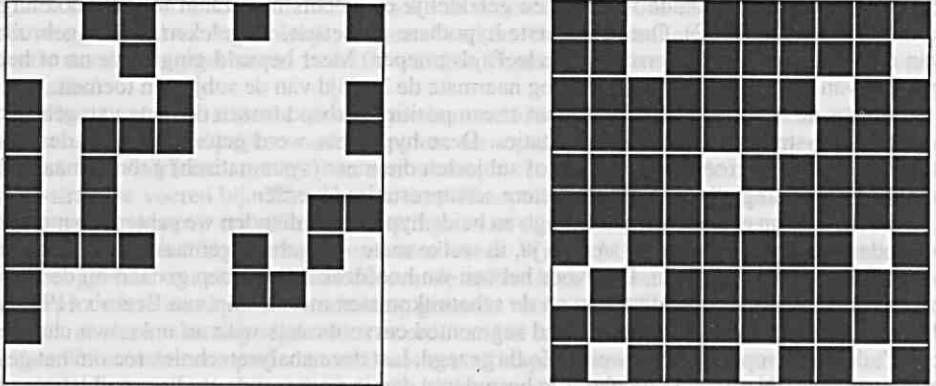
Op het moment dat de tweedeklassers onderzocht werden, hadden zij reeds uitvoerig leren optellen en aftrekken in het getalengebied tot 20, het getalengebied 20-100 systematisch verkend en al wat ervaring opgedaan met optellingen en aftrekkingen met getallen beneden 50 (zoals $24 + 10 = .$ en $36 - 8 = .$). Ten slotte hadden zij ook al een introductie gehad in de betekenis van de operaties 'vermenigvuldigen' en 'delen' en ook al geoefend op 'de tafel van 2 en van 3'.

Materiaal

De schattaak werd aangeboden met behulp van een IBM-compatibel 486 computer, uitgerust met een VGA grafische kaart en met een schermresolutie van 640 X 480. De stimuli waren vierkantige roosters bestaand uit 10 X 10 kleine vierkante hokjes. Elk hokje was ofwel gevuld met een groen vierkant blokje ofwel leeg (d.w.z. met dezelfde zwarte kleur als de achtergrond van het scherm). Zowel de hoogte als de breedte van het rooster was 200 pixels

(± 9 cm op het scherm). De vierkante eenheden waren 20 op 20 pixels. De omtrek van het 10 X 10 rooster was zichtbaar en rood van kleur. De taak van het subject bestond erin het aantal (gekleurde, vierkantige) blokjes te schatten (zie figuur 1). Er werden 100 schattaken aangeboden, gaande van 1 tot 100 blokjes. De volgorde waarin deze 100 taken werden aangeboden, werd gerandomiseerd door de computer. Ook de plaatsing van de blokjes in het rooster geschiedde 'ad random'. De computer registreerde zowel de antwoorden als de reactietijden.

Figuur 1: Twee voorbeelden van items uit de schattaak



Procedure

Alle subjecten werden individueel onderzocht. De proefleider legde hen vooraf uit dat ze 100 taken zouden krijgen en dat de opdracht er telkens in bestond een goede schatting te maken van het aantal blokjes. Een 'goede schatting' werd omschreven als 'iets tussen een pure gok en het resultaat van heel precies tellen en/of berekenen' (Sowder, 1992; Treffers & De Moor, 1990; Verschaffel & De Corte, 1996). Om de subjecten te verduidelijken wat onder een 'goede schatting' verstaan werd en om hen vooraf enige ervaring te laten opdoen met de aard van de schattaak en met de concrete werkwijze (zie verder), kregen zij na deze korte uitleg 10 voorbeeld-items, telkens gevolgd door directe feedback op het scherm over de nauwkeurigheid van hun schatting. Wanneer de schatting niet meer dan 10 % van het preciese aantal blokjes afweek, werd zij als 'goed' bestempeld; bij een afwijking van meer dan 10 % werd de schatting als 'fout' beschouwd. Zodra het echte experiment begon, werd deze feedback over de juistheid van de schatting echter achterwege gelaten.

Om te vermijden dat de subjecten puur tellend zouden tewerkgaan en hen tot schatten aan te zetten, was elke schattaak slechts gedurende maximaal 20 seconden zichtbaar. Zodra men de uitkomst van de schatting wilde meedelen, duwde men op de spatiebalk. Op dat moment stopte de klok, verdween het aangeboden rooster en kwam een nieuw scherm te voorschijn waarop het geschatte aantal blokjes via het toetsenbord kon worden ingetikt. Wanneer binnen de 20 seconden geen reactie kwam, verdween het scherm met de schattaak automatisch en kwam het andere scherm met daarop de vraag een schatting te maken, te voorschijn.

Op het eind van het onderzoek volgde een kort interview waarin ingegaan werd op de wijze waarop de deelnemers de taak hadden aangepakt. Het interview begon met een algemene vraag naar de gevolgde aanpakstrategie voor het geheel van de schattaken. Wanneer een subject het gebruik van de MIN-strategie niet spontaan rapporteerde, toonde de proefleider in een tweede fase drie kaarten met daarop respectievelijk een schattaak met weinig blokjes (nl. 27), één met veel blokjes (nl. 87) en één met ongeveer evenveel blokjes als lege hokjes (nl. 54); bij elke kaart werd gevraagd om zo gedetailleerd mogelijk te beschrijven hoe men in dit geval zou tewerkgaan om het aantal blokjes te schatten.

Er waren enkele kleine verschillen tussen de afnameprocedures bij de drie groepen. Voor de tweede- en zesdeklassers waren er korte onderbrekingen voorzien na 25, 50 en 75 items, terwijl de studenten de 100 items zonder onderbreking afwerkten. Verder werd de tweede-klassers - in tegenstelling tot de zesdeklassers en de studenten - niet gevraagd om zelf het antwoord in te tikken; zij gaven hun antwoord mondeling en het intikken gebeurde door de proef-leider. Vooronderzoek had immers uitgewezen dat het zelf intikken van het antwoord bij deze jonge kinderen tot minder betrouwbare reactietijden leidde (Dhert, 1997). Ten slotte kregen de tweedeklassers vooraf een uitvoeriger uitleg over schatten dan de andere twee groepen, omwille van de onvertrouwdheid van deze jonge kinderen met schatten.

De bespreking van de onderzoeksresultaten valt uiteen in drie delen. In een eerste deel wordt de ontwikkeling van de ADD- naar de MIN-strategie belicht door op de reactietijden en de schatuitkomsten van alle subjecten uit de drie leeftijdsgroepen zowel een lineaire als een tweefasige regressie toe te passen. In deel twee volgen de overige resultaten in verband met de ontwikkeling van de stereotiepe ADD-strategie naar de adaptieve MIN-strategie, die gebaseerd zijn op de analyse van de zelfrapporteringen, de systematische schatfouten en de zgn. 'voltreffers' in de drie leeftijdsgroepen. In een derde en laatste deel worden de gegevens over het gebruik van de ADD- en de MIN-strategie gerelateerd aan de schatprestaties. Dit betekent dat de onderzoeksgegevens uit de eerste twee delen betrekking hebben op de eerste hypothese, terwijl de resultaten uit het derde deel op de tweede hypothese slaan.

'FIT' VAN DE REACTIETIJDEN EN DE AFWIJINGEN MET HET LINEAIR MODEL EN HET SEGMENTATIEMODEL

Een eerste analyse van de ontwikkeling van de MIN-strategie vond plaats door met behulp van regressie-analyse voor alle subjecten uit de drie leeftijdsgroepen na te gaan in hoeverre het patroon van de reactietijden en van de afwijkingen ten opzichte van het exact antwoord overeenkomt met wat theoretisch te verwachten valt bij toepassing van de ADD- en de MIN-strategie.

Om tot deze analyse te kunnen overgaan, werden vooraf een aantal 'outliers' - d.w.z. schatfouten die extreem ver verwijderd lagen van het juiste aantal blokjes en bovendien het gevolg waren van een duidelijk te traceren foutief schatproces - uitgesloten. (Omdat deze systematische schatfouten bijkomend licht werpen op het gebruik van de MIN-strategie worden ze in het tweede deel van het resultatenstuk uitvoerig besproken). Tevens verwijderden we enkele extreme schatfouten die vrijwel zeker als tikfouten te beschouwen waren (bijv. een subject had bij het item met 87 blokjes het getal '887' ingetikt) of die ontstonden doordat een subject per abuis onmiddellijk (d.w.z. binnen de 50 ms) na het verschijnen van een nieuw item op de spatietoets had gedrukt. Al deze eliminaties leidden tot een reductie van de gegevens van de 20 (universiteits)studenten en de 20 zesdeklassers van 2000 tot respectievelijk 1855 en 1970, en van die van de 10 tweedeklassers van 1000 tot 992.

Beem's (1995a, 1995b) programma voor het analyseren van procedureverandering bij cognitieve taken

Zoals eerder vermeld, pasten we zowel op de reactietijden als op de afwijkingen - d.w.z. het (absoluut) verschil tussen het gegeven aantal blokjes en het door het subject geschatte aantal - twee statistische modellen toe: een lineair model en een tweefasig regressiemodel (ook segmentatiemodel genoemd).

Lineaire modellen zijn alom bekend en worden veelvuldig toegepast. De relatie tussen onafhankelijke en afhankelijke variabelen (in dit onderzoek respectievelijk het aantal blokjes en de reactietijd of de grootte van de afwijking) wordt beschreven door een lineaire regressievergelijking van de vorm $y=a+bx$. De parameters a en b die bepaald moeten worden wijzen op de intercept en de helling van de regressielijn.

Het segmentatiemodel is minder bekend en minder toegepast, hoewel het volgens Beem (1995a, 1995b; Ippel & Beem, 1987) uitermate geschikt is voor het bestuderen van strategie-veranderingen zoals deze die in onderhavig onderzoek bestudeerd worden. Recent beschreef Beem (1995a, 1995b) een algoritme voor het gebruik van dit segmentatiemodel. Zoals de naam 'two-phase segmented curve' laat vermoeden, wordt de relatie tussen de afhankelijke en onafhankelijke variabelen hier niet beschreven door één maar door twee regressievergelijkingen, die slaan op twee verschillende deelgebieden van de onafhankelijke variabelen. Dit model kan beschreven worden als

$$y = a_1 + b_1 x \quad (x < s)$$

$$y = a_2 + b_2 x \quad (x > s).$$

Voor waarden van de onafhankelijke variabele tot en met s wordt de eerste regressievergelijking gebruikt, terwijl voor waarden hoger dan s een tweede regressievergelijking - met een andere intercept en een andere helling - gebruikt wordt. De parameter s wordt het keerpunt genoemd. Het programma van Beem (1995a, 1995b) berekent al deze parameters met behulp van een algoritme gebaseerd op de 'kleinste kwadraten'. Meer bepaald gaat het altijd op zoek naar het globale minimum van de som van de kwadraten van de fouten. Er bestaan twee varianten van het model, namelijk het 'join' model en het 'jump' model. Het 'join' model veronderstelt dat het model continu is in s ; dit betekent dat de waarde van y in s dezelfde is voor de twee regressievergelijkingen. Wanneer het model niet continu is in s , is er sprake van een 'jump' model. In onderhavige studie wordt met het 'jump' model gewerkt. Omdat uit de rationele analyse gebleken is dat de aftrekprocedure complexer is dan de optelprocedure, valt te verwachten dat zowel de reactietijden als de afwijkingen bij de aftrekprocedure (iets) hoger zullen liggen dan bij de optelprocedure. Ter illustratie vergelijken we de items met 25 en 75 blokjes. Bij beide items zal iemand die de MIN-strategie toepast eenzelfde aantal blokjes, namelijk 25, moeten tellen of schatten: in het eerste geval gaat het om de 25 aangeboden blokjes, terwijl het in het tweede geval de 25 lege vakjes in het rooster betreft. Terwijl het schatten of tellen van deze 25 elementen in beide gevallen in principe evenveel tijd in beslag neemt, komt er in het laatste geval minstens één extra stap bij, namelijk het uitvoeren van de aftrekking $100 - 25 =$. In theorie zal deze bijkomende stap resulteren in een langere reactietijd en zal daardoor ook de kans op een minder nauwkeurige schatting toenemen.

Bij subjecten die de ADD-strategie toepassen valt - zowel wat reactietijden als afwijkingen betreft - een goede 'fit' met het lineair model te verwachten; immers, bij toepassing van deze strategie zullen reactietijd en afwijking min of meer lineair stijgen naarmate het aantal te schatten blokjes toeneemt. Bij gebruikers van de MIN-strategie daarentegen valt een grotere 'fit' met het segmentatiemodel te verwachten; immers, bij hen zullen reactietijd en afwijking stijgen zolang er gebruik gemaakt wordt van de optelprocedure, maar vanaf een bepaald aantal blokjes zal er overgestapt worden op de aftrekprocedure en zullen reactietijden en afwijkingen als gevolg daarvan opnieuw beginnen te dalen. De kernvraag is dus of het segmentatiemodel een betere beschrijving en verklaring biedt voor de reactietijden en afwijkingen dan het lineair model. De maat die we gebruiken om de 'fit' met beide modellen aan te geven is r^2 (of de proportie van de verklaarde variantie).

Bij de interpretatie van de uitkomsten van deze 'fits' dient men echter wel te beseffen dat een hoge r^2 voor het segmentatiemodel nog geen 100% garantie biedt dat het subject de MIN-strategie heeft toegepast; daarnaast zijn ook de geschatte waarden van de verschillende parameters van belang. Vooral met de parameters van de helling, b_1 en b_2 , moet rekening gehouden worden. De MIN-strategie voorspelt immers een positieve b_1 en een negatieve b_2 . Neem bijvoorbeeld het denkbeeldig geval van een subject dat de ADD-strategie toegepast heeft en wiens reactietijden daardoor lineair stijgen tot ze op een bepaald moment (bijv. vanaf het item met 50 blokjes) de limiet van de 20 seconden bereiken en vanaf daar tot aan het item met 100 blokjes rond deze limiet blijven schommelen. In dit geval zal het tweefasig segmentatiemodel een (veel) betere verklaring voor de reactietijden van dit subject opleveren dan het lineair

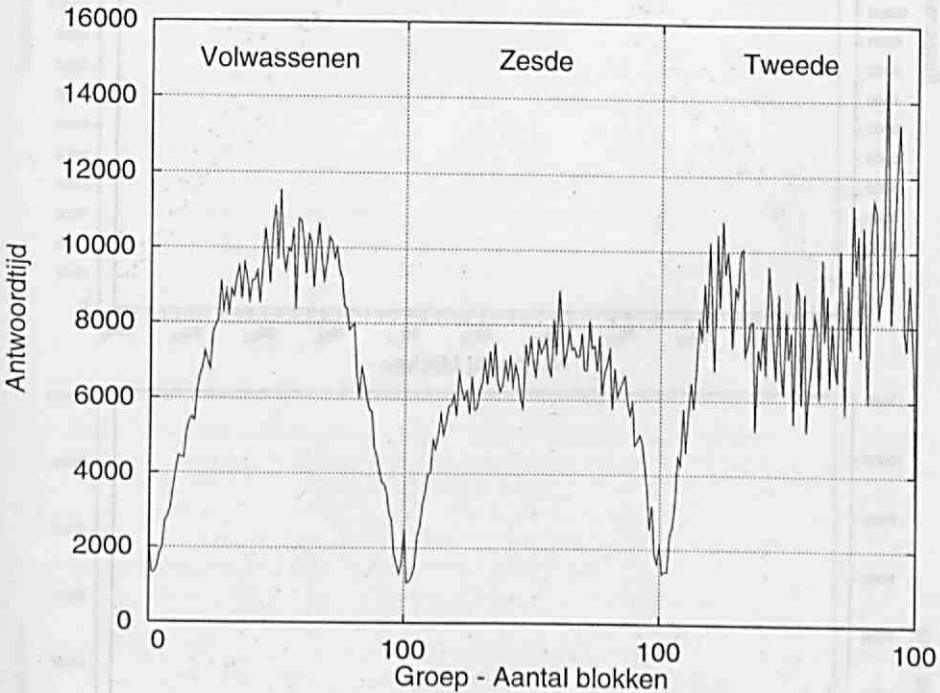
model, doch de afwezigheid van een negatieve b_2 , vormt een zeer sterke indicatie dat de MIN-strategie hier niet toegepast werd. Naast het bekijken van al deze parameters kan het trouwens ook erg nuttig zijn om de (grafieken met de) ruwe data aan een meer globale, kwalitatieve inspectie te onderwerpen.

Hierna volgt de bespreking van de resultaten in verband met de "fits" (r^2) van de reactietijden en de afwijkingen van alle subjecten met het lineair model en met het segmentatiemodel.

Reactietijden

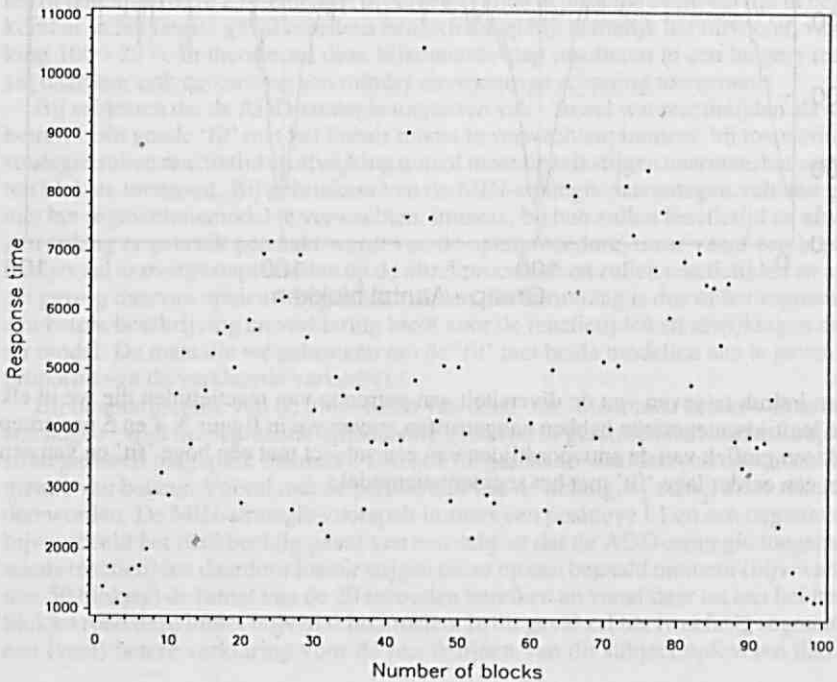
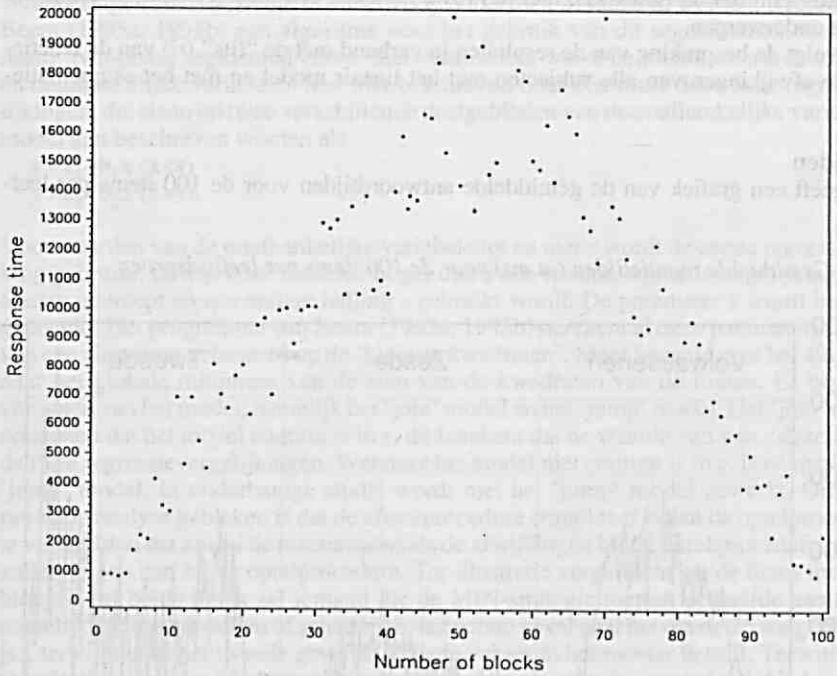
Figuur 2 geeft een grafiek van de gemiddelde antwoordtijden voor de 100 items per leeftijdsgroep.

Figuur 2: Gemiddelde reactietijden (in ms) voor de 100 items per leeftijdsgroep

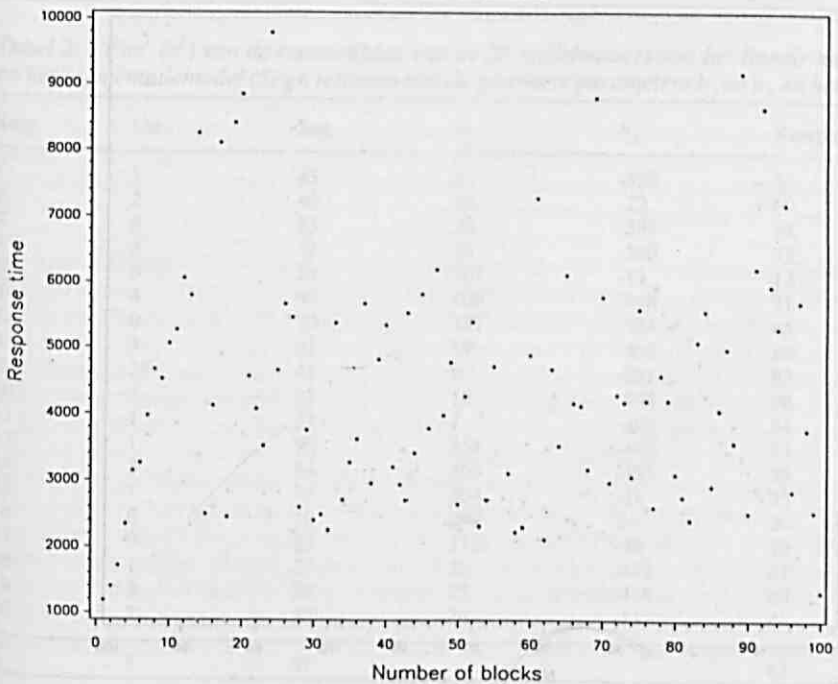
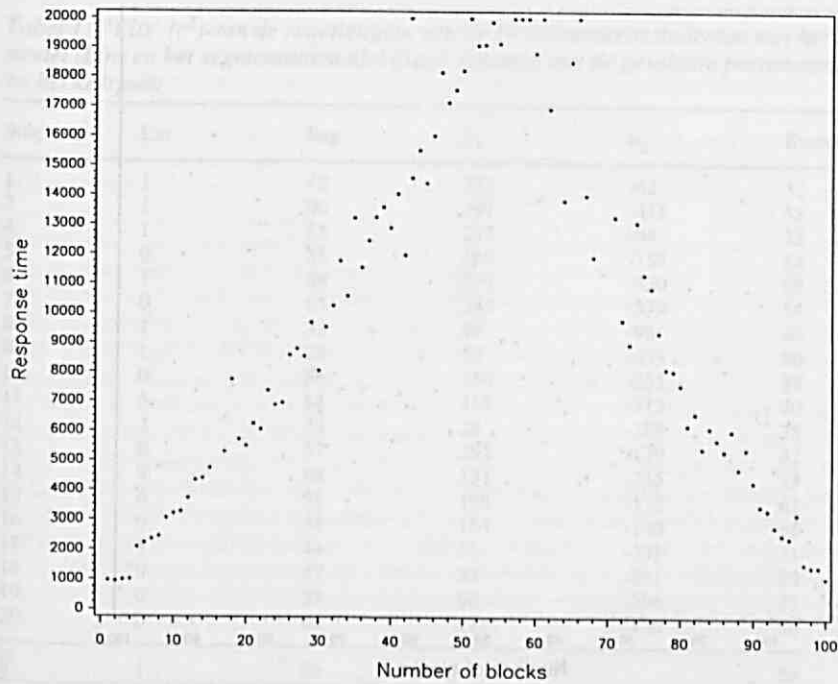


Om ook een indruk te geven van de diversiteit aan patronen van reactietijden die we in elk van de drie leeftijdscategorieën hebben aangetroffen, geven we in figuur 3, 4 en 5 per groep telkens ook een grafiek van de antwoordtijden van een subject met een hoge 'fit' en van een subject met een eerder lage 'fit' met het segmentatiemodel.

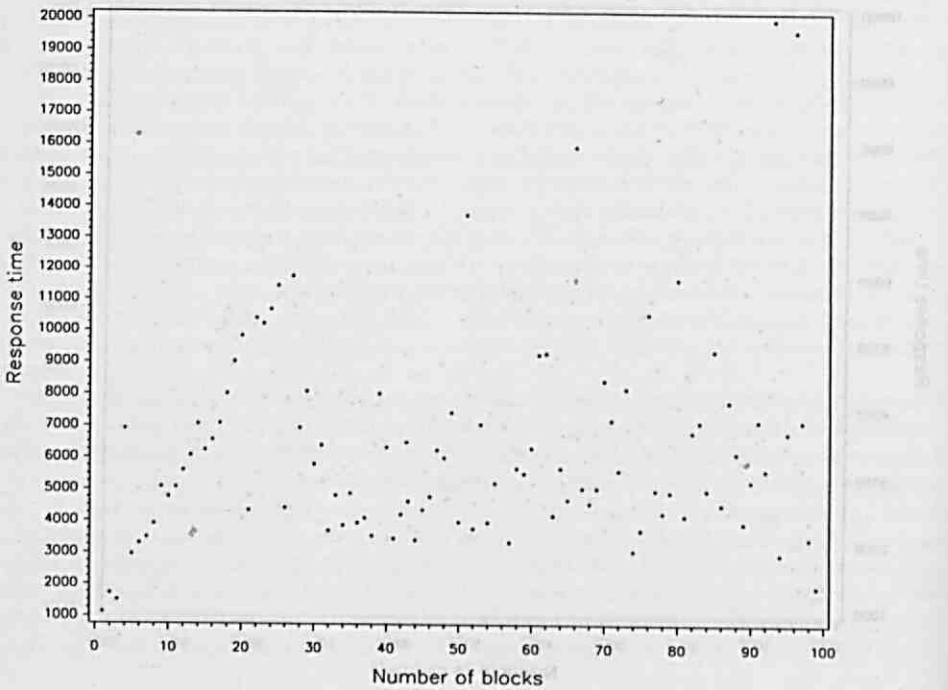
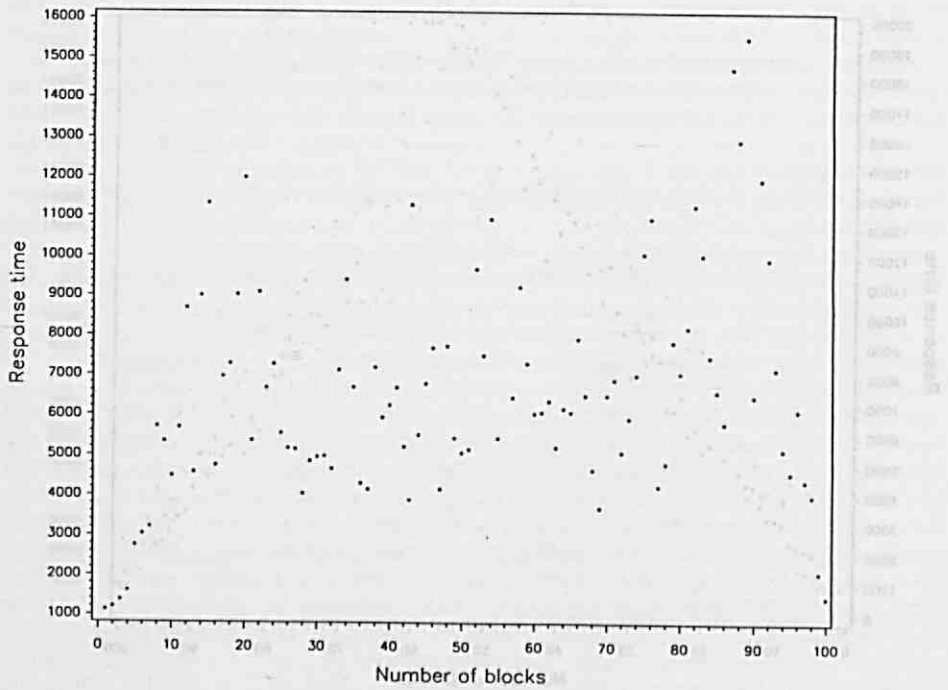
Figuur 3: Reactietijden (in ms) voor (a) subject 3 and (b) subject 12 uit de groep van universiteitsstudenten



Figuur 4: Reactietijden (in ms) voor (a) subject 12 and (b) subject 16 uit de groep van de zesdeklassers



Figuur 5: Reactietijden (in ms) voor (a) subject 4 and (b) subject 1 uit de groep van de tweedeklassers



De belangrijkste statistische gegevens in verband met de 'fit' van de reactietijden van alle subjecten met het lineair model en het segmentatiemodel zijn te vinden in tabel 1, 2 en 3.

Tabel 1: 'Fits' (r^2) van de reactietijden van de 19 universiteitsstudenten met het lineair model (Lin) en het segmentatiemodel (Seg), tezamen met de geschatte parameters b_1 en b_2 en het keerpunt

Subj	Lin	Seg	b_1	b_2	Keerpunt
1	1	42	234	-42	32
3	1	90	297	-411	58
4	1	55	297	-94	32
5	0	55	166	-158	54
6	1	88	379	-520	58
7	0	85	343	-379	54
8	1	32	59	-98	46
9	1	22	10	-339	86
10	0	86	356	-551	59
11	5	84	215	-313	70
12	2	33	28	-286	75
13	0	57	292	-129	42
14	2	68	131	-355	73
15	0	51	168	-172	61
16	0	38	164	-140	51
17	1	44	55	-331	71
18	0	47	93	-391	70
19	0	33	60	-306	77
20	0	62	173	-156	50
\bar{X}	1	56			59

Tabel 2: 'Fits' (r^2) van de reactietijden van de 20 zesdeklassers met het lineair model (Lin) en het segmentatiemodel (Seg), tezamen met de geschatte parameters b_1 en b_2 en het keerpunt

Subj	Lin	Seg	b_1	b_2	Keerpunt
1	1	45	52	-425	65
2	2	46	35	-22	60
3	6	25	24	-393	94
4	2	33	74	-580	73
5	0	24	501	-13	12
6	4	55	109	-448	71
7	0	70	187	-344	65
8	3	21	19	-494	86
9	15	41	6	-271	83
10	7	23	18	-494	90
11	1	33	7	-489	84
12	1	96	354	-449	83
14	3	84	366	-493	59
15	1	33	304	-31	35
16	0	18	297	5	20
17	0	32	1133	-24	16
18	1	27	25	-473	81
19	3	22	27	-458	85
20	2	57	76	-81	46
\bar{X}	3	41			63

Tabel 3: 'Fits' (r^2) van de reactietijden van de 10 tweedeklassers met het lineair model (Lin) en het segmentatiemodel (Seg), tezamen met de geschatte parameters b_1 en b_2 en het keerpunt

Subj	Lin	Seg	b_1	b_2	Keerpunt
1	7	39	31	-1006	86
2	13	24	52	-374	72
3	1	31	330	99	49
4	3	20	319	35	27
5	2	16	-1.116	-761	87
6	10	26	47	-1501	87
7	7	15	597	32	19
8	2	39	329	26	32
9	26	46	629	81	32
10	11	18	87	98	68
\bar{X}	8	27			56

Alvorens de resultaten per groep van naderbij te bekijken, beklemtonen we eerst en vooral dat deze 'fit' voor het segmentatiemodel voor elk subject afzonderlijk hoger is dan voor het lineair model. Gebruiken we de procedure voor het toetsen van het bestaan van een keerpunt - zoals dit door Beem (1995b) wordt beschreven - dan stellen we vast dat de 'nulhypothese' - d.w.z. de hypothese dat er geen keerpunt is of, anders gezegd, dat de data eerder beantwoorden aan een lineair model - voor elk subject verworpen kan worden (op het $\alpha = 0.05$ niveau).

Ten tweede neemt de gemiddelde r^2 voor het segmentatiemodel toe met de leeftijd (nl. van 27 voor de tweedeklassers naar 41 voor de zesdeklassers en 56 voor de studenten) terwijl die voor het lineaire model afneemt (van 8 voor de tweedeklassers naar 3 voor de zesdeklassers en 1 voor de studenten). Dit wordt weerspiegeld in figuur 2: terwijl we bij de studenten en - in mindere mate - bij de zesdeklassers het globale patroon van reactietijden terugvinden dat te verwachten valt bij gebruik van de MIN-strategie, ziet dit patroon er bij de tweedeklassers anders uit.

Deze algemene bevindingen liggen in de lijn van de eerste hypothese, nl. dat naarmate de leeftijd van de subjecten toeneemt er ook meer sporen gevonden worden van (systematisch) gebruik van de MIN-strategie. Maar ze suggereren tevens dat ook in de jongste leeftijdsgroep de ADD-strategie niet de dominante strategie is. Hierna worden de resultaten voor de drie groepen meer in detail besproken.

Studenten. De gemiddelde 'fit' van de studenten met het segmentatiemodel bedraagt 56. Zoals uit tabel 1 blijkt, zijn er maar enkele studenten die een zeer hoge 'fit' (van meer dan 80, zoals bijv. subject 3 uit figuur 3a) behalen, maar ook zeer lage 'fits' (d.w.z. van minder dan 50, zoals subject 12 uit figuur 3 b) komen maar weinig voor (zie tabel 1).

Alle b_1 -waarden zijn positief, wat impliceert dat - wat de eerste regressielijn betreft - de voorspelde antwoordtijd stijgt bij toename van het aantal blokjes. Alle b_2 -waarden daarentegen zijn negatief, hetgeen betekent dat vanaf het keerpunt de voorspelde antwoordtijd vermindert naarmate het aantal te schatten blokjes toeneemt.

De gemiddelde waarde van het keerpunt - d.w.z. de plaats waar de eerste regressievergelijking stopt en de tweede begint - is 59 (zie tabel 1). Dat dit keerpunt niet precies op het middelste item (d.w.z. het item met 50 blokjes) gelegen is, maar hoger ligt, klopt met onze rationele taakanalyse. Daaruit is namelijk gebleken dat de aftrekprocedure ingewikkelder is dan de optelprocedure. Op grond daarvan is het niet verwonderlijk dat gebruikers van de MIN-strategie de optelprocedure niet inruilen voor de aftrekprocedure van zodra het aantal blokjes het aantal lege vakjes overschrijdt, maar iets later. Verder valt op dat de meeste individuele

keerpunten relatief dicht bij dit gemiddelde liggen; zowel keerpunten kleiner dan 50 als keerpunten groter dan 80 komen bij de studenten vrijwel niet voor.

Uit een aanvullende, kwalitatieve inspectie van de grafieken van alle studenten (waarvan er in figuur 3 twee voorbeelden gegeven worden) blijkt dat vooral bij de items met (zeer) weinig en met (zeer) veel blokjes reactietijden behaald worden die goed stroken met het patroon dat bij gebruikmaking van de MIN-strategie te verwachten valt; bij de items uit de middelste regio daarentegen is de overeenstemming met de voorspellingen met de MIN-strategie vaak veel minder groot. Hoewel de data uit deze regio vaak nogal chaotisch zijn, valt het toch op dat een beduidend aantal items uit deze middelste regio veel lagere reactietijden uitlokken dan bij de MIN-strategie (of de ADD-strategie) te voorspellen valt (zie subject 12 uit figuur 3b).

Zesdeklassers. De resultaten van de zesdeklassers lijken goed op die van de studenten, hoewel de gemiddelde r^2 voor het segmentatiemodel (41) zoals gezegd beduidend lager ligt en de gemiddelde r^2 voor het lineair model (3) iets hoger dan in de oudste leeftijdscategorie (zie tabel 2). In vergelijking met de studenten waren er ook minder zesdeklassers met een zeer hoge 'fit' (van meer dan 80) (zoals subject 12 uit figuur 4a), en meer zesdeklassers met een eerder lage 'fit' (van minder dan 50) (zoals subject 16 in figuur 4b).

Verder zijn sommige b_2 -waarden erg laag en voor één zesdeklasser (nl. subject 16) is deze waarde zelfs positief. Dit wijst op een minder sterke daling van de reactietijden voorbij het keerpunt bij de leerlingen uit deze leeftijdsgroep.

Het gemiddelde van het keerpunt is 63, hetgeen hoger is dan bij de studenten (59). Dit suggereert dat zesdeklassers langer gebruikmaken van de optelprocedure en minder snel overstappen op aftrekprocedure.

Uit de individuele grafieken met reactietijden blijkt dat er bij de zesdeklassers nog vaker dan bij de studenten sprake was van erg chaotische data voor de items in de middelste regio (zie bijv. de grafiek van subject 16 in figuur 4b).

Samengevat: hoewel de resultaten van de zesdeklassers in elk geval beter overeenstemmen met de MIN-strategie dan met de ADD-strategie, is deze overeenkomst al bij al toch geringer dan bij de studenten.

Tweedeklassers. Het patroon van de antwoorttijden van de tweedeklassers verschilde sterk van dat van de twee overige groepen (zie figuur 2). In vergelijking daarmee is de gemiddelde r^2 voor het segmentatiemodel relatief laag (27), terwijl de r^2 voor het lineair model eerder hoog is (8). Bij geen enkele leerling stijgt de 'fit' met dit laatste model uit tot boven de 50.

Slechts vier van de tien tweedeklassers (nr. 1, 2, 5 en 6) hadden een negatieve b_2 -waarde. Dit betekent dat enkel bij deze vier leerlingen sprake was van een daling van de reactietijden na het keerpunt. Bij drie van hen (nr. 1, 5 en 6) lag dit keerpunt overigens erg hoog. Dit suggereert dat de tweedeklassers die (ook) gebruikmaakten van de aftrekprocedure, daar pas toe overgingen wanneer het rooster vrijwel geheel gevuld was met blokjes. De grafiek van subject 1, die in figuur 5a gegeven wordt, is typisch voor het patroon van reactietijden bij deze vier tweedeklassers.

De zes overige tweedeklassers hadden over het algemeen lage positieve b_2 -waarden én een zeer laag keerpunt. De grafieken van de reactietijden van deze leerlingen laten zien hoe dit komt: bij de items met (zeer) kleine aantallen stijgt de reactietijd min of meer lineair naarmate het aantal blokjes toeneemt; vanaf een bepaald punt echter neemt de reactietijd niet verder toe en is er intengendeel zelfs sprake van een forse terugval. Bij sommige leerlingen blijft dit zo voor de rest van de opgaven (zoals bij subject 4 uit figuur 5b), maar bij anderen vinden we helemaal op het einde van de reeks van 100 items nog een kleine serie reactietijden waarin een lineair dalend patroon te herkennen is. Deze reeks is echter te klein - in vergelijking met de rest van de reactietijden - om in de regressie-analyse tot uiting te kunnen komen. Men zou deze gegevens als volgt kunnen interpreteren: bij de items met zeer weinig blokjes wordt de optelprocedure toegepast, al betreft het hier eerder *precies* tellen dan *schattend* optellen;

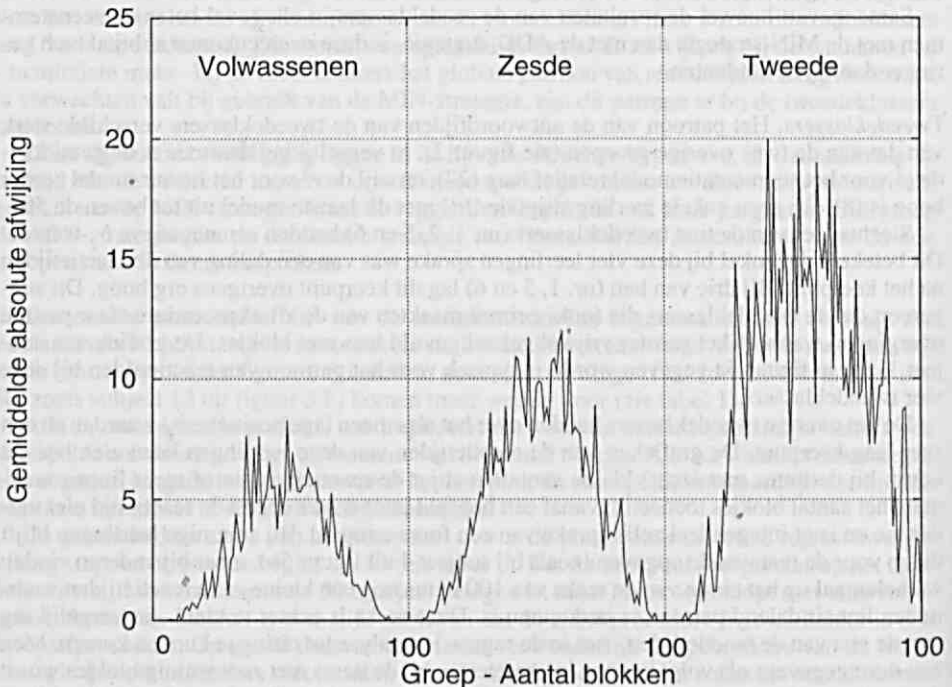
zodra het echter onmogelijk wordt om alle blokjes binnen de toegestane tijd (precies) te tellen, valt men terug op een soort van snelle, oppervlakkige schatstrategie waarvan de tijdsduur niet langer afhankelijk is van het aantal blokjes; enkel bij de items waarbij het rooster vrijwel compleet opgevuld is met blokjes, lijken sommige leerlingen over te schakelen op de aftrekprocedure (of de 'exact tellen' variant daarvan).

Samengevat: de reactietijdpatronen van de tweedeklassers bevatten globaal genomen aanzienlijk minder evidentie voor het (systematisch) gebruik van de MIN-strategie dan die van de twee overige leeftijdsgroepen. Als er sporen van de aftrekprocedure teruggevonden worden, blijven die in elk geval beperkt tot die items die het gebruik van deze strategie erg sterk uitlokken. Daarbij mag men echter niet vergeten dat de 'fit' met het segmentatiemodel al bij al nog beduidend hoger was dan met het lineair model. Dit betekent dat de ADD-strategie evenmin een goede verklaring biedt voor de manier waarop de kinderen uit de jongste leeftijdsgroep deze schatopdracht hebben aangepakt.

Patronen van de afwijkingen voor de verschillende groepen

De ADD- en de MIN-strategie voorspellen ook een sterk verschillend patroon van afwijkingen - d.w.z. (absolute) verschillen tussen het aantal aangeboden blokjes en de schatting daarvan door het subject - voor de 100 items. Terwijl bij toepassing van de ADD-strategie te verwachten valt dat deze afwijkingen alsmear zullen toenemen naarmate het aantal blokjes stijgt, zullen bij de MIN-strategie de grootste afwijkingen te vinden zijn in de middelste regio. Daarom berekend we de 'fit' van de (absolute) afwijkingen met het lineair model en met het segmentatiemodel op dezelfde manier voor de reactietijden gebeurde.

Figuur 6: Gemiddelde (absolute) afwijkingen voor de 100 items per leeftijdsgroep



Figuur 6 bevat een grafiek van de gemiddelde (absolute) afwijkingen bij alle 100 items voor elk van de drie groepen. In tegenstelling tot wat gevonden werd bij de reactietijden (zie figuur 2), verschillen deze patronen van afwijkingen in de drie groepen niet zo erg van elkaar: telkens verschijnt een patroon dat in hoge mate overeenkomt met wat te verwachten valt bij de MIN-strategie, nl. geen of kleine afwijkingen voor de items met (zeer) kleine en (zeer) grote aantallen blokjes en relatief grote(re) afwijkingen voor de aantallen uit de middelste regio. Figuur 6 laat echter ook zien dat - ondanks de gelijkenissen in het globale patroon van de drie groepen - de grootte van de afwijkingen sterk verschilt van groep tot groep: terwijl de curve bij de studenten stijgt tot een waarde van ± 5 (voor de items uit de middelste regio), stijgt die bij de zesdeklassers tot ± 10 en bij de tweedeklassers tot 15 à 20.

De continuïteit in het patroon van de afwijkingen van de drie groepen komt ook in de statistieken tot uiting (voor een gedetailleerde bespreking van deze statistieken zie Dhert, 1997). Het segmentatiemodel leverde telkens een veel betere 'fit' op dan het lineair model: voor de studenten, de zesde- en de tweedeklassers was de gemiddelde 'fit' met het segmentatiemodel respectievelijk gelijk aan 41, 45 en 40% tegenover 0, 0 en 7% voor het lineair model. Het gemiddeld keerpunt was in de drie groepen respectievelijk gelegen op 52, 52 en 59. Verder waren alle b1-waarden positief en alle b2-waarden negatief, op één uitzondering (een tweedeklasser) na.

Samengevat: de patronen van afwijkingen bevestigen de conclusie die we op het einde van de bespreking van de reactietijden hebben gemaakt, nl. dat de MIN-strategie de wijze waarop de schattaak in alle groepen werd aangepakt veel beter beschrijft dan de ADD-strategie. Verrassend is evenwel dat de daling van de afwijkingen in alle groepen zo vroeg begint (nl. bij de items met 50 à 60 blokjes) en zich in alle groepen zo sterk doorzet. Dit lijkt er immers op te wijzen dat ook in de twee jongste leeftijdsgroepen meer gebruik gemaakt is van 100 als referentiepunt dan uit het geheel van de daarvoor besproken resultaten was gebleken. In het tweede resultaatdeel worden onderzoeksgegevens vermeld op basis van de zelfrapporteringen, de systematische schatfouten en de zgn. 'voltreffers' in de drie leeftijdsgroepen, die bijkomend licht werpen op de ontwikkeling van de adaptieve MIN-strategie.

RESULTATEN GEBASEERD OP ANALYSE VAN ZELFRAPPORTERINGEN, SYSTEMATISCHE SCHATFOUTEN EN 'VOLTREFFERS'

Zelfrapporteringen. Zoals vermeld, werden de subjecten achteraf gevraagd om verslag uit te brengen over hun aanpakstrategie. Steunend op deze zelfrapporteringsgegevens werden de subjecten ingedeeld in één van de volgende drie categorieën: (a) subjecten die spontaan verklaarden dat ze de MIN-strategie hadden toegepast; (b) subjecten die pas naar aanleiding van de confrontatie met de drie voorbeelditems vermeldden dat ze bij sommige opgaven de optelprocedure hadden toegepast en bij andere de aftrekprocedure; (c) subjecten die noch spontaan, noch naar aanleiding van de voorbeelditems op enigerlei wijze melding maakten van de MIN-strategie.

Uit de analyse van deze zelfrapporteringen bleek dat bijna alle subjecten beweerden de MIN-strategie (ten minste in een aantal gevallen) te hebben gebruikt. Dertien van de 20 (universiteits)studenten antwoordden spontaan dat ze de MIN-strategie hadden toegepast en bij zes andere kwam dit tot uiting in hun reactie op de voorbeelditems; in de uitleg van één student was van de MIN-strategie geen spoor te bekennen. De ene helft van de zesdeklassers verklaarde spontaan dat ze de MIN-strategie hadden gebruikt, en bij de andere helft bleek dit uit hun reactie op de voorbeelditems. En zelfs bij de tweedeklassers bevatten negen van de tien interviews aanwijzingen voor het gebruik van de MIN-strategie: drie kinderen rapporteerden het gebruik van deze strategie spontaan bij het begin van het interview en bij zes andere kwam dit tot uiting in hun verklaringen naar aanleiding van de voorbeelditems. Hoewel we ons terdege bewust zijn van de methodologische problemen rond de volledigheid en de

betrouwbaarheid van dit soort van retrospectieve verslagen (Ericsson & Simon, 1984), vormen ze toch een bijkomende indicatie dat reeds op 7- à 8-jarige leeftijd bij het schatten van hoeveelheden niet op een vaste, uniforme manier tewerkgegaan wordt, maar dat daarbij - al dan niet bewust - adaptief gebruikgemaakt wordt van verschillende procedures afhankelijk van bepaalde kenmerken van de schattaak. De geconstateerde toename van het aantal *spontane* zelfrapporteringen van toepassing van de MIN-strategie in functie van de leeftijd, vormt een bijkomende aanwijzing dat oudere subjecten deze strategie vaker en bewuster gebruiken dan jongere subjecten.

Systematische schatfouten. Een rationele taakanalyse suggereert dat de MIN-strategie kan leiden tot systematische schatfouten die bij de ADD-strategie niet zullen optreden. We denken daarbij vooral aan de volgende twee soorten fouten:

- (1) fouten als gevolg van een verkeerde bepaling van het totaal aantal hokjes in het rooster (bijv. $12 \times 12 = 144$ in plaats van $10 \times 10 = 100$), wat aanleiding geeft tot systematische over- of onderschattingen van het aantal blokjes bij alle items waarvoor beroep gedaan wordt op de aftrekprocedure;
- (2) fouten die voortvloeien uit het feit dat men, éénmaal een schatting gemaakt, vergeten is of zij betrekking heeft op het aantal blokjes of op het aantal lege hokjes, wat kan resulteren in foutieve schattingen die meer in de nabijheid liggen van het aantal lege hokjes dan van het aantal gekleurde blokjes (en die we daarom 'complementaire fouten' noemen).

Een gedetailleerde analyse van de antwoorden op de 100 beurten in de drie leeftijdsgroepen toonde aan dat beide soorten systematische fouten inderdaad voorkwamen, en dit vooral bij de (universiteits)studenten.

De eerste systematische fout werd teruggevonden bij één student. Zoals bij de meeste andere subjecten uit deze leeftijdsgroep, nam de afwijking van zijn schattingen ten opzichte van het exact aantal blokjes vanaf het item met één blokje tot het item met 70 blokjes (zeer) geleidelijk toe. Maar vanaf dat steeg deze afwijking plots op spectaculaire wijze tot 70, en dit bleef zo voor bijna alle overige items. De meest voor de hand liggende verklaring voor dit merkwaardig patroon van afwijkingen is dat deze student om de één of andere (onverklaarbare) reden een verkeerde schatting van de grootte van het rooster had gemaakt (wellicht $13 \times 13 = 169$ in plaats van $10 \times 10 = 100$) en van deze verkeerde schatting van de roostergrootte gebruikgemaakt had telkens wanneer hij de aftrekprocedure had toegepast (d.w.z. voor alle items vanaf 70 tot 100 blokjes). Noch bij de zesdeklassers, noch bij de derdeklassers vonden wij een spoor van deze systematische schatfout terug.

De tweede foutencategorie kwam vaker voor. Deze ontstaat zoals gezegd wanneer een gebruiker van de MIN-strategie aan het eind van zijn schatting vergeten is of deze schatting nu deel uitmaakte van een optel- dan wel van een aftrekprocedure, en daardoor ofwel nalaat om het geschatte aantal lege hokjes af te trekken van het totaal aantal hokjes (in het geval van een aftrekprocedure), ofwel het aantal geschatte blokjes verkeerdelijk aftrekt van het totale aantal (wanneer het een optelprocedure betrof). Deze vergetelheid resulteert in een eindresultaat dat (veel) dichter bij het aantal lege vakjes ligt dan bij het aantal blokjes. Bij de operationalisering van dergelijke 'complementaire' fouten hanteerden we de volgende twee criteria:

- (a) de schatting wijkt ten minste twee standaarddeviaties af van de gemiddelde afwijking van een subject voor de 100 items én
- (b) de schatting voldoet bovendien aan de 'complementariteitsregel', die bepaalt dat de som van deze schatting en het werkelijk aantal blokjes ongeveer gelijk is aan (meer bepaald: 100 ± 5) (noot 2).

Toepassing van deze criteria leidde tot een totaal van 36 (= 1.9 %) complementaire fouten voor de studenten, 28 (= 1.4 %) voor de zesdeklassers en 6 (= 0.6 %) voor de tweedeklassers. Een ANOVA toonde aan dat het leeftijdseffect significant was, $F(2,46) = 11.01$; $M_s_e = 1.67$; $p < .0462$. Een bijkomende Tukey test ($\alpha = 0.05$) toonde aan dat de studenten significant meer fouten van dit type maakten dan de tweedeklassers.

Samenvattend kunnen we stellen dat ook de analyse van de systematische fouten aanwijzingen bevat voor het gebruik van de MIN-strategie in de drie leeftijdsgroepen, evenals voor het stijgend gebruik ervan afhankelijk van de leeftijd.

'*Voltreffers*'. Er werd ook gekeken naar het voorkomen van 'voltreffers', d.w.z. schatuitkomsten die perfect samenvallen met het werkelijk aantal blokjes. We kunnen ervan uitgaan dat dergelijke 'voltreffers' (meestal) tot stand gekomen zijn door middel van exact tellen, eerder dan via een benaderende schatting. De rationele taakanalyse laat vermoeden dat gebruikers van de ADD-strategie dergelijke 'voltreffers' vooral zullen realiseren bij items met (heel) weinig blokjes, terwijl bij gebruikers van de MIN-strategie dergelijke 'voltreffers' niet enkel zullen voorkomen bij items met (zeer) weinig blokjes maar eveneens bij items met (zeer) veel blokjes. Daarom voorspelden we dat het verschil in 'voltreffers' tussen beurten met kleine en met grote aantallen zou afnemen naarmate de leeftijd van de subjecten stijgt.

Om deze voorspelling te kunnen toetsen, telden we voor elk subject (a) het aantal *opeenvolgende* 'voltreffers' te beginnen met het item met één blokje en ook (b) het aantal *opeenvolgende* 'voltreffers' vanaf het item met 100 blokjes. Wanneer een subject bijvoorbeeld bij de items met één, twee, drie en vier blokjes telkens het juiste aantal als antwoord gaf, maar bij het item met vijf blokjes het getal '6' intikte, kreeg dit subject voor de 'kleine aantallen' een score van 4; wanneer ditzelfde subject met 'voltreffers' reageerde op de items met 100 en met 99 blokjes, maar op het item met 98 blokjes het getal '95' intikte, kreeg het voor 'grote aantallen' een score van 2.

Bij 'kleine aantallen' was het gemiddeld aantal opeenvolgende exacte schattingen voor de studenten, de zesdeklassers en de tweedeklassers respectievelijk gelijk aan 16.68, 9.25 en 13.3. Voor 'grote aantallen' was het gemiddeld aantal opeenvolgende 'voltreffers' voor deze drie groepen respectievelijk gelijk aan 14.05, 9.55 en 3.6. Een vergelijking van het aantal opeenvolgende 'voltreffers' bij items met een klein en met een groot aantal blokjes, liet zien dat noch voor de studenten noch voor de zesdeklassers de lengte van deze reeks bij de kleine en de grote aantallen significant van elkaar verschilden (16.68 versus 14.05 voor de studenten en 9.25 versus 9.55 voor de zesdeklassers). Voor de tweedeklassers daarentegen was dit verschil (13.3 versus 3.6) wel significant [$t(9) = -3.1978; p < .0109$].

Ook deze resultaten bieden dus bijkomende steun voor het toenemend gebruik van de MIN-strategie naarmate de leeftijd vordert. Maar het feit dat ook de tweedeklassers gemiddeld genomen meer dan drie opeenvolgende 'voltreffers' hadden bij de 'grote aantallen' suggereert dat ook bij hen de MIN-strategie zeker niet geheel afwezig was.

De bevindingen uit het eerste en het tweede resultatenstuk samenvattend kunnen we stellen dat zij over het algemeen genomen overtuigende steun verlenen aan de hypothese dat het gebruik van de MIN-strategie toeneemt met de leeftijd; enkel in het patroon van de afwijkingen waren geen sterke aanwijzingen te vinden van het stijgend gebruik van de MIN-strategie met de leeftijd. Ook voor het tweede luik van de eerste hypothese, nl. dat het gebruik van de ADD-strategie zal afnemen met de leeftijd, werden meerdere aanwijzingen gevonden, maar het meest opvallende onderzoeksresultaat in dat verband was toch de verrassend geringe 'fit' van de data met de voorspellingen vanuit de ADD-strategie, zelfs in de jongste leeftijdsgroep.

EFFECTIVITEIT VAN DE MIN-STRATEGIE

Om de tweede onderzoekshypothese (over de samenhang van het gebruik van de MIN-strategie met de nauwkeurigheid van de schattingen) te kunnen toetsen, berekenden we voor de drie leeftijdsgroepen bijzonderlijk correlaties tussen de 'fit' van de reactietijden en de afwijkingen van de subjecten met het segmentatiemodel enerzijds en de volgende twee maten van nauwkeurigheid van hun schatprestaties anderzijds: (a) het gemiddelde van alle

(absolute) afwijkingen ten opzichte van het exacte aantal blokjes per subject en (b) het aantal schattingen dat minder dan 10 % afwijkt van het exact aantal blokjes. In feite beschouwen wij deze laatste maat als de belangrijkste, omdat bij de introductie op de schattaak dit criterium van 10 % toegestane afwijking ook gehanteerd werd om feedback te geven over de kwaliteit van de schattingen bij de 10 voorbeelditems.

Steunend op de veronderstelling dat toepassing van de MIN-strategie de sleutel vormt tot het vaardig aanpakken van de schattaak waarmee in onderhavige studie is gewerkt, verwachtten we (a) een *negatieve* correlatie tussen de 'fit' met het segmentatiemodel en de eerste bovenvermelde maat van nauwkeurigheid en (b) een *positieve* correlatie met de tweede maat van nauwkeurigheid.

Reactietijden. Wat de reactietijden betreft liggen de correlaties van de twee oudste leeftijdsgroepen duidelijk in de lijn van de verwachtingen: bij de studenten en de zesdeklassers waren de correlaties met de eerste maat voor de kwaliteit van de schatprestaties respectievelijk gelijk aan $-.76$ en $-.78$, terwijl voor de tweede maat positieve correlaties van $.77$ en $.81$ gevonden werden. Al deze correlaties zijn significant ($p < .05$). Bij de tweedeklassers daarentegen werd geen significante correlatie gevonden ($.12$ voor de eerste maat en $.25$ voor de tweede).

Afwijkingen. Bij de afwijkingen verschijnt hetzelfde patroon. De correlatie tussen de 'fits' met het segmentatiemodel en de eerste maat van schatvaardigheid is $-.50$ voor de volwassenen en $-.58$ voor de zesdeklassers en bij de tweede maat bedraagt de correlatie respectievelijk $.47$ en $.49$. Al deze correlaties zijn significant ($p < .05$). Ook hier werden geen significante correlaties gevonden bij de tweedeklassers ($-.11$ voor de eerste maat en $.52$ voor de tweede maat).

Samengevat is met betrekking tot de tweede onderzoekshypothese vastgesteld dat er bij de studenten en de zesdeklassers inderdaad een sterke samenhang bestaat tussen de mate waarin men gebruik maakt van de MIN-strategie enerzijds en de schatprestaties anderzijds. Bij de tweedeklassers daarentegen werd er geen verband gevonden. Dit laatste is niet zo verwonderlijk, rekening houdend met het relatief gering gebruik van deze strategie in deze leeftijdsgroep.

DISCUSSIE

Een rationele taakanalyse laat zien dat bij het schatten van aantallen (blokjes) die in een vast en vierkantig rooster worden aangeboden, het best gebruikgemaakt wordt van een adaptieve aanpakstrategie. Deze strategie bestaat hierin dat men het aantal blokjes de ene keer schat door schattingen van groepen blokjes bij elkaar op te tellen (= optelprocedure) en de andere keer door het (geschatte) aantal lege vakjes in het rooster af te trekken van het (geschatte) totaal aantal hokjes (= aftrekprocedure), afhankelijk van welk van beide procedures het cognitief systeem het minst belast (vandaar de naam MIN-strategie).

In onderhavig onderzoek werd de ontwikkeling van deze adaptieve aanpakstrategie empirisch bestudeerd. Aan deze studie participeerden 50 subjecten: 20 (universiteits)studenten, 20 zesdeklassers en 10 tweedeklassers. Hoe zij deze schattaken aanpakten, werd achterhaald via de analyse en triangulatie van diverse soorten van onderzoeksgegevens, nl. hun antwoorden en reactietijden voor 100 schatopgaven en hun zelfrapporteringen achteraf over de manier waarop ze deze opgaven hadden aangepakt. Twee hypothesen lagen aan de basis van deze studie. Ten eerste voorspelden we dat het gebruik van de ADD-strategie zou afnemen en het gebruik van de MIN-strategie zou toenemen met de leeftijd en ten tweede verwachtten we een positief verband tussen de mate van het gebruik van de MIN-strategie en de schatprestaties.

Wat de eerste hypothese betreft, blijkt uit de resultaten dat het gebruik van de adaptieve MIN-strategie inderdaad stijgt met de leeftijd. Dit blijkt zowel uit de analyse van de individuele reactietijdpatronen met behulp van een computerprogramma dat door Beem (1995a, 1995b) speciaal ontworpen werd voor het detecteren van strategieveranderingen bij cognitieve taken, als uit de bijkomende analyses gebaseerd op de zelfrapporteringen, de systematische fouten en de 'voltreffers' in de drie leeftijdsgroepen. Alleen in het patroon van de afwijkingen waren geen duidelijke aanwijzingen te vinden van het stijgend gebruik van de MIN-strategie in functie van de leeftijd. Hoewel de tweedeklassers dus het minst blijken te geven van het gebruik van de MIN-strategie, vonden we bij hen toch ook reeds de eerste sporen daarvan. Doch het gebruik van de aftrekprocedure bij deze leerlingen bleef beperkt tot die items waarbij het rooster vrijwel geheel gevuld was. In termen van Lemaire en Siegler's (1995) model van 'strategic change' is het verschil tussen de diverse leeftijdsgroepen dus niet zozeer een kwestie van het gebruik van nieuwe procedures, maar van het frequenter, vlotter én adaptiever gebruik van de meest efficiënte procedure, m.n. de aftrekprocedure. Hoe het komt dat (vooral) de jonge kinderen de aftrekprocedure niet frequenter gebruikten, kan op basis van de beschikbare data niet worden uitgemaakt. Dit vergt diepgaander onderzoek, waarin tevens systematische informatie wordt ingewonnen over de ontwikkeling bij de subjecten van allerlei conceptuele, procedurele en strategische kennisaspecten die volgens onze rationele taakanalyse aan de MIN-strategie ten grondslag liggen. Voor het andere luik van de eerste hypothese, nl. dat het gebruik van de ADD-strategie zal afnemen met de leeftijd, werden eveneens meerdere aanwijzingen gevonden, maar het meest opvallende onderzoeksresultaat in dat verband was toch de verrassend geringe 'fit' van de data met de voorspellingen vanuit de ADD-strategie, zelfs in de jongste leeftijdsgroep. Het is dus niet zo dat jonge subjecten de schat-taak aanpakken via de inefficiënte ADD-strategie en deze strategie geleidelijk inruilen voor de meer efficiënte MIN-strategie (zie verder).

Wat de tweede hypothese betreft, vonden we zowel bij de studenten als bij de zesdeklassers de verwachte (sterke) samenhang tussen de mate waarin er door de subjecten gebruikgemaakt werd van de MIN-strategie enerzijds en de kwaliteit van hun schatuitkomsten anderzijds. Dit wijst erop dat het flexibel en functioneel afwisselen van verschillende procedures inderdaad een belangrijke eigenschap van goede schatters is. Ook dit stemt overeen met wat gevonden werd in andere studies over de relatie tussen adaptief strategiegebruik en leerprestaties (zie bijv. Geary & Burlingham-Dubree, 1989). Dat deze samenhang bij de tweedeklassers niet werd gevonden, heeft wellicht te maken met het relatief gering gebruik van de MIN-strategie in deze leeftijdsgroep.

Samengevat kunnen we stellen dat de resultaten van onderhavige studie globaal genomen in de lijn liggen van de bevindingen van Newman en Berger (1984) bij een verwante schat-taak en van andere onderzoekers bij andersoortige cognitieve taken (bijv. Baroody & Ginsburg, 1986; Geary & Burlingham-Dubree, 1989; Hope & Sherill, 1987; Lefevre, Greenham & Waheed, 1993; Siegler, 1988; Siegler & Jenkins, 1989; Lemaire & Siegler, 1995), nl. dat subjecten adaptief tewerkgaan om de cognitieve inspanning bij (reken)taken te minimaliseren, dat het verwerven van een dispositie tot adaptief strategiegebruik een langzaam proces is dat reeds op jonge leeftijd aanvangt, en dat er een samenhang bestaat tussen de mate van adaptief strategiegebruik enerzijds en de leerprestaties anderzijds. De waarde van onderhavige studie bestaat er vooral in dat zij vanuit een nieuw taakgebied - nl. het schatten van aantallen - bijkomende empirische evidentie aanreikt voor rol van 'strategic change' ter verklaring van inter- en intra-individuele verschillen in prestaties op leertaken. Een ander waardevol punt van deze studie is dat zij laat zien hoe de techniek van segmentatie-analyse een meer precieze en verantwoorde meting mogelijk maakt van de mate waarin en de efficiëntie waarmee er adaptief tewerkgegaan wordt dan in de meeste andere studies op dit terrein het geval is, waardoor ook het verband tussen adaptief strategiegebruik en leerprestaties beter geanalyseerd kan worden.

Er rijzen echter ook nog heel wat vragen van theoretische, methodologische en onder-wijspraktische aard die om verder onderzoek vragen.

Vanuit theoretisch oogpunt wijzen we erop dat verscheidene subjecten wellicht nog andere schatprocedures gehanteerd hebben dan de optel- en de aftrekprocedure waarop de analyse toegespitst was. Zo suggereren de grafieken met reactietijden van nogal wat studenten en zesdeklassers dat zij bij de items met middelmatige aantallen blokjes geregeld een soort van (zeer) snelle, oppervlakkige schatprocedure hanteerden waarop het feitelijk aantal blokjes nauwelijks invloed lijkt te hebben (zie bijv. de grafieken in figuur 3b en 4b). Wellicht gaat het hier om een 'primitieve' manier van schatten waarbij men het aantal blokjes (of het aantal lege hokjes) a.h.w. in één oogopslag ruwweg bepaalt, in plaats van dit aantal stapsgewijs (en daardoor ook preciezer) te achterhalen in het kader van een optel- of aftrekprocedure. Terwijl studenten en zesdeklassers deze primitieve, ruwe schatprocedure vermoedelijk enkel hanteerden bij items uit de middelste regio (en dus niet voor items met weinig of veel blokjes), pasten tweedeklassers deze primitieve, ruwe schatprocedure wellicht veel vaker toe, nl. vanaf het item waarvan ze het preciese aantal blokjes niet langer (precies) tellend konden bepalen binnen de tijdslimiet van 20 sec. tot aan het item waarbij het rooster vrijwel geheel met blokjes was gevuld (zie figuur 5b). Verder onderzoek naar het bestaan van deze hypothetische schatprocedure, de relatie ervan met de optel- en de aftrekprocedure, de wijze waarop zij zich ontwikkelt en de samenhang ervan met schatprestaties is noodzakelijk. Daarbij kan de techniek van oogbewegingsregistratie (in combinatie met andere technieken) wellicht een nuttige rol vervullen (Verschaffel, De Corte, Gielen & Struyf, 1994).

Een volgende theoretisch erg belangrijke vraag is hoe gebruikers van de MIN-strategie bij een bepaald item voor de optel- dan wel voor de aftrekprocedure (of voor nog een andere werkwijze) opteren (zie ook: Siegler & Jenkins, 1989). In onze rationele taakanalyse hebben wij gesteld dat in deze beslissing meerdere elementen een rol spelen, zoals de verhouding van het aantal blokjes tot het totaal aantal hokjes in het rooster en de complexiteit van de procedures waartussen er te kiezen valt. Of beter gezegd: *de schatting door het subject* van de verhouding van het aantal blokjes tot het totaal aantal hokjes in het rooster en *zijn inschatting* van de complexiteit van de procedures waaruit gekozen kan worden. Om beter zicht te krijgen op de wijze waarop en de mate van bewustheid waarmee dit complex beslissingsproces verloopt, is diepgaander onderzoek vereist. Daarin kan wellicht beter gewerkt worden met een uitgebreider en gevarieerder aanbod van schattaken dan in onderhavige studie het geval was. Immers, nu dienden subjecten steeds hoeveelheden te schatten die aangeboden werden in een vast 10 X 10 rooster. Een rationele taakanalyse wijst echter uit dat dit een schattaak is die sterk in het voordeel speelt van de aftrekprocedure. Er zijn echter andere soorten van (verwante) taken te bedenken waarbij te verwachten valt dat (efficiënte) schatters de overstap van de optel- naar de aftrekprocedure langer zullen uitstellen. Neem bijvoorbeeld een reeks items waarvan het rooster geen eenvoudige 10 X 10 structuur heeft, maar een andere vorm aanneemt zoals een andere vierkante structuur (bijv. 8 X 8 of 12 X 12) of - erger nog - een rechthoekige structuur (vb. 7 X 8 of 7 X 13); of denk aan een serie items die niet in een vast rooster aangeboden worden maar in roosters met afwisselende groottes (bijv. de ene keer 10 X 10, de andere keer 9 X 12, nog een andere keer 13 X 13, enz.). Een rationele taakanalyse laat vermoeden dat het adaptief gebruik van de MIN-strategie bij deze schattaken veel grotere eisen stelt aan de conceptuele, procedurele en - vooral - strategische kennis van de schatter dan in onderhavige studie het geval was.

Vanuit theoretisch oogpunt is het ten slotte van groot belang te achterhalen hoe de adaptieve MIN-strategie precies ontstaat, zich verder ontwikkelt en uitbreidt, en welke interne en externe condities daarop van invloed zijn. Om deze vragen te kunnen beantwoorden moet het cross-sectioneel design, waarbij product- en procesgegevens van subjecten uit verschillende leeftijdsgroepen met elkaar vergeleken worden, aangevuld worden met een design waarbij data over schatprestaties en -strategieën van eenzelfde subject op verschillende momenten met elkaar vergeleken worden. In principe zou het mogelijk geweest zijn om de beschikbare onderzoeksgegevens vanuit dit oogpunt nog verder te ontrafelen (bijv. door een vergelijking te maken tussen de individuele schatprestaties en -procedures bij de aanvang, in het midden

en op het einde van het experiment), maar omwille van een aantal aspecten van de huidige proefopzet (zoals het feit dat de subjecten voor het eigenlijke experiment reeds een tiental oefen-beurten gekregen hadden én het feit dat de aanbieding van de 100 schattaken helemaal 'at random' geschiedde), hebben we daarvan afgezien.

Vanuit methodologisch oogpunt staan we eerst stil bij enkele problemen in verband met de wijze waarop de data in onderhavige studie verzameld werden. Zo is er ten eerste het probleem van de tijdslimiet. Zoals vermeld, kregen de subjecten elk item maximaal 20 sec. op het scherm te zien. Met een dergelijke, vaste tijdslimiet wilden we voorkomen dat men de blokjes stevast precies zou tellen in plaats van te schatten. Doch enerzijds bleek dat sommige subjecten bij bepaalde items (uit de middelste regio) meer dan 20 sec. nodig hadden om een (goede) schatting te maken. Daardoor leenden de reactietijdpatronen van deze subjecten zich niet zo goed voor de geplande regressie-analyses. Aan de andere kant was de tijdslimiet van 20 sec. bij andere items (uit de kleinste en grootste regio) zo ruim dat de subjecten de hoeveelheden exact tellend konden bepalen in plaats van via een benaderende schatting. Het lijkt daarom aangewezen om er in verder onderzoek op een andere, meer verfijnde manier voor te zorgen dat de subjecten de taak niet al tellend maar al schattend aanpakken. Een tweede probleem in verband met de gegevensverzameling betreft de wijze waarop de reactietijd bepaald werd. Zoals gezegd werd de subjecten gevraagd om op de spatiebalk te drukken zodra ze hun schatting hadden gemaakt; op dat moment verdween het scherm met het rooster en verscheen er een nieuw scherm waarop het subject z'n schatuitkomst kon intikken. Dit moment waarop op de spatiebalk gedrukt werd werd gebruikt om de reactietijd te bepalen. Doch, tijdens het individueel interview legden enkele studenten uit dat zij vaak reeds op de spatiebalk gedrukt hadden voor hun schatproces helemaal beëindigd was. Meer bepaald beweerden zij dat zij het verschil tussen hun schatting van het aantal lege hokjes en het totaal aantal hokjes ($100 - x = .$) berekenden *nadat* ze op de spatiebalk hadden gedrukt, hetgeen tot een systematische onderschatting van hun reactietijden bij de aftrekprocedure kan hebben geleid. Dit probleem kan voorkomen worden door subjecten hun schat-uitkomsten mondeling te laten geven en de reactietijd op basis daarvan te berekenen.

Vragen van methodologische aard rijzen er ook bij de statistische analyse van de individuele patronen van reactietijden en afwijkingen met behulp van Beem's (1995a, 1995b) segmentatiemodel. Zoals eerder gezegd verdeelden Newman en Berger (1984) het gehele getalengebied waarmee in hun studie gewerkt werd in drie groepen (d.w.z. items met kleine, middelmatige en grote aantallen) en vergeleken zij het gemiddelde voor deze groepen. Het hoofdprobleem met een dergelijke categorische benadering is dat de keuze van het aantal groepen waarin het hele gebied verdeeld wordt arbitrair bepaald wordt en de resultaten dus zullen verschillen naargelang de groepen waarin men het getalengebied indeelt. Dit probleem stelt zich bij regressie-analyse niet. Bovendien gaat Beem's (1995a, 1995b) model zelf op zoek naar het keerpunt in plaats van - zoals Newman en Berger (1984) doen - a priori aan te nemen dat dit zich ergens in het midden van het getallenbereik zal bevinden. Beem's (1995a, 1995b) model is ons inziens ook te verkiezen boven een kwadratisch model, omdat er twee verschillende regressievergelijkingen gevonden kunnen worden die niet per se continu moeten zijn in het keerpunt; het kwadratisch model staat dit niet toe. Hoewel we het segmentatiemodel dus als een belangrijke methodologische aanwinst voor het onderzoek op het gebied van 'adaptive strategy choice' opvatten, stootten we toch ook op een aantal problemen die om verdere uitwerking vragen. Het belangrijkste is dat Beem's (1995a, 1995b) model slechts één keerpunt toelaat, terwijl de inspectie van sommige grafieken van reactietijden suggereert dat er bij de betreffende subjecten geen sprake is van twee maar van drie verschillende procedures (met bijhorende regressievergelijkingen).

Ten slotte gaan we kort in op de onderwijspraktische implicaties van deze studie. In de huidige wereldwijde vernieuwing van het reken/wiskundeonderwijs wordt sterk de klemtoon gelegd op het (leren) schatten en schattend rekenen. Zo komen in de nieuwe leerplannen voor

de lagere en middelbare school doelstellingen voor als: het begrijpen van de (relatieve) grootte van aantallen, het hanteren van betekenisvolle referentiepunten voor discrete en niet discrete hoeveelheden, het beheersen van strategieën voor het tellen en schatten van (grote) aantallen, enz. (National Council of Teachers of Mathematics, 1989; Treffers & De Moor, 1990; Verschaffel & De Corte, 1996). De resultaten van deze studie suggereren dat reeds in de onderbouw van de basisschool kan worden begonnen met het opdoen van ervaringen die gericht zijn op de ontwikkeling van allerlei inzichten, vaardigheden en houdingen die aan de basis liggen van 'number sense' en 'estimation'. Maar ze wijzen ook op de complexe, veelzijdige aard van deze onderdelen van een wiskundige dispositie. Zo hebben we vastgesteld dat het inzichtelijk en efficiënt gebruik van de MIN-strategie allerlei conceptuele en procedurele wiskundige kennis vereist zoals inzicht in de basisbewerkingen, parate kennis van eenvoudige optellingen en aftrekkingen, de 'tafels van vermenigvuldiging', inzicht in de plaatswaarde van getallen, mentale procedures voor het optellen en aftrekken in het getallengebied tot 100, enz. tezamen met metacognitieve kennis en vaardigheden. De laatste jaren zijn al heel wat inspanningen gedaan om instructiemateriaal te ontwikkelen dat bijdraagt tot de ontwikkeling van de 'feeling' voor getallen en het schattend rekenen bij leerlingen van de basisschool (Sowder, 1992; Treffers & De Moor, 1990; Verschaffel & De Corte, 1996). In de nabije toekomst zullen wij de vakdidactische waarde onderzoeken van een computerondersteunde leeromgeving gebaseerd op de schattaak die in deze studie gebruikt is.

NOTEN

1. In de rest van deze tekst zullen wij gemakshalve de term 'procedure' hanteren om te verwijzen naar de verschillende concrete werkwijzen voor een bepaald schatitem (bijv. de optel- en de aftrekprocedure), terwijl de term 'strategie' gereserveerd wordt voor de verschillende globale aanpakwijzen van de schattaak in z'n geheel (nl. de ADD- en de MIN-strategie).
2. Wanneer de schatting drie maal afweek van de standaarddeviatie van het gemiddelde van de fouten van een proefpersoon, werd het tweede criterium minder strikt gehanteerd, ± 5 werd dan vervangen door ± 10 .

REFERENTIES

- Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 75-112). Hillsdale, NJ: Erlbaum.
- Beem, A. L. (1995a). A program for fitting two-phase segmented curve models with an unknown change point, with an application to the analysis of strategy shifts in a cognitive task. *Behavior Research Methods, Instruments, & Computers*, 27, 392-399.
- Beem, A. L. (1995b). *Segcurve: A program for fitting two-phase segmented curve models with an unknown change point* [Program manual]. Leiden, The Netherlands: Leiden University, Centre for the Study of Education and Instruction.
- De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. C. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 491-549). New York: Macmillan.
- Dhert, N. (1997). *Ontwikkeling van een adaptieve strategie voor het schatten van hoeveelheden*. (Niet-gepubliceerde licentiaatsverhandeling.) Leuven: Centrum voor Instructiepsychologie en -Technologie, K.U. Leuven.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Fuson, K. C. (1992). Research on whole number addition and subtraction. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243-275). New York: Macmillan.
- Geary, D. C. & Brown, S. C. (1991). Cognitive addition: strategy choice and speed-of-processing difference in gifted, normal and mathematically disabled children. *Developmental Psychology*, 27, 398-406.
- Geary, D. C., & Burlingham-Dubree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology*, 47, 175-192.

- Hope, J.A., & Sherill, J.M. (1987). Characteristics of unskilled and skilled mental calculators. *Journal for Research in Mathematics Education*, 18, 98-111.
- Ippel, M. J., & Beem, A. L. (1987). A theory of antagonistic strategies. In E. De Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction: European research in an international context* (pp. 111-121). Oxford: Pergamon.
- Lefevre, J.-A., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction*, 11, 95-132.
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124, 83-97.
- National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Newman, R. S., & Berger, C. F. (1984). Children's numerical estimation: Flexibility in the use of counting. *Journal of Educational Psychology*, 76, 55-64.
- Schoenfeld, A. H. (1992). Learning to think mathematically: problem solving, metacognition, and sense-making in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.
- Siegler, R. S. (1988). Individual differences in strategy choices. *Child Development*, 59, 833-851.
- Siegler, R. S., & Jenkins, E. A. (1989). *How do children discover new strategies?* Hillsdale, NJ: Erlbaum.
- Sowder, J. (1992). Estimation and number sense. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371-389). New York: Macmillan.
- Treffers, A., & de Moor, E. (1990). *Proeve van een nationaal programma voor het reken-wiskundeonderwijs op de basisschool*. Deel 2. Basisvaardigheden en cijferen. Tilburg: Zwijssen.
- Verschaffel, L., & De Corte, E. (1996). Number and arithmetic. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education: Part 1* (pp. 99-137). Dordrecht: Kluwer.
- Verschaffel, L., De Corte, E., Gielen, I., & Struyf, E. (1994). Clever rearrangement strategies in children's mental arithmetic: a confrontation of eye-movement data and verbal protocols. In H. Van Luit (Ed.), *Research on mathematics learning and instruction in (special) primary schools* (pp. 153-180). Doetinchem/Rapallo: Graviant.

AUTEURSNOTEN

Lieven Verschaffel is onderzoeksdirecteur bij het Vlaams Fonds voor Wetenschappelijk Onderzoek (F.W.O.).

Dit onderzoek kadert in de Geconcerteerde Onderzoeksactie (G.O.A.) nr. 93/1 van de Katholieke Universiteit Leuven.

De auteurs bedanken Wim Fias voor zijn hulp bij het opstellen van het computerprogramma waarmee de onderzoeksgegevens verzameld en geanalyseerd werden.

Manuscript ontvangen: 10-01-'97

Definitief aanvaard: 01-07-'97

Notities en Commentaren

INDICATIESTELLING ONDERWIJSACHTERSTANDEN: EEN COMMISSIE RAPPORTEERT

G. Wim Meijnen

SCO-Kohnstamm Instituut, Wibautstraat 4, 1091 GM Amsterdam

Schoolloopbanen dienen zich overeenkomstig heersende normen volgens meritocratische principes te ontwikkelen. Prestaties in een voorafgaand onderwijstraject behoren het criterium te zijn voor toegang tot een vervoltraject. Wanneer fysieke, financiële, sociale of culturele belemmeringen de oorzaak zijn van het niet kunnen volgen van bepaalde onderwijstrajecten, wordt dat als een maatschappelijk probleem ervaren. De overheid is reeds in een ver verleden begonnen met het creëren van specifieke, extra middelen vergende, voorzieningen voor kinderen met fysieke handicaps. Scholen voor blinde en dove leerlingen dateren al van voor de eeuwwisseling (Doornbos en Stevens, 1987). Daarna volgt de inrichting van scholen voor kinderen met mentale handicaps zoals een (zeer) lage intelligentie, nu scholen voor (zeer) moeilijk lerende kinderen ((z)mlk) genoemd. Na de Tweede Wereldoorlog zijn er scholen voor kinderen met leer- en opvoedingsmoeilijkheden bijgekomen (lom-scholen).

Het te besteden bedrag per leerling is op mlk- en lomscholen ongeveer drie keer zo hoog als op een reguliere basisschool. De oorzaken van de problemen waarmee deze kinderen worstelen worden in de eerste plaats in het kind zelf gezocht dan wel in de directe leefomgeving (fysieke belemmeringen, downsyndroom, verstoorde gezinsrelaties, etc.). Voordat een kind verwezen wordt naar het Speciaal Onderwijs (SO) wordt het aan allerlei diagnoseprocedures onderworpen om foutieve beslissingen zo veel mogelijk te voorkomen. Psychologische tests voor zowel cognitieve, linguïstische als sociaal-emotionele vaardigheden spelen daarbij een belangrijke rol.

Het Speciaal Onderwijs herbergt inmiddels meer dan 110.000 leerlingen en de overheid doet er alles aan om deze groei te beteugelen: het Weer Samen Naar School (WSNS) beleid. De veronderstelling is dat een groot deel van de huidige MLK- en LOM-leerlingen ook in het reguliere onderwijs kunnen worden opgevangen als daar meer adaptieve vormen van instructie worden toegepast.

In de loop der tijd is ook vast komen te staan dat de maatschappelijke positie van gezinnen samengaat met condities die de opvoeding beïnvloeden. In de hogere sociale strata zijn deze condities zodanig dat ze *gemiddeld* tot betere prestaties in het onderwijs leiden. Opvallend is dat elk stratum zijn eigen normaalverdeling van prestaties kent. In de hoogste strata komen veel hoog- en weinig laagpresteerders voor terwijl dat in de laagste strata omgekeerd is. De gemiddelden nemen per stratum continu toe, zonder een enkele inversie. Zo vond Meijnen bij een groep zesjarige kinderen in de jaren zeventig een gemiddelde van 96.5 voor cognitieve vaardigheden bij de laagste statusgroep en 112.5 bij de hoogste (Meijnen, 1977). De standaarddeviaties waren per sociale laag ongeveer gelijk.

Meer recent laten de resultaten uit het PRIMA-cohort onderzoek eenzelfde beeld zien (tabel 1). De indeling naar sociale strata is niet helemaal vergelijkbaar met die van Meijnen omdat nu voor enkele allochtone groepen specifieke posities zijn gereserveerd, maar de resultaten met betrekking tot de prestaties blijven dezelfde: de gemiddelde taalscores klimmen op met het statusniveau en de standaarddeviaties per stratum ontploen elkaar niet veel. Dat geldt voor alle jaargroepen. Eenzelfde tabel valt te construeren voor het vak rekenen. De verschillen tussen de sociale strata zijn dan weliswaar kleiner maar de lineaire samenhang met de statushiërarchie is ook nu overduidelijk aanwezig.

Tabel 1: Taalscores naar jaargroep en sociaal-economische/culturele herkomst (N=27749)

SES-groepen	groep 2		groep 4		groep 6		groep 8		N
	gem	sd	gem	sd	gem	sd	gem	sd	
1 LBO Tu/Marok	939	28	993	30	1037	25	1078	29	1599
2 LBO ov.allocht	949	29	1009	32	1053	29	1093	35	1092
3 LBO autocht	964	32	1024	34	1064	31	1110	35	8427
4 MBO	973	36	1035	36	1075	32	1123	35	10331
5 HBO/WO	981	33	1043	35	1085	35	1134	35	6300
Ned. gemiddeld	970	35	1030	37	1071	34	1117	37	27749

Legenda:

- SES-categorie 1: beide ouders hebben maximaal een LBO-opleiding en zijn van Turkse of Marokkaanse herkomst;
- SES-categorie 2: beide ouders hebben maximaal een LBO-opleiding en zijn afkomstig uit andere landen dan Nederland, Turkije of Marokko (overige allochtonen);
- SES-categorie 3: beide ouders hebben maximaal een LBO-opleiding en zijn van Nederlandse afkomst;
- SES-categorie 4: de hoogst opgeleide ouder, autochtoon of allochtoon, heeft maximaal een MBO-opleiding;
- SES-categorie 5: de hoogst opgeleide ouder, autochtoon of allochtoon, heeft een opleiding op HBO- of WO-niveau.

Bron: Jungbluth e.a. (1996)

Hoewel niet alle 'kansrijke' leerlingen uit de lagere strata doorstromen naar HAVO of VWO zijn de verschillen in gemiddeld prestatieniveau aan het eind van het basisonderwijs, de belangrijkste oorzaak van de *ondervertegenwoordiging* van leerlingen uit de lagere sociale strata in de bovenstroom van het onderwijsstelsel (HAVO/VWO en HBO/Universiteit). De conclusie die daaruit getrokken kan worden is dat de condities in de lagere strata remmend zijn voor *alle* daarin opgroeiende kinderen. Potentiële VWO-leerlingen worden HAVO-leerlingen en potentiële LBO-leerlingen komen in het Speciaal Onderwijs terecht. Uiteraard ligt aan deze redenering de assumptie ten grondslag dat de werking van 'ongunstige' condities zich niet beperkt tot een eventueel ongunstiger genenpotentieel. Gaat men daar wel vanuit dan valt aan bovenstaande verdeling van gemiddelde prestaties door sociale stimulering uiteraard weinig te veranderen.

De overheid tracht deze *collectief* op te lopen achterstand te pareren door een premieering gebaseerd op basis van groepslidmaatschap (collectieve stimulering). Elk kind uit de lagere strata, ongeacht zijn fenotypisch prestatieprofiel, werd en wordt extra bedeeld in termen van meer formatie (de 'gewichtenregeling') en aan gebieden met een concentratie van deze kinderen werden en worden extra gelden verstrekt op basis van de Onderwijsvoorrangswet. Met de extra formatie en de gebiedsgelden dienen maatregelen te worden getroffen, zowel binnen- als buitenschools, die de onderwijsachterstanden van deze groepen moeten verminderen.

Het Onderwijsvoorrangbeleid (OVb) is dus gebaseerd op de filosofie dat kinderen uit bepaalde strata, *ongeacht hun potentiële talent*, collectief door hun afkomst belemmeringen ondervinden. Het streven is derhalve niet alleen te voorkomen dat veel kinderen uit de laagste strata naar het SO worden verwezen of doubleren maar ook dat meer kinderen uit deze strata het HAVO of het VWO gaan bezoeken.

Het streven naar *zorgverbreding* daarentegen, in het bijzonder als het om potentiële leerlingen voor LOM of MLK gaat (ca. 80% van het speciaal onderwijs) is gericht op leerlingen die het in het reguliere onderwijs niet kunnen bolwerken en (zeer) lage prestaties leveren.

Het Speciaal Onderwijs is als systeem zelfs in het leven geroepen voor kinderen die het reguliere onderwijs niet met vrucht kunnen volgen.

Uiteraard is er in de praktijk een zekere overlap van leerlingen die respectievelijk onder het zorgverbredings- en het OVB-beleid vallen. Zo komen onder MLK-leerlingen disproportioneel veel leerlingen uit de lagere strata voor. Identiek zijn de groepen qua sociale herkomst echter allerminst. De beleidsfilosofieën die achter respectievelijk Zorgverbreding en OVB schuilgaan zijn dan ook totaal verschillend.

Groot was dus de verrassing voor ingewijden toen Staatssecretaris Netelenbos op 28 februari 1996 een Commissie Indicatiestelling Onderwijsachterstanden installeerde met als opdracht na te gaan of er tests zijn te ontwikkelen waarmee op jonge leeftijd al kan worden voorspeld of een leerling achterstand zal oplopen ten opzichte van leeftijdgenoten. Indien dat het geval is, zou kunnen worden overwogen de gewichtengelden niet meer toe te kennen op basis van sociale herkomst maar op basis van de resultaten van een entreetoets. Over de ommezwaai met betrekking tot de vigerende beleidsfilosofie inzake het OVB, werd in de opdracht met geen woord gerept.

De Commissie Indicatiestelling Onderwijsachterstanden, naar haar voorzitter ook wel de Commissie-Kohnstamm genoemd, kwam in december 1996 met haar eindrapport, getiteld: *'Zo onvoorspelbaar als het leven zelf'*. Kort gezegd komt de Commissie tot de conclusie dat het niet goed mogelijk is voor de huidige praktijk, waarbij het aan heldere (tussen)doelen ontbreekt, tests te ontwikkelen die een veel kleinere foutenmarge opleveren bij toedeling van middelen dan de huidige indicatiestelling die gebaseerd is op de opleiding en de etniciteit van de ouders. In het rapport wordt testonderzoek uit Nederland, de Verenigde Staten en het Verenigd Koninkrijk geanalyseerd met betrekking tot de voorspellende waarde. De conclusie is dat veel van de tests redelijke resultaten laten zien maar dat er te veel 'fouten' gemaakt zullen worden om differentieële toewijzing van hulp te rechtvaardigen. Bovendien ontbreekt het aan landelijk geformuleerde tussendoelen aan het einde van de onderbouw zodat een valide criteriummaat ontbreekt. Pas als deze in operationele definities zijn vastgelegd, kan volgens de Commissie onderzoek worden gedaan naar wat de beste indicatoren van zulke latere vaardigheden zijn. Haar centrale aanbeveling vloeit daar dan ook logisch uit voort: het formuleren van tussendoelen voor het eind van groep 3. Deze tussendoelen moeten de minimale vaardigheden aangeven waarover leerlingen moeten beschikken. De tussendoelen zouden als ijkpunt kunnen fungeren voor het vaststellen van achterstanden.

Uit het rapport blijkt overigens dat de Commissie de ommezwaai met betrekking tot de beleidsfilosofie moeiteloos heeft gemaakt. Zo schrijft zij (p. 7): 'Het onderwijsachterstandenbeleid zou erop gericht kunnen worden dat zo veel mogelijk kinderen *ten minste* een algemeen als *voldoende* gedefinieerd ontwikkelingspeil of prestatieniveau bereiken.' Deze opmerking wordt gemaakt nadat de zin van het vergelijken van groepsgemiddelden wordt gekritiseerd omdat evaluatieonderzoek laat zien dat van een inhaalmanoeuvre toch geen sprake is.

Een heldere beleidsanalyse die de oorzaken blootlegt van het veronderstelde falen van de huidige indicatiestelling, ontbreekt helaas geheel. Wanneer dat wel was gebeurd dan was de Commissie waarschijnlijk tot de conclusie gekomen dat de vrijblijvende toekenning van de middelen (geen voorschriften voor besteding, geen verantwoording met sancties) waarschijnlijk de belangrijkste oorzaak is. We stuiten dan op het leerstuk van de vrijheid van onderwijs en hoe de centrale overheid daarmee wil of kan omgaan. De Staatssecretaris omzeilde dit meeromvattende probleem en koos voor een, impliciete, theorie dat als middelen worden toegekend op basis van valide voorspellende tests, deze wel effectief zullen worden aangewend. Vraag: hoe komt ze daarbij? Die vraag hoefde de Commissie niet te beantwoorden. In plaats van zich grondig te vergewissen van de validiteit van deze assumptie, neemt de Commissie deze veronderstelling als werkhypothese over en doet vervolgens wat haar gevraagd is: nagaan of er tests met een voldoende prognosticerend vermogen bestaan. Zoals gezegd, de Commissie komt dan tot de conclusie dat dit bij de huidige onderwijspraktijk niet het geval is en derhalve een indicatiestelling op basis van testuitslagen bij de entree van het basis-

onderwijs niet is aan te bevelen. Praktische bezwaren voorkomen derhalve een alsnog noodzakelijke legitimering van een nieuwe filosofie voor een achterstandenbeleid.

Staan er dan geen aantrekkelijke voorstellen in het rapport, geredeneerd vanuit de filosofie waarop het huidige Onderwijsvoorrrangsbeleid is gebaseerd? Die voorstellen zijn er zeker. Omdat er, zoals gezegd, een overlap bestaat tussen de populatie van OVB-leerlingen en de populatie kinderen die tot de laagste prestatiecategorieën in het basisonderwijs behoren, is het formuleren van tussendoelen, dat wil zeggen doelen voor het eind van groep 3 die *alle* leerlingen moeten beheersen, een goede suggestie. Deze suggestie zou echter ook heel goed los van de indicatiestelling c.q. de toewijzing van extra middelen gedaan kunnen worden. Een dergelijke formule geldt bijvoorbeeld al in het voortgezet onderwijs. Daar bestaat geen gewichtenregeling maar wel het systeem van kerndoelen voor de basisvorming met het oogmerk de gelijkheid van kansen te bevorderen.

Anders gezegd: over middelentoewijzing, centraal geïnitieerd innovatiebeleid en realisatie van gewenste effecten, en in onderlinge samenhang, zijn tal van intelligente analyses te schrijven, ook als het gaat om het bestrijden van achterstandssituaties. Natuurlijk hoeft dit traject niet in elk adviesproject tot in de finesses te worden afgelegd. En uiteraard dient het zwaartepunt van het advies het meest geprofileerd te worden. Van een commissie mag echter wel worden verwacht dat zij duidelijk het speelveld markeert waarop zij wenst te spelen en waarbinnen zij serieus genomen wenst te worden. In dit geval had de Commissie te kennen kunnen geven dat zij zich niet verantwoordelijk voelde voor de (verandering) in de beleidsfilosofie en zich wenste te beperken tot een advies over de kern van de opdracht: hoe zit het met de prognostische waarde van tests voor (zeer) jonge kinderen? Indien de Commissie ook de verandering in beleidsfilosofie wenste te legitimeren, moet worden vastgesteld dat de diagnose van uitgangspunten en uitvoering van het beleid inzake OVB tekortschiet en dat de verantwoording van die wijziging, te weten van collectieve naar individuele emancipatie, volstrekt onvoldoende is.

Literatuur

- Commissie Indicatiestelling Onderwijsachterstanden (1996). *Zo onvoorspelbaar als het leven zelf*. Zoetermeer: Ministerie van O.C.& W.
- Doombos, K. & Stevens, L.M. (1987). *De groei van het Speciaal Onderwijs (A)*. Den Haag: Staatsuitgeverij.
- Jungbluth, P., Langen, A. van, Peetsma, Th. & Vierke, H. (1996). *Leerlinggegevens basisonderwijs en speciaal onderwijs*. Amsterdam/Nijmegen: SCO-KI/ITS.
- Meijnen, G.W. (1977). *Maatschappelijke achtergronden van intellectuele ontwikkeling*. Groningen: Wolters Noordhoff.

COMMISSIE KWALITATIEVE ASPECTEN VAN GROEPSGROOTTE IN HET BASISONDERWIJS

(1996 ~ *Klassenverkleining, advies over de betekenis van klassenverkleining voor beter basisonderwijs*).

J. F. M. Claessen

Open Universiteit, Valkenburgerweg 167, 6419 AT Heerlen

1. Inleiding

In oktober 1996 verscheen het spraakmakende rapport van de Commissie Kwalitatieve Aspecten van Groeps grootte in het Basisonderwijs. Bespreking van het advies van de Commissie, naar de voorzitter ook wel de Commissie-Van Eijndhoven genoemd, is om meerdere redenen wenselijk. Het is al weer enige tijd geleden dat in het onderwijs een advies verscheen dat zoveel publiciteit oepiep. Ook voor de politiek bleek het advies een hot item te zijn. Uitvoering van het advies heeft financiële en organisatorische implicaties van een ongekende omvang. De kroniek van de politieke discussie en besluitvorming over het advies is uiterst leerzaam.

2. Hoofdlijnen van het advies

De instelling van de Commissie in februari 1996 was het politieke antwoord van Staatssecretaris Netelenbos op het rapport van de Inspectie van het onderwijs over groeps-grootte in het basisonderwijs (1995).

Het rapport is opgebouwd uit drie delen. Deel één behandelt de opdracht voor de Commissie, de gebruikte bronnen en de uitgevoerde analyses (p. 7-79). Deel twee doet verslag van de raadpleging van een aantal deskundigen door de Commissie (p. 81-115). Het laatste deel geeft de standpunten en de adviezen van de Commissie weer (p. 117-150). Als bijlagen van het advies zijn onder meer opgenomen: een onderzoek door de onderwijskundigen uit de Commissie naar klassengrootte en het functioneren van leerlingen en leerkrachten op basis van de PRIMA-cohort-gegevens, een nadere analyse door de onderwijsinspectie naar schoolkenmerken en groeps-grootte in het onderwijs alsmede een berekening van de meerkosten van klassenverkleining door OCenW.

De Commissie kreeg de opdracht een verkenning uit te voeren naar de kwalitatieve aspecten van klassengrootte naar aanleiding van het inspectie-rapport. "Op basis van de resultaten van de verkenning brengt de Commissie advies uit over eventueel noodzakelijke regelgeving met betrekking tot een minimum danwel een maximumgroeps-grootte en andere mogelijke beleidsactiviteiten..."

Hoofdstuk twee besteedt aandacht aan de oorzaken van grote klassen. Klassengrootte is daarbij gedefinieerd als het aantal leerlingen waarmee leerkrachten in hun werk te maken hebben. Er zijn twee oorzaken voor grote klassen: het overheidsbeleid en het schoolbeleid. De gemiddelde klassengrootte in Nederland van 25,7 met grote uitschieters naar boven en beneden, ligt internationaal aan de hoge kant. In landen als Denemarken, Italië, Noorwegen en Zwitserland bevat de gemiddelde klas minder dan 20 leerlingen. Spanje (28,8) en Nieuw Zeeland (29,7) behoren tot de koplopers. De grootte van de klassen in de landen van de Europese Unie laat overigens een dalende lijn zien over de afgelopen 30 jaar (Education Training Youth, 1996). Klassengrootte is uiteraard nauw gerelateerd aan het uitgavenniveau van de overheid. De investering van de Nederlandse rijksoverheid in het basisonderwijs is internationaal gezien dan ook matig. De Commissie koppelt deze vaststelling aan de bevinding dat internationale peilingen uitwijzen dat de prestaties van Nederlandse kinderen op cruciale basisvaardigheden (IEA-studies) nog wel te wensen overlaten. Wat voorts opvalt in het beleid van de Nederlandse rijksoverheid is dat er geen maximale klassengrootte wordt aangegeven. Landen als Denemarken; Duitsland, Finland, Frankrijk, Griekenland, Italië, Noorwegen en Schotland kennen wel een dergelijke regeling. Een derde verklaringsgrond voor grote klassen in Nederland is gelegen in het formatiebeleid van de overheid: grote basisscholen in Nederland en basisscholen met relatief veel 'ongewogen' leerlingen krijgen relatief minder formatie. Behalve regelgeving door de rijksoverheid, zijn ook de beleidskeuzes van basisscholen zelf van invloed op de klassengrootte. Zo kunnen scholen kiezen voor het vermijden van combinatieklassen, voor het inzetten van formatie om de schooldirectie vrij te roosteren, om extra ruimte te scheppen voor zorgverbreding etc.

Hoofdstuk 3 vormt in de opbouw van het rapport een intermezzo. Er is een literatuur-inventarisatie gedaan naar de eisen waaraan effectief onderwijs aan jonge kinderen zou moeten voldoen. "De veronderstelling hierbij luidt dat in grote klassen minder effectief kan worden onderwezen. Dit vermindert de gelegenheid om te leren en de betrokkenheid van de leerling bij het leren." De Commissie merkt daarbij op dat het gaat om 'plausibele veronderstellingen' en constateert tegelijkertijd dat 'in veel onderzoek geen of slechts matige samenhangen tussen klassengrootte en het functioneren van leerlingen worden gepresenteerd'.

De empirische 'bewijslast' voor de relatie tussen klassengrootte en onderwijsprestaties is te vinden in hoofdstuk 4. Vertrekpunt is de veel geciteerde meta-analyse van Glass & Smith (1978): '40% van de 725 studies toont voordelen voor de grotere klassen en 60% voor de kleinere; ...we zien bovendien dat er alleen relevante verschillen zijn in extreme situaties: een klas van 26 moet minstens teruggebracht worden tot 15 leerlingen, wil er enige winst boeken

worden; ...het heeft er alle schijn van dat alleen substantiële reducties in de klassengrootte tot betere prestaties leiden'. De studie van Glass & Smith is nog eens kritisch bekeken door Slavin (1989). Hij komt tot de nog verder gaande conclusie 'dat de kleinere klassen (14 tot 17 leerlingen of 20 tot 23 leerlingen) het amper beter doen dan de grotere klassen; ..en waar de kleinere klassen het beter doen, blijken op de lange duur de effecten weg te ebben'. Onderzoek naar de relatie tussen klassengrootte en leerprestaties blijkt overigens vol methodologische voetangels en klemmen te zitten (Akerhielm, 1995). Veelal wordt bij de samenstelling van de klassen rekening gehouden met het vaardigheidsniveau van de leerlingen: in de grotere klassen de vaardiger leerlingen, in de kleinere klassen de minder vaardige leerlingen, zodat deze meer begeleiding kunnen krijgen.

Voorstanders van klassenverkleining beroepen zich veelal op het Prime-Timeonderzoek en het STAR-experiment. Prime-Time is een beleidsprogramma in Indiana (VS). Door middel van het verkleinen van klassen werd beoogd het kennisniveau van de leerlingen te verhogen. Vanaf 1984/1985 werden de jaargroepen vanaf groep 3 state-wide verkleind van gemiddeld 26.4 naar 19.1 leerlingen. Vergelijking van vóór en ná de klassenverkleining wees uit dat de kleinere klassen het duidelijk beter deden. Het STAR-experiment (Student Teacher Achievement Ratio) is gesitueerd in de staat Tennessee (VS). Met recht kan hier gesproken worden van een experiment. Binnen 79 geselecteerde grote scholen werden voor groep 2, 3, 4 en 5 binnen elke school steeds drie experimentele condities gehanteerd: een klas van 13-17 leerlingen, een klas van 22-25 leerlingen en een klas van 22-25 leerlingen met een full-time onderwijsassistent. Zowel voor de vakken taal als rekenen zijn de resultaten in het voordeel van de kleinere klassen. Grotere klassen met of zonder onderwijs-assistent doen het slechter. Vervolgonderzoek wees uit dat ook na het vijfde jaar de voorsprong van de kinderen uit de kleinere klassen beklifde.

Op zoek naar empirisch materiaal uit Nederland is door de onderwijskundigen uit de Commissie een heranalyse uitgevoerd op gegevens uit het PRIMA-cohort (een aselechte steekproef van 416 basisscholen waarbij gegevens verzameld zijn in groep 2, 4, 6 en 8). Via een multi-level-analyse werd de relatie tussen klassengrootte en leerlingprestaties vastgesteld na correctie voor de achtergrondkenmerken van leerlingen (o.a. IQ) en de samenstelling van de klas (o.a. leerlinggewicht). Zowel voor rekenen als taal blijven de onderwijsprestaties van leerlingen in groep twee achter vanaf klassen met een groeps-grootte 25-29 leerlingen. Bij de hogere (4, 6 en 8) jaargroepen is voor rekenen geen sprake van een voordeel voor kleinere klassen. Integendeel! Voor taal blijven in de hogere jaargroepen de leerlingprestaties pas achter in klassen met 35 of meer leerlingen.

Klassengrootte is door de Commissie ook nog in verband gebracht met andere variabelen: het affectief functioneren van leerlingen (een meta-evaluatie van Smith & Glass uit 1980 geeft in 85% van de onderzochte studies positieve effecten voor kleinere klassen); het oordeel van ouders, leerkrachten en schoolbesturen over de voordelen van kleinere klassen (een Brits opinie-onderzoek door Bennet (1996) leverde de te verwachten resultaten: kleinere klassen zijn in tal van opzichten beter); het welbevinden en instructiegedrag van leraren (kleinere klassen scores (Glass c.s., 1982) hoger). In de heranalyse van het PRIMA-cohort werd de relatie tussen klassengrootte en arbeidssatisfactie overigens niet bevestigd.

Het vijfde hoofdstuk doet verslag van de raadpleging van 17 deskundigen door de Commissie. Aan hen zijn vragen voorgelegd als "heeft de huidige klassengrootte volgens u directe effecten op de kwaliteit van het basisonderwijs? Heeft de klassengrootte volgens u gevolgen voor de prestaties van leerlingen en voor verwijzing naar speciaal onderwijs? In hoeverre vormt volgens u de huidige klassengrootte een probleem voor de organisatie van de basisschool?" etc. De antwoorden van de deskundigen vormen een bonte en veelal uiteenlopende caleidoscoop. In de bewoordingen van de Commissie: "De deskundigen hebben uitgesproken opvattingen en veel creatieve suggesties. Zij worstelen, als zovelen met de complexiteit van het vraagstuk".

Hoofdstuk 6 van het rapport markeert de overgang naar het advies van de Commissie. Naast

argumenten uit onderzoek en de raadpleging van deskundigen spelen ook de eigen opvattingen van de Commissie een rol. Ten aanzien van de klassengrootte stelt de Commissie "dat klassengrootte er als factor wel degelijk toe doet. Grotere klassen hebben negatieve gevolgen voor in elk geval de leerprestaties van leerlingen, verhogen de aanmelding voor het speciaal onderwijs en verminderen de kwaliteit van het didactisch handelen van leerkrachten. Om die redenen is zij voorstander van maximering van de klassengrootte, zowel in de onderbouw als de bovenbouw".

De uitwerking van dit beleid krijgt zijn beslag in hoofdstuk 7. "Uitgaande van de conditionele betekenis van de omvang van klassen en het specifieke belang daarvan voor de onderbouw, adviseert de Commissie beleid te voeren dat resulteert in een maximering van de klassengrootte. Daarbij wordt, om onderwijskundige redenen, een aanzienlijk lager maximum voorgesteld in de onderbouw dan in de bovenbouw; voor de onderbouw 25 leerlingen en voor de bovenbouw 34 leerlingen per klas. Dat impliceert een toekenning van de ongewogen basisformatie van één leraar op 20 leerlingen in de onderbouw en 28 in de bovenbouw." De maximering van de klassengrootte wordt niet aan de scholen zelf overgelaten, maar wordt gezien als een taak van de rijksoverheid naar analogie van de leerplicht, vastgestelde kerndoelen etc. De meerkosten van deze personele investering liggen in de orde van grootte van 923 miljoen, waarbij de OVB-middelen, de WSNS-middelen en de middelen voor taakrealisatie van de directie en voor het schoolprofielbudget onaangetast blijven. In de materiële sfeer (schoolgebouwen, stichtingskosten en eerste inrichting) is een extra budget vereist van ongeveer 225 miljoen. Voor de realisatie stelt de Commissie een meerjarig implementatie-traject voor.

Om de effecten van deze klassenverkleining maximaal te benutten, is flankerend beleid nodig in de sfeer van onder meer verdere professionalisering van leerkrachten, verantwoording van de gekozen klassengrootte door scholen, aansluiting op het voorschoolse traject.

3. Bespreking

Een advies dat dwingt tot keuzes

Het rapport van de Commissie verdient zonder enige twijfel waardering. Een dergelijk advies samenstellen in een tijdsbestek van ruim zes maanden is een knappe prestatie. Het empirisch gedeelte is duidelijk gescheiden van het advies, zodat elke lezer zijn of haar eigen conclusies kan trekken. Lof verdient ook de openhartigheid van de Commissie. Van een selectieve weergave van onderzoeksresultaten is geen sprake. Integendeel: ook bevindingen, meningen en argumenten die indruisen tegen het uiteindelijke advies tot klassenverkleining komen ruimschoots aan de orde.

Klassengrootte, overheidsuitgaven en onderwijsprestaties in internationaal perspectief

In hoofdstuk twee brengt de Commissie internationale onderzoeksgegevens over klassengrootte, overheidsuitgaven en onderwijsprestaties (IEA-onderzoeken over begrijpend lezen, reken- en wiskunde-onderwijs) met elkaar in verband. Nederland kent een relatief hoge klassengrootte, het uitgaveniveau voor het basisonderwijs is relatief laag en de onderwijsprestaties zijn over het geheel genomen matig. Een dergelijke associatie leidt tot een voorspelbare conclusie. Nadere bestudering van het gepresenteerde cijfermateriaal noopt evenwel tot meer terughoudendheid. De relatie tussen klassengrootte en onderwijsprestaties is niet zo eenduidig als verwacht. Finland bijvoorbeeld heeft met een klassengrootte van 24,6 een gemiddelde dat in de buurt ligt van Nederland (25,7). Toch is Finland de trotse koploper wanneer het gaat om de gemiddelde leerlingsscore voor begrijpend lezen van 9-jarigen in 27 landen (met Nederland overigens op de 21ste plaats). Landen als Ierland, Nieuw Zeeland en Spanje kennen een klassengrootte die ruim boven Nederland ligt, maar scoren desondanks fors hoger voor begrijpend lezen. De relatie in internationaal opzicht tussen overheidsuitgaven en onderwijsprestaties kent daarentegen meer symmetrie: een relatief laag niveau van uitgaven gaat gepaard met relatief lage onderwijsprestaties (zoals bijvoorbeeld in Nederland).

Het verhogen van de uitgaven door klassenverkleining is dus geen waarborg voor hogere onderwijsprestaties.

De lessen uit het Prime-Time-onderzoek en het STAR-experiment

In de speurtocht naar voordelen van kleinere klassen spelen het Prime-Time-onderzoek en STAR-experiment begrijpelijkerwijze een belangrijke rol. In het eerste onderzoek kon een grootschalige vergelijking gemaakt worden van de effecten van klassenverkleining voor en na de implementatie van Prime-Time. Het STAR-experiment bood de unieke kans tot een experimentele opzet.

Desondanks plaatst ook de Commissie al de nodige kanttekeningen bij de lessen uit beide projecten: dreiging van het Hawthorne-effect (aandacht van de onderzoekers voor leraren in kleinere klassen kan leiden tot autonome effecten) en het John Henry-effect (de leerkrachten in de controleconditie, d.w.z. met de grote klassen, gaan onder invloed van het experiment extra hun best doen); deelname aan het STAR-experiment vond plaats op basis van vrijwilligheid en voldoende grootte van de school; in welk dilemma verkeren leerkrachten met grotere klassen die weten dat er een experiment naar de betekenis van kleinere klassen wordt gedaan? Aan deze kritische opmerkingen willen we nog de volgende opvallende observatie toevoegen. In het STAR-experiment werden drie experimentele condities gehanteerd: een klas van 13-17 leerlingen, een klas van 22-25 leerlingen en een klas van 22-25 leerlingen met naast de reguliere leerkracht een fulltime onderwijsassistent. In tegenstelling tot wat men zou verwachten bleek de inzet van deze fulltime onderwijs-assistent geen enkele toegevoegde waarde te hebben! Hadden de leraren met de grotere klassen in het experiment al bij voorbaat het hoofd in de schoot gelegd? Een pikant detail in het geheel is de geplande, grootschalige inzet van onderwijsassistenten in Nederland. Voor de komende jaren is hiervoor een bedrag vrijgemaakt van 100 miljoen. Een verstandige investering?

Overigens zijn ook in de Verenigde Staten de resultaten van Prime-Time en STAR niet onomstreden. Zo betoogt Tomlinson (1988, 1989) dat, gelet op de gigantische kosten van Prime Time en STAR, "better and more durable achievements effects might be expected far more practically and, in most instances, more cost-effectively through cross-age peer tutoring and better use of instructional technology, graded homework, more flexible instructional practice and grouping methods, improved proficiency at academic studying, year-round schooling, better curricula and textbooks, and perhaps even improved education and training of teachers. Some of these practices cost little or nothing to provide, and in contrast to smaller classes, most of them improve the teachers' instructional skills" (1989, p. 269).

De heranalyse op de data van het PRIMA-cohort

De beschikbaarheid van data uit dit zeer omvangrijke en zorgvuldig samengestelde cohort was voor de Commissie aanleiding tot een heranalyse wat betreft de relatie tussen klassengrootte en onderwijsprestaties. Bij deze secundaire analyses werd gecorrigeerd voor achtergrondkenmerken van leerlingen en de samenstelling van de klas. Dit effect van juist minder vaardige leerlingen in kleinere klassen blijkt namelijk nogal wat onderzoek naar de effecten van klassengrootte te verstoren, kortom een onderzoek volgens het boekje.

Hoewel de resultaten uit de heranalyse maar een zeer beperkte ondersteuning geven aan de voordelen van kleinere klassen (zie de eerder gegeven samenvatting in paragraaf 2) leunt de Commissie toch zwaar op deze bevindingen. Wie zich nader verdiept in de bijlage waarin de heranalyse beschreven wordt, gaat zich evenwel afvragen of de rapportage en interpretatie van de analyses óók wel volgens het boekje waren. Bij rekenvaardigheid scoren leerlingen van groep twee in klassen van 5-9, 10-14 en 15-19 leerlingen evengoed als leerlingen met een klassengrootte van 20-24 leerlingen (geen significant verschil). Er is dus geen sprake van voordelen van kleinere klassen. Vanaf klassen met een grootte van 25-29, 30-34 en 35-39 leerlingen blijven de prestaties dan inderdaad significant achter bij die van leerlingen in een klas met 20-24 leerlingen. Van een monotoon oplopend verband tussen klassengrootte en

reken score is daarbij evenwel geen sprake. Leerlingen in een klas van 30-34 leerlingen doen het bijvoorbeeld beter dan leerlingen in een klas van 25-29! Men zou leerlingen uit een klas met 25-29 leerlingen dus een dienst bewijzen door hen in een klas onder te brengen met 30-34 leerlingen.

In de hogere leerjaren (4, 6 en 8) roept de relatie tussen klassengrootte en rekenprestaties nog meer vragen op. Het zijn nu juist de kleinere klassen (5-9, 10-14 en 15-19 leerlingen) waar de prestaties achterblijven. Koplopers zijn leerlingen afkomstig uit klassen met 30-34 leerlingen!

Qua taalprestaties in groep twee zijn er geen significante verschillen tussen klassen met 5-9, 10-14, 15-19, 20-24 en 30-34 leerlingen. Klassen met 25-29 en 35-39 leerlingen blijven wel iets achter. In de jaargroepen 4, 6 en 8 zijn er vrijwel geen verschillen naar klassengrootte. Alleen bij meer dan 35 leerlingen is er sprake van enige achterstand.

Kijkend naar het cijfermateriaal, kan niet anders geconcludeerd worden dan dat de heranalyse van het PRIMA-cohort eerder contra dan pro klassenverkleining wijst.

Klassengrootte en de samenhang met zittenblijven en de verwijzing naar speciaal onderwijs
De commissie is van mening dat klassenverkleining gunstig uitpakt in termen van minder zittenblijven en minder verwijzingen naar het speciaal onderwijs. Deze uitspraak is verbazingwekkend. De Commissie citeert zelf immers onderzoek waaruit blijkt dat zittenblijven meer voorkomt in kleinere klassen, ook na controle voor achtergrondkenmerken van leerlingen. De Commissie relateert dit op zichzelf volstrekt duidelijke onderzoeksgegevens met de stelling dat in kleinere klassen leerkrachten gevoeliger zijn voor problemen van individuele leerlingen, waardoor de kans op doubleren toeneemt. Anders gezegd: wanneer leerkrachten in grotere klassen dezelfde gevoeligheid voor problemen van individuele leerlingen zouden hebben, zou ook in grotere klassen zittenblijven met dezelfde frequentie voorkomen.

Met betrekking tot de verwijzing naar speciaal onderwijs beroept de Commissie zich op inspectieonderzoek, waaruit blijkt dat uitstroom naar speciaal onderwijs iets frequenter zou voorkomen op scholen met grotere klassen. Onderzoek van Petersen en Oudenhoven (1996) leverde geen bevestiging op van een grotere uitstroom naar het speciaal onderwijs in grotere klassen.

De verwachting van de Commissie dat klassenverkleining leidt tot minder zittenblijven en tot een beperking van de uitstroom naar het speciaal onderwijs is dus op zijn minst discutabel.

Het gewicht van deskundigen

Conform haar opdracht heeft de Commissie ten behoeve van haar standpuntbepaling een aantal deskundigen geraadpleegd. Het gaat daarbij om het beperkte aantal van 17 personen, waarbij een verantwoording van de selectie ontbreekt. De verslaglegging van deze raadpleging roept vraagtekens op: wie zegt wat? Wat is de basis van hun uitspraken? Waarom wordt soms wel geluisterd naar de mening van de deskundigen en soms niet? De indruk die blijft hangen na lezing van de tekst is er een van selectief gebruik.

Onderwijsprestaties en klassenverkleining

Hoe klein moeten klassen nu zijn alvorens er een positief effect optreedt richting onderwijsprestaties? Glass & Smith (1978) laten zien dat relevante verschillen zich pas voordoen in extreme situaties: een klas van 26 moet minstens teruggebracht worden tot 15 leerlingen wil er enige winst geboekt worden. Slavin (1989), die de analyses van Glass & Smith nog eens opnieuw onder de loep nam, komt tot de conclusie dat de kleinere klassen (14 tot 17 leerlingen of 20 tot 23 leerlingen) het amper beter doen dan de grotere klassen. Hij pleit voor een nog verder gaande reductie. In het STAR-experiment werden positieve effecten voor kleinere klassen gevonden bij een gemiddelde klassengrootte van 15 leerlingen.

Tot welke adviezen komt nu de Commissie? Een maximum voor de onderbouw van 25 leerlingen met een formatietoedeling van één leraar op 20 leerlingen en in de bovenbouw een

maximum van 34 leerlingen met een formatiesleutel van één leraar op 28 leerlingen. Deze cijfers blijven in aanzienlijke mate achter bij de kengetallen uit literatuur en experimenten. Voor betere onderwijsprestaties zou de klassengrootte minimaal terug moeten naar 15 leerlingen!

De Commissie weet dat natuurlijk ook, maar realiseert zich vermoedelijk tevens dat dit in financieel opzicht een utopie is (iedere leerling minder kost + 200 miljoen). Met de nu voorgestelde klassenverkleining is evenwel geen verbetering van de onderwijsprestaties te verwachten.

Klassenverkleining: top-down of bottom-up?

Uitgaande van de noodzaak van klassenverkleining stond de Commissie voor de vraag in welke mate deze klassenverkleining via regelgeving afgedwongen moet worden. Onder de geraadpleegde deskundigen was dit punt een van de meest omstreden onderwerpen. De Commissie kiest duidelijk voor een top-down benadering met het voorschrijven van een maximumklassengrootte zowel voor de onderbouw als de bovenbouw.

Met deze benadering kiest de Commissie duidelijk voor een trendbreuk met het huidige beleid, gericht op het streven naar meer autonomie voor scholen. Men mag echter verwachten dat ook zonder een prescriptieve benadering de scholen de extra formatie primair zullen inzetten voor klassenverkleining. De scholen kunnen daarbij rekening houden met hun eigen specifieke omstandigheden, hun eigen onderwijskundig beleid en hun personeelsbeleid. Dwang van boven af zal door de scholen ervaren worden als beknotting van hun mogelijkheden. Met bijvoorbeeld slechts één leerling boven het maximale aantal zal nu gesplitst moeten worden of een combinatieklas gevormd moeten worden. Bovendien moeten scholen de mogelijkheid hebben om de extra formatie ook in te zetten voor bijvoorbeeld zorgverbreding, taakverlichting, scholing, professionalisering, consultatie, interne begeleiding, activiteiten in de sfeer van lokaal onderwijsbeleid etc. Met handhaving van de bestaande autonomie voor scholen inzake formatie-inzet bespaart de overheid zich bovendien veel ellende en energie in de sfeer van controle.

Ook de gedetailleerde wijze waarop de scholen rekenschap moeten gaan afleggen over de inzet van de toebedeelde formatie roept vraagtekens op. Zo adviseert de Commissie om 'scholen te verplichten in het jaarverslag rekenschap af te leggen over de inzet van de toebedeelde formatie en het hieruit resulterende kwantitatieve en kwalitatieve rendement op leerlingniveau in relatie tot de schoolpopulatie. De resultaten worden gepresenteerd in een tabel, waarin de positie van de school ten opzichte van qua leerlingpopulatie vergelijkbare scholen wordt weergegeven. De resultaten dienen te worden uitgesplitst naar leerlinggewichten;...in het jaarverslag dienen daarnaast verplichte paragrafen te komen ten aanzien van de inrichting van klassen, zoals klassengrootte, combinatieklas en doorbreking van het klassenverband voor specifieke activiteiten; zoals de inzet van de leerkrachten in relatie tot hun deskundigheid van de onder- en bovenbouw en een hieraan gekoppeld scholingsplan, etc.'. Een dergelijke opsomming van gedetailleerde instructies is bestuurlijke overkill.

Kroniek van de politieke gedachtevorming over klassenverkleining

In het advies van de Commissie dat in oktober 1996 werd gepubliceerd, is als bijlage 6 een motie van Wallage c.s. opgenomen. Deze motie draagt de datum van 19 september 1996 en werd ingediend tijdens de algemene politieke beschouwingen. In deze motie, ingediend door Wallage, Bolkenstein en Wolffensperger, wordt de regering uitgenodigd om (mede op basis van de nog niet gepubliceerde bevindingen van de Commissie) een meerjarenplan, inclusief financiering, op te stellen voor verkleining van de groeps grootte. Een dergelijke voortvarendheid van de politiek mag uitzonderlijk genoemd worden. Nog voor publicatie van het advies heeft de politiek de hoofdlijnen ervan al omarmd. Vanwaar deze slagvaardigheid?

Het idee om klassen te verkleinen is uiteraard niet nieuw. Overwegingen in die richting werden evenwel al in de kiem gesmoord door wetenschappelijke twijfels over de opbrengst

van zo'n uiterst kostbare operatie. Met het verschijnen van het inspectierapport over klas-sengrootte werd de politiek danig opgeschrikt. De economie vertoonde een opgaande lijn, zo-dat er enige ruimte ontstond om ook in het onderwijs weer investeringen te doen. De resulta-ten van het dure OVB-beleid bleken jaar in, jaar uit tegen te vallen. Het klimaat om een nieuwe richting in te slaan was dus gunstig. In de politieke discussie over het rapport van de inspectie kwam de suggestie van klassenverkleining al op tafel, maar staatssecretaris Netelenbos wist de gemoederen nog te bezweren door 100 miljoen toe te zeggen voor de aanstelling van onderwijsassistenten ('meer handen in de klas'). De discussie over eventuele klassenverklei-ning werd doorgeschoven tot het rapport van de Commissie op tafel zou liggen.

In de zomer van 1996, wanneer de Commissie voortvarend werkt aan haar rapport, duikt dan opens de successory van het STAR-experiment in de pers op (met waarschijnlijk Robert Sikkes van de Volkskrant als gangmaker). Tennessee is het beloofde land. Klassen-verkleining blijkt wel degelijk zoden aan de dijk te zetten. Het zwaan-kleef-aan-effect treedt op en liet uiteraard ook de politici niet onberoerd. Al tijdens de politieke beschouwingen in najaar 1996 tekende zich een kamerbrede meerderheid af voor klassenverkleining met als resultaat de motie van Wallage. Met zoveel politieke wind in de zeilen gaat staatssecretaris Netelenbos uiteraard gezwind overstag. Tijdens de behandeling van de onderwijsbegroting zeg-de zij toe om al in april 1997 met concrete voorstellen te komen voor het schooljaar 1997-1998.

4. Conclusies

Met de voorstellen van de Commissie wordt zonder enige twijfel een stevige impuls gegeven aan de ontwikkeling van het basisonderwijs; positieve ontwikkeling van de werkgelegenheid (op termijn ruim 11 duizend extra leraren); 'vele handen maken licht werk'; verlichting van de werkdruk voor leraren; een warm onthaal van de voorstellen bij leraren, directies, bestu-ren en ouders; een uitgavenniveau voor het basisonderwijs waarmee Nederland weer in de pas gaat lopen van de landen van de Europese Unie; ruimte voor professionalisering van le-raren; versterking van de aansluiting tussen basisonderwijs en voorschoolse traject etc.

De verwachting van de Commissie dat de voorgestelde klassenverkleining ook zal voeren tot verbetering van de onderwijsprestaties, tot vermindering van de uitstroom naar het specia-aal onderwijs en tot reductie van het zittenblijven, is voorshands illusoir en mist een gezon-de empirische basis. Voor politici hoeft dit wellicht geen struikelblok te zijn, maar de onder-wijskunde als een op empirie gestoelde wetenschap mag niet meer beloven dan zij kan waarmaken.

Aan de technische uitwerking van de voorstellen (maxima in de wet vastleggen, het door-breken van de lijn van deregulering; scholen niet zelf laten beslissen over de inzet en verde-ling van de toegewezen formatie, gedetailleerd rekenschap moeten afleggen in het jaarver-slag etc.) kleven de nodige bezwaren. De verleiding om bij zo'n majeure operatie tot verdichting van regelgeving te komen is groot, maar dient weerstaan te worden. Scholen moe-ten zelf kunnen blijven beslissen op welke wijze ze de toegewezen formatie willen inzetten. Klassenverkleining is daarbij een voor de hand liggende optie, maar ook andere opties kun-nen, afhankelijk van de specifieke omstandigheden van een school, de voorkeur verdienen.

5. Literatuur

- Akerhielm, K. (1995). Does Class Size Matter? *Economics of Education Review*, 3, 229-241.
- Bennet, N. (1996). Class Size in Primary Schools: perceptions of headteachers, chairs of governors, teachers and parents. *British Educational Research Journal*, 1, 33-55.
- Education Training Youth (1996). *Key data on education in the European Union*. Luxembourg: Office for Official Publications of the European Communities.
- Glass, G. V. & M. L. Smith (1978). *Meta-analysis of Research on Class Size and Achievement*. San Francisco: Far West Laboratory.
- Glass, G. V., Cahen, L. S., Smith, M. L. & N. N. Filby (1982). *School Class Size*. London: Sage Publications.
- Inspectie van het onderwijs (1995). *Groepsgrootte in het basisonderwijs*. De Meern: Inspectie van het onder-wijs.

- Petersen, B. en D. Oudenhoven (1996). Waar komen de regionale verschillen in verwijzing vandaan? *Pedagogische Studiën*: 6, 434-440.
- Slavin, R. E. (1989). Achievement effects of substantial reductions in class size. In: R. E. Slavin (Ed) *School and classroom organization*. Hillsdale: Lawrence Erlbaum.
- Smith, M. L. & G. V. Glass (1980). Meta-analysis of Research on Class Size and Its Relationship tot Attitudes and Instruction. *American Educational Research*: 4, 419-433.
- Tomlinson, T. M. (1988). *Class Size and Public Policy: Politics and Panaceas*. Office of Educational Research and Improvement U.S. Departement of Education.
- Tomlinson, T. M. (1989). Class Size and Public Policy: Politics and Panaceas. *Educational Policy* 3, 261-273.

Heerlen, 24 januari 1997

Boekbesprekingen

R. Rijmenans, V. Geudens, H. Coucke, H. van den Bergh & F. Daems
Effectiviteit van Vlaamse secundaire scholen
Universiteit Antwerpen, 1996, ISBN 90-5728-003-5

Het rapport *Effectiviteit van Vlaamse secundaire scholen* geeft een uitgebreid overzicht van een onderzoek dat eind 1992 startte op initiatief van het ministerie. Uitgebreid, want naast de hoofdtekst van 231 bladzijden is er nog een apart deel bijlagen van 369 bladzijden. Volgens de auteurs is hun onderzoek het eerste Vlaamse effectiviteitsonderzoek, hoewel in de literatuurlijst wel verwezen wordt naar andere onderzoeken die op zijn minst kunnen meedingen naar die titel.

Het onderzoek

Aan het onderzoek werkten 2743 leerlingen uit het tweede en vierde leerjaar van 102 scholen mee, evenals 158 leraren Nederlands en 100 directieleden. De leerlingen maakten een IQ-test, een schrijftoets en een leestoets en vulden een vragenlijst in waarin naar achtergrondkenmerken werd gevraagd, de mening over de school en over het vak en de vrijetijdsbesteding. De leraren Nederlands vulden een vragenlijst in over contacten binnen de school en hun wijze van lesgeven. Ook gaven ze in logboeken over een aantal lessen aan hoe ze de tijd hadden besteed. Een deel van de vragen was specifiek voor het leesonderwijs, een deel voor het schrijfonderwijs. De directie van de scholen ten slotte vulde een vragenlijst in over de context van de school en het beleid ten aanzien van het vak Nederlands.

Alle gegevens zijn begin 1994 tegelijkertijd verzameld, op het moment dat de leerlingen ongeveer halverwege het schooljaar waren. Het onderzoek was niet longitudinaal van opzet en kende geen herhaalde metingen. Effecten zijn geanalyseerd door middel van multilevelprocedures, steeds apart voor de drie onderscheiden Vlaamse schooltypen ASO, BSO en TSO en de klassen 2 en 4.

De resultaten

Vanwege het grote aantal afzonderlijke analyses is het correct weergeven van de effecten bijna onbegonnen werk. Dat geldt des te sterker omdat variabelen zelden een effect hebben dat in alle analyses op dezelfde wijze terugkeert; een resultaat dat vanuit de gedachten over effectieve school- en klaskenmerken verwacht zou mogen worden.

De analyses leveren echter het beeld dat ook uit vergelijkbaar Nederlands onderzoek bekend is: een variabele die in het ene schooltype effect heeft, heeft dat weer niet in een ander schooltype; een variabele die in leerjaar 2 een effect heeft, heeft dat niet vanzelfsprekend ook in leerjaar 4 en een variabele die in de ene analyse een positief effect heeft kan in de volgende analyse weer een negatief effect hebben. Dat treedt bijvoorbeeld op bij enkele variabelen die tijdsbesteding in de lessen meten.

Voorzover effecten enigszins consequent over leerjaren en schooltypen heen optreden, gaat het bijna altijd om tamelijk voor de hand liggende effecten die maar weinig te maken hebben met wat zich feitelijk in de lessen afspeelt. Zo rapporteren de onderzoekers positieve samenhangen tussen rapportcijfer en prestaties en tussen plezier in lezen en lees- en schrijfpredaties. Negatieve samenhangen zijn er tussen veel spijbelen en prestaties en tussen thuis geen Nederlands spreken en prestaties.

Bijdrage aan het effectiviteitsonderzoek

Het effectiviteitsonderzoek is al vaak vergeleken met een ongerichte visexpeditie. We gooien eens wat netten uit, denkt de effectiviteitsonderzoeker, en allicht vangen we wat aardigs. Ook

dit onderzoek is een voorbeeld van zo'n ruim opgezette expeditie, die onherroepelijk leidt tot de vangst van een paar aardige vissen waarvan helaas niet meteen duidelijk is tot welke soort ze eigenlijk horen en verder tot een hoop guppies die meteen weer vrijgelaten kunnen worden. In die zin levert het onderzoek dan ook geen echt nieuwe bevindingen op.

Waarom zoveel variabelen onderzocht moesten worden leggen de auteurs nergens uit. Toen het onderzoek eind 1992 startte, waren, gelet op de stand van zaken in het toenmalige effectiviteitsonderzoek, ook andere keuzes mogelijk geweest, bijvoorbeeld keuzes voor een beperkt aantal kernvariabelen op meer theoretische gronden. Ook een meer hypothese-toetsende opzet was mogelijk geweest. En als men dat niet de aangewezen weg vond omdat er in Vlaanderen nog zo weinig was onderzocht, dan zou een krachtige data-reductie op theoretische gronden in de analyse-fase van het onderzoek altijd nog mogelijk zijn geweest. Maar ook dat is niet gebeurd. De onderzoekers zeggen wel dat ze zich baseren op een conceptueel effectiviteitsmodel, maar gaan vervolgens zonder uit te leggen waarom andere en vooral veel meer variabelen meten dan in dat model staan vermeld.

Daar komt nog bij dat de opzet van het onderzoek niet ideaal is in effectiviteitstermen: er is geen herhaalde meting en alle variabelen, zowel de onafhankelijke als de afhankelijke, zijn op eenzelfde moment halverwege het schooljaar gemeten. Controleren voor eerdere prestaties is dus niet mogelijk en het wordt zo wel erg lastig om iets over effecten van de leraren en de school te zeggen. Als toch al bekend was dat een herhaalde meting er niet in zat, was een heel andere, meer kleinschalige opzet van het onderzoek misschien ook te overwegen geweest, waarbij men bijvoorbeeld wat lessen had kunnen observeren.

Het onderzoek blijft nu, zowel in de opzet als in de resultaten, erg ongericht. Jammer genoeg nemen de onderzoekers niet de moeite om hun bevindingen aan het eind van het verslag even te vergelijken met ander recent onderzoek. Daardoor blijven de resultaten meer in de lucht hangen dan nodig was.

Beperking

Dat de onderzoekers zichzelf nergens beperkingen opleggen, is het grootste probleem in het onderzoek zelf maar ook in de verslaggeving. Die is zo gedetailleerd dat de lezer al snel door de bomen het bos niet meer ziet, om nu maar even in de beeldspraak van een expeditie op het land te vervallen. De onderzoekers lijken letterlijk alles het vermelden waard te vinden, ook dingen die doodnormaal zijn: "eerst gaven we aan elke leerling een uniek identificatienummer. Dat schreven we zowel op de leerlingenlijsten die we van de school gekregen hadden, als op de toetsen en de vragenlijst die voor die leerling bestemd waren. Per leerling vulden we vervolgens een envelop met daarin het toetsmateriaal in de volgorde van de afname (...). Op de envelop werd het identificatienummer herhaald en schreven we eveneens de naam van de leerling die alleen gebruikt zou worden bij het uitdelen van de enveloppen" (p. 63).

Men verliest zich hier en op andere plaatsen in het beschrijven van procedures, zonder stil te staan bij belangrijker onderzoeksmatige overwegingen. Wat betekent het bijvoorbeeld dat een deel van de klassen door iemand van de school zelf werd getest en een deel door een onafhankelijke testleider? Daaraan wordt bijna geen woord vuil gemaakt.

De overvloedige verslaggeving is ook bij de resultaten niet geschuwd, zodat de frequenties de volhardende lezer hoofdstukken lang om de oren blijven vliegen. Is hij dan eindelijk bij de multilevel-analyses aangeland, dan moet hij zich eerst nog door een obligate uitleg over het wezen en het zijn van multilevel-analyse heenbijten alvorens de resultaten eindelijk in beeld komen. En dan nog maar een deel van de resultaten, want voor de regressiecoëfficiënten moet het bijlagendeel geraadpleegd worden. Dit alles maakt het lezen van het verslag behoorlijk vermoeiend.

Leerwinst

De onderzoekers konden geen herhaalde meting uitvoeren en daardoor konden ze ook niet bepalen wat de effecten van school- en klaskenmerken zijn op de leerwinst van leerlingen.

Leerwinst dan in de gangbare betekenis van prestaties bij de tweede meting gecorrigeerd voor de prestaties van diezelfde leerlingen bij de eerste meting. Toch nemen de onderzoekers het woord leerwinst in de mond, maar ze geven er een heel eigen draai aan. Leerwinst zien zij als een construct op schoolniveau, namelijk de correlatie tussen de variantie in klas 2 en in klas 4. Deze definitie is op zijn zachtst gezegd zeer ongebruikelijk en lijkt hooguit iets te zeggen over de stabiliteit van de school over leerjaren heen. De onderzoekers hadden, als ze per se deze index wilden gebruiken, dat eerst eens goed moeten verantwoorden en ze hadden zeker voor een andere benaming moeten kiezen.

Leger

Het enige leuke aan de uitputtende verslaggeving is dat er zo nu en dan verrassende zaken aan het licht komen. Zo blijkt bij de beroepsclassificatie van ouders een aparte categorie voor het leger gereserveerd te zijn, waarin zowel de lage als de hoge legerfuncties zitten. Waarom dit zo is, en hoe zo'n categorie zich in de analyses verhoudt tot alle overige die gezamenlijk een rangorde van hoog naar laag vormen, blijft helaas in nevelen gehuld. Ook is het verrassend om te lezen dat sommige Vlaamse ouders hun kinderen met de dood bedreigen als ze spijbelen van school. Die kinderen zijn op hun beurt behoorlijk vindingrijk als ze aangeven waarom ze wel eens spijbelen: 'stakingen van het openbaar vervoer' (gaat iedereen dan met bus of trein naar school?) en 'betogingen tegen racisme' worden massaal genoemd. Even consequent doorredeneren, en voor je het weet correleren de salariëring van buschauffeurs en machinisten en niet te vergeten de politieke bewustheid van leerlingen negatief met taalprestaties. Toch nog weer wat nieuwe variabelen voor het effectiviteitsonderzoek ontdekt!

Gerry Reezigt

Marc Vermeulen

Human capital in the hinterland. An analysis of causes and consequences of regional variation in educational participation

(Tilburg, Tilburg University Press, 1996) ISBN 90.361-9976-x

De vraag naar regionale gegevens neemt de laatste jaren snel toe. Het blijkt dat voor een goed bestuur de plaatselijke overheden regionale informatie nodig hebben, met inbegrip van de regionale verschillen. De roep om een nieuwe volkstelling is mede een gevolg van de wens om over regionaal gedetailleerde statistische informatie te kunnen beschikken. Op zich is de grote vraag naar regionale informatie in een land als Nederland verbazingwekkend, gezien het feit dat Nederland een klein land is dat voor een belangrijk deel door een centrale overheid wordt bestuurd. Gezien de sterke bemoeienis van de centrale overheid met het onderwijs en het streven naar het creëren van gelijke kansen, is het des te opmerkelijker dat er regionale verschillen bestaan in onderwijsdeelname. Vanuit deze achtergrond concentreert Marc Vermeulen zich in zijn studie op de vraag welke factoren regionale verschillen in onderwijsdeelname verklaren en wat de gevolgen zijn van deze regionale verschillen in onderwijsdeelname voor de regionale economische ontwikkeling.

Onderzoek naar onderwijsdeelname is tot nu toe vooral op micro-niveau (op het niveau van de leerlingen) of op macro-niveau (tijdreeksen binnen één land of tijdreeksen die tussen landen worden vergeleken). Op het micro-niveau vindt school- en beroepsloopbaanonderzoek plaats, waarin de kenmerken van jongeren en hun ouders in verband worden gebracht met het succes in het onderwijs en op de arbeidsmarkt. Op het macro-niveau zijn vooral de relaties tussen kenmerken van de samenleving en de onderwijsexpansie en die tussen de onderwijsexpansie en de economische ontwikkeling onderwerp van onderzoek. Het onderzoek van Vermeulen is een logische aanvulling hierop: het verklaren van de regionale verschillen in

onderwijsdeelname en de gevolgen daarvan voor de regionale economische ontwikkeling.

In het eerste deel probeert Vermeulen de regionale verschillen te verklaren in de participatiegraad in verschillende vormen van voortgezet en hoger onderwijs. De verklaring wordt gezocht in de economische *human capital*-theorie, en de meer sociologisch georiënteerde statusverwervings- en emancipatietheorieën, aangevuld met enige empirisch relevante factoren die niet uit deze theorieën kunnen worden afgeleid. Voor alle vormen van onderwijs blijkt dat de invloed van de samenstelling van de regionale populatie naar onderwijsniveau de belangrijkste is. Dit wijst erop dat het niveau van onderwijsdeelname het best verklaard wordt uit de statusverwervingstheorie, die stelt dat mensen opleidingen volgen om hun status te verhogen dan wel te handhaven. De economische factoren leveren op zijn best aanvullende verklaringen voor de verschillende onderwijsdeelname. Dit is op zichzelf een verrassende conclusie, omdat nog steeds wordt volgehouden dat arbeidsmarktperspectieven van schoolgaanden van groot belang zijn voor hun studiekeuze.

In het tweede deel worden de gevolgen van de verschillen in onderwijsdeelname voor de regionale macro-economische ontwikkeling geschetst. Dit deel heeft een exploratief karakter, omdat de relaties tussen regionale onderwijsdeelname en economische ontwikkelingen uiterst complex zijn en er nog relatief weinig onderzoek naar is verricht. Er worden daarom slechts drie relatief afgeronde onderwerpen onderzocht.

In de eerste plaats wordt de relatie tussen kenmerken van de onderwijsdeelname in de regio en het innovatie-potentieel van de regionale economie onderzocht. De gegevens bevestigen de conclusies van eerder onderzoek: de invloed van onderwijsdeelname of het opleidingsniveau van de bevolking hebben nauwelijks invloed op de gezondheid en innovatieve kracht van de regionale economie. Een reden daarvoor kan zijn dat de sterke regionale economieën geschoolde arbeidskrachten van buiten hun regio aantrekken.

In de tweede plaats wordt daarom de geografische allocatie op de arbeidsmarkt van schoolverlaters (forenzen en verhuizen voor een baan) onderzocht. Het blijkt dat schoolverlaters naarmate hun opleidingsniveau hoger is vaker een baan buiten hun eigen regio vinden. Van de investeringen van de overheid in het regionale onderwijs wordt dus door economisch sterkere regio's de vruchten geplukt.

In de derde plaats wordt mogelijke verdringing van lager door hoger opgeleiden op de regionale arbeidsmarkten onderzocht. Eerst wordt daarvoor een verdringingsmaat ontworpen die beoogt overproductie van het onderwijs te meten. Idealiter zou Vermeulen graag een vergelijking maken tussen de niveaus van bekwaamheden die benodigd zijn voor de beroepsstructuur en de feitelijke opleidingsniveaus van de beroepsbevolking. Als de opleidingsniveaus gemiddeld hoger zijn dan de beroepsniveaus, is sprake van 'overproductie' van het onderwijs. Dat is volgens Vermeulen niet mogelijk en vele malen herhaalt hij het argument dat deze informatie in Nederland niet beschikbaar zou zijn. Helaas heeft hij daarin geen gelijk. Sinds 1993 is de Standaard Beroepenclassificatie 1992 beschikbaar (CBS, 1993), waarin beroepen worden ingedeeld naar de meest geëigende opleiding. Deze classificatie is reeds met succes toegepast voor het meten van overscholing op micro-niveau, en niets staat toepassing op regionaal niveau in de weg. Nu moet Vermeulen zich behelpen met een econometrische maat, die nogal gekunsteld aandoet en waarvan de betekenis verre van helder is. Een tweede punt van kritiek dat ik op deze analyse heb is dat verdringing op één moment in de tijd wordt onderzocht. Verdringing is per definitie een dynamisch proces dat uitsluitend goed kan worden onderzocht met behulp van tijdreeksen: op dezelfde beroepsniveaus groeit het aandeel (te) hoog opgeleiden. De conclusie dat deelname aan het HBO de verdringing bevordert, terwijl deelname aan universitair onderwijs dat niet doet, zou ik graag nog eens met behulp van een betere verdringingsmaat bevestigd willen zien.

Al met al is het een redelijk geslaagde toevoeging aan de reeds bestaande literatuur over de oorzaken en gevolgen van onderwijsdeelname. De grote verdienste van Vermeulen ligt vooral in het verzamelen en analyseren van het statistische materiaal, waarbij hij de meeste van de methodologische problemen innovatief en creatief oplost. Waar hij helaas niet in slaagt

is een goed doordacht overzicht te geven van de theorieën die van belang zijn voor zijn vraagstellingen en welke hypothesen uit deze theorieën kunnen worden afgeleid. Het doet nogal hap-snap aan en de relaties tussen de theorieën, de reeds gevonden empirische resultaten, en de te toetsen hypothesen moeten uit een voortkabbellende tekst worden geplukt. Verder is hij gedwongen om theorieën die op micro-niveau zijn ontworpen toe te passen op macro-gegevens, hetgeen, zoals hij ook zelf stelt, de nodige complicaties oplevert. Weliswaar probeert hij inzichten vanuit een sociologische en economische invalshoek te combineren, maar tot uitgewerkte theoretische modellen, tegenspraken tussen de theorieën en hypothesen om deze tegenspraken te onderzoeken komt het niet.

Het is overigens opvallend dat door sociologen veel vaker van economische theorieën gebruik wordt gemaakt dan economen inzichten uit de sociologie gebruiken. Dit verschil is in zekere zin tegengesteld aan het patroon van geografische allocatie van hoger opgeleiden. Sterke lokale economieën rekruteren vaak hoger opgeleiden uit zwakkere lokale economieën. Het verschil in status van disciplines heeft ook gevolgen voor de mate waarin leentjebuur wordt gespeeld bij de ander, maar dan omgekeerd: de disciplines met een lagere status gaan vaker te rade bij disciplines met een hogere status. Als je alleen op de studie van Vermeulen zou afgaan is het misplaatste arrogantie van economen: de economische theorieën verklaren aanzienlijk minder goed de relaties tussen onderwijsdeelname en economie dan de sociologische.

Referenties:

CBS, 1993, *Standaard beroepenclassificatie 1992 ('s-Gravenhage: SDU)*
Bart Bakker. *Kijkduinstraat 36, 2014 DE Haarlem*

Maaïke Hajer

Leren in een tweede taal. Interactie in vakonderwijs aan een meertalige mavo-klas
Groningen, Wolters Noordhoff 1996.

ISBN 90 01 36830 1

Er is de laatste jaren veel belangstelling voor het leren in een tweede taal, met name in het vakonderwijs waar de leerstofoverdracht grotendeels verbaal (via doceren, teksten, met lezen en schrijfopdrachten etc.) plaatsvindt. Dit vanuit het besef dat we hier met een cruciaal punt te maken hebben als het gaat om onderwijskansen en schoolsucces van tweedetaalleerlingen. De recente dissertatie van Maaïke Hajer (1996) biedt naast een uitvoerige en nauwgezette case study van het vakonderwijs in het voortgezet onderwijs een uitgebreide theoretische studie van dit aandachtsgebied, dat in met name de angelsaksische wereld veel is bestudeerd en onderzocht.

Hajers titel van het theoriehoofdstuk: 'Leren in een tweede taal: probleem of kans?' geeft precies aan waar het in de (onderzoeks)literatuur steeds om gaat: enerzijds is er veel evidentie dat tweedetaalleerders serieuze problemen ondervinden op school, o.a. doordat ze moeten leren in de (voor hen tweede) taal die ze minder goed beheersen dan hun klasgenoten, anderzijds biedt dit onderwijs juist ook kansen voor taalverwerving. Wat de probleemkant betreft: leerlingen die van huis uit een andere taal spreken blijken ook na jarenlang Nederlandstalig onderwijs nog steeds problemen te hebben met de schooltaal en met schoolboekteksten. Met name blijken deze leerlingen over een geringere woordkennis te beschikken. Hajer geeft in dit eerste hoofdstuk een uitgebreid overzicht van het Nederlandse en Vlaamse onderzoek naar taalvaardigheid van allochtone leerlingen in het Nederlandstalige onderwijs, dat overigens schaars is als het gaat om onderinstromende leerlingen in het voortgezet onderwijs. Dit onderzoek heeft de aandacht gevestigd op het probleem van de 'schoolse taalvaardigheid', het leren in de zaakvakken en de rol van taal bij het vakonderwijs in het algemeen. Hajers onderzoek is een interessante aanvulling op dit (veld)experimentele onderzoek; zij wil in een intensieve case study inzicht verkrijgen in het complex van factoren die het leersucces van de

leerlingen bepalen, uitgaande van een dynamische visie op de schoolse context waarin interactie en 'betekenisonderhandeling' centraal staan.

Hiermee plaatst zij zich ook aan de andere kant van de tegenstelling 'probleem of kans'. Zij wil ook zicht krijgen op de kansen die het vakonderwijs biedt aan tweedetaalleerders om de taal beter te verwerven. De vakinhoudelijke besprekingen in de klas vormen ook een natuurlijke context voor het gebruik van de taal waardoor de (tweede)taalverwerving geoptimaliseerd zou kunnen worden, het zgn. 'content based language learning' (cbll). Deze laatste optie is interessant, voor zowel het NT2-onderwijs als voor het onderwijs in het algemeen. Juist in de klasse-interactie bij het vakonderwijs liggen kansen voor cbll, tenminste als er sprake is van een actieve inbreng van leerlingen in de interactie, in 'betekenisonderhandelingen' waardoor hun vakkennis en daarmee samenhangend hun taalvaardigheid in de vakken wordt vergroot. De notie 'betekenisonderhandeling' staat hierbij centraal. De achterliggende veronderstelling is dat onder optimale interactiecondities leerlingen taalverwervingskansen worden geboden, maar dat er problemen ontstaan bij een (kwantitatief en/of kwalitatief) gebrekkige klasse-interactie. Hajers onderzoek naar klasse-interactie is opgezet vanuit dit tweeledige perspectief op tweedetaalverwerving.

In het onderzoek worden vaklessen beschreven opdat inzichtelijk wordt in hoeverre er gunstige condities voor tweedetaalverwerving worden gecreëerd. De vragen die Hajer zich stelt zijn: (1) welke talige interactiesituaties doen zich voor, (2) in welke mate leveren leerlingen inbreng in de lessen en (3) welke kenmerken hebben de uitingen van leerlingen en docenten? Ook kijkt Hajer naar de ontwikkeling van centrale begrippen in de interactie (4) en het oordeel van docent en leerlingen over het taalgebruik in de vaklessen (5). De opzet is die van een intensieve case-study; ze beperkt zich tot het vakonderwijs in een 3 MAVO-klas van 26 leerlingen waarin het grootste deel van de leerlingen allochtoon is, door Hajer consequent aangeduid als tweedetaalleerders: 15 Marokkaanse, 1 Italiaanse, 1 Surinaamse, 1 Pakistaanse en 5 Turkse leerlingen. Behalve de Surinaamse leerling (en de 3 Nederlanders in de klas) hebben deze leerlingen het Nederlands als tweede taal geleerd vanaf de kleuterschool of wat later in de basisschoolperiode. De vaklessen die Hajer in deze klas observeert en opneemt betreffen aardrijkskunde (10 lessen), biologie (10), scheikunde (12), wiskunde (17) en natuurkunde (7). Daarnaast verzamelde ze per vak afgenomen proefwerken, kopieën van de schriften, de leerboeken. In de loop van de lessenreeksen werden de leerlingen geënquêteerd en in duo's geïnterviewd en na afloop van de lessenreeks is met de docenten een individueel gesprek gevoerd.

De opnamen van de lessen vormt de voornaamste bron van gegevens. Deze zijn getranscribeerd en geschikt gemaakt voor analyse volgens het CHAT-systeem (Codes for the Human Analysis of Transcripts), waardoor ook met behulp van de computer bepaalde analyses konden worden gemaakt (zoals de bepaling van de lengte van beurten en uitingen alsmede woordfrequenties). De uiting (als syntactische of prododische eenheid) is de belangrijkste analyse-eenheid, waarmee de inbreng van sprekers wordt bepaald. Teneinde de vergelijkbaarheid tussen de verschillende vaklessen mogelijk te maken, werden de lessen ingedeeld in fasen, waarbinnen weer lesfragmenten worden getypeerd naar setting (klassikaal, groepswerk of individueel werk) en de onderwijskundige functie van het fragment: introductie op de stof, inhoudelijke toelichting op de stof, demonstraties, dictaat, instructies, klassikale verwerking en niet-klassikale doe-opdrachten en verwerking. Binnen de lesfragmenten zijn nog weer communicatievormen getypeerd: docentbetoog, triades docent/leerling, onderwijsleergesprek, leerling geïnitieerde interactie, 'externe tekst dialoog' (wanneer uit het schoolboek wordt geciteerd), dialogen tussen leerlingen onderling en tussen docent en een enkele leerling. Deze indelingen werden in de transcripten gecodeerd, zodat bij de verdere analyses van de protocollen vergelijkingen binnen dezelfde categorieën mogelijk werd. Het gaat in de verdere analyse dan om karakterisering van functies van beurtreeksen (bv. directief of informerend) en van uitingen (bv. de vraag- of terugkoppelingsfunctie). Na de uitgebreide kwalitatief beschrijvende analyse volgt een kwantitatieve analyse per vak. Naast deze structureel/functionele analyse wordt ten slotte ook de inhoudelijke dimensie onderzocht door vaststelling van

de 'centrale begrippen' die bij de vijf vaklesreeksen werden behandeld (bij aardrijkskunde was dat bijvoorbeeld het thema 'water').

De conclusies van Hajer na de uitgebreide analyses: er is een groot verschil in didactische praktijk tussen de verschillende docenten. De docenten hebben elk hun eigen opvattingen over wat een goede doceerstijl is, ook met het oog op de allochtone leerlingen. De docente die zich het meest bewust is van de taalproblematiek laat de meest interactieve doceerstijl zien, en de leerlingen presteren bij haar goed. Andere docenten passen hun doceerstijl aan door de eisen bewust te verlagen, bijvoorbeeld door ingewikkelde opdrachten over te slaan, terug te vallen op feitenkennis en/of door het leerboek aan de kant te leggen en de meest basale stof te dicteren. De leerlingen blijken vooral handig in overlevingsstrategieën: ze passen hun leergedrag aan aan de doceerstijl van een docent. Ze weten dat je bij de één een voldoende kunt halen door dictaten uit het hoofd te leren terwijl je bij de ander mee moet doen in de lessen omdat hier de stof 'goed wordt uitgelegd'.

Het zal duidelijk zijn dat de scientieuzere dataverzameling en analyse een immense klus is geweest, waardoor we nu beschikken over een unieke documentatie van wat er zich afspeelt op het microniveau van een klas. De hamvraag bij de beoordeling van de waarde van dit documentaire materiaal is natuurlijk in hoeverre hier een representatief beeld wordt geschetst. Het proefschrift laat zich hier en daar lezen als een spannende reportage, waar de meeste lezers waarschijnlijk veel in herkennen. Wie herinnert zich niet de leraar die z'n lessen nooit voorbereide en die voor de vuist weg z'n lessen afdraaide, blind en doof voor pogingen van leerlingen er iets van te begrijpen. Of de leraar die in grote rust en orde dictaten gaf, waarbij een inbreng van de leerlingen alleen op gezette momenten en dan in beperkte mate mogelijk was? Of de leraar die niet snel genoeg door de stof heen kwam, maar in gesprek met de klas oog had voor de begripsproblemen van leerlingen? Een 3-mavoklas waarvan de meeste leerlingen tweede-taalverwerfers zijn is niet representatief voor 'een klas', in een gemiddelde klas zitten maar een paar zulke leerlingen. Het merendeel van de leerlingen uit deze klas blijkt na twee jaar gestrand te zijn, slechts een klein aantal behaalt het mavo-diploma. Het ging Hajer echter om een beeld van 'didactische praktijk in het zaakvakonderwijs' in een (toch niet extreme) situatie waarin deze op de proef wordt gesteld. De didactische praktijk die we aantreffen in deze klas is waarschijnlijk niet uitzonderlijk in het Nederlandse voortgezet onderwijs. Het verband met de (begrips)problemen die deze praktijk voor met name allochtone leerlingen met zich meebrengt is onthullend, zeker nu we weten hebben van hun grote schooluitval.

Hilde Hacquebord
Vakgroep Taal en Communicatie, Faculteit der Letteren
Rijksuniversiteit Groningen

H.B.G. Ganzeboom & W.C. Ultee

De sociale segmentatie van Nederland in 2015

Nr. 96 van de reeks Voorstudies en achtergronden van de Wetenschappelijke Raad voor het Regeringsbeleid.

Sdu Uitgevers, Den Haag, 1996. ISBN 90-399-1275-0

De rapporten van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) hebben in de loop van de tijd steeds meer een politiek en steeds minder een wetenschappelijk karakter gekregen. Het recent verschenen WRR-rapport *Tweedeling in perspectief* illustreert dit verschijnsel goed, omdat het vergelijkbaar is met een in 1977 verschenen WRR-rapport getiteld *Over sociale ongelijkheid - een beleidsgerichte probleemverkenning*. Deze verschuiving maakt WRR-rapporten voor het wetenschappelijk bedrijf minder interessant, behalve als men geïnteresseerd is in actuele beleidstheorieën.

Deze verwording van de WRR betekent echter niet dat alle producten van de WRR voor

de wetenschap minder interessant zijn geworden. Met name de voorstudies die de WRR laat uitvoeren kunnen interessant zijn, omdat deze vaak wel het evenwicht hebben behouden tussen de politieke wenselijkheid en de objectiviteit van de wetenschap.

Voor voorbeeld van een dergelijke nuttige voorstudie is het rapport over de sociale segmentatie van Nederland in 2015, onder redactie van de sociologen Ganzeboom en Ultee. Voor sommige onderwijskundigen zullen dit wellicht onbekende namen zijn, maar zij behoren beiden tot de prominente onderzoekers op het terrein van de sociale ongelijkheid, zowel in Nederland als daarbuiten. Dat maakt dit rapport voor onderwijskundigen interessant en rechtvaardigt deze bespreking: het rapport kan hun een gezaghebbend beeld van de plaats van het onderwijs binnen de sociale segmentatie van Nederland geven.

Hoewel het rapport een toekomstverkenning is (de derde probleemstelling luidt: welke ontwikkelingen zijn er tot 2015 te verwachten wat betreft mate en inrichting van de sociale segmentatie), is het in hoofdzaak toch een sterk empirisch werkstuk geworden. Dat komt omdat de redacteurs terecht hebben afgezien van de gebruikelijke methode van toekomstverkenning, waarbij met name wordt vertrouwd op de voorzienende blikken van geraadpleegde deskundigen en vervolgens verschillende deels tegenstrijdige 'scenario's' naar buiten worden gebracht. Zij menen dat deze gebruikelijke werkwijze met wetenschappelijk onderzoek niet zoveel heeft te maken: de hierboven geschetste verschuiving in het karakter van de WRR-rapporten ondersteunt hun standpunt. Daarom beginnen de redacteurs met twee veel meer beschrijvende probleemstellingen: 1. In welke mate is er sprake van sociale segmentatie en in hoeverre vindt deze plaats langs lijnen van opleiding en inkomen? 2. Welke langjarige trends zijn er voor het verleden vast te stellen en hoe kunnen deze verklaard worden? De redacteurs hopen vanuit deze kennis omtrent heden en verleden de toekomstige ontwikkelingen verantwoord te kunnen schetsen.

De redacteurs, tezamen met een keur van auteurs uit twee onderzoekscholen ICS en NETHUR, behandelen deze drie probleemstellingen in negen hoofdstukken waarin verschillende aspecten van de segmentatie aan de orde komen: onderwijsexpansie en onderwijskansen; inkomensongelijkheid van huishoudens; trouwpatronen, afgemeten aan opleidingsniveaus van partners; onderwijs en arbeidsmarkt; bewoningsspatronen; vrijetijdsbesteding; politieke scheidslijnen; ethnocentrisme en rechts-extremisme; rechtvaardiging van inkomensverschillen.

Hoewel sommige hoofdstukken voor onderwijskundigen interessanter zijn dan andere, is het opvallend dat in bijna alle hoofdstukken onderwijs als een belangrijke verklarende variabele optreedt. Dat betekent voor de onderwijskunde dat zij niet eindigt bij het gedrag van leerlingen of leerkrachten, de deuren van de school of zelfs de grenzen van het onderwijsstelsel, maar dat de onderwijskunde een echte interdisciplinaire wetenschap behoort te zijn. De conclusie van Ultee in het afsluitend hoofdstuk belicht nog eens het belang van het onderwijs in onze huidige samenleving: "In deze studie is duidelijk geworden dat in Nederland met name de opleiding het overheersend segmentatiecriterium is geworden. Ondanks de informatisering van de maatschappij, terugtrekkende overheden en flexibilisering van markten verwachten we echter geen sterke toename van segmentatie als gevolg van opleidingsverschillen tussen mensen. De reden hiervoor is dat de verhouding tussen het aanbod van mensen een bepaalde opleiding en de vraag naar juist mensen met die opleiding zodanig is dat zich meer mensen met de hoogste opleiding aanbieden dan er vraag naar mensen met de hoogste opleiding is."

Dit citaat laat een motief zien dat in dit rapport een grote rol speelt, namelijk dat de groei van het onderwijsstelsel sterker is geweest dan de toename in het niveau van de beschikbare functies. Deze toestroom naar het onderwijs heeft volgens Ultee als onbedoeld gevolg in twee opzichten tot geringere ongelijkheid geleid. Ten eerste is door de onderwijsexpansie de ongelijkheid in onderwijskansen tussen de verschillende milieus teruggelopen. Interessant is dat Ganzeboom in zijn hoofdstuk laat zien dat dit onbedoelde gevolg van de onderwijsexpansie belangrijker voor de verkleining van de onderwijsongelijkheid is geweest dan allerlei onderwijsexperimenten rond selectie en toelating in het voortgezet onderwijs. Ten tweede leidt de

onderwijsexpansie tot geringere inkomensverschillen. De Graaf laat in zijn hoofdstuk zien dat deze geringere inkomensverschillen passen in een ontwikkeling van een afnemende marginale waarde van opleidingen. De snelle onderwijsexpansie en de veel minder snelle regradatie van de werkgelegenheid heeft er blijkens zijn analyse ervoor gezorgd dat vooral de lagere diploma's minder waard zijn geworden. Het ziet er echter volgens De Graaf naar uit dat nu ook de hogere onderwijskwalificaties aan de beurt zijn voor ontwaarding. Om evenveel succes te bereiken op de arbeidsmarkt is een steeds hoger opleidingsniveau nodig. De logica van de huidige samenleving is dat deze onderwijsexpansie de jongere generaties ertoe dwingt alleen met het hoogste onderwijsniveau genoeg te nemen. Deze dwang tot hoger scholen werkt averechts uit volgens De Graaf, want de kans op succes in de arbeidsmarkt daalt omdat steeds meer mensen deze strategie van een zo hoog mogelijk onderwijsniveau volgen.

Ook het hoofdstuk van Jansen en Mulder over de samenhang van opleidingsniveau van partners is voor onderwijskundigen belangrijk. Beide auteurs laten zien dat de kans op huwelijks-homogamie (een gelijk onderwijsniveau van beide partners) vele malen groter is dan de kans op huwelijks-heterogamie (een ongelijk onderwijsniveau van beide partners). Maar nog interessanter is dat het verschijnsel in de periode 1950 tot 1990 niet systematisch veranderd is, ondanks de enorme onderwijsexpansie, het inlopen van de onderwijsachterstand door vrouwen en de veranderingen in de machtsbalans tussen mannen en vrouwen. Wel vinden de auteurs fluctuaties, maar deze blijken trendloos en onsystematisch te zijn.

Kortom, het rapport geeft onderwijskundigen een empirisch goed gefundeerd beeld van de rol van het onderwijs bij de opbouw van de maatschappelijke ongelijkheid en geeft ze daarvoor een ruimer uitzicht op het belang van onderwijs in de gehele samenleving.

Uiteraard blijft er voor elke boekbespreker wel wat te mopperen over. Zo bestaat er een grote overlap tussen dit rapport onder redactie van Ganzeboom en Ultee en een iets eerder verschenen boek onder redactie van Dronkers en Ultee, waarin ongeveer dezelfde onderwerpen aan de orde komen. De lezer die het boek reeds kent, zou zich door deze overlap bekocht kunnen voelen bij de aanschaf van dit rapport. Ook kiest Ganzeboom bij zijn bespreking van het onderzoek van trends in het ongelijkheidsonderzoek met behulp van schoolloopbaan-cohorten een te gemakkelijke tegenstander: een SCP-rapport uit 1992. Op dat terrein is sinds 1992 meer en vooral beter gepubliceerd dan dit samenvattende SCP-rapport, waardoor een groot aantal van de terechte bezwaren van Ganzeboom allang achterhaald zijn. Ten slotte laat dit rapport ook zien dat onderzoekscholen niet altijd een zegen zijn voor de wetenschaps-beoefening; het kan leiden tot sluitingsprocessen ten opzichte van de tradities en de publicaties buiten de school waartoe men zelf behoort.

Deze bespreker heeft ten slotte een grote onbeantwoorde vraag, meer gericht aan de WRR dan aan de auteurs van deze voorstudie. Hoe kon de WRR gezien de inhoud van deze voorstudie komen tot haar optimistisch advies over de toekomstige tweedeling van de samenleving en het grote belang van werk daarbij?

*J. Dronkers
SCO/Kohnstamm Instituut,
Universiteit van Amsterdam*

Inhoud

Drop-out of startkwalificatie / **pag. 211**

W. de Witt en H. Dekkers

Op welk niveau verder studeren na VWO of HAVO; veranderingen tussen 1982 en 1991 / **pag. 228**

H.D. Webbink

Invloed van het vakkenpakket VWO op het studieresultaat in de propedeuse WO / **pag. 238**

E.P.W.A. Jansen

Delfts blauw en rose; studieloopbanen van mannen en vrouwen aan een technische universiteit / **pag. 249**

E. van Eck, P. Koopman en M. Robijns

Centreren in multilevelanalyse: implicaties van twee centeringsmethoden voor het bestuderen van schooleffectiviteit / **pag. 264**

M.C. Opendakker en J. Van Damme

Notities en Commentaren

H. Oosterbeek en D. Webbink - Is verkleining van klassen een goede investering? / **pag. 291**

R.J. Bosker - De empirische bewijslast voor de extra uitgaven in het primaire onderwijs / **pag. 294**

H.A.L. Kiers en W.K.B. Hofstee - Gewogen loting in combinatie met rechtstreekse toelating vanaf examencijfer 8 / **pag. 298**

Boekbesprekingen

J.B. Kuhlemeier - Taalvaardigheid, taalactiviteiten en taalattituden

Door G. Rijlaarsdam / **pag. 302**

VOR-bulletin / **pag. 309**

Publikaties in jaargang 1997 / **pag. 316**

Lijst van redactionele medewerkers in 1977 / **pag. 319**

Drop-out of startkwalificatie

Omvang, achtergronden en gevolgen van voortijdig schoolverlaten in de eerste vijf jaren voortgezet onderwijs

Walter de Wit & Hetty Dekkers

ABSTRACT

Als onderdeel van de evaluatie van het Onderwijsvoorrangsbeleid is een groep van 5000 leerlingen gevolgd in het Voortgezet Onderwijs. Bij deze groep wordt apart onderzoek uitgevoerd naar voortijdig schoolverlaten. In het onderzoek is het schoolverlaten binnen vier respectievelijk vijf jaar Voortgezet Onderwijs nauwkeurig nagegaan; ook de samenstelling van de groep naar achtergrondkenmerken als etnische herkomst, sekse, sociaal-economische achtergrond en voorafgaande schoolloopbaan kenmerken is vastgesteld. Daarnaast is bij de leerlingen zelf nagegaan wat de redenen en achtergronden zijn die hen deden besluiten voortijdig iedere vorm van (regulier) onderwijs af te wijzen. Ook de situatie na het schoolverlaten was onderwerp van onderzoek. De groep voortijdig schoolverlaters blijkt relatief veel allochtone leerlingen te bevatten, vooral bij de ongediplomeerde drop-out. Voor schoolverlaten binnen vier jaar worden eerder schoolinterne redenen opgegeven, voor latere drop-out meer 'pull' vanuit arbeidsmarkt of gezin (voor meisjes). Praktijkgerichte opleidingen lijken voor een groot deel van de betreffende jongeren soelaas te kunnen bieden.

1. INTRODUCTIE

Voortijdig schoolverlaten is traditioneel vooral benaderd vanuit het perspectief van sociale ongelijkheid en sociale integratie. Al lang is bekend dat het met name arbeiderskinderen en allochtone kinderen zijn die het onderwijs voortijdig, dat wilde zeggen zonder diploma, de rug toekeren. Sinds het begin van de jaren tachtig is beleidsmatig de aandacht voor voortijdig schoolverlaters enigszins verschoven van het probleem van sociale ongelijkheid naar het probleem van arbeidsmarkt- en kwalificatiebehoeften. Dit leidde tot een geheel nieuwe definitie van voortijdig schoolverlaten, gebaseerd op het begrip 'startkwalificatie'; het minimum-kwalificatieniveau waaraan alle leerlingen zouden moeten voldoen. In de schoolverlatersonderzoeken binnen de evaluatie van het Onderwijsvoorrangsgebied (OVb) zijn beide perspectieven opgenomen. In dit artikel beschrijven we de resultaten van de onderzoeken naar het voortijdig schoolverlaten binnen de eerste vijf jaar van het voortgezet onderwijs.

In het kader van de evaluatie van het Onderwijsvoorrangsbeleid wordt regelmatig onderzoek uitgevoerd onder ruim 5000 leerlingen in het voortgezet onderwijs; deze leerlingen maken deel uit van het cohort dat eerder werd gevolgd in het basisonderwijs. Sinds 1993 is binnen die VO-onderzoeken ook aandacht voor drop-outs en worden ook deze gevolgd. Het eerste onderzoek richtte zich op de drop-out binnen de eerste 4 jaar voortgezet onderwijs. Geconstateerd werd dat de problematiek minder ernstig was dan aanvankelijk werd gedacht, zowel wat betreft de omvang ervan als de situatie waarin de drop-outs belandden (De Wit & Dekkers, 1994). In vervolg hierop is onderzoek uitgevoerd naar schoolverlaten binnen 5 jaar voort-

gezet onderwijs. Het betreft hier een meer gedifferentieerde groep, omdat er ook schoolverlaters met een LBO- of MAVO-diploma toe behoren. Gegevens over de omvang van het voortijdig schoolverlaten, uitgesplitst naar onder meer etniciteit en sekse, werden gecombineerd met gegevens over de oorzaken en achtergronden ervan en de situatie na het schoolverlaten. We bespreken een aantal van de belangrijkste resultaten en conclusies. We beschrijven achtereenvolgens de definitie van voortijdig schoolverlaten (§2), het theoretisch kader (§3), de probleemstelling (§4), de onderzoeksopzet en analyse (§5), de resultaten (§6) en een slotbeschouwing (§7).

2. DEFINIËRING VAN VOORTIJDIG SCHOOLVERLATEN

Diverse auteurs hebben erop gewezen dat het moeilijk is een definitie te geven van een voortijdig schoolverlater of drop-out (Gaustad, 1991). In 1993 is in de Verenigde Staten de National Drop-out Statistics Review Task Force opgericht, met als doel tot een goede definitie te komen (Edirisooriya, 1995). Er worden voornamelijk twee belangrijke perspectieven onderscheiden: sociale ongelijkheid en onderwijskwalificatie. Het is echter ook mogelijk voortijdig schoolverlaten te definiëren in termen van wat werkgevers acceptabele minimum-niveaus achten voor onderwijs en opleiding van jonge mensen die de arbeidsmarkt instromen (Hannan, Hövels, Van den Berg & White, 1995). Zo variëren drop-out cijfers naargelang de invalshoek. Rumberger (1987) beweert dat in de Verenigde Staten niemand weet hoe hoog het aantal drop-outs is, en dat er geen standaard methode is om dat te berekenen. Veel gehanteerd zijn leeftijd-cohorten en gegevens over doorstroom (bijvoorbeeld entree en na vier jaar eindexamen high-school).

Van het begrip 'voortijdig schoolverlaten' zijn ook in Nederland verschillende definities in omloop. De meer traditionele definitie van een voortijdig schoolverlater is: een leerling die zonder *diploma* de eerste fase van het voortgezet onderwijs verlaat. Belangrijkste criterium is hierbij dus het behalen van het diploma.

Van vrij recente datum is de koppeling van voortijdig schoolverlaten aan de *kwalificatienorm*. De overheid stelt dat ieder individu een (inspannings)verplichting heeft om zich te scholen tot het niveau van de primaire startkwalificatie. Voortijdig schoolverlaters zijn in deze betekenis jongeren die zonder primaire startkwalificatie het onderwijs hebben verlaten. De startkwalificatie staat gelijk aan minimaal een beroepsopleiding op eindniveau primair Leerlingwezen of kort-MBO, twee MBO-jaren voltooid, of een HAVO- of VWO-diploma behaald. Buiten de definitie vallen dus de leerlingen die zonder diploma VBO/AVO/VWO overstappen naar een andere vorm van kwalificerend vervolgonderwijs (LW/MBO); leerlingen met LBO- of MAVO-diploma die verder geen onderwijs volgen vallen er wel onder.

In dit artikel is cijfermateriaal opgenomen over diverse definities, maar is prioriteit gelegd bij de huidige overheidsdefinitie. Dit betekent dat de groep voortijdig schoolverlaters zowel jongeren omvat die ongediplomeerd het onderwijs hebben verlaten, als jongeren die na het behalen van het VBO- of MAVO-diploma het onderwijs verlieten. Er is dus sprake van ongediplomeerde en gediplomeerde voortijdig schoolverlaters.

3. OORZAKEN VAN VOORTIJDIG SCHOOLVERLATEN: THEORETISCH KADER

Voortijdig schoolverlaten is een veelzijdig fenomeen; er zijn veel factoren die direct of indirect met het verschijnsel samenhangen. Het is niet altijd eenvoudig oorzaak en gevolg te onderscheiden. Ook bestaan er nog geen volledig adequate modellen om drop-out te voorspellen.

Diverse auteurs onderscheiden verschillende categorieën factoren die samenhangen met

voortijdig schoolverlaten. Meestal betreffen de categorieën demografische factoren, persoonlijke factoren, gezinsachtergrond, leeftijdgenoten, schoolfactoren, en soms meer macrofactoren zoals bijvoorbeeld de situatie op de arbeidsmarkt (Rumberger, 1987, Coley, 1995, Gaustad, 1991, de Vries, 1993). Pogingen om samenhangende modellen te ontwikkelen voor de beschrijving van de factoren of processen die leiden tot voortijdig schoolverlaten zijn nog niet voldoende succesvol. Overigens betreft de literatuur altijd onderzoek naar schoolverlaten voordat een diploma is behaald, en dus niet de (Nederlandse) startkwalificatiedefinitie.

In deze paragraaf beschrijven we de (combinaties van) factoren die onderdeel waren van dit onderzoek.

Factoren

Factoren op leerlingniveau

Rumberger (1987) geeft in zijn review over high school drop-out eenzelfde classificatie als wij hierboven gaven, en noemt als persoonlijke kenmerken die samenhangen met voortijdig schoolverlaten: lager zelfvertrouwen, minder gevoel van controle over het eigen leven, een negatieve houding tegenover school, lage onderwijs- en beroepsaspiraties en de wens om te trouwen. Coley (1995) concludeert uit analyses op de NELS 1988 (National Educational Longitudinal Study) data met betrekking tot attitudes en verwachtingen van leerlingen dat drop-outs van mening zijn dat ze geen controle hebben over hun eigen leven, dat veel van toeval en geluk afhangt, en dat er altijd iets voorviel dat hen verhinderde door te gaan. Ondanks hun schoolverlaten was toch 85 procent van de drop-outs optimistisch over het alsnog verwerven van ten minste een high school diploma. Nederlands onderzoek toont aan dat voortijdig schoolverlaten ook gerelateerd is aan capaciteiten van leerlingen. Scores op intelligentie- en prestatietoetsen van voortijdig schoolverlaters blijven gemiddeld achter bij die van leerlingen die de opleiding voltooien (Bosker & Hofman, 1994; De Vries 1993). Ook de motivatie en schoolbeleving van leerlingen lijken een grote rol te spelen: toekomstige schoolverlaters hechten minder belang aan school dan blijvers en hebben minder hoge aspiraties met betrekking tot het te bereiken onderwijs- en beroepsniveau. Zij oordelen bovendien negatiever over hun eigen schoolbekwaamheid en vinden het vervelender om naar school te gaan. Dit uit zich gedragsmatig in een geringere inzet tijdens de lessen en een minimale tijdsinvestering in huiswerk (Hofman, 1993; De Vries 1993). Demografische individuele factoren die samenhangen met drop-out betreffen etnische en gender factoren: leerlingen uit etnische minderheden en jongens verlaten vaker voortijdig de school. Leeftijd is een andere factor die van belang kan zijn bij voortijdig schoolverlaten. Een laatste persoonlijke factor betreft de mate van schoolverzuim. Frequente afwezigheid gaat behalve met voortijdig schoolverlaten bovendien vaak gepaard met slechte prestaties (Hofman 1993; De Vries 1993).

Factoren op gezinsniveau

In de onderzoeksliteratuur wordt het gezin (en in samenhang hiermee milieu) aangemerkt als belangrijkste tussenschakel ter verklaring van spijbelen en voortijdig schoolverlaten. De thuissituatie van voortijdig schoolverlaters wordt gekenmerkt door grote gezinnen, éénoudergezinnen, werkloosheid van de vader en materiële belemmeringen. Gaustad (1991) noemt sociaal milieu de belangrijkste factor in dit verband. 'Ouders van uitvallers zijn gemiddeld lager opgeleid, hebben lager gestelde aspiraties en toekomstverwachtingen ten aanzien van hun kind en zijn minder bij het onderwijs betrokken dan ouders van blijvers.' Bij uitvallers is er bovendien vaker sprake van een discrepantie tussen de communicatie, opvoeding en disciplineren in de thuissituatie en de waarden en normen van de school (De Vries 1993).

Vrijwel alle hier genoemde verklaringsgronden gelden in versterkte mate voor allochtone gezinnen. Met name de discrepantie tussen thuiscultuur en schoolcultuur wat betreft sociale omgang kan van invloed zijn op het schoolwelbevinden van leerlingen. Daarnaast spelen taalproblemen, de mate van oriëntatie op integratie en met name voor meisjes, traditionele beroepspectieven een rol (Dekkers, 1993; De Vries 1993).

Factoren op peergroup-niveau

Een belangrijke factor met betrekking tot de peergroup van leerlingen is de mate van verzuim. Vrienden en/of vriendinnen van uitvallers verzuimen veel vaker dan vriend(inn)en van blijvers (Rumberger, 1987). Voor de binding met de school is het van belang of de peergroup bestaat uit mede-leerlingen van dezelfde school. Voortijdig schoolverlaters gaan vaker om met vrienden van buiten de eigen school (Hofman, 1993). Er zijn aanwijzingen dat spijbelen met mede-leerlingen van dezelfde school eerder een schoolbindende functie heeft en dat vooral spijbelen in buitenschoolse vriendennetwerken voorafgaat aan voortijdig schoolverlaten (Hövels & Bock, 1991). Onder meer Hofman (1993) constateert dat een multi-etnische samenstelling van, en een niet-Nederlands taalgebruik binnen de peergroup samenhangen met voortijdig schoolverlaten. Waarschijnlijk zijn deze laatste peergroupkenmerken echter rechtstreeks terug te voeren op het hoge percentage uitvallers onder allochtone leerlingen.

Factoren op schoolniveau

Het is belangrijk te verhelderen welke schoolfactoren relevant zijn in het kader van voortijdig schoolverlaten, omdat vooral deze door praktijk en beleid kunnen worden beïnvloed. Verschillende auteurs wijzen erop dat slechte schoolprestaties en gedragsproblemen op school, inclusief spijbelen en disciplineproblemen zoals genoemd onder persoonlijke factoren, samenhangen met voortijdig schoolverlaten. Er is echter weinig aandacht besteed aan de invloed van de school zelf: de organisatie, het management en de leerkrachten. Er blijken grote verschillen te bestaan tussen scholen onderling, zelfs na correctie voor het sociaal milieu van de instromende leerlingen. Wu (1992) rapporteert in dit verband dat tot nu toe meer naar voortijdig schoolverlaten is gekeken als een individueel probleem dan als een schoolprobleem, terwijl de mate waarin leerlingen juist op bepaalde scholen uitvallen sterk gerelateerd is aan omgevings- en organisatiekenmerken van die scholen. Analyse van de NELS 1988 data leidt hem tot de conclusie dat de omvang van de drop-out van een bepaalde school is gerelateerd aan het opleidingsprogramma en de uitvoering daarvan. De gevonden relatie is overigens meer associatief dan causaal. Niet duidelijk is welke specifieke factoren hieraan bijdragen. In ieder geval is er minder uitval op scholen waar leerlingen (al dan niet individueel en zowel binnen als buiten het lesrooster) extra worden begeleid. Daarnaast speelt het schoolklimaat een belangrijke rol: leerlingen vallen minder uit op scholen met een 'warm' schoolklimaat, waar minder sancties worden toegepast wegens wangedrag van leerlingen, waar zich minder conflicten (al dan niet gevoed door etnische tegenstellingen) voordoen en waar prestaties positief worden gewaardeerd (De Vries, 1993). Ten slotte zijn er aanwijzingen dat een hoge frequentie van leerkracht-leerling interacties, zoals vragen stellen en feedback geven, samengaat met minder uitval (Hofman, 1993).

Combinaties van factoren*Push-pull*

Verschillende auteurs combineren een aantal van de bovengenoemde factoren tot push- en pull-factoren die van invloed zijn op voortijdig schoolverlaten. Jordan et al (1994) geven als theoretische visie: dropping out of school as a cumulative process arising out of a series of events and experiences which impact on youth into their transition into adolescence'. Zij spreken over push-factoren als het om factoren binnen de school (internal) gaat, en om pull-factoren als het om externe factoren gaat, zoals familie, wijk, leeftijdgenoten (peergroup), werk etc.. Andere auteurs categoriseren push en pull meer in termen van de leerling. Bij afstotingsmechanismen onderscheidt De Vries (1993) drie subgroepen leerlingen: leerlingen die de school verlaten wegens een gebrek aan capaciteiten, tegenvallende leerprestaties en een (allochtone) leerlingen die de school verlaten als gevolg van culturele danwel sociale isolatie en/of discriminatie. Bij aantrekkingsmechanismen is er sprake van factoren van buiten het onderwijs die een aanzuigende werking hebben op leerlingen. Een grote zuigkracht kan uit-

gaan van de arbeidsmarkt, met name van de lagere en ongeschoolde segmenten (Hövels & Bock, 1991). Ook het stichten van een gezin of economische ondersteuning van de ouders kunnen tot zogenaamde pull-factoren worden gerekend.

De eerdergenoemde Amerikaanse NELS data bevatten een lijst van 21 redenen om de school te verlaten. School-gerelateerde redenen om de school te verlaten (push) werden veel vaker dan andere categorieën genoemd; die andere betroffen familie/milieu, werk, veiligheid, vrienden, mobiliteit en schorsing. De schoolgerelateerde items waren: hekel aan school, hekel aan leerkrachten en het zich niet thuis voelen en falen op school. Mannelijke drop-outs noemden vaker dan vrouwelijke werk als reden. Meisjes daarentegen noemden vaker familieomstandigheden zoals zwangerschap, ouderschap, of zorg voor een familielid als reden om de school voortijdig te verlaten (Jordan et al. (1994)).

Modellen

Al bijna tien jaar geleden wees Rumberger (1987) erop dat een meer omvattend, causaal model nodig was, dat 'should successfully identify the full range of proximal and distal influences, the interrelationships among them, and their long-term, cumulative effect'. Aandachtspunten moesten volgens hem hierbij zijn: de onderliggende processen (bijvoorbeeld een toenemende afname van betrokkenheid met school, door sociale of academische redenen), de onderlinge samenhang tussen de verschillende variabelen (oorzakelijk danwel associatief), de lange termijn, cumulatieve effecten van de verschillende invloeden (gezinsachtergrond, eerdere schoolprestaties) en de verschillende typen drop-out (de 'typische' drop-out bestaat niet).

Finn (1989) schetst twee modellen voor een beter begrip van voortijdig schoolverlaten als een ontwikkelingsproces. Volgen het klassieke 'frustration-selfesteem' model leiden slechte schoolprestaties tot een beschadigd zelfbeeld wat weer leidt tot gedragsproblemen. Het 'participation-identification' model behelst de betrokkenheid van de student bij de school. Participatie in schoolactiviteiten wordt gezien als essentieel voor positieve resultaten, inclusief het gevoel van de leerling erbij te horen en de waardering van schoolgerelateerde doelen.

Noch door Rumberger, noch door Finn of enig ander auteur is dit modellen denken uitgewerkt in empirisch onderzoek. We moeten concluderen dat er tot nu toe geen valide samenhangend verklaringsmodel bestaat om voortijdig schoolverlaten te beschrijven of analyseren. Tot nu toe zijn alleen heel algemene ideeën gepresenteerd, uitgaande van de verschillende categorieën zoals hierboven beschreven.

Verdere (school)loopbaan

Er is weinig bekend over het verdere verloop van de school- en arbeidsloopbanen van voortijdig schoolverlaters. Wat is een leerling gaan doen, hoe beoordeelt hij zijn schoolverlaten en hoe ziet hij zijn toekomst? Diverse auteurs vragen zich af of voortijdig schoolverlaten niet als veel te negatief wordt gezien en of het belang van diploma's niet wordt overgewaardeerd. Jordan et al. (1994) analyseerden de verdere onderwijsplannen van voortijdig schoolverlaters. De grote meerderheid van de drop-outs was nog steeds van plan in de toekomst een high-school-diploma te halen, maar ze kozen een andere route om dat te bereiken; de routes verschilden afhankelijk van etniciteit en sekse. Ook Chuang (1997) concludeert dat een aantal leerlingen weer terugstroomt in een of andere vorm van onderwijs, vooral de jongere leerlingen en degenen die betere leerresultaten hadden in het voortgezet onderwijs.

Hanssen et al (1995) merken op dat '... we tend to see failure in causal terms ('because of' statements), and success in terms of individual agency ('in order to' statements)'. Zij pleiten voor een invalshoek waarin plaats is voor individuele intentionaliteit en weloverwogen keuzen als verklaring voor voortijdig schoolverlaten. Zij waarschuwen ook voor te grote nadruk op tussen-groep vergelijkingen, terwijl binnen-groep verschillen mogelijk tot meer inzicht kunnen leiden.

Andere auteurs benadrukken de negatieve effecten van voortijdig schoolverlaten. Rum-

berger (1987) claimt dat individuele drop-outs er slecht aan toe zijn omdat velen moeilijk vast en/of goed betaald werk kunnen vinden, niet alleen als ze van school afgaan, maar gedurende hun hele leven.

Schwartz (1995) vat Coley (1995) samen die ongeveer hetzelfde beweert, namelijk dat het verschil tussen drop-outs en beter opgeleiden in salaris steeds groter worden omdat er sprake is van een cumulatief proces.

Relevant in relatie tot de verdere carrière van voortijdig schoolverlaters is de opvang en de begeleiding van de drop-out tijdens en direct na het schoolverlaten. Wie ondersteunt hen daarbij? Wie begeleidt hen richting eventuele scholing en arbeid? In de literatuur zijn weinig aanknopingspunten te vinden met betrekking tot dit aspect.

4. ONDERZOEKSVRAGEN

Het theoretisch framework, in combinatie met het feit dat een aantal data al beschikbaar waren uit het onderzoek in het kader van de landelijke evaluatie OVB (LEO) in het primair en secundair onderwijs, leidt voor het onderzoek tot een kwantitatief en een kwalitatief gedeelte.

Van de leerlingen uit het schoolverlatersonderzoek zijn veel gegevens al voorhanden, omdat ze deel uitmaken van het LEO-cohort: leerlingen die vanaf groep 8 in het schooljaar 1988/1989 (cohort 88-8) jaarlijks in het voortgezet onderwijs worden gevolgd. In het kader van het Onderwijsvoorangsbeleid worden aan de doelgroepen wegingsfactoren toegekend (OVB-wegingsfactor). De belangrijkste doelgroepen zijn de Nederlandse arbeiderskinderen, zij hebben wegingsfactor 1.25 en de allochtone arbeiderskinderen, met een wegingsfactor 1.90. Leerlingen die niet tot de doelgroepen behoren, tellen als 1.00.

In het betreffende databestand zitten zowel persoonlijke achtergrondgegevens als schoolloopbaan kenmerken. Behalve dat via deze landelijke steekproef van ruim 5000 leerlingen nauwkeurig het schoolverlaten binnen een jaargroep kan worden vastgesteld, is deze dus te relateren aan (een combinatie van) de achtergrondgegevens.

Dit leidt tot de eerste kwantitatieve onderzoeksvraag:

Vraag 1: Hoe vaak komt het voor dat leerlingen binnen vijf jaar het voortgezet onderwijs voortijdig verlaten? En hoe groot zijn de percentages uitgesplitst naar een aantal leerling-, gezins- en school(loopbaan)kenmerken (*seks, herkomstland, OVB-wegingsfactor en opleidingsniveau van de ouders, schooltype, jaar van uitval, advies basisonderwijs*)?

Naast het vaststellen van percentages schoolverlaters binnen diverse categorieën leerlingen, is het ook mogelijk om de voortijdig schoolverlaters als uitgangspunt te nemen en de samenstelling van deze groep te beschrijven, zowel naar een aantal persoons- en gezinskenmerken, als naar enkele schoolloopbaan kenmerken (waaronder diploma-bezit).

Vraag 2: Hoe is de samenstelling van de groep voortijdig schoolverlaters? Om wat voor leerlingen gaat het en wat voor een schoolloopbaan hebben ze achter de rug (vgl. *kenmerken vraag 1*)?

En in hoeverre onderscheidt deze groep zich van de groep leerlingen die *niet* tot de schoolverlaters behoren?

Uit de statistische analyses binnen het kwantitatieve deelonderzoek ontstaat een nog vrij oppervlakkig beeld van de achtergronden en (een aantal) oorzaken van het voortijdig schoolverlaten. In het tweede deel van het schoolverlatersonderzoek, het kwalitatieve deelonderzoek, zijn we dieper ingegaan op deze achtergronden en oorzaken. De afgelopen tien jaar is veel gepubliceerd over dit onderwerp. Opvallend echter is het gemis aan onderzoek naar de perceptie

van de voortijdig schoolverlaters zelf, en naar de feitelijke situatie waarin deze na het schoolverlaten terecht komen. De voortijdig schoolverlater wordt in de literatuur meestal gekenschetst als een leerling die in het onderwijs volledig is mislukt en vanuit allerlei negatieve omstandigheden de school verlaat. Voortijdig schoolverlaten wordt beschouwd als een negatieve, passieve en ongewenste keuze. Het voortijdig schoolverlaten zou ook gezien kunnen worden als een actief handelen van de leerling om de eigen, als negatief ervaren, situatie te verbeteren. Wanneer een leerling de school verlaat om een ander soort (deeltijd)onderwijs te gaan volgen of om op zoek te gaan naar werk is er nauwelijks reden om hem of haar te beschouwen als een leerling die 'uit de boot is gevallen' of 'mislukt' is.

In het kwalitatieve deelonderzoek wordt de schoolverlater zelf aan het woord gelaten over de directe redenen van het schoolverlaten en over aspecten die er (bewust of onbewust) een rol bij hebben gespeeld; de onderwerpen zijn grotendeels ontleend aan de literatuur, deels heuristisch. Dit kwalitatieve onderzoeksgedeelte levert (meer) inzicht op in de (complexiteit van de) processen die ten grondslag liggen aan het schoolverlaten. Bovendien levert het indicaties op omtrent mogelijke verschillen in die achtergronden tussen diverse categorieën leerlingen. Gezien onder meer de kleinere aantallen kunnen we niet uitgaan van representativiteit van deze gegevens.

Globaal gaat het om de volgende onderzoeksvragen:

- Vraag 3: Wat zien de vijfdejaars schoolverlaters zelf als redenen die tot schoolverlaten hebben geleid?
- Vraag 4: Welke specifieke achtergronden in de persoonlijke, gezins- en schoolsituatie zijn van belang voor de verklaring van het voortijdig schoolverlaten?
- Vraag 5: Hoe is het verdere carrièreverloop van de voortijdig schoolverlaters?
- Vraag 6: Wat zijn hun toekomstplannen en -verwachtingen? En wat zijn hun wensen ten aanzien van arbeid en scholing?

Als aanvullend aandachtspunt (vraag 7) formuleren we dat bij alle onderzoeksvragen de resultaten worden vergeleken met die van de drop-out binnen vier jaar voortgezet onderwijs.

5. ONDERZOEKSOPZET, ANALYSE EN REPRESENTATIVITEIT

Opzet

In het schooljaar 1988/1989 vond de eerste meting plaats van de longitudinale onderzoeken in het basisonderwijs in het kader van de landelijke OVB-evaluatie. In deze cohortonderzoeken bestaat de totale steekproef uit twee delen: een referentiesteekproef en een aanvullende steekproef. De referentiesteekproef omvat een toevalsselectie van 3,5 procent van alle Nederlandse basisscholen. Deze steekproef telt echter te weinig scholen die voor de OVB-evaluatie van belang zijn. Er zitten nauwelijks scholen in met veel allochtone kinderen en/of Nederlandse arbeiderskinderen, de twee belangrijkste doelgroepen van het OVB. Vandaar dat de referentiesteekproef is aangevuld met scholen waar die leerlingen oververtegenwoordigd zijn. Van alle leerlingen die in 1988/1989 in groep 8 van het basisonderwijs zijn getoetst is ongeveer de helft geselecteerd voor verder onderzoek in het voortgezet onderwijs. Zij doen dus mee in het schoolloopbanenonderzoek. Bij de selectie van die leerlingen is uitgegaan van de steekproefopzet van de OVB-cohortonderzoeken. Er is dus weer sprake van een referentiesteekproef en een aanvullende steekproef met scholen waar relatief veel leerlingen uit de OVB-doelgroepen op zitten. Het totale aantal geselecteerde leerlingen bedroeg ruim 5000, waaronder ruim 1500 allochtone leerlingen (vgl. Mulder, 1991). In dit schoolloopbanenonderzoek is nauwkeurig bijgehouden of leerlingen op dezelfde school bleven, naar een nieuwe school gingen of dat ze vrijwillig of noodgedwongen (bv. ziekte of emigratie) het onderwijs hebben verlaten. Jaarlijks zijn aan de vo-scholen gegevens opgevraagd over de onderwijs-

positie van de leerlingen. Inmiddels is jaarlijks gerapporteerd over de eerste vijf jaren in het voortgezet onderwijs (o.a. Suhre, De Wit & Mulder, 1995).

Het *schoolverlatersonderzoek* vloeit voort uit bovengenoemd OVB-schoolloopbanenonderzoek. In het schoolverlatersonderzoek wordt specifiek ingegaan op de totale groep schoolverlaters uit het cohort. In het vierde jaar VO werd het eerste schoolverlatersonderzoek uitgevoerd (De Wit & Dekkers, 1994). In het daaropvolgende jaar is een vervolgonderzoek uitgevoerd naar de nieuwe schoolverlaters. De groep omvat alle leerlingen uit het oorspronkelijke OVB-cohort die tussen het eerste schooljaar (1989/1990) en maart/april van het vijfde schooljaar het voortgezet onderwijs hebben verlaten. Eind 1994 is ten behoeve van dit tweede schoolverlatersonderzoek van iedere leerling nogmaals via een extra telefonische dataverzameling nagegaan of hij/zij daadwerkelijk voortijdig schoolverlater was. De opzet van beide schoolverlatersonderzoeken is in grote lijnen dezelfde. Het betreft steeds twee deelonderzoeken; een kwantitatief en een kwalitatief. We zullen kort op beide deelonderzoeken ingaan.

Methoden, technieken en analyses

Gekozen is voor een combinatie van kwantitatieve en kwalitatieve onderzoeksmethoden. Bij de keuze van deze methoden hebben we ons primair laten leiden door de onderzoeksvragen. Voordeel van deze methodologische triangulatie is de vergroting van de betrouwbaarheid en geldigheid van de onderzoeksresultaten (vgl. Teunissen, 1988; Weeda, 1986).

Kwantitatieve deelonderzoek

In het kwantitatieve onderzoek stellen we de omvang van de groep voortijdig schoolverlaters vast. Middels vragenlijsten zijn in het schoolloopbanenonderzoek bij scholen gegevens verzameld over de onderwijssituatie van de leerlingen in het cohort. Bij de leerlingen die niet meer op een school bleken te zitten, zijn zo veel mogelijk telefonische gegevens verzameld op het privé-adres. Voor het vaststellen van de omvang van het voortijdig schoolverlaten zijn eenvoudige statistische analyses uitgevoerd op de databestanden van het schoolloopbanen- en schoolverlatersonderzoek. Met behulp van soortgelijke beschrijvende technieken (frequenties, kruistabellen en variantieanalyses) wordt een beeld gegeven van de samenstelling van de groep voortijdig schoolverlaters op de in het longitudinale databestand aanwezige kenmerken en de verschillen hierin met leerlingen die niet van school zijn gegaan.

Kwalitatieve deelonderzoek

In het kwalitatieve deelonderzoek gaan we dieper in op de achtergronden en oorzaken van het voortijdig schoolverlaten (onderzoeksvragen 3 t/m 6). Deze informatie is op twee manieren verzameld: via een telefonische en schriftelijke enquête en via diepte-interviews.

Telefonische en schriftelijke enquêtes

Zo veel mogelijk voortijdig schoolverlaters zijn thuis opgebeld voor afname van een telefonische vragenlijst. In die lijst is gevraagd naar de redenen van uitval, de huidige situatie en de toekomstplannen van de respondent. De vragen zijn gesteld aan de hand van een gestructureerde vragenlijst met veelal gesloten en enkele open vragen. Deze enquête heeft plaatsgevonden in de periode november en december 1994. Een relatief grote groep leerlingen was telefonisch niet bereikbaar (geheim, afgesloten of onjuist nummer of geen telefoon). Deze leerlingen hebben we een schriftelijke versie van de vragenlijst toegestuurd. Voor het samenstellen van die schriftelijke versie is de telefonische vragenlijst zo betrouwbaar mogelijk omgezet, met zo veel mogelijk dezelfde vragen en antwoordmogelijkheden. Met beide vragenlijsten zijn eerst proef-interviews afgenomen. Met behulp van vooral beschrijvende analysetechnieken worden de interviewresultaten weergegeven.

Diepte-interviews

Bij de jongeren die daaraan mee wilden werken zijn diepte-interviews afgenomen, waarin onder andere is doorgevraagd over de onderwerpen die in de telefonische/schriftelijke enquête aan de orde waren gekomen. De interviews zijn afgenomen bij leerlingen thuis en duurden ongeveer een uur. In een aantal gevallen was één of beide ouders van de schoolverlater aanwezig. Dit kan in een aantal gevallen nadelig zijn geweest, met name wanneer het over onderwerpen ging als het gedrag op school, problemen in de gezinssituatie of criminele activiteiten. Ons inziens viel deze negatieve invloed echter erg mee; na een korte gewenning spraken de (ex-)leerlingen veelal vrijuit over alle gespreksonderwerpen. Bovendien leverde de aanwezigheid van ouders soms ook voordelen op; de onderzoeker kon meer inzicht krijgen in de gezinssituatie en -relaties en meer informatie over bepaalde onderwerpen (bv. vrijstelling Leerplicht) en soms werd door de ouder aan de respondent enige emotionele ondersteuning gegeven indien onderwerpen ter sprake kwamen die gevoelig lagen. We hebben gebruikgemaakt van gestandaardiseerde interviews, in de vorm van een afname met een topic-lijst. Dit betekent dat de onderwerpen vooraf gespecificeerd waren, maar dat de onderzoeker enige vrijheid had bij het bepalen van de volgorde van de vragen en de wijze waarop ze werden aangeboden. De topics zijn gegenereerd uit een literatuuroverzoek naar actuele, relevante publicaties over voortijdig schoolverlaten (zie §3).

Respons en representativiteit

In het OVB-schoolloopbanenonderzoek is sprake een jaarlijkse toename in de non-respons van leerlingen. Niettemin is deze uitval dankzij de zeer intensieve dataverzameling zo gering mogelijk gehouden. In het vijfde jaar VO bedraagt de non-response ten opzichte van de eerste meting 13 procent. Naast permanente uitval (uit het onderzoek) is ook sprake van tijdelijke uitval, in de gevallen dat leerlingen in eerste instantie spoorloos raken, maar later op een andere school worden teruggevonden. Allochtone leerlingen zijn oververtegenwoordigd in de non-respons. Non-responsanalyses tonen aan dat dit geen grote gevolgen heeft voor de representativiteit van de steekproef (Suhre, De Wit & Mulder, 1995, 2.3). Bij uitspraken over verschillen tussen herkomstlanden dient hier echter wel rekening mee te worden gehouden.

In het eerste schoolverlatersonderzoek (de Wit & Dekkers, 1994) kwamen 298 leerlingen naar voren die gedurende de eerste vier jaren het reguliere onderwijs verlaten zouden hebben, ofwel na een eerste zoektocht 'spoorloos' bleken. Hiervan bleken er door nadere dataverzameling in het schoolverlatersonderzoek zo'n 200 als drop-out te kunnen worden beschouwd (de overigen waren geëmigreerd, opnieuw het onderwijs ingestroomd en dergelijke). In het vijfde jaar heeft wederom een groep leerlingen het onderwijs verlaten of is ongediplomeerd overstapt op vervolg- of beroepsonderwijs. Binnen het schoolverlatersonderzoek is deze groep van 371 leerlingen verder opgespoord en zo veel mogelijk persoonlijk benaderd voor het vaststellen van de onderwijs- of arbeidsmarktsituatie.

Het schoolverlatersonderzoek omvat drie onderdelen (kwantitatieve analyses, telefonische en schriftelijke enquête, diepte-interviews) waarin telkens verschillende groepen leerlingen zijn betrokken. In ieder deel is sprake van een groep deelnemers en een groep niet-deelnemers, ofwel uitvallers. Om na te gaan of de uitval tot vertekeningen in de resultaten heeft geleid, hebben we uitvalsanalyses uitgevoerd waarin we bekijken of de uitval selectief is geweest en in hoeverre de 'respons'-groep als representatief mag worden beschouwd. We vatten kort de responsgegevens en de uitkomsten van de non-responsanalyses samen. Voor een meer uitvoerige beschrijving verwijzen we naar het betreffende onderzoeksrapport (de Wit & Dekkers, 1995). De non-response in de kwantitatieve analyses is enigszins selectief. De groep bevat meer 1.90 leerlingen, en met name Marokkaanse leerlingen, van wie we niet precies konden nagaan of ze echt iedere vorm van onderwijs hadden verlaten. Dit geldt ook voor de respons op de telefonische en schriftelijke enquête. De respons op de mondelinge interviews toonde geen verschillen.

We plaatsen hierbij nog een enkele kanttekening. We hebben de respons op een beperkt aantal kenmerken vergeleken met de non-respons. Het is mogelijk en zelfs waarschijnlijk dat de responsgroep toch een wat specifieke groep is op andere, moeilijk te controleren, kenmerken. De leerlingen in meer extreme omstandigheden (gevangenis, inrichting, zwervend, Raad van Kinderbescherming) zijn bijvoorbeeld in mindere mate in de responsgroep vertegenwoordigd. We moeten dus de nodige terughoudendheid aannemen ten aanzien van de representativiteit van de resultaten van de enquêtes.

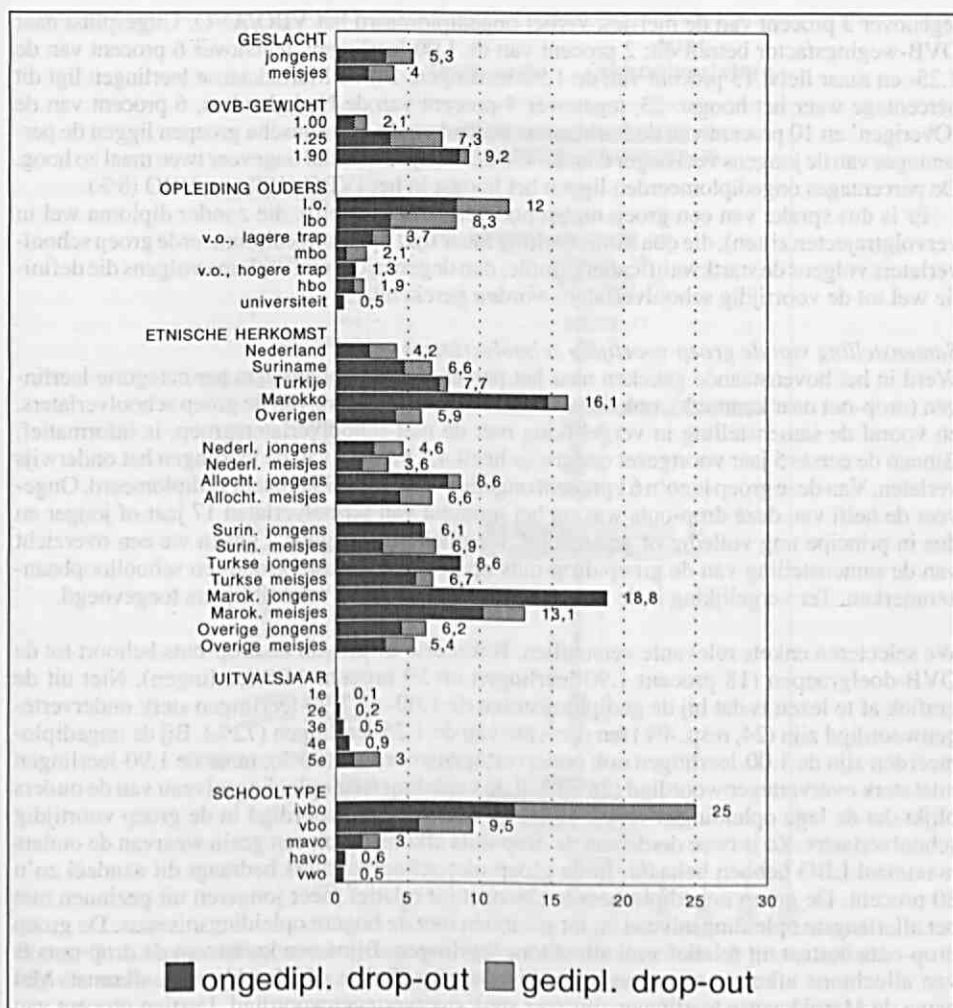
6. RESULTATEN

De omvang van het voortijdig schoolverlaten

Binnen vijf jaar voortgezet onderwijs blijkt zo'n 4,7% van de totale leerlingenpopulatie het onderwijs, de startkwalificatie-definitie volgend, voortijdig te hebben verlaten. Dit percentage omvat 2,9% ongediplomeerde en 1,8% gediplomeerde onderwijsverlaters. In vergelijking met de gegevens van een jaar eerder zijn de percentages fors toegenomen (in grove lijnen verdubbeld), hetgeen mede samenhangt met het aflopen van de leerplicht voor veel leerlingen. Figuur 1 toont de omvang van het voortijdig schoolverlaten, uitgesplitst voor een aantal achtergrond- en schoolkenmerken. In de figuur vermelden we het totaalpercentage en we onderscheiden gediplomeerd en ongediplomeerd schoolverlaten, beide in dit verband ook drop-out genoemd, in de betekenis van 'uit het onderwijs vallen'.

Het percentage voortijdig schoolverlaters ligt bij jongens iets hoger dan bij meisjes. Het is met name de ongediplomeerde drop-out die hoger is bij jongens, het gediplomeerd voortijdig schoolverlaten ligt gelijk. De uitsplitsing naar OVB-gewicht laat zien dat de 1.90- en 1.25-leerlingen vaker het onderwijs voortijdig hebben verlaten dan de 1.00-leerlingen. Bij de 1.90-leerlingen betreft het bovendien voor het overgrote deel ongediplomeerde schoolverlaters. De drop-out van 1.25-leerlingen omvat ook relatief veel gediplomeerden. Uit de uitsplitsing naar opleidingsniveau van de ouders komt een sterke milieu-ongelijkheid naar voren. De drop-out is veel hoger in de laagste opleidingsniveaus. Met name in de laagste opleidingscategorie - ouders met maximaal basisonderwijs - is de drop-out erg hoog (12%). Uit de uitsplitsing naar herkomstland komt de relatief slechte positie van allochtone leerlingen opnieuw naar voren. In alle vier de allochtone herkomstlandgroepen is de drop-out beduidend hoger dan bij de Nederlanders. Uitschieter in negatieve zin zijn de Marokkaanse leerlingen; ruim 16 procent heeft het voortgezet onderwijs verlaten. Uit de nadere uitsplitsing naar sekse blijkt dat de drop-out van allochtone jongens iets hoger ligt dan die van de meisjes. Alleen bij de Surinaamse leerlingen is de drop-out van de meisjes iets hoger dan van de jongens (hoewel we dit vanwege de kleinere aantallen met de nodige voorzichtigheid moeten interpreteren). Daarnaast omvat de drop-out van allochtone jongens beduidend vaker ongediplomeerden dan die van allochtone meisjes. Bij de Turkse, Marokkaanse en Surinaamse jongens bestaat (vrijwel) de gehele drop-out uit ongediplomeerden. Bij de meisjes uit deze groepen is ook sprake van een deel gediplomeerden die stopten met vervolgonderwijs. De categorieën met de hoogste drop-out zijn de Marokkaanse jongens (19%) en meisjes (13%). Hoewel in diverse publicaties wordt gewezen op een hoge drop-out van met name Turkse en Marokkaanse meisjes in de eerste fase van het voortgezet onderwijs, nuanceren onze resultaten dit beeld; veel meer allochtone jongens hebben het voortgezet onderwijs voortijdig verlaten. Wel blijkt de migratiecijfers en de uitval bij de overgang van basis- naar voortgezet onderwijs bij Turkse en Marokkaanse meisjes groter dan bij de jongens.

Uit de verdeling naar uitvaljaar blijkt dat het drop-out-percentage jaarlijks toeneemt. Er is sprake van een zeer ongelijke verdeling van de voortijdig schoolverlaters over de diverse schooltypen: hoe lager het schooltype des te hoger de drop-out. Het IVBO springt eruit in



Figuur 1: Voortijdig schoolverlaten binnen 5 jaar in het voortgezet onderwijs, uitgesplitst naar achtergrond- en school(loopbaan)kenmerken (percentages)

negatieve zin, met een drop-out-percentage van 25 procent. Hierbij moet worden bedacht dat het zowel ongediplomeerde als gediplomeerde schoolverlaters betreft en dat leerlingen in het IVBO, in tegenstelling tot leerlingen in andere schooltypen, niet meer de mogelijkheid hebben om af te stromen naar een lager schooltype. Ook het VBO heeft een veel hoger drop-out-percentage dan de hogere schooltypen.

In aanvulling op de percentages voortijdig schoolverlaters hebben we de totaal-percentages ongediplomeerde VBO/AVO/VWO-verlaters berekend, dus inclusief de doorstromers naar vervolgonderwijs; het gaat hier dus om schoolverlaters volgens de oude (en internationale) definitie, maar schoolblijvers volgens de startkwalificatiedefinitie. De verschillen naar sekse, OVB-wegingsfactor, milieu en etniciteit blijken beduidend groter dan bij de percentages voortijdig schoolverlaters volgens de startkwalificatiedefinitie het geval was. Zes procent van de jongens,

tegenover 3 procent van de meisjes, verliet ongediplomeerd het VBO/AVO. Uitgesplitst naar OVB-wegingsfactor betreft dit: 2 procent van de 1.00-leerlingen, tegenover 6 procent van de 1.25- en maar liefst 13 procent van de 1.90-leerlingen. Bij de Marokkaanse leerlingen ligt dit percentage weer het hoogst: 23, tegenover 4 procent van de Nederlanders, 6 procent van de 'Overigen' en 10 procent van de Surinamers en Turken. In alle etnische groepen liggen de percentages van de jongens veel hoger dan die van de meisjes, meestal ongeveer twee maal zo hoog. De percentages ongediplomeerden liggen het hoogst in het IVBO (19%) en VBO (8%).

Er is dus sprake van een groep niet-schoolverlaters (namelijk die zonder diploma wel in vervolgotrajecten zitten), die qua samenstelling meer lijkt op de ongediplomeerde groep schoolverlaters volgens de startkwalificatiedefinitie, dan degenen die met diploma volgens die definitie wel tot de voortijdig schoolverlaters worden gerekend.

Samenstelling van de groep voortijdig schoolverlaters

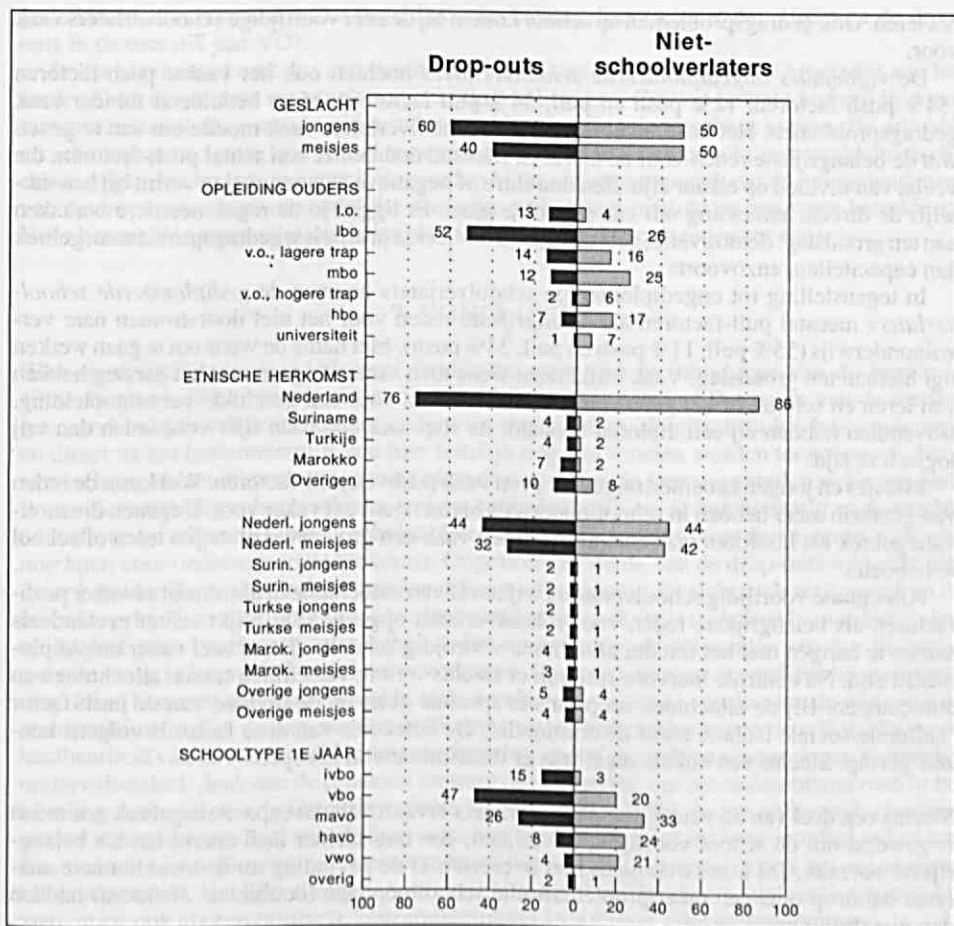
Werd in het bovenstaande gekeken naar het percentage schoolverlaters per categorie leerlingen (drop-out naar kenmerk), ook de samenstelling per kenmerk van de groep schoolverlaters, en vooral de samenstelling in vergelijking met de niet-schoolverlatersgroep, is informatief. Binnen de eerste 5 jaar voortgezet onderwijs heeft 4,7 procent van de leerlingen het onderwijs verlaten. Van deze groep is zo'n 61 procent ongediplomeerd en 39 procent gediplomeerd. Ongeveer de helft van deze drop-outs was op het moment van schoolverlaten 17 jaar of jonger en dus in principe nog volledig of gedeeltelijk leerplichtig. In figuur 2 geven we een overzicht van de samenstelling van de groep drop-outs op een aantal achtergrond- en schoolloopbaan-kenmerken. Ter vergelijking is de samenstelling van de groep schoolblijvers toegevoegd.

We selecteren enkele relevante verschillen. Ruim drie kwart van de drop-outs behoort tot de OVB-doelgroepen (18 procent 1.90-leerlingen en 59 procent 1.25-leerlingen). Niet uit de grafiek af te lezen is dat bij de gediplomeerden de 1.00- en 1.90-leerlingen sterk ondervertegenwoordigd zijn (24, resp. 4%) ten opzichte van de 1.25-leerlingen (72%). Bij de ongediplomeerden zijn de 1.00-leerlingen ook ondervertegenwoordigd (23%), maar de 1.90-leerlingen juist sterk oververtegenwoordigd (28%). Uit de verdeling naar opleidingsniveau van de ouders blijkt dat de lage opleidingsniveaus sterk zijn oververtegenwoordigd in de groep voortijdig schoolverlaters. Zo is twee derde van de drop-outs afkomstig uit een gezin waarvan de ouders maximaal LBO hebben behaald. In de groep niet-schoolverlaters bedraagt dit aandeel zo'n 30 procent. De groep ongediplomeerden bestaat uit relatief meer jongeren uit gezinnen met het allerlaagste opleidingsniveau en uit gezinnen met de hogere opleidingsniveaus. De groep drop-outs bestaat uit relatief veel allochtone leerlingen. Bijna een kwart van de drop-outs is van allochtone afkomst, waarvan bijna de helft van Turkse of Marokkaanse afkomst. Met name de Marokkaanse leerlingen zijn zeer sterk oververtegenwoordigd. Dertien procent van de groep drop-outs is een allochtone jongen, 11 procent een allochtoon meisje. Turkse en Marokkaanse meisjes, over wiens uitval een zeer negatief beeld bestaat, vormen slechts een relatief klein gedeelte van deze groep vroegtijdige schoolverlaters: zo'n vijf procent. Binnen de groep ongediplomeerde drop-outs zijn allochtone jongens sterk oververtegenwoordigd (22 versus 11%), binnen de groep gediplomeerden juist ondervertegenwoordigd (4 versus 8%)!

Zo'n 88 procent van de drop-outs begon het voortgezet onderwijs op IVBO-, VBO- of MAVO-niveau. Met name het IVBO-niveau is daarbij sterk oververtegenwoordigd. Twaalf procent van de drop-outs begon in het eerste jaar op het HAVO of VWO. Velen zijn afgestroomd naar lagere schooltypen; uiteindelijk zat zo'n 7 procent van de drop-outs in het laatste jaar op het HAVO of VWO. *

Het proces van schoolverlaten

In de enquête en diepte-interviews zijn de voortijdig schoolverlaters zelf aan het woord gelaten over de directe redenen van het schoolverlaten en over aspecten die er (bewust of onbe-



Figuur 2: Samenstelling van de groep voortijdig schoolverlaters en de groep schoolblijvers naar achtergrond- en schoolloopbaanmerken (percentages)

wust) een rol bij hebben gespeeld. Deze aspecten hebben betrekking op hun cognitieve en sociale eigenschappen, de gezinsachtergrond, de vriendenkring en de door hen bezochte school. Voortijdig schoolverlaten kan voortkomen uit een complex van motieven, aspiraties en voorgeschiedenissen; in de literatuur worden vaak twee categorieën van factoren onderscheiden: push- en pull-factoren.

Uit de enquête en diepte-interviews blijkt dat voortijdig schoolverlaters een grote diversiteit aan redenen noemen voor het schoolverlaten. Bovendien speelt bij het grootste deel van de leerlingen een complex van oorzaken een rol. Wat is de belangrijkste reden die voortijdig schoolverlaters noemen voor het schoolverlaten? We zien een belangrijk verschil tussen: - de leerplichtige jongeren die in de eerste 4 jaar VO uitvallen, de zogenaamde zeer voortijdig schoolverlaters; - de (deels nog partieel of volledig leerplichtige) ongediplomeerde schoolverlaters in het vijfde jaar VO en - de (deels nog partieel of volledig leerplichtige) gediplomeerde schoolverlaters.

De zeer voortijdig schoolverlaters noemen vooral push-factoren als belangrijkste reden voor schoolverlaten, zoals problemen met leerkrachten en geen motivatie meer hebben voor

het leren. Ook gedragsproblemen op school komen bij de zeer voortijdige schoolverlaters vaak voor.

De *vijfdejaars ongediplomeerde schoolverlaters* noemen ook het vaakst push-factoren (54% push-factoren, 12% push en pull, 34% pull-factoren). Maar beduidend minder vaak gedragsproblemen. Het kost ongediplomeerde schoolverlaters vaak moeite om aan te geven wat de belangrijkste reden voor de uitval is. Meestal noemen ze een aantal push-factoren, die veelal van invloed op elkaar zijn. Een doublure of negatieve examenuitslag vormt bij hen dikwijls de directe aanleiding om van school te gaan. Er liggen in de regel meerdere oorzaken aan ten grondslag: demotivatie, gebrekkige inzet, slechte prestaties, gedragsproblemen, gebrek aan capaciteiten, enzovoorts.

In tegenstelling tot ongediplomeerde schoolverlaters noemen de *gediplomeerde schoolverlaters* meestal pull-factoren als belangrijkste reden voor het niet doorstromen naar vervolgonderwijs (55% pull, 11% push en pull, 34% push). Met name de wens om te gaan werken ligt hieraan ten grondslag. Vaak blijkt deze wens toch samen te gaan met het genoeg hebben van leren en school, of het geen keuze kunnen maken voor een geschikte vervolgopleiding. Bovendien hebben zij een diploma behaald; de stap naar een baan lijkt voor velen dan vrij logisch te zijn.

Meisjes en jongens noemen ongeveer even vaak push- en pull-factoren. Wel komt de reden van geen zin meer hebben in school/onderwijs bij meisjes veel vaker voor. Degenen die motivatiegebrek als hoofdoorzaak aangaven zeiden vaak een grote weerzin tegen leren of school te hebben.

Allochtone voortijdig schoolverlaters wijken af van autochtone in die zin dat ze vaker push-factoren als belangrijkste reden voor schoolverlaten opgeven. Dit blijkt echter grotendeels samen te hangen met het feit dat allochtone voortijdig schoolverlaters veel vaker ongediplomeerd zijn. Na controle hiervoor resteren er slechts weinig verschillen tussen allochtonen en autochtonen. Bij de allochtone drop-outs is speciaal gelet op de invloed van de push-factor 'culturele/sociale isolatie en/of discriminatie'. De betekenis van deze factor is volgens henzelf gering. Slechts een enkele maal was er discriminatie in het spel.

Slechts een deel van de voortijdig schoolverlaters ervaart zelf dat capaciteitsgebrek een reden is geweest om de school voortijdig te verlaten; een nog kleiner deel erkent het als belangrijkste oorzaak. Dit komt overigens niet overeen met de bevinding uit de kwantitatieve analyses dat drop-outs een meer problematische schoolloopbaan (doubleren, afstroom) hadden dan niet drop-outs.

Veel drop-outs maken in hun oordeel over het onderwijs een onderscheid tussen bepaalde opleidingen of vakken. Opleidingen en vakken met een theoretisch karakter achten zij weinig waardevol, in tegenstelling tot praktijkgerichte scholing. Uit het onderzoek blijkt ook dat factoren op gezinsniveau van groot belang zijn. De ouders van de meeste voortijdig schoolverlaters zijn laag opgeleid en drop-outs lijken relatief vaker in een problematische gezins-situatie te verkeren. De vriendengroep speelt in het algemeen slechts een ondergeschikte rol bij het schoolverlaten (8%). Opvallend is dat bij allochtone leerlingen de vriendengroep beduidend vaker van invloed lijkt te zijn (26%).

Situatie na het schoolverlaten

Een overgrote meerderheid (zo'n 90%) van de drop-outs in de eerste vijf jaren VO heeft na het schoolverlaten betaalde arbeid verricht en/of een andere vorm van scholing gevolgd. Dit geldt zowel voor de gediplomeerde als de ongediplomeerde voortijdig schoolverlaters. Bij de meesten is geen sprake geweest van een langere periode van inactiviteit, buiten de normale vakanties. Een grote meerderheid van de jongeren is tevreden met de situatie en heeft geen spijt van het schoolverlaten. Een tevredenheid die vermoedelijk samenhangt met de opluchting na een moeilijke periode van voortijdig schoolverlaten (wat met name geldt voor de ongediplomeerden). Een kleine minderheid verkeert in zeer problematische omstandigheden, zoals

gedetineerd, zwerfend of onder psychiatrische behandeling (minder dan 10% van de drop-outs in de eerste 4 jaar VO).

In het algemeen komt dus een beeld naar voren van een gemotiveerde groep die aan het werk wil en in mindere mate ook nog scholing wil volgen. Op wat langere termijn lijkt een wat groter aantal drop-outs in een minder rooskleurige situatie terecht te komen. De arbeids- en scholingsparticipatie lijkt wat af te nemen en het aandeel jongeren in zeer problematische omstandigheden iets toe te nemen. Bovendien blijkt uit het onderzoek dat de gevonden banen van de drop-outs veelal van tijdelijke aard zijn, met weinig uitzicht op een vaste betrekking. Voor het arbeidsperspectief op de langere termijn is dit dan ook reden tot zorg.

7. CONCLUSIES

'Een startkwalificatie voor iedereen', luidde het motto bij de introductie van dit begrip in 1990. Binnen vijf jaar voortgezet onderwijs blijkt echter al zo'n 4,7 procent van de leerlingen zonder startkwalificatie uit het onderwijs weggevallen (de uitval uit het basisonderwijs en direct na het basisonderwijs zou hier feitelijk nog aan moeten worden toegevoegd). Naar verwachting neemt dit percentage in de twee daaropvolgende jaren nog sterk toe. Een gedeelte van de voortijdig schoolverlaters zou op latere leeftijd nog de beleidsmatig nagestreefde educatieve kwalificatie kunnen behalen. Een aanzienlijk gedeelte van de jongeren staat ook nog open voor onderwijs in de toekomst. Ongeveer een derde van de drop-outs lijkt echt niet meer bereidwillig tot het volgen van kwalificerend onderwijs en richt zich uitsluitend op de arbeidsmarkt. Veel voortijdig schoolverlaters zullen dan ook in de toekomst binnen het onderwijsstelsel geen startkwalificatie behalen. Dit onderstreept de relevantie van een discussie over de vraag in hoeverre kwalificaties verkregen (kunnen) worden via praktische werkervaring. En in hoeverre dergelijke kwalificaties gerelateerd dienen te worden aan de notie van de startkwalificatie. Tot dusver worden oplossingen voor het probleem van de reikwijdte en de haalbaarheid van de startkwalificatie-omschrijving vooral gezocht in aanpassingen binnen het onderwijsstelsel; denk aan de (plannen omtrent de) invoering van het assistentenniveau in het Leerlingwezen en de arbeidsmarktgerichte leerweg in het VBO. In het onderzoek constateren we echter dat een deel van de drop-outs niet meer binnen het onderwijsstelsel wil of kan functioneren; kwalificering via werkervaring lijkt voor deze groep dan ook het enige resterende alternatief. Het lijkt zinnig deze mogelijkheid niet alleen als een 'deficiet' oplossing te zien, maar ook als een positieve keuzemogelijkheid voor (ook allochtone) jongeren.

Uit de onderzoeksresultaten wordt duidelijk dat er niet één type voortijdig schoolverlater is. Dit is ook in de onderzoeksrapporten (o.a. de Wit & Dekkers 1996) tot uitdrukking gebracht in een 'typologie' van drop-outs. Het grootste gedeelte van de voortijdig schoolverlaters - met name de ongediplomeerden - verliet het onderwijs vooral vanwege factoren die binnen het onderwijs speelden (push-factoren), zoals gedrags-, motivatie- en in mindere mate capaciteitsproblemen. De overigen - vooral de gediplomeerden - verlieten het onderwijs vooral omdat ze wilden gaan werken of geld gaan verdienen (pull-factor). Hieruit blijkt al dat *schoolfactoren en leerlingkenmerken* van belang zijn voor het voortijdig schoolverlaten. Veel drop-outs maken in hun oordeel over het onderwijs een onderscheid tussen bepaalde opleidingen of vakken. Opleidingen en vakken met een theoretisch karakter achten zij weinig waardevol, in tegenstelling tot praktijkgerichte scholing.

De *milieu-ongelijkheid* in het voortijdig schoolverlaten is groot. Jongeren uit gezinnen waarvan de ouders een laag of zeer laag opleidingsniveau hebben, zijn sterk oververtegenwoordigd. Dit betekent echter niet dat er geen schoolverlaters zijn uit de hogere 'milieus'; zo heeft van 8 procent van de drop-outs minimaal een van de ouders HBO of universiteit voltooid.

Ook is sprake van een sterke *etnische ongelijkheid*; de drop-out van allochtone leerlingen is bijna twee maal zo hoog als die van autochtone leerlingen. Uitschieter in negatieve zin zijn de Marokkaanse jongeren. Bezien we het percentage leerlingen dat zonder diploma het VBO/AVO verlaat dan ligt dit bij allochtonen bijna drie maal zo hoog als bij autochtonen. Allochtone leerlingen zijn zeer sterk oververtegenwoordigd bij de ongediplomeerde drop-out, bij de gediplomeerde drop-out zijn ze juist ondervertegenwoordigd.

We constateren bovendien een ongunstigere arbeidsparticipatie van de allochtone schoolverlaters en een hogere mate van ontevredenheid over het schoolverlaten. Voldoende redenen voor extra aandacht voor deze groep, met name inspanningen om hen aan een diploma te helpen en/of aan een noodzakelijke werkplek.

Jongens verlaten wat vaker dan *meisjes* voortijdig het onderwijs, met name in de eerste vier jaren. Dit lijkt mede samen te hangen met persoonlijkheidstypen. Bij jongens spelen gedragsproblemen een belangrijke rol bij het schoolverlaten. Bij meisjes zijn het veelal algemene motivatieproblemen. Mannelijke schoolverlaters lijken in het algemeen te zeer naar buiten gericht, te weinig volgzzaam, te ongedurig, ook wel te speels om zich aan te passen aan het stramien van een reguliere opleiding. Dit uit zich in een later stadium in een hekel aan het schoolsysteem op zich en in een oriëntatie die veel meer gericht is op de maatschappelijke omgeving dan op het onderwijs. Meisjes lijken daarentegen wat hun attitudes betreft beter toegerust te zijn om te voldoen aan de voorwaarden voor het volgen van onderwijs. Wanneer zij afhaken gaat het vaker om een lager aspiratieniveau. Opvallende constatering in het onderzoek is dat ouders bij meisjes beduidend vaker instemmen met schoolverlaten dan bij jongens het geval is. Dit geldt met name voor het gediplomeerd schoolverlaten. Bij de uitval van ongediplomeerde *jongens* spelen vooral gedragsproblemen een rol, naast moeite met de theoretische vakken. Zij willen nog wel onderwijs volgen, maar dan anders. Bij *meisjes* zijn het veelal algemene motivatieproblemen; onderwijs hoeft voor hen niet meer zo. Onderwijsmaatregelen als de arbeidsmarktgerichte leerweg en opleidingen op assistentniveau zullen - uitgaande van een ongewijzigde leerplicht - dan ook vooral jongens aantrekken en kunnen leiden tot een toenemende sekse-ongelijkheid.

Met betrekking tot de eerder beschreven theoretische achtergrond ten slotte kunnen we de resultaten van dit onderzoek als volgt samenvatten. Persoonlijke factoren spelen een rol, met name motivatie en gebrek aan cognitieve bekwaamheid. De zeer voortijdig schoolverlaters hadden ook vaak gedragsproblemen op school. De schoolverlaters met diploma wilden liever gaan werken dan nog kwalificerend onderwijs volgen; er is hier sprake van positieve motivatie om niet verder te gaan met onderwijs (op school). Ouderlijk milieu blijkt ook een belangrijke factor. Leerlingen die zonder kwalificatie de school verlaten, hebben ouders die zelf ook een laag onderwijsniveau hebben; bovendien behoren ze relatief vaak tot etnische minderheidsgroepen. Ook sekse doet ertoe, zowel in kwantitatieve zin als in redenen om de school te verlaten.

De invloed van de peergroep blijkt van minder belang, maar is belangrijker voor leerlingen uit etnische minderheden. Met betrekking tot schoolfactoren toont dit onderzoek aan dat het van belang is adequate, met name meer praktische cursussen aan te bieden aan risico-leerlingen.

Meestal is voortijdig schoolverlaten te wijten aan een combinatie van factoren. We vonden duidelijk push- en pull-combinaties. Voor het modeldenken voorzag dit onderzoek slechts in een eerste stap, door verschillende soorten factoren binnen een onderzoeksdesign te bestuderen. Het feit dat de data gedeeltelijk kwantitatief, gedeeltelijk kwalitatief waren maakt het onmogelijk een model te toetsen dat alle belangrijke factoren omvat. De resultaten van dit onderzoek bieden echter wel aanknopingspunten om dat in toekomstig onderzoek te doen.

LITERATUUR

- Bosker, R.J. & W.H.A. Hofman (1994). School effects on drop out. A multi-levellogistic approach to assessing school-level correlates of drop out of ethnic minorities. *Tijdschrift voor Onderwijsresearch*, 1 (19), 50-64.
- Chuang, H.L. (1997). High school youths' drop out and re-enrollment behaviour. *Economics of education review*, 16 (2), 171-186.
- Coley, R.J. (1995). *Dreams deferred: High school drop-outs in the United States*. Princeton: Educational Testing Service, Policy Information Center.
- Dekkers, H. (1993). *Allochtone meisjes in het VBO*. Nijmegen: ITS.
- Dekkers, H. & Wit, de H. (1995). Zeer voortijdig schoolverlaten in het voortgezet onderwijs. Omvang en achtergronden van schoolverlaten van allochtone en autochtone leerlingen in de eerste vier jaar voortgezet onderwijs. *Migrantenstudies*, 11.3.
- Edirisooriya, G. & F.H. Lawrence (1995). Outcomes of a cohort/drop-out tracking initiative in an urban school system. A Paper Presented at the *Annual Meeting of the American Educational Research Association* in San Francisco, April 18-22.
- Finn, J.D. (1989). Withdrawing From School. *Review of Educational Research*, 59, No.2, 117-142.
- Gaustad, J. (1991). *Identifying Potential Drop-outs*. ERIC Digest. ERIC Clearinghouse on Educational Management, Eugene, Oreg.
- Hannan, D., Hövels, B., Berg, S. van den & M. White (1995). 'Early Leavers' from Education and Training in Ireland, the Netherlands and the United Kingdom. *European Journal of Education*, 30, 325-346.
- Hansen, D.A., Fisher, J. & V.A. Johnson (1995). The hidden persistence of immigrant 'Drop-outs': distortions, blank spots and blind spots in research on schooling careers. *International Journal of Educational Research*, 23, No.1, 83-105.
- Harkink, B., Oyen, M., Dekkers, H. (1995). *Zeer voortijdig schoolverlaters*. De situatie van autochtone en allochtone jongeren na het schoolverlaten. Nijmegen: ITS.
- Hofman, W. (1993). *Effectief onderwijs aan allochtone leerlingen*. Delft: Eburon.
- Hövels, B. & B. Bock (1991). *Achtergronden, oorzaken en perspectieven van voortijdig schoolverlaten*. (Backgrounds, causes and perspectives of early drop-out.) Uit serie: Zonder beroepkwalificatie uit het onderwijs. Deel 3. Nijmegen: ITS.
- Jordan, W.J., Lara, J. & J.M. McPartland (1994). *Exploring the Complexity of Early Drop-out Causal Structures*. Center for Research on Effective Schooling for Disadvantaged Students. The Johns Hopkins University, Baltimore, Maryland. Ministerie van Onderwijs en Wetenschappen (1993). *Een goed voorbereide start. Notitie voortijdig schoolverlaten*. Zoetermeer.
- Mulder, L. (1991) *De overgang van basis- naar voortgezet onderwijs van de OVB-doelgroepen*. Nijmegen: ITS.
- Rumberger, R.W. (1987). High School Drop-outs: A Review of Issues and Evidence. *Review of Educational Research*, 57, No. 2, 101-121.
- Schwartz, W. (1995). *School drop-outs: new information about an old problem*. ERIC Clearinghouse on Urban Education. New York, N.Y.
- Suhre, C., W. de Wit & L. Mulder (1995). *De onderwijspositie van de OVB-doelgroep leerlingen vijf jaar na het verlaten van het basisonderwijs*. (Educational results of disadvantaged students five years after leaving primary education.) Nijmegen/Groningen: ITS/GION.
- Teunissen, J. (1988). *Etnische relaties in het basisonderwijs*. Wageningen.
- Vries, G. de (1993). *Schoolverzuim en schooluitval in het voortgezet onderwijs. Een literatuurstudie*. Amsterdam/Lisse: Swets & Zeitlinger.
- Vries, G. de (1993). *Het pedagogisch regiem*. Amsterdam: Meulenhof.
- Weeda, W.C. (1986). *Kwalitatief of kwantitatief. Een oneigenlijk methodologisch dilemma*. Lisse.
- Wit, W. de & H. Dekkers (1994). *Schoolverlaten in de eerste vier jaar voortgezet onderwijs*. Nijmegen: ITS.
- Wit, W. de & H. Dekkers (1995). *Schoolverlaten zonder startkwalificatie. Schoolverlaten binnen vijf jaar voortgezet onderwijs*. Nijmegen: ITS.
- Wit, W. de & Dekkers, H. (1996). *Geen goed voorbereide start? Voortijdig schoolverlaten van allochtone en autochtone jongeren in de eerste vijf jaar voortgezet onderwijs*. Nijmegen: ITS.
- Wu, S-C. (1992). *Education and Learning in Schools with High Drop-out Rates*. National Center for Education Statistics (ED), Washington, D.C.

Op welk niveau verder studeren na VWO of HAVO; veranderingen tussen 1982 en 1991

H.D. Webbink

Stichting voor Economisch Onderzoek, Universiteit van Amsterdam¹

ABSTRACT

In this article the choice of the level of education has been analyzed for students in the final year of secondary education. The determinants of the choice of the level of education are derived from the vertical sorting theory: students with higher initial capability levels will choose higher levels of education. By estimating the vertical sorting model for samples of different years (1982 and 1991) changes over time are identified. The model is estimated for two levels of secondary education (Vwo and Havo). The quality of students from the highest type of secondary education (Vwo) entering University has diminished since 1982. The main indicator is the average score on the final exam. Most of the changes between 1982 and 1991 are caused by changes in preferences and not by changes in the characteristics of the students. The quality of students from the lowest type of secondary education (Havo) entering the highest level of vocational education (Hbo) has risen.

1. INLEIDING

In het Ontwerp-HOOP 1994 wordt aangestuurd op een compacter en efficiënter stelsel van Hoger Onderwijs. Te veel studenten kiezen leerwegen die ofwel te lang zijn ofwel niet passen bij hun aanleg en capaciteiten hetgeen, volgens het Ministerie van Onderwijs en Wetenschappen, leidt tot een niet optimale aanwending van de beperkte financiële middelen. De aansluiting tussen voortgezet en hoger onderwijs dient daarom beter te worden.

In dit artikel wordt nagegaan in hoeverre er in de periode 1982-1991 veranderingen zijn opgetreden in het keuzegedrag van eindexamenkandidaten van Havo en Vwo, die de bovenstaande roep om sterkere selectie en verwijzing kunnen rechtvaardigen. Het gaat dan om de keuze van het niveau van de vervolgopleiding. Daarbij staat de vraag centraal of de 'geschiktheid' van de leerlingen die kiezen voor de hoogste vervolgopleiding, voor Vwo'ers is dit het Wo en voor Havo'ers is dit het Hbo, is veranderd.

Hartog (1992) geeft een model voor 'verticale sortering' in het onderwijs. Door scholing worden initiële capaciteiten van individuen bruikbaar gemaakt voor de arbeidsmarkt. Scholing wordt aldus gezien als intermediair voor het afstemmen van werknemers en banen. Anders gezegd, scholing transformeert initiële capaciteiten tot productieve capaciteiten op de arbeidsmarkt. Bij 'verticale sortering' gaat het om de vraag of individuen met een hoger initieel niveau van capaciteiten ook kiezen voor meer (hoger) onderwijs. Als dit het geval is, dan bestaat er een positieve samenhang tussen het niveau van initiële capaciteiten van leerlingen en de lengte (het niveau) van het onderwijs. Een belangrijke consequentie hiervan zou zijn dat onderwijs leidt tot een vergroting van de verschillen tussen individuen.

In het navolgende wordt de niveaukeuze van Havo- en Vwo-leerlingen geanalyseerd vanuit dit perspectief. Hiervoor wordt gebruikgemaakt van datamateriaal uit 1991 en uit 1982. De centrale vragen bij de analyse zijn:

- 1) is er in 1991 sprake van verticale sortering bij de niveaukeuze?
- 2) is de verticale sortering veranderd tussen 1982 en 1991?

2. DE DATA

Voor het analyseren van de niveaukeuze van *Vwo-leerlingen* maken we gebruik van datamateriaal uit 1991 en uit 1982. Het datamateriaal voor 1991 is ontleend aan het project 'Verder Studeren', dit is een longitudinaal onderzoek onder scholieren en studenten dat gestart is in 1991 en een looptijd heeft van vijf jaar². Het project wordt in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen uitgevoerd door drie instituten van de Universiteit van Amsterdam: de Stichting voor Economisch Onderzoek, het SCO-Kohnstamm instituut en de vakgroep Micro-economie. De Vwo-leerlingen zijn geselecteerd uit twee bestanden: het bestand van eindexamenkandidaten (het zogenaamde pré-HO-panel) en het bestand van eerstejaars en ouderejaars studenten (het HO-panel). De gegevens voor 1982 zijn ontleend aan het project 'De vraag naar Hoger Onderwijs' (Kodde en Ritzen, 1986). Uit deze bestanden zijn alle Vwo'ers geselecteerd die zijn gaan deelnemen aan het HO en meegedaan hebben aan de derde belangstellingsenquête.

Voor de analyse geeft dit vier verschillende groepen, te weten: directe instromers in 1991, eerstejaars en ouderejaars in 1991 en directe instromers in 1982. Interessant hierbij is dat de groepen verschillen naar ervaring in het Hoger onderwijs. De directe instromers in 1982 en in 1991 hebben nog geen ervaring in het Hoger onderwijs. Bij de eerstejaars 1991 bevinden zich ook om- en toezwaaiers die al enige HO-ervaring hebben; uit eerder onderzoek is bekend dat ongeveer 40% van de eerstejaars in het Wo en Hbo niet direct afkomstig is uit het voortgezet onderwijs. Uiteraard hebben de ouderejaars de meeste HO-ervaring.

Bij de *Havo-leerlingen* wordt de keuze tussen Hbo en Mbo geanalyseerd. Hiervoor wordt eveneens gebruikgemaakt van de bestanden uit 1982 en 1991, maar alleen van het pré-HO-panel uit 1991. Het HO-panel is in dit geval niet bruikbaar omdat het geen Mbo'ers bevat. Dit betekent dat voor de analyse van de Havo'ers twee groepen resteren: directe instromers in 1982 en in 1991.

3. SELECTIE VAN VARIABELEN

In het verticale sorteringsmodel zijn de initiële capaciteiten van leerlingen zeer belangrijk. Aangezien deze niet direct gemeten zijn hebben wij een groot aantal variabelen geselecteerd als indicatoren van de geschiktheid van de student alsmede een aantal persoonlijke en achtergrondkenmerken. De variabelen kunnen ingedeeld worden in vier groepen: studieprestaties, motieven, achtergrondkenmerken en persoonlijke kenmerken. De variabelen in de eerste twee groepen worden vaak gebruikt als indicatoren van geschiktheid van leerlingen. In de eerste groep gaat het om objectieve kenmerken. Er zijn twee variabelen gemaakt voor de eind-examenresultaten, te weten het resultaat voor de exacte vakken plus economie en het resultaat voor talen plus aardrijkskunde en geschiedenis. Daarnaast zijn het schooladvies aan het eind van de lagere school en het zittenblijven meegenomen. Bij de tweede groep gaat het om subjectieve kenmerken zoals motivatie. In deze groep zijn een aantal keuzemotieven rond het latere werk en andere motieven voor de keuze van het niveau opgenomen. De groep persoonlijke kenmerken bestaat uit de variabelen geslacht en leeftijd, de groep achtergrondkenmerken bestaat uit de opleiding van vader en moeder en het ouderlijk inkomen.

Voor alle hierboven genoemde groepen Havo- en Vwo-leerlingen zijn modellen geschat van de keuze van het niveau van het vervolgonderwijs. Hiervoor wordt gebruikgemaakt van het logitmodel (zie o.a. Cramer (1991)). De afhankelijke variabele is de niveaukeuze en als onafhankelijke variabelen zijn in eerste instantie alle hiervoor genoemde variabelen meegenomen. Een aantal variabelen bleek in geen van de modellen van invloed. Bij de presentatie van de schattingsresultaten zijn deze variabelen derhalve buiten beschouwing gelaten. Het model voor 1982 wijkt enigszins af omdat toen niet het schooladvies gevraagd is en ook niet alle motieven. De presentatie van de multivariate modellen wordt voorafgegaan door een tabel waarin de bivariate samenhangen van de hiervoor genoemde variabelen met de keuze van het niveau zijn gegeven. De bedoeling van deze tabel is om een eerste indruk te geven van de verbanden tussen de geselecteerde variabelen en de niveaukeuze.

4. DE KEUZE TUSSEN WO EN HBO VOOR VWO'ERS

In Tabel 1 zijn voor vier groepen Vwo-leerlingen de bivariate samenhangen van alle in de vorige paragraaf genoemde variabelen met de niveaukeuze gegeven. De eerste twee groepen zijn direct vergelijkbaar, het gaat daarbij om leerlingen die direct na het eindexamen doormoeten naar Hbo of Wo. De twee laatste groepen zijn ouder, zij kunnen al enige jaren gestudeerd hebben in het hoger onderwijs.

Tabel 1: Bivariate samenhangen met de keuze van Vwo-leerlingen tussen Hbo en Wo

	1982			1991								
	direct			direct			eerste jaars			ouderejaars		
	Hbo	Wo	p	Hbo	Wo	p	Hbo	Wo	p	Hbo	Wo	p
vrouw	0,55	0,36	**	0,66	0,59	n.s.	0,55	0,48	n.s.	0,52	0,49	n.s.
leeftijd	18,5	18,5	n.s.	18,3	18,4	n.s.	19,4	18,8	**	21,4	21,7	n.s.
score talen	6,8	7,0	**	6,7	7	**	6,7	7	**	6,7	7,1	**
score exacte vakken	6,4	6,8	**	6,3	6,7	**	6,4	6,8	**	6,3	6,9	**
schooladvies (1-7)				6	6,1	n.s.	6,1	6,2	*	6,2	6,2	n.s.
zittenblijven	0,3	0,23	*	0,21	0,16	n.s.	0,32	0,26	n.s.	0,38	0,24	**
opleiding ouders (1-5)	3	3,3	**	3	3,5	**	3,1	3,6	**	3,1	3,5	**
inkomen ouders motieven (0-10)	3860	3790	n.s.	3860	4870	**	4140	4920	**	4110	5030	**
interesse	8,2	8,6	**	9,2	9,1	n.s.	8,6	8,6	n.s.	8,6	8,8	n.s.
zelfstandig werk				7,1	6,7	n.s.	6,5	5,7	**	6,2	5,9	n.s.
leidinggevende functie				6,4	6	n.s.	5,7	4,9	**	5,2	5,1	n.s.
kans op baan	6,7	6,4	n.s.	7,7	7,3	n.s.	7,3	6,4	**	6,6	6,1	n.s.
bepaald beroep	8,4	7,7	**	8,4	7,3	**	7,7	6,4	**	7,5	6,2	**
instelling dichtbij	3,4	2,6	**	2,9	2	**	2,8	2,1	**	2,4	1,6	**
studie niet moeilijk				4,6	3,8	**	4	3,2	**	3,5	2,8	**
curcusduur				5,9	4,8	**	5,6	4,1	**	4,9	3,3	**
baan in buitenland				3,5	4,2	n.s.	3,2	3,6	n.s.	2,4	3,1	*
aandeel in totaal	37,4	62,6		28,8	71,2		30,3	69,7		18,1	81,9	
aantal	290	486		92	227		159	524		133	603	

Per variabele is steeds de gemiddelde waarde gegeven voor de groep leerlingen die gekozen heeft voor het Hbo en de groep die gekozen heeft voor het Wo en of het verschil tussen deze twee groepen significant is of niet. In 1982 bestaat bijvoorbeeld de groep die kiest voor het Hbo voor 55 procent uit meisjes terwijl dit voor de Wo-groep 36 procent is en dit

verschil is significant. De gemiddelde leeftijd in beide groepen van 1982 is 18,5 jaar.

Een eerste blik op Tabel 1 leert dat er een groot aantal significante verschillen bestaat tussen Vwo'ers die kiezen voor het Wo en Vwo'ers die kiezen voor het Hbo. Meisjes lijken vaker voor het Hbo te kiezen terwijl leerlingen met betere onderwijsprestaties in de vorm van hogere cijfers, een hoger schooladvies en minder zittenblijven, vaker voor het Wo kiezen. Ook lijken er duidelijke milieu-effecten te bestaan waarbij de groep, die voor het Wo gekozen heeft, gemiddeld hoger opgeleide en rijkere ouders heeft. In hoeverre deze verschillen ook in het multivariate model gevonden worden is te zien in Tabel 2.

Tabel 2: Geschatte effecten op keuze voor Wo door Vwo'ers

	1982	1991		
	direct	direct	eerstejaars	ouderejaars
vrouw	-18,6**	-6,2	-5,6	0,9
leeftijd	-3,8	5,0	-1,0	1,1*
score talen	12,4**	5,6	9,9**	5,4**
score exacte vakken	7,7**	4,6	3,9*	7,7**
schooladvies (1-7)		2,4	0,1	-3,0**
opleiding vader t.o.v. middelbaar				
opleiding vader laag	-3,7	-14,5	-2,8	-2,0
opleiding vader hoog	9,3*	7,4	12,0**	6,3*
inkomen ouders		0,03	0,0	0,02**
motieven (0-10)				
interesse	3,1**	-2,8	-0,2	1,4
kans op baan	0,0	-0,5	-1,6*	-0,2
bepaald beroep	-1,5**	-6,3**	-2,3**	-1,4**
instelling dichtbij	-3,2*	-1,4	0,1	-0,7
studie niet moeilijk		-2,1	-0,3	0,1
curcusduur		-1,5	-1,6**	-1,3**
baan in buitenland		0,9	0,9	1,3**
likelhood	-454,1	-157,0	-313,2	-285,2
lik. ratio test	117,58**	68,61**	114,92	125,16**
aantal Hbo	290	92	159	133
aantal Wo	486	226	524	603
totaal	776	318	683	736

*significant op 5%-niveau;

** significant op 1%-niveau

Gegeven zijn de schattingsresultaten van het model van de keuze tussen Wo en Hbo voor vier verschillende groepen leerlingen. Bij de variabele 'inkomen ouders' is de quasi-elasticiteit gegeven: deze geeft aan met hoeveel procentpunten de kans op de keuze voor het Wo verandert als het inkomen met 1 procent toeneemt. Bij de andere variabelen is steeds de afgeleide gegeven: deze geeft aan met hoeveel procentpunten de kans op het Wo verandert als de variabele met 1 eenheid toeneemt. Een positief effect betekent dat bij een toename van de betreffende variabele de kans stijgt dat een leerling kiest voor het Wo.

Geen invloed werd gevonden van het behoren tot een culturele minderheid, de opleiding van de moeder, het zittenblijven in het voorgaande onderwijs en de motieven 'zelfstandig werk' en 'leidinggevende functie'. Deze variabelen zijn bij het schatten van de modellen van Tabel 2 niet meegenomen.

De invloed van eindexamencijfers

Naarmate leerlingen hogere eindexamencijfers hebben, stijgt de kans dat zij voor het Wo kiezen. De richting van dit effect wordt voor alle vier groepen leerlingen gevonden. Opvallend is dat dit effect voor de directe instromers in 1991 niet significant is en wel significant bij de andere drie groepen. De effecten voor de drie groepen uit 1991 verschillen niet sterk. Wel lijken de effecten bij de ouderejaars sterker te zijn dan bij de directe instromers uit het Vwo. Dit zou erop kunnen wijzen dat hoge eindexamencijfers doorwerken in de selectie van studenten binnen het Hoger onderwijs, dat wil zeggen, tijdens het doorlopen van het Hoger Onderwijs hebben leerlingen met lagere eindexamencijfers een grotere kans om uit te vallen. De invloed van de score voor de exacte vakken is het sterkst bij de ouderejaars. Het meest opvallend is echter het verschil tussen 1982 en 1991; in 1982 was het effect van de eindexamencijfers op de keuze tussen Wo en Hbo aanmerkelijk sterker dan thans.

De invloed van de sociale achtergrond

Leerlingen waarvan de vader een hoge opleiding heeft gevolgd kiezen vaker voor het Wo en leerlingen met een lager opgeleide vader gaan vaker naar het Hbo. Dit duidt erop dat cultureel kapitaal de stap naar het Wo vergemakkelijkt. Bij de ouderejaars studenten blijkt ook het ouderlijk inkomen van invloed op de niveaukeuze; kinderen uit rijkere gezinnen kiezen vaker voor het Wo.

De invloed van de keuzemotieven

Het keuzemotief 'door deze studie te volgen kan ik een bepaald beroep uitoefenen' wordt door leerlingen die voor het Hbo gekozen hebben belangrijker geacht dan door leerlingen die voor het Wo gekozen hebben. Dit geldt voor alle vier groepen leerlingen. Bij de eerstejaars en ouderejaars worden tevens verschillen gevonden op de motieven 'studie afronden binnen cursusduur' en 'baan in buitenland'. Hbo'ers vinden het afronden van de studie binnen de formele cursusduur belangrijker en hechten minder aan een baan in het buitenland.

De persoonlijke kenmerken geslacht en leeftijd hebben geen sterke invloed op de niveaukeuze in 1991. In 1982 kozen jongens duidelijk vaker voor het Wo, in 1991 wordt alleen een zwak geslachtseffect gevonden bij de eerstejaars. Merkwaardig is dat bij de ouderejaars een negatief effect van het schooladvies aan het eind van de lagere school wordt gevonden. Naarmate het advies hoger is kiest men eerder voor het Hbo. Het leeftijdseffect bij deze groep is moeilijk te interpreteren omdat de leeftijd ten tijde van de keuze tussen Wo en Hbo niet bekend is.

Veranderingen tussen 1982 en 1991

Een aantal verschillen tussen 1982 en 1991 kunnen gevonden worden door het vergelijken van de directe instromers in Tabel 2. In 1982 kozen jongens vaker voor het Wo, in 1991 is dit effect niet meer significant. Een milieu-effect wordt in beide jaren gevonden en ook zijn de effecten van de keuzemotieven in beide jaren redelijk vergelijkbaar. Het belangrijkste verschil tussen de jaren is het effect van de eindexamencijfers. In 1982 wordt een sterk significant effect van beide cijfers gevonden, in 1991 zijn de effecten niet meer significant. Een punt meer voor de talen of de exacte vakken heeft in 1982 bijna twee keer zoveel effect op de keuze voor het Wo als in 1991. Geconcludeerd kan worden dat de verticale sortering op grond van eindexamencijfers duidelijk is afgenomen.

Het is mogelijk om de totale verandering tussen 1982 en 1991 uiteen te leggen in een gedeelte veroorzaakt door een verandering in de samenstelling van de populatie en een gedeelte veroorzaakt door veranderingen van de voorkeuren/preferenties van de populatie. Gomulka & Stern (1990) geven de volgende uitdrukking voor de decompositie van een verandering in de tijd bij een binomiale afhankelijke variabele:

$$y^{91} - y^{82} = (P(\beta^{91}, X^{91}) - P(\beta^{82}, X^{91})) + (P(\beta^{82}, X^{91}) - P(\beta^{82}, X^{82}))$$

Links staat de verandering in de deelname aan het Wo tussen 1982 en 1991, en $P(\beta^i, X^j)$ is het gemiddelde over de steekproef X^j van de voorspelde kans op deelname aan het Wo met de parameters van jaar i . De eerste term tussen haakjes aan de rechter kant geeft het effect van een verandering in de parameters, de tweede term tussen haakjes geeft het effect van een verandering van de verdeling van de steekproefkenmerken.

In Tabel 3 zijn de resultaten gegeven van het voorspellen van de instroom in het Wo in jaar j (conditioneel op deelname aan het HO) gegeven de parameters van jaar i ($i, j = 1982, 1991$). Aangezien niet alle variabelen uit het model van directe instromers in 1991 ook in het 1982-bestand voorkomen is het model van de directe instromers in 1991 opnieuw geschat overeenkomstig het model van 1982. Deze resultaten zijn gebruikt voor de berekeningen in Tabel 3³.

Tabel 3: De voorspelde deelname aan het Wo met rij-jaar kenmerken gegeven kolom-jaar-parameters

kenmerken steekproef	parameters	
	1982	1991
1982	62,6	68,0
1991	63,5	70,8

We zien dat de verschillen tussen de kolommen veel groter zijn dan de verschillen tussen de rijen. Dit betekent dat de stijging in de deelname aan het Wo vooral veroorzaakt wordt door de verandering van de parameters, dat wil zeggen de verandering in het keuzegedrag. De veranderingen in de samenstelling van de populatie tussen 1982 en 1991 zijn slechts in geringe mate van invloed. Tussen 1982 en 1991 zijn de preferenties van de Vwo-leerlingen zodanig veranderd dat meer leerlingen voor een studie in het Wo kiezen. Hierbij speelt de afname van de selectie op grond van de eindexamencijfers een belangrijke rol. Een deel van de kwalitatief zwakkere Vwo'ers die in 1982 voor het Hbo gekozen zouden hebben kiest in 1991 voor het Wo. Dit betekent dat tussen 1982 en 1991 de verticale sortering van Vwo-eindexamenkandidaten is afgenomen.

5. DE KEUZE TUSSEN MBO EN HBO VOOR HAVO'ERS

Na het eindexamen kiezen vrijwel alle Havo-leerlingen die verder studeren voor een vervolgopleiding in het Mbo, Hbo of voor het Vwo. De keuze tussen Mbo en Hbo wordt hieronder geanalyseerd vanuit het perspectief van de 'verticale sortering'. Aangezien het niveau van het Vwo ten opzichte van het Mbo en het Hbo niet duidelijk is, wordt de keuzemogelijkheid Vwo bij onderstaande analyse buiten beschouwing gelaten.

Allereerst zijn in Tabel 4 (volgende pagina) de bivariate samenhangen van de geselecteerde variabelen met de keuze tussen Hbo en Mbo gegeven. Ook hier zijn per variabele de gemiddelde waarden berekend voor de groep die voor het Mbo kiest en voor de groep die voor het Hbo kiest.

In 1982 kiezen meisjes vaker voor het Mbo, dit verschil wordt in 1991 niet gevonden. In beide jaren lijken leerlingen die kiezen voor het Hbo hogere eindexamencijfers behaald te hebben hoewel de verschillen niet steeds significant zijn. Ook lijken Hbo'ers gemiddeld hoger opgeleide ouders te hebben. De schattingsresultaten van het multivariate model staan in Tabel 5. Bij de variabele 'inkomen ouders' is de quasi-elasticiteit gegeven, bij de overige variabelen zijn de afgeleiden gegeven. De waarden zijn berekend in het steekproefgemiddelde. Een positief effect betekent meer kans op een keuze voor het Hbo bij een toename van de betreffende variabele.

Tabel 4: Bivariate samenhangen met de keuze van Havo'ers tussen Mbo en Hbo

	1982			1991		
	Hbo	Mbo	p	Hbo	Mbo	p
vrouw	0,46	0,57	*	0,61	0,61	n.s.
leeftijd	18,1	18,1	n.s.	18	17,8	*
score talen	6,7	6,6	n.s.	6,7	6,5	*
score exacte vakken	6,5	6,2	**	6,6	6,1	**
schooladvies				4,6	4,8	n.s.
zittenblijven	0,48	0,53	n.s.	0,54	0,47	n.s.
opleiding ouders (1-5)	2,8	2,6	n.s.	2,9	2,6	
inkomen ouders	3530	3480	n.s.	3780	3860	n.s.
motieven (0-10)						
interesse	8,4	8,5	n.s.	9	8,5	*
zelfstandig werk				6,5	6,8	n.s.
leidinggevende functie				6,6	6,5	n.s.
kans op baan	7,5	6,9	*	8	8,5	n.s.
bepaald beroep	8,7	8,8	n.s.	8,3	8,4	n.s.
instelling dichtbij	2,9	3,3	n.s.	3,5	3,9	n.s.
studie niet moeilijk				4,2	5,5	**
cursusduur				5,6	6,6	**
baan in buitenland				3,6	3,5	n.s.
aandeel in totaal	69,1	30,9		59,3	40,7	
aantal	351	157		127	87	

Tabel 5: Geschatte effecten op keuze voor Hbo door Havo'ers

	1982	1991
geslacht	-10,7 *	-0,7
leeftijd	5,0	19,2 *
score talen	2,0	6,3
score exacte vakken	7,4 **	17,7**
schooladvies		-4,3
opleiding moeder t.o.v. middelbaar		
opleiding moeder laag	-2,9	-1,9
opleiding moeder hoog	26,2 *	56,1**
inkomen ouders		-0,01
motieven (0-10)		
interesse	0,1	6,3 *
kans op baan	-1,9 *	-3,8
bepaald beroep uitoefenen	-0,5	2,0
instelling dichtbij	-0,3	1,1
studie niet moeilijk		-4,6 *
cursusduur		-2,0
baan in buitenland		1,6
zittenblijven	10,6 *	-16,6
likelichood	-295,9	-112,3
lik. ratio test	36,5 **	64,7 **
aantal Mbo	157	87
aantal Hbo	351	127
totaal	508	214

* significant op 5 %-niveau; ** significant op 1 %-niveau.

Ook in het multivariate model wordt gevonden dat jongens in 1982 vaker voor het Hbo kozen en dat dit geslachtseffect in 1991 verdwenen is. Daarentegen blijkt de leeftijd alleen in 1991 van invloed te zijn. Oudere Havo-leerlingen kiezen vaker voor het Hbo. Uit ander onderzoek is bekend dat dit vooral 'stapelaars' zijn⁴.

De opleiding van de vader bleek niet van invloed op de keuze tussen Mbo en Hbo, dit in tegenstelling tot de niveaukeuze bij Vwo-leerlingen. In 1991 wordt wel een sterk milieu-effect gevonden via de opleiding van de moeder, dit effect is in 1982 aanmerkelijk zwakker.

Twee motieven zijn in 1991 significant van invloed op de keuze tussen Mbo en Hbo. Leerlingen die voor het Hbo gekozen hebben zijn sterker geïnteresseerd in het onderwerp van de studie en verwachten een hogere moeilijkheidsgraad bij hun studie dan degenen die voor het Mbo gekozen hebben. Beide motieven hebben direct te maken met verticale sortering; interesse verwijst naar motivatie en de moeilijkheidsgraad spreekt voor zich. In 1982 was er geen verschil in interesse in het onderwerp van de studie, wel hechten leerlingen die voor het Mbo kozen toen meer aan de kans op een baan. Zittenblijven vermindert de kans op een keuze voor het Hbo. In 1982 daalt de kans op een keuze voor het Hbo met 10 procent als de leerling in het voorgaande onderwijs is blijven zitten. In 1991 neemt de kans nog sterker af namelijk met 17 procent, dit effect is evenwel niet significant.

De invloed van eindexamencijfers

Hogere eindexamencijfers vergroten de kans dat een leerling kiest voor het Hbo. De effecten van de taalcijfers zijn niet significant maar het effect van de scores op de exacte vakken wel. Voor zowel talen als voor de exacte vakken zijn de effecten in 1991 aanzienlijk sterker. In 1991 is sprake van een sterkere selectie op grond van eindexamencijfers.

Veranderingen tussen 1982 en 1991

In Tabel 5 zagen we al een belangrijke verandering tussen 1982 en 1991: tussen 1982 en 1991 is de selectie op grond van eindexamencijfers toegenomen. Interessant is nu om de verandering in het totale effect van de parameters tussen 1982 en 1991 te bepalen. Hiervoor wordt op dezelfde wijze als bij de Vwo-leerlingen de totale verandering tussen 1982 en 1991 uiteengelegd in een gedeelte veroorzaakt door een verandering in de samenstelling van de populatie en een gedeelte veroorzaakt door veranderingen van de preferenties van de populatie. Tabel 6 geeft de resultaten van deze decompositie. Aangezien het model van 1982 afwijkt van dat van 1991 in Tabel 5 is het model van 1991 opnieuw geschat volgens de specificatie van 1982.

Tabel 6: De voorspelde deelname aan het Hbo met rij-jaar kenmerken gegeven kolom-jaar parameters

kenmerken steekproef	parameters	
	1982	1991
1982	69,1	62,9
1991	65,5	59,2

De totale daling van de deelname aan het Hbo met 10 procent wordt in iets sterkere mate veroorzaakt door veranderingen in de voorkeuren van de populatie dan door veranderingen in de samenstelling van de populatie. Vooral de eerste verandering is zeer interessant. Een deel van de Havo-leerlingen, die in 1982 nog voor het Hbo kozen, kiezen in 1991 voor het Mbo. Gezien het sterkere effect van de 'verticale sortering' variabelen eindexamencijfers, zittenblijven en een aantal motieven in 1991 kan geconcludeerd worden dat de verticale sortering in 1991 sterker is dan in 1982. De sombere verhalen over de overgang Havo-Hbo hebben kennelijk geleid tot een voorzichtiger niveaukeuze bij de Havo-leerlingen. Ook lijkt de toegenomen (zelf)selectie

tie van Havo-leerlingen samen te gaan met een toename van de invloed van de sociale achtergrond van de leerlingen. Een sterkere meritocratische selectie blijkt dus gelijktijdig op te treden met een sterkere sociale selectie.

Gevolgen voor het Hbo

Het is interessant om de ontwikkelingen bij de Havo-leerlingen te vergelijken met die bij de Vwo-leerlingen. Bij de Havo-leerlingen zien we een toename van de verticale sortering terwijl bij de Vwo-leerlingen juist sprake is van een afname van de verticale sortering. Het Hbo ondervindt de gevolgen van beide veranderingen. Enerzijds krijgt het Hbo in 1991 vergeleken met 1982 kwalitatief betere Havo-leerlingen. Anderzijds kiest een deel van de Vwo-leerlingen die in 1982 voor het Hbo gekozen zouden hebben nu voor het Wo. Het Hbo krijgt hierdoor waarschijnlijk een wat meer heterogene maar kwalitatief slechtere Vwo-instroom. De Vwo-instroom in het Hbo zal meer heterogeen zijn vanwege de afname van het effect van de eindexamencijfers, maar ook kwalitatief slechter omdat de betere Vwo-leerlingen, uit de groep die in 1982 naar het Hbo zou zijn gegaan, nu naar het Wo zullen gaan. Voor het Hbo resulteert dus een kwalitatief betere Havo-instroom maar een kwalitatief slechtere Vwo-instroom. Waarschijnlijk leidt dit tot een meer homogene directe instroom.

6. CONCLUSIES

De keuze van het niveau van de vervolgopleiding door Vwo- en Havo-leerlingen is in de voorgaande analyses benaderd vanuit het perspectief van 'verticale sortering': Kiezen individuen met een hoger niveau van initiële capaciteiten voor een hogere vervolgopleiding?

Vwo

Bij de Vwo-leerlingen zijn vier groepen geanalyseerd te weten: directe doorstromers in 1982 en 1991, eerstejaars en ouderejaars in 1991. De belangrijkste bevindingen zijn:

- leerlingen met hogere cijfers op het eindexamen kiezen vaker voor het Wo; de sterkte van dit effect is tussen 1982 en 1991 duidelijk afgenomen;
- leerlingen met een hoger opgeleide vader kiezen vaker voor het Wo;
- leerlingen die voor het Hbo kiezen zijn beroepsgerichter;
- de verandering in de deelname aan het Wo tussen 1982 en 1991 wordt voornamelijk veroorzaakt door een verandering in de preferenties van de leerlingen en slechts in geringe mate door veranderingen in de samenstelling van de populatie;
- tussen 1982 en 1991 is de 'verticale sortering' afgenomen, dat betekent dat de instroom in het Wo vanuit het Vwo in 1991 kwalitatief minder is dan in 1982.

Havo

Bij de keuze tussen Mbo en Hbo van Havo-abituriënten is de verticale sortering toegenomen. Het effect van de eindexamencijfers, het zittenblijven en een aantal motieven is in 1991 sterker. Ook wordt in 1991 een sterke invloed gevonden van de opleiding van de moeder, dit effect is duidelijk sterker dan in 1982. Een toename van de verticale sortering gaat kennelijk samen met een toename van de sociale selectie. De totale verandering tussen 1982 en 1991 wordt in iets sterkere mate veroorzaakt door veranderingen in de preferenties van de leerlingen dan door veranderingen in de samenstelling van de populatie.

Gevolgen voor Hbo

Het Hbo ondervindt de directe gevolgen van zowel de veranderingen in de Havo-instroom als in de Vwo-instroom. Enerzijds krijgt het Hbo vergeleken met 1982 in 1991 kwalitatief bete-

re Havo-leerlingen maar anderzijds ook kwalitatief slechtere Vwo-leerlingen. Waarschijnlijk ontstaat hierdoor een meer homogene directe instroom in het Hbo.

Dient er sterker geselecteerd te worden?

In de inleiding is de vraag gesteld of de roep om sterkere selectie en verwijzing gerechtvaardigd is. De bovenstaande analyses laten zien dat de zelfselectie van Vwo'ers die voor het Wo kiezen is afgenomen in de onderzochte periode. Wanneer het keuzegedrag in 1982 als meer wenselijk beschouwd wordt dan het gedrag in 1991 zou dit resultaat kunnen worden opgevat als een ondersteuning van de door het ministerie voorgestelde sterkere selectie voor het Wo. Voor het Hbo geldt dit evenwel niet.

LITERATUUR

- Cramer, J.S. (1991), *The LOGIT-model: an introduction for economists*, Edward Arnold, London
- Gomulka, J. and N. Stem, The employment of married women in the United Kingdom 1970-1983, in: *Economica*, 57, 171-99
- Hartog, J. (1992), *Capabilities, allocation and earnings*, Kluwer
- Jong U. de, P. Koopman en J. Roeleveld (1991), *Snelwegen en slingerpaden in en om het Hoger Onderwijs*; eindrapport project 'Studieloopbanen in het Hoger Onderwijs', Reeks Achtergrondstudies Hoger Onderwijs en Wetenschappelijk Onderzoek, no. 10, Ministerie van Onderwijs en Wetenschappen, Den Haag
- Jong U. de, H. Oosterbeek, J. Roeleveld en H.D. Webbink (1992), *Voornemens van eindexamenkandidaten 1991*, Ministerie van Onderwijs en Wetenschappen, Serie Verder Studeren, een panelstudie onder scholieren en studenten, Den Haag
- Kodde, D.A. en J.M.M. Ritzen (1986), *Vraag naar Hoger Onderwijs. Eindrapport*, Ministerie van Onderwijs en Wetenschappen, Serie beleidsgerichte studies Hoger Onderwijs en Wetenschappelijk Onderzoek, nr. 6, Den Haag
- Webbink, H. D., U. de Jong, H. Oosterbeek en J. Roeleveld (1993), *Studiekeuze van scholieren en studenten in 1991*, Ministerie van Onderwijs en Wetenschappen, Serie Verder Studeren deel 2, een panelstudie onder scholieren en studenten, Den Haag

1. Werkte tijdens het schrijven van dit artikel als senior onderzoeker bij de Stichting voor Economisch Onderzoek. Werkt momenteel als onderwijs econoom bij de Onderwijsraad.

2. Voor een uitvoerige beschrijving van de opzet en doelstelling van het project alsmede van de steekproef-trekking en de representativiteit zie de Jong e.a. (1992).

3. De motieven werden in 1982 op een andere wijze gevraagd. Men kon kiezen uit vier antwoordcategorieën: (zeer) onbelangrijk, (zeer) belangrijk. Deze categorieën zijn met de respectievelijke waarde 0; 2,5; 7,5; 10 getransformeerd naar de in 1991 gebruikte 0-10 schaal.

4. De Jong, e.a. (1991).

Ontvangen 21-03-97

Definitief aanvaard 05-11-97

Invloed van het vakkenpakket VWO op het studieresultaat in de propedeuse WO

E.P.W.A. Jansen¹

GION, afdeling COWOG, Rijksuniversiteit Groningen

ABSTRACT

The second stage of secondary education will be restructured according to the 'profielnota's'. The policy of the ministry of education is to formulate transfer sets, specific clusters of subjects in secondary education, that should guarantee a better transfer between secondary and university education. On the universities this should result in better completion rates in a shorter time period for the first-year examination and it should result in less drop-out. The analysis of students' progress of five cohorts of students from eight fields of study at the University of Groningen does not support the expectation that the proposed transfer sets contribute to more study success. The Mean Grade Point Average at secondary education seems to be the best predictor for the time students need to complete their first-year examination. Probably, the Mean Grade Point Average reflects a more general level as well as the ability to study successfully on a university.

INLEIDING

In april 1991 is door de staatssecretaris van Onderwijs en Wetenschappen de nota 'Profiel van de tweede fase voortgezet onderwijs' uitgebracht, gevolgd door een vervolgnota in mei 1992 op basis van commentaren op de eerste uitgebrachte nota. In de nota 'Profiel' worden voorstellen gedaan voor de inrichting van de tweede fase van het voortgezet onderwijs. Het gaat hierbij uiteraard over het hele scala aan mogelijkheden van voortgezet onderwijs, van vervolg op lbo of MAVO tot de inrichting van de hogere leerjaren van het VWO en HAVO. In de nota worden voor het VWO expliciet drie knelpunten genoemd, waaronder de aansluiting tussen VWO en WO.

Profielen en de aansluitingsproblematiek

In de nota 'Profiel' wordt gesteld dat het VWO zijn voorbereidende functie op het wetenschappelijk onderwijs behoudt. Om studie-uitval en -vertraging in het WO te reduceren zou een betere aansluiting tussen VWO en WO noodzakelijk zijn. Dit heeft zowel inhoudelijke consequenties, die tot uiting komen in de voorgestelde doorstroomprofielen, als ook consequenties voor de inrichting van het onderwijs, het idee van de school als studiehuis met de docent als begeleider. Zo zou binnen het VWO meer aandacht moeten komen voor vaardigheden die binnen het WO vereist zijn, zoals studievaardigheden, studieplanning en schrijfvaardigheden. Om vaardigheden systematisch in het onderwijs in te bouwen is het volgens de vervolgnota juist noodzakelijk om een inhoudelijk *samenhangend* onderwijsaanbod te omschrijven. De huidige vakkenpakketkeuze zou één van de oorzaken zijn dat het systematisch inbouwen van vaardigheden ontbreekt. Het invoeren van profielen, waardoor een inhoudelijk samenhangend vakkenpakket zou ontstaan en het bepalen van de relatieve omvang van vakken in termen van studielast in plaats van 'lessen', waardoor meer ruimte voor variatie in werkvoor-

men ontstaat, zou voor een betere aansluiting van het VWO op het WO moeten zorgen.

In de nota wordt gesproken over vier profielen, namelijk

- natuur en techniek
- natuur en gezondheidszorg
- economie en maatschappij
- cultuur en maatschappij.

Bij de invulling is sprake van vakken en deelvakken en er is sprake van een verplicht en vrij profieldeel. In de uitwerking op hoofdlijnen die de Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs begin dit jaar heeft gepresenteerd, wordt de omvang en globale inhoud van de vier profielen weergegeven.

In een reactie op de vervolgnota profiel onderstreept de VSNU het belang van vaardigheden in de vernieuwing en herstructurering van de bovenbouw van het VWO. Juist omdat de VSNU zo'n groot gewicht toekent aan vaardigheden zoals taal- en studievaardigheden, plannen van eigen gedrag, methodisch aanpakken van problemen en het omgaan met verschillende bronnen, is zij erop tegen dat er deelvakken worden ingevoerd. Vaardigheden zijn niet los van een bepaalde vakinhoud aan te leren. De kenmerkende aanpak van een discipline kan alleen worden overgebracht aan studenten als het vak een redelijke omvang heeft. Verder pleit de VSNU voor twee algemeen geformuleerde pakketten, een bèta-pakket en een alfa/gamma pakket. De instromende studenten dienen over een brede basis te beschikken van zowel kennis als vaardigheden. De VSNU verwacht niet dat de door de Stuurgroep voorgestelde profielen een einde zullen maken aan de deficiëntieproblematiek. Zij verwacht eerder een toename van deficiënte studenten, omdat is gebleken dat het voor leerlingen moeilijk is in zo'n vroeg stadium een definitieve studiekeuze te maken.

Misschien om tegemoet te komen aan de kritiek van onder meer de VSNU geeft de Stuurgroep in de uitwerking op hoofdlijnen aan dat door een bepaalde invulling van het vrije deel het heel goed mogelijk is om twee profielen als eindexamenpakket te hebben.

In deze bijdrage gaan we na of de kans op het succesvol afronden van de propedeuse samenhangt met het vakkenpakket uit het VWO en in hoeverre studiesucces in de propedeuse samenhangt met de voorgestelde profielen.

Aansluiting VWO-WO

Het is duidelijk dat het aansluitingsprobleem tussen VWO en WO uit twee componenten bestaat. Enerzijds is er mogelijk sprake van een aansluitingsprobleem qua vakinhoud i.e. vakkenpakket, anderzijds is er sprake van een aansluitingsprobleem op het gebied van vaardigheden. In deze bijdrage beperken we ons tot de vakkenpakketten in relatie tot de studievoortgang in de propedeuse.

Hoewel de samenstelling van het vakkenpakket vaak genoemd wordt in relatie tot studiesucces in het WO, is er binnen het WO weinig systematisch onderzoek naar gedaan. De VSNU heeft geïnventariseerd of in lopend of reeds uitgevoerd onderzoek melding gemaakt wordt van een relatie tussen VWO-vakken en/of eindexamencijfers enerzijds en studiesucces in de propedeuse anderzijds. Er was weinig materiaal beschikbaar (Smit, 1994). Enige resultaten zijn gemeld met betrekking tot toegepaste wiskunde (UT) en bouwkunde (TUD), psychologie en geschiedenis. Hierbij is gekeken naar het verband tussen VWO-wiskunde en studiesucces in de propedeuse WO. Voor de technische studies blijkt een duidelijk verband tussen het cijfer wis- en natuurkunde VWO en studiesucces in de propedeuse. Bij psychologie bleek dat VWO-ers met een wiskunde deficiëntie een aanmerkelijk lagere slaagkans hadden voor de propedeuse dan de VWO-ers met wiskunde A. Bij geschiedenis lijkt bèta-kennis een contra-indicatie voor studiesucces. In geen van de genoemde studies is echter nagegaan wat het totale vakkenpakket of een bepaalde combinatie van vakken bijdraagt aan de slaagkans voor de propedeuse.

PROBLEEMSTELLING

De probleemstelling voor dit onderzoek kan in de volgende deelvragen uiteengelegd worden:

- In hoeverre passen studenten met een VWO-vooropleiding die nu aan de universiteit studeren in één van de voorgestelde profielen?
- Hebben studenten die in het VWO een pakket hebben gekozen dat overeenkomt met de (voorlopige) profielen uit de nota 'Profiel' de propedeuse in een 'aansluitende studierichting' vaker en/of sneller behaald dan studenten die in het VWO niet een dergelijk pakket hebben gevolgd?
- Welke vakken of combinatie van vakken uit het VWO dragen bij tot een hogere studiesnelheid in de propedeuse van het WO?

METHODE

Om een antwoord te kunnen geven op de in de probleemstelling genoemde vragen zijn de studievoortgangsgegevens van studenten van de cohorten 1987 t/m 1991 van acht studierichtingen² aan de RUG geanalyseerd. Hierbij zijn alleen studenten met een VWO-vooropleiding die zijn gaan studeren aan de RUG in de analyses opgenomen. De gegevens zijn ontleend aan het centrale studieresultatenregistratiesysteem BASIS. Via het door het COWOG ontwikkeld programma ProgRESS (Programma voor Registratie en Evaluatie van Studieresultaten van Studenten) zijn de gegevens overgebracht naar een plat bestand, waarop verdere analyses met SPSS zijn uitgevoerd. De analyses hebben betrekking op de inhoud van het vakkenpakket, dat wil zeggen op het wel of niet hebben gevolgd van het vak of de vakken op het VWO, als wel op de behaalde eindexamencijfers op het vak of de vakken.

Het behalen van de propedeuse en de snelheid waarmee de propedeuse behaald is (aantal maanden) zijn als criteriumvariabelen gebruikt. De propedeuse is onder andere bedoeld om studenten te selecteren op geschiktheid en het ligt derhalve voor de hand om geen direct effect van vakkenpakket op het VWO te verwachten op het behalen van het doctoraal-examen.

Naast het al dan niet behalen van de propedeuse is ook nagegaan of er samenhangen gevonden konden worden tussen (de cijfers op) vakken op het VWO en de gerealiseerde studielast (het behaalde aantal studiepunten) na één jaar studie.

Beperking

Het is zaak hier een beperking te noemen die aan de hier gepresenteerde gegevens kleeft. Het gaat alleen over de in de profielen genoemde vakken. Hierbij speelt het probleem dat er in de profielen sprake is van deelvakken en het probleem dat ook de vakinhouden in een aantal gevallen drastisch veranderd wordt, waardoor niet echt een zuivere vergelijking mogelijk is tussen de vakken die de studenten die nu studeren hebben gevolgd in het VWO en de vakken die de nieuwe generatie studenten onder het profielenregime zullen volgen.

RESULTATEN

Beschrijving populatie

De cohorten 1987 tot en met 1991 van de volgende acht studierichtingen zijn in het onderzoek betrokken: Engels, Geschiedenis, PAO (pedagogiek, andragogie en onderwijskunde), Psychologie, Geneeskunde, Farmacie, Economie en Bedrijfskunde. Ter beschrijving van de populatie zijn in de tabellen 1 en 2 per studierichting de gemiddelde eindexamencijfers, het gemiddeld aantal b-vakken en het percentage studenten zonder b-vakken weergegeven. Onder b-vakken verstaan we wiskunde B, natuurkunde, scheikunde en biologie. Het aantal b-vakken

is hier als kenmerk opgenomen omdat uit onderzoek is gebleken dat b-vakken in het vakkenpakket een voorspeller voor studiesucces zijn (zie bijvoorbeeld de Jong e.a. 1991).

Tabel 1: Gemiddeld eindexamencijfer per studierichting (cohorten '87 t/m '91)

studierichting	gemiddeld eindexamencijfer
Engels	6.7
Geschiedenis	6.8
PAO	6.6
Psychologie	6.7
Geneeskunde	6.9
Farmacie	6.8
Economie	6.7
Bedrijfskunde	6.8

Tabel 2: b-vakken in vakkenpakket per studierichting (wiskunde B, Scheikunde, Natuurkunde, Biologie)

studierichting	gemiddeld aantal b-vakken	% studenten zonder b-vakken	indien wel b-vakken gemiddeld aantal
Engels	.6	63	1.7
Geschiedenis	.7	58	1.6
PAO	1.5	31	2.2
Psychologie	1.4	35	2.1
Geneeskunde	3.4	0	3.4
Farmacie	3.5	0	3.5
Economie	1.4	34	2.1
Bedrijfskunde	1.3	37	2.1

We zien dat het gemiddelde eindexamencijfer varieert tussen de 6.6 en 6.9. Het gemiddeld aantal b-vakken varieert tussen de .6 en 3.5, en het percentage studenten zonder b-vakken tussen de 0 en 63%.

Profielen

De vier profielen zijn als volgt gedefinieerd voor de analyses:

- profiel 1 (cultuur en maatschappij): economie 1, wiskunde A, geschiedenis of aardrijkskunde, moderne vreemde taal;
- profiel 2 (natuur en gezondheid): natuurkunde, scheikunde, wiskunde B, biologie;
- profiel 3 (economie en maatschappij): economie 1, economie 2, wiskunde A, aardrijkskunde of geschiedenis;
- profiel 4 (natuur en techniek): natuurkunde, scheikunde, wiskunde B en geen biologie.

De vraag kan gesteld worden of wiskunde A bij de profielen 1 en 3 expliciet opgenomen moet worden, omdat wiskunde A voor een deel in het gemeenschappelijke deel van de tweede fase VWO opgenomen is. Maar ook dan is in de profielen voorzien in een 'aanvullend' deel wiskunde A. Daarom is ervoor gekozen in deze bijdrage de profielen te definiëren met wiskunde A. Een definiëring van de profielen zonder wiskunde A leverde geen noemenswaardige verschuivingen op in de verdeling van studenten over de profielen.

Onduidelijk is voor het wo welke profielen aansluiten op welke studierichting. Bij farmacie

bijvoorbeeld speelt de vraag of farmacie zou moeten aansluiten bij profiel 2, natuur en gezondheid, of bij profiel 4, natuur en techniek. Omdat er volgens de gebruikte definiëring in de onderzoekspopulatie geen zuivere profiel 4 studenten zitten, speelt dit probleem bij de analyses geen rol en is als uitgangspunt gehanteerd dat profiel 2 het 'juiste' voorbereidingstraject weergeeft. Bij psychologie en PAO is het moeilijker te beoordelen welk profiel het meest op hen van toepassing is. Psychologie stelt zich vaak op als dichter hangend tegen de bèta-wetenschappen aan dan tegen de sociale wetenschappen. Dit zou pleiten voor profiel 2. Op de VSNU-themadag Aansluiting VWO-WO, gehouden op 7 juni 1994, werden de sociale wetenschappen gebracht onder profiel 3, evenals economie en bedrijfskunde. Profiel 1 zou moeten voorbereiden op alfa-studierichtingen zoals geschiedenis en Engels.

Een overzicht van de percentages studenten in onze onderzoeksgroep die in het VWO een vakkenpakket hebben gevolgd, dat overeenkomt met de voorgestelde profielen, zoals hierboven gedefinieerd, zien we in tabel 3.

Tabel 3: % studenten met vakkenpakket overeenkomstig met de voorgestelde profielen (profielen gedefinieerd met wiskunde A)

profiel	studierichting							
	ENG	GESCH	PAO	PSY	GK	FAR	ECO	BDK
1	8	10	9	12	1	0	6	8
2	2	1	11	6	50	56	4	4
3	50	62	36	34	7	7	38	34
4								
geen	40	27	44	48	42	37	51	50
n	773	901	609	1088	1057	476	3308	2093

Het eerste dat opvalt is dat bij de meeste geanalyseerde studierichtingen meer dan 40% van de studenten geen vakkenpakket heeft dat binnen één van de profielen past. Van hen die dat wel hebben, studeert bij PAO, psychologie, bedrijfskunde en economie ongeveer een derde en bij farmacie en geneeskunde bijna de helft van de studenten in het juiste profiel volgens het profielenidee. Bij psychologie ligt de keuze van de vakken door studenten meer in een van de maatschappijprofielen dan in het natuur en gezondheidsprofiel. Bij geschiedenis en Engels zien we dat de onderzochte studentpopulatie voor de helft of meer vakken heeft die overeenkomen met profiel 3 en niet met profiel 1 dat als een meer alfa-gericht profiel kan worden gezien.

Voor de verschillende studierichtingen zijn we nagegaan of er een samenhang bestond tussen het 'juiste' of 'niet-juiste of geen' profiel en het behalen van de propedeuse binnen 1 jaar of langer dan een jaar. Het behalen van de propedeuse binnen 1 jaar is gedefinieerd als 14 maanden of minder, om administratieve 'fouten' uit te sluiten. Het blijkt dat studenten vaak eerder aan alle verplichtingen hebben voldaan dan de datum die uiteindelijk als propedeusedatum geregistreerd wordt. Langer dan een jaar is gedefinieerd als 15 maanden of meer. Dus ook studenten die langer dan twee jaar over de propedeuse hebben gedaan zijn in deze groep opgenomen. Gezien de kleine aantallen waar het hier om gaat, was het niet zinvol deze laatste groep als aparte groep op te nemen.

Het verband is in een kruistabel getoetst met een Chi-kwadraattoets. Voor de studierichting farmacie kan een significant positief verband aangetoond worden en voor PAO een significant negatief verband. Bij de andere studierichtingen kon geen significant verband worden aangetoond.

Tabel 4: Verband tussen 'juiste' profiel en het behalen van de propedeuse in 1 jaar (gedefinieerd als 14 maanden of minder) of langer dan 1 jaar (gedefinieerd als 15 maanden of meer), getoetst met een Chi²-toets

profiel studierichting	juist % geslaagden prop. 1 jaar	niet juist % geslaagden prop. 1 jaar	Chi-kwadraat- toets p-waarde
Engels	73	65	< .33 n.s.
Geschiedenis	63	58	< .54 n.s.
PAO	67	76	< .10
Psychologie	44	52	< .17 n.s.
Geneeskunde	68	66	< .71 n.s.
Farmacie	65	35	< .00
Economie	16	15	< .84 n.s.
Bedrijfskunde	14	13	< .82 n.s.

Kijken we naar tabel 5 dan zien we bij vijf studierichtingen een significant verband tussen het 'juiste' profiel en het al dan niet behaald hebben van de propedeuse, ongeacht de tijd die de student er voor gebruikt heeft. Voor psychologie en economie is het verband negatief, voor de studierichtingen engels, PAO en geneeskunde is het verband positief.

Tabel 5: Verband tussen 'juiste' profiel en het al dan niet behalen van de propedeuse, getoetst met een Chi²-toets

profiel studierichting	juist % geslaagden propedeuse	niet juist % geslaagden propedeuse	p-waarde Chi-kwadraat
Engels	77	86	<.08
Eeschiedenis	75	68	<.23 n.s.
PAO	71	80	<.10
Psychologie	51	66	<.00
Geneeskunde	92	86	<.09
Farmacie	79	74	<.56 n.s.
Economie	38	45	<.02
Bedrijfskunde	72	73	<.70 n.s.

Vakken(pakket)

De profielen geven dus geen eenduidige aanwijzing voor mogelijk studiesucces in de propedeuse. Daarom zijn we nagegaan of mogelijk bepaalde vakken samenhang vertonen met het aantal maanden waarin de propedeuse is behaald. In tabel 6 staan de correlaties tussen de cijfers op de afzonderlijke vakken en de studiesnelheid, uitgedrukt in het aantal maanden waarin de propedeuse is behaald. Een vak is alleen opgenomen indien een kwart of meer van de studenten dat vak in het VWO gevolgd had.

Tabel 6: Correlatie tussen aantal maanden waarin de propedeuse is behaald en het cijfer op VWO-vakken en het gemiddelde eindexamencijfer

VWO-vak	studierichting							
	ENG	GESCH	PAO	PSY	GK	FAR	ECO	BDK
eng	.27	.12	<u>.09</u>	.14	.18	<i>.10</i>	.14	.16
fra	.27	<u>.14</u>		.19			.27	.25
dui	.18	.11	.27	.20			.29	.23
ned	.25	.16	.29	.12	.21	.21	.18	.21
aard	<u>.18</u>	<u>.15</u>		<u>.19</u>			.26	
gesch	.22	.16	<i>.03</i>	.24			.30	.22
bio			.19	.17	.23	.24		
nat				.20	.33	.33	.35	.34
scheik				.22	.30	.33	.31	.28
eco1	<u>.15</u>	<u>.11</u>	.24	.19	.29		.34	.34
eco2	<i>.03</i>	.15					.43	.37
wiskA	.19	.17	.15	<u>.13</u>	.27	.27	.31	.32
wiskB					.24	.30	.33	.27
gem. cijfer	.35	.24	.29	.26	.37	.37	.43	.39

* De cursief gedrukte correlaties zijn niet significant, de onderstreepte correlaties zijn significant op .05 niveau en de overige correlaties zijn significant op .01 niveau.

** VWO-vakken zijn alleen meegenomen indien ten minste een kwart van de studenten uit de betreffende studierichting in dat vak eindexamen had gedaan.

Bij bijna alle vakken blijken de behaalde eindexamencijfers significant samen te hangen met studiesnelheid. Gezien de hoge correlatie met het gemiddeld eindexamencijfer lijkt studiesucces meer een kwestie van niveau dan van vakkenpakket. Bij de studierichtingen bedrijfskunde en economie speelt economie 2 in het VWO een belangrijke rol. Dit VWO-vak heeft een bijna evenhoge correlatie met studiesnelheid als het gemiddelde eindexamencijfer. Verder is opvallend dat wiskunde A in het VWO bij psychologiestudenten laag of niet significant correleert met het aantal maanden dat studenten nodig hadden om hun propedeuse te behalen. Eveneens noemenswaardig is het feit dat het VWO-vak geschiedenis geen significante correlatie laat zien bij de studierichting PAO, maar wel redelijk hoog correleert bij de 'aanverwante' studierichting psychologie.

We zijn ook nagegaan of er een samenhang is tussen de cijfers op de VWO-vakken en de gerealiseerde studielast na één jaar. De correlaties geven ongeveer hetzelfde beeld te zien als die in tabel 6, behalve dat bij psychologie en PAO de correlatie met het gemiddelde eindexamencijfer niet meer significant is.

Waarschijnlijk hangen in dat geval specifieke kennis en vaardigheden uit bepaalde VWO-vakken samen met kennis en vaardigheden die in bepaalde propedeusevakken gevraagd worden.

Alfa-, bèta-, gammavakken

Vanuit de literatuur is met name bekend dat bètavakken in het vakkenpakket het studiesucces in de propedeuse gunstig beïnvloeden. Ook in dit onderzoek zijn we nagegaan of het aantal bètavakken of het gemiddelde cijfer daarop samenhangt met het aantal maanden waarin de propedeuse is behaald (studiesnelheid). We hebben dit ook gedaan voor de alfavakken en gammavakken. Alfavakken zijn gedefinieerd als Frans, Duits en geschiedenis; bètavakken als wiskunde B, natuurkunde, scheikunde en biologie; gammavakken als wiskunde A, economie I en economie II. De gemiddelde cijfers op deze clusters van vakken correleren bij alle studierichtingen significant met studiesnelheid, behalve bij farmacie waar het gemiddelde cijfer op de alfavakken, voor zover studenten die gevolgd hebben, niet significant correleert met studiesnelheid. Het aantal bèta-vakken correleert significant met studiesnelheid bij psychologie, geneeskunde, en economie. Het aantal alfavakken en gammavakken geeft bij geen enkele studierichting een significante correlatie te zien met studiesnelheid.

Wat voorspelt studiesnelheid?

De laatste analyses betreffen een meervoudige regressieanalyse waarbij onderzocht is welke vakken waarin door ten minste een kwart van de studenten uit de betreffende studierichting eindexamen is gedaan, studiesnelheid het beste voorspellen. Indien een student een vak niet in zijn/haar eindexamenpakket heeft gehad, is voor dat betreffende vak een score 0 toegerekend. Naast de afzonderlijke vakken is ook het gemiddelde eindexamencijfer in de analyse meegenomen. Zoals te verwachten was, is het gemiddelde eindexamencijfer de beste voorspeller bij alle studierichtingen en verklaart tussen de 5 en 18% van de variantie in het aantal maanden dat een student doet over zijn/haar propedeuse. De cijfers op de afzonderlijke vakken dragen verder weinig bij aan een verhoging van de verklaarde variantie.

Laten we het gemiddelde eindexamencijfer weg als onafhankelijke variabele in de regressieanalyse dan zien we in tabel 7 voor de verschillende studierichtingen de volgende combinaties van vakken als voorspellers naar voren komen. In de tabel worden de gestandaardiseerde regressiecoëfficiënten weergegeven van de significante VWO-vakken in de regressieanalyse. Hierbij is gebruik gemaakt van de stapsgewijze methode met defaultswaarden om een variabele in een volgende stap op te nemen.

Tabel 7: Gestandaardiseerde regressiecoëfficiënten, afhankelijke variabele aantal maanden gebruikt voor het behalen van de propedeuse

VWO-vak	studierichting							
	ENG	GESCH	PAO	GK	PSY	FAR	ECO	BDK
eng	-.20			-.09			-.07	
fra					-.11		-.10	-.14
ned	-.19	-.16	-.27	-.16	-.09	-.16	-.11	-.15
dui		-.10					-.15	-.16
bio			-.14					
scheik						-.31	-.20	-.15
nat				-.10	-.20		-.13	-.12
wiskA					-.09			-.11
wiskB				-.16			-.11	-.10
eco1							-.10	-.16
eco2							-.25	-.19
gesch							-.07	
aard								
R ²	.10	.04	.09	.10	.05	.14	.13	.14

We zien dat de verklaarde variantie op grond van de combinaties van vakken voor de verschillende studierichtingen varieert tussen de 4% en 14%.

Nemen we de gerealiseerde studielast na 1 jaar als afhankelijke variabele dan komt het gemiddelde eindexamencijfer wederom op de eerste plaats en levert in combinatie met één tot drie VWO-vakken een verklaarde variantie tussen de 3% bij psychologie en 27% bij economie.

Zonder het gemiddelde eindexamencijfer levert de meervoudige regressieanalyse met de cijfers op de VWO-vakken als onafhankelijke en gerealiseerde studielast na 1 jaar als afhankelijke variabele als resultaat dat het cijfer voor Nederlands behalve bij psychologie bij alle studierichtingen op de eerste of tweede plaats komt.

Tabel 8: Gestandaardiseerde regressiecoëfficiënten, afhankelijke variabele gerealiseerde studielast na 1 jaar *

VWO-vak	studierichting							
	ENG	GESCH	PAO	GK	PSY	FAR	ECO	BDK
eng	.18		-.10					
fra					.09		.13	.18
ned	.15	.20	.25	.16	.12	.21	.11	.20
dui					.07		.18	.14
bio				.12	.13			
scheik				.12		.27	.19	.15
nat				.07			.20	.16
wiskA						.10	.04	.08
wiskB				.10		.15	.19	.11
eco1					-.11		.14	.16
eco2							.21	.16
gesch		.15					.07	
aard							.07	
R ² .18	.07	.07	.06	.10	.07	.19	.17	.17

* Bij de studierichting Engels waren alleen gegevens beschikbaar voor de cohorten 1990 en 1991.

Bij sommige studierichtingen komen andere combinaties van vakken naar voren dan in de regressie-analyse met het aantal maanden waarin de propedeuse is behaald (zie tabel 8). Dit wijst waarschijnlijk naar een direct verband tussen de leerstof van het betreffende VWO-vak en de leerstof van een bepaald propedeusevak. Ook bij deze afhankelijke variabele is de verklaarde variantie tussen de 6% en 19% laag te noemen.

DISCUSSIE

Bij de discussies over de herinrichting van de tweede fase van het Voortgezet Onderwijs is het HBO van meet af aan veel intensiever betrokken geweest dan het WO. Een mogelijke verklaring hiervoor is dat de problematiek rond de aansluiting tussen het HAVO en HBO al sinds langere tijd pregnanter naar voren is gekomen. Bezwaren lagen bij het HBO bij het niet aansluiten van de inhoud en de zwaarte van het vakkenpakket. De profilering binnen het HAVO is ook duidelijker gericht op de verschillende sectoren binnen het HBO waar zij toegang toe

dienen te verschaffen (zie bijvoorbeeld Verhoeven en Van Zoelen, 1993). De discussies over aansluitingsproblemen tussen VWO en WO spelen zich vaak meer af op het terrein van ontbreken van bepaalde studievaardigheden, een zelfstandige en wetenschappelijke studiehouding en dergelijke en minder op het ontbreken van specifieke kennis vanuit het VWO. Een één op één relatie tussen een bepaald profiel en een bepaald cluster van studierichtingen in het WO wordt niet adequaat geacht. Volgens Van Lieshout op de VSNU-themadag Aansluiting VWO-WO is een strikte voorbereiding via een eigen profiel voor clusters van studierichtingen binnen het WO niet noodzakelijk. De voorgestelde profielen voor de tweede fase van het VWO lijken op grond van de hier gepresenteerde gegevens niet direct samen te hangen met studiesucces. In recent onderzoek kwamen ook Den Boer en Guldemond (1996) tot deze conclusie. Prins (1997) verwierp op grond van zijn onderzoek de hypothese dat studieuitval lager is bij studenten met een vakkenpakket dat inhoudelijk aansluit op het opleidingstype.

Van Lieshout stelde ook de vraag of er twee zinvolle profielen 'cultuur en maatschappij' en 'economie en maatschappij' voor het VWO kunnen worden ontwikkeld. De keuze van een meerderheid van de studenten uit de studierichtingen geschiedenis en Engels ondersteunen dit vermoeden dat het moeilijk zal zijn een zinvol onderscheid te maken tussen profiel 1 en 3.

Zoals ook al uit andere onderzoeken naar voren is gekomen vertonen de cijfers op de bèta-vakken een relatief grote samenhang met studiesnelheid bij de studierichtingen farmacie, geneeskunde, bedrijfskunde en economie. Bij psychologie is de correlatie met wiskunde A laag. Dit is in tegenspraak met de vaak genoemde onmisbaarheid van wiskunde A voor het wetslagen van de studie psychologie. Verder is het opvallend dat het cijfer voor economie 2 bij bedrijfskunde en economie een belangrijke plaats inneemt. Dit pleit voor het opnemen van beide economievakken in het profiel, zoals ook is voorgesteld door de Stuurgroep.

Het gemiddelde eindexamencijfer is de beste voorspeller. Dit is ook gevonden door Jansen (1996) in een onderzoek naar factoren die van invloed zijn op studievoortgang. Daaruit zou de voorzichtige conclusie getrokken kunnen worden dat het meer een kwestie is van een algemeen niveau dan van specifieke vakken. Zo vond Prins (1997) geen significante invloed van de samenstelling van het VWO-pakket op studie-uitval. Mogelijk dat een gemiddeld eindexamencijfer ook iets zegt over de probleemaanpak en vaardigheden van de student. Onderzoek heeft uitgewezen dat bijvoorbeeld planningsvaardigheden vaak van meer belang zijn dan algemene studievaardigheden (zie bijvoorbeeld Schouwenburg, 1994). Verder blijkt dat de inzet van studenten een bepalende factor is voor studiesucces. De Groot (1979) heeft leerwinst geformuleerd als inspanning maal begaafdheid, waarbij een lagere begaafdheid tot op zekere hoogte gecompenseerd kan worden door een hogere inspanning. Uiteraard is voor iedere studierichting een minimale vakkennis vanuit het VWO noodzakelijk. Het is dan ook een goede zaak dat in de profielnota's het heil niet alleen gezocht wordt in de vakinhouden. Een goede studiehouding, planningsvaardigheden en regelmatige studie-inzet zijn minstens zo belangrijk. Van het studiehuis-concept (zie bijvoorbeeld Wijnen, 1997) mogen we mogelijk meer invloed verwachten op studiesucces in het WO dan van de vakinhoudelijke invloed van de profielen. Zouden de voorgestelde maatregelen in de profielnota's op dit terrein positief uitwerken, dan is er voor de studenten al veel winst geboekt.

LITERATUUR

- Boer, P. den en H. Guldemond (1996). *De effecten van vakkenpakketkeuze en omwegen in het voortgezet onderwijs op het studiesucces in het hoger onderwijs*. GION, Rijksuniversiteit Groningen.
- Groot, A.D. de (1979). Studielast en normstudent: Ontwerp van een akkoordtheorie, I. Algemeen model. *Tijdschrift voor Onderwijsresearch*, 5, p. 9-28.
- Jansen, E.P.W.A. (1996). *Curriculumorganisatie en studievoortgang. Een onderzoek onder zes studierichtingen aan de Rijksuniversiteit Groningen*. GION, Studies over opvoeding en onderwijs, 1, Rijksuniversiteit Groningen.

- Jong, U. de, P. Koopman en J. Roeleveld (1991). *Snelwegen en slingerpaden in en om het Hoger Onderwijs*. Eindrapport Project 'Studieloopbanen in het Hoger Onderwijs'. Ministerie van Onderwijs en Wetenschappen, Zoetermeer.
- Ministerie van Onderwijs en Wetenschappen (1991). *Profiel van de tweede fase voortgezet onderwijs*, voorstellen voor een betere toerusting van scholen en leerlingen in de tweede fase voortgezet onderwijs. Ministerie van Onderwijs en Wetenschappen, Zoetermeer.
- Ministerie van Onderwijs en Wetenschappen (1992). *Profiel van de tweede fase voortgezet onderwijs*, vervolgnota. Ministerie van Onderwijs en Wetenschappen, Zoetermeer.
- Prins, J. (1997). *Studieuitval in het wetenschappelijk onderwijs. Studentkenmerken en opleidingskenmerken als verklaring voor studieuitval*. Nijmegen University Press.
- Schouwenburg, H.C. (1994). *Uitstelgedrag bij studenten*. Proefschrift Rijksuniversiteit Groningen.
- Smit, N.J. (1994). De invloed van VWO-wiskunde op studiesucces in de propedeuse. *Onderzoek van Onderwijs*, 23-1, p. 3-5.
- Stuurgroep Profiel Tweede Fase Voortgezet Onderwijs (1994). *Tweede fase, scharnier tussen basisvorming en hoger onderwijs*. Een uitwerking op hoofdlijnen van de nota's profiel van de tweede fase van het voortgezet onderwijs. Den Haag.
- Verhoeven, M.J.E. en A.A.J. van Zoelen (1993). *Het HBO geprofileerd*. Onderzoek naar verbetering van de aansluiting HAVO-HBO door vorming van instroomprofielen. LICOR, Rijksuniversiteit Leiden.
- Wijnen, W.H.F.W. (1997). Studiehuis: een buikbaar concept voor het onderwijs. *Tijdschrift voor Hoger Onderwijs*, 15, 1, p. 47-54.

1. E.P.W.A. Jansen, GION, afdeling COWOG, Rijksuniversiteit Groningen, Postbus 335, 9700 AH Groningen, e-mail: e.p.w.a.jansen@ppsw.rug.nl

2. In deze bijdrage spreken we nog van studierichtingen in plaats van opleidingen, omdat het onderzoek betrekking heeft op cohorten van vóór de invoering van de MUB.

Ontvangen 21-03-97

Definitief aanvaard 15-11-97

Delfts blauw en rose; studieloopbanen van mannen en vrouwen aan een technische universiteit

E. van Eck, P. Koopman en M. Robijns

SUMMARY

The main focus of this article are the differences between the educational careers of female and male students in a technical university. Educational careers are characterized in terms of educational choices and student performance. The educational choices are gender-related in this field; women are underrepresented in technical university and the percentages of female students vary strongly between the different technical disciplines. We found no significant differences in study results, however.

Furthermore we tried to identify the influence of personal characteristics, characteristics of the academic and the social environment and characteristics of study content and teaching on educational careers. Significant determinants of the educational careers of male students are examination results in secondary education in mathematics and science as well as in the other subjects, perceived doability of the study and judgment of the quality of teaching and the accessibility of the university lecturers. The educational careers of female students can hardly be predicted on the basis of the career model used. We discuss possible explanations of this phenomenon.

1. INLEIDING

Vrouwelijke studenten zijn van oudsher sterk ondervertegenwoordigd in het technisch onderwijs op alle niveaus; aan de TU Delft bijvoorbeeld bedroeg in 1976 het percentage vrouwelijke studenten slechts vier procent. Sindsdien is het aandeel vrouwelijke studenten bij die universiteit gestaag toegenomen tot 17% in 1993, een tendens die zich ook in de instroom van eerstejaars voortzet. Het aandeel vrouwelijke studenten varieert echter sterk per opleiding. In het algemeen blijken er weinig verschillen in studieresultaten tussen vrouwelijke en mannelijke studenten; de percentages die na twee jaar de propedeuse halen en na zes jaar het ingenieursexamen zijn ongeveer even hoog. Er zijn echter wel indicaties dat de studieloopbanen van vrouwelijke studenten anders verlopen, bijvoorbeeld dat zij andere studiewegen/afstudeer-richtingen kiezen, dat zij andere studieproblemen hebben en dat andere factoren van invloed zijn op studiesucces. In dit artikel gaan we na of de loopbanen van vrouwelijke en mannelijke studenten aan technische universiteiten seksespecifiek verlopen en welke factoren van invloed zijn op de studieloopbanen. Daarbij baseren we ons op gegevens uit de studentenadministratie en op materiaal dat met een survey is verzameld bij studenten van het cohort '88/89. Het onderzoek is gesitueerd binnen de TU Delft; deze universiteit biedt van de Nederlandse technische universiteiten verreweg de meeste technische wetenschappelijke opleidingen.

2. THEORETISCH KADER EN ONDERZOEKSMODEL

Studieloopbanen kunnen gekarakteriseerd worden aan de hand van een aantal beslissingen: over deelname, over voortzetten, onderbreken of voortijdig beëindigen, over studietempo en

inzet, en over de keuze van vakken en richtingen binnen de studie (De Jong, Koopman & Roeleveld, 1991).

Determinanten van het verloop van studieloopbanen worden meestal gedefinieerd op drie niveaus. We doelen hier op respectievelijk kenmerken van de student, de studie en van de sociale omgeving. Voor een overzicht van de factoren die hieronder geordend worden verwijzen wij naar Bijleveld (1993). In onderzoek naar studieloopbanen in het hoger onderwijs is het gebruikelijk de determinanten te ordenen met behulp van het model van Tinto (De Jong, 1992; Dronkers, 1976; Kempkens, 1987; Bijleveld, 1993). Centraal in dit model staat het socialisatieproces op basis waarvan de student kan functioneren binnen de studie-omgeving en in de sociale context. Daarbij onderscheidt Tinto achtergrondkenmerken van de student, doelen en verplichtingen, studie- en studentervaringen en de mate van integratie in de academische en sociale omgeving (Tinto, 1986; 1987; De Jong, Koopman & Roeleveld, 1991).

Ook in dit onderzoek speelt het model van Tinto een rol. Omdat dit model echter geen specifieke focus heeft op de loopbaan van vrouwen in het *technisch* wetenschappelijk onderwijs is een aanvullende literatuurstudie verricht. Voordat we het onderzoeksmodel schetsen dat in onze studie uiteindelijk benut is, gaan we in het onderstaande eerst in op de resultaten van de literatuurstudie. De literatuurstudie is primair gebaseerd op onderzoek naar seksespecifieke studieloopbanen in het hoger onderwijs en op onderzoek naar meisjes en exacte vakken in andere sectoren van het onderwijs. Over de specifieke loopbanen van vrouwen in het technisch wetenschappelijk onderwijs is uit onderzoek, behoudens enkele kleinschalige studies binnen specifieke opleidingen, weinig systematisch bekend.

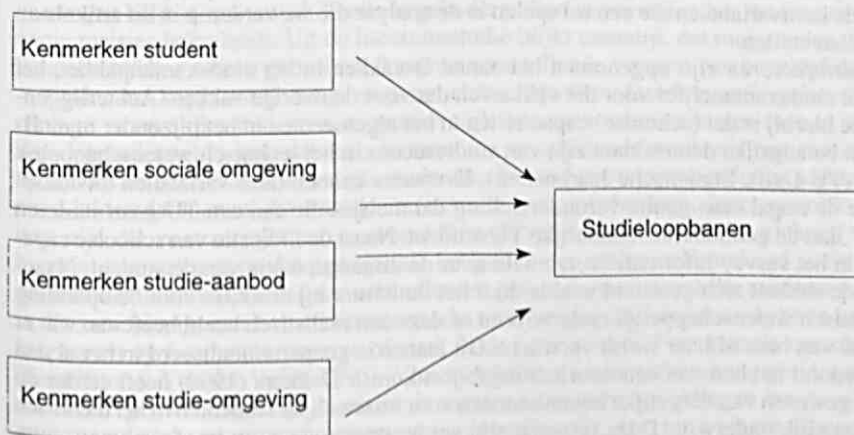
Uit de literatuur over seksespecifieke studieloopbanen en studierendement in het hoger onderwijs blijkt dat enkele factoren in het bijzonder voor vrouwelijke studenten belangrijk zijn. Op het individuele niveau van de *vrouwelijke student* wijst men op mogelijke invloed van verschillen in technische voorkennis en op de houding van vrouwen ten aanzien van techniek (Ten Dam, Van Eck & Volman, 1992). Daarnaast wordt erop gewezen dat vrouwen vaak niet exclusief gericht zouden zijn op studie en carrière maar ook op een taak in het gezin. Naast dit dubbele toekomstperspectief zouden vrouwen eveneens te maken hebben met een spanningsverhouding tussen de vrouwenrol en de studentrol binnen de universiteit (Verwayen-Leijh, 1985; Saharso & Westerbeek, 1983; Reerink, 1988). Verder is aandacht besteed aan het verschijnsel ongewenste intimiteiten, dit is echter niet empirisch in verband gebracht met studieloopbanen (Oude Avenhuis, 1990). Op het niveau van het *studie-aanbod* wordt verondersteld dat inhoud en didactiek van technisch en exact onderwijs slecht zijn afgestemd op kenmerken van vrouwen en meisjes die de opleiding binnenkomen. Voor vrouwelijke studenten zou met name contextrijkheid van de studie-inhoud (en daarmee zicht op de toepasbaarheid in de dagelijkse praktijk of in een toekomstig werktrein) van belang zijn. Vrouwen zouden de invulling van 'techniek' vaak te beperkt vinden. Ook zou de gehanteerde didactiek beter zijn afgestemd op de leerstijlen van jongens dan op die van meisjes (Ten Dam, Van Eck & Volman, 1992; Brouwer, Groenendaal, Bosch & Marinussen, 1990). In de *omgeving* zijn eveneens factoren geïdentificeerd die mogelijk van invloed zijn op studieloopbanen van vrouwen in het technisch onderwijs: het (nagenoeg) ontbreken van rolmodellen, de mannelijke opleidingscultuur, houding en gedrag van docenten, begeleiders en medestudenten, de uitzonderingspositie van vrouwen in de opleiding (Everts & Van Oost, 1985; Marinussen, Rinck & Udo, 1987; Alting & Udo, 1990; Turkenburg & Maenhout, 1990).

In onderzoek naar studieloopbanen in het hoger onderwijs staat - zoals hiervoor is besproken het model van Tinto centraal. Factoren die samenhangen met de loopbaan van meisjes in exacte vakken worden veelal niet met het model van Tinto maar met het model van Parsons/Eccles (Eccles, 1987) inzichtelijk gemaakt. Het model van Parsons/Eccles is primair een psychologisch georiënteerd model dat processen betreffende studiekeuzen en -succes beschouwt als een resultaat van afwegingen op individueel niveau. Keuzen en prestaties op individueel niveau

worden enerzijds gerelateerd met *de succesverwachting* ten aanzien van een bepaalde activiteit en anderzijds met *de subjectieve waarde* van deze activiteit (zie Van Eck & Volman, 1990).

In het onderzoek dat we presenteren in dit artikel vormt het model van Tinto het uitgangspunt. Verondersteld wordt dat factoren in de omgeving van de student de student beïnvloeden en zo ook de processen die het verloop van de studieloopbaan bepalen. Deze factoren in de omgeving van de student kunnen worden ondergebracht in de hiervoor genoemde indeling studie-aanbod, studie-omgeving en sociale omgeving. In het onderzoek hebben we gebruik gemaakt van het volgende model:

Figuur 1: Het onderzoeksmodel



De inzichten van Parsons/Eccles hebben een belangrijke rol gespeeld bij de invulling van de variabelenclusters. Voordat we hierop nader ingaan, schetsen we in het onderstaande eerst de opzet van het onderzoek.

3. DE OPZET VAN HET ONDERZOEK

In dit artikel staan de volgende onderzoeksvragen centraal:

In hoeverre verlopen de studieloopbanen van mannen en vrouwen aan de Technische Universiteit Delft verschillend en welke factoren zijn daarop van invloed?

Alvorens deze vragen te beantwoorden schetsen we eerst de onderdelen waaruit het onderzoek is opgebouwd.

Het onderzoek naar de studieloopbanen van mannen en vrouwen aan de Technische Universiteit Delft is opgebouwd uit drie componenten: een analyse van het studie-aanbod en de studie-omgeving binnen de diverse opleidingen; een reconstructie van studieloopbanen van de cohorten 1988/1989 en 1991/1992 en een survey onder vrouwelijke en mannelijke studenten van deze twee cohorten. In dit artikel beperken we ons tot de gegevens van het cohort 1988/89.

De analyse van het studie-aanbod en de studie-omgeving richtte zich op de veertien opleidingen die in 1988/89 aan de TUD gevolgd konden worden. Van elke opleiding is op basis van kenmerken van het studie-aanbod en de studie-omgeving een kwantificeerbare opleidingskarakteristiek opgesteld. Op basis van de literatuurstudie zijn als relevante factoren onder-

scheiden: onderwijsstructuur en ordening, didactiek, beoordelingssysteem, studiebelasting en supportsystemen, de studie-omgeving op basis van de topics beleid en klimaat. Ten einde de betrouwbaarheid van de opleidingskarakteristiek te optimaliseren is deze opgesteld op basis van twee typen informatiebronnen (Yin, 1984). Het betreft respectievelijk een analyse van schriftelijk materiaal (o.a. de onderwijsverslagen van een aantal achtereenvolgende studiejaar) en een interviewronde waarin een sleutelpersoon van elke opleiding bevraged is op basis van een specifiek voor dit doel ontwikkeld instrument. De reconstructie van de studieloopbanen is uitgevoerd op basis van gegevens uit de studentenadministratie over studenten van cohort '88/89 en daarvan diegenen die hebben deelgenomen aan het derde deelonderzoek, het survey.

Het derde onderzoeksdeel bestaat uit een schriftelijk survey onder de studenten van het cohort 1988/89. Hiermee heeft dataverzameling plaatsgevonden over de variabelen die de studieloopbaan kunnen determineren. Voor elk van de in het onderzoeksmodel gedefinieerde niveaus (student, sociale omgeving, studie-omgeving, en studie-aanbod) presenteren we hieronder de kernvariabelen die een rol spelen in de analyse die we verderop in dit artikel aan de orde zullen stellen.

Als *studentfactoren* zijn opgenomen het aantal B-vakken in het eindexamenpakket, het gemiddelde eindexamencijfer voor die vakken en dat voor de overige vakken. Achterliggende gedachte hierbij is dat (schoolse) capaciteiten in het algemeen en in het bijzonder in de B-vakken een belangrijke determinant zijn van studiesucces in het technisch wetenschappelijk onderwijs (TU Delft, Statistische Jaarboeken). Eveneens kunnen deze variabelen uitsluitel geven over de nogal eens geuite veronderstelling dat meisjes die aan een TU gaan studeren 'beter zijn' dan de gemiddelde mannelijke TU-student. Naast de indicatie van schoolse capaciteiten is in het survey informatie verzameld over de uitgangspositie van de student. Nagegaan is of de student zich gesteund voelde door het thuisfront bij de keuze voor de opleiding in het technisch wetenschappelijk onderwijs en of deze een realistisch beeld heeft van wat er in de studie van hem of haar wordt verwacht. Dit laatste is geoperationaliseerd in het al dan niet bijgewoond hebben van een voorlichtingsbijeenkomst. Dekkers (1990) heeft eerder op het belang gewezen van dergelijke bijeenkomsten voor vrouwelijke studenten in het technisch wetenschappelijk onderwijs. Deze factoren zijn verder opgenomen omdat steun en acceptatie belangrijke determinanten blijken van studiesucces bij vrouwelijke techniekstudenten. Meisjes in technische studies wijzen vaker dan jongens op het belang van steun uit hun (directe) omgeving bij de studiekeuze, zij ervaren vaker dat een dergelijke studie hen is afgeraden. Daarnaast is duidelijk dat de studiekeus van meisjes voor een technisch-wetenschappelijke studie een meer tentatief karakter heeft. Zij benutten de eerste fase van de studie om na te gaan of zij een reëel beeld hadden van hun studie en of hun keuze inderdaad de juiste is (voor een uitgebreider onderbouwing met bronvermelding zie Van Eck, Koopman en Robijns, 1995).

Onder *sociale omgeving* zijn twee factoren in de loopbaananalyses opgenomen: ervaren steun in de studie-omgeving en de mate waarin de woonsituatie van de student gunstig wordt geacht voor de studievoortgang. Het zal duidelijk zijn dat de eerder besproken factor 'ondervonden steun bij de studiekeuze' niet alleen relevant is in de context van het ouderlijk huis, maar ook in de leefsituatie van de student ten tijde van de studie. Deze kan bepalend zijn voor de mate waarin de student zich geïsoleerd of juist sociaal geïntegreerd voelt: onderhoudt deze contacten met studiegenoten (m/v), met ouderejaars (m/v) e.d.? Contacten met andere studenten kunnen verschillende functies vervullen: deze studenten kunnen dienen als rolmodel, als bron voor het vergelijken van eigen prestaties, ambities, e.d. en men kan er morele en daadwerkelijke steun bij de studie aan ontfen. De woonsituatie van de student kan daarbij bepalend zijn voor de steun die de student ervaart bij het 'volhouden' van de studie. Verondersteld wordt dat de deelname aan activiteiten van studentenverenigingen en vrijetijdsclubs, alsook het wonen in een studentenhuus of -flat de integratie van studenten bevordert en vanuit dit perspectief gunstig kunnen zijn voor studiesucces (Engelfriet & Van der Veldt, 1993 en Vermeulen, 1993 in: Van Eck, Koopman en Robijns, 1995). We veronderstellen dat deze factoren in de sociale omgeving met name voor meisjes - gezien hun minderheidspositie - van belang zijn voor studiesucces.

Als kenmerken van het *studie-aanbod* wijzen we in het kader van dit artikel met name op twee schalen die in het survey bevraagd zijn; 'studeerbaarheid' en 'sfeer/omgang docent-student binnen de opleiding'. De studeerbaarheid van de studieprogramma's is in kaart gebracht op basis van een schaal (α .88) uit onderzoek van De Jong, Webbink & Roeleveld (1996). De schaal valt uiteen in vier factoren die een goede studievoortgang kunnen belemmeren: de organisatie van afzonderlijke vakken, de organisatie van de studie algemeen, de zwaarte van de studie en 'onduidelijkheid over eisen' (Voor een uitgebreide bespreking zie Van Eck, Koopman en Robijns, 1995). De sfeer binnen de opleiding is in kaart gebracht aan de hand van een schaal die de volgende componenten omvat (α .81): 'effectieve feedback' en de afstand tussen docent en student. Deze componenten zijn respectievelijk bevraagd vanuit de veronderstelling dat vrouwelijke studenten in deze richtingen relatief onzekerder zouden zijn over hun studieprestaties, dus meer gebaat zouden zijn bij adequate en snelle feedback. De afstand tussen docent en student is bevraagd vanuit de veronderstelling dat de mate waarin docenten het vermogen hebben om verschillende begeleidersrollen te vervullen het studiesucces van met name meisjes beïnvloedt. Uit de literatuurstudie blijkt namelijk dat meisjes doorgaans meer behoefte hebben aan en gebruik maken van verschillende vormen van begeleiding.

Onder het kopje *studie-omgeving*, ten slotte, zijn twee factoren ondergebracht: het percentage vrouwelijke studenten, respectievelijk docenten waar studenten mee te maken hebben. Verondersteld wordt dat deze factoren via verschillende mechanismen van invloed zijn op het studiesucces van vrouwelijke studenten. Een oververtegenwoordiging van mannen kan ertoe leiden dat mannelijke waarden en normen gaan overheersen, dat er onvoldoende vrouwen zijn die voor meisjes een voorbeeldfunctie vervullen, dat meisjes zichzelf als buitenbeentje gaan ervaren en dat hun sociale integratie minder vanzelfsprekend verloopt. Uit de onderzoeksliteratuur komt naar voren dat het percentage vrouwelijke studenten een factor is die samenhangt met het studiesucces van vrouwelijke studenten: zij hebben ongunstiger loopbanen op faculteiten waar ze of sterk in de minderheid of sterk in de meerderheid zijn (Engelfriet & Van der Veldt, 1993). Vrouwelijke studenten verkeren in opleidingen waar ze ondervertegenwoordigd zijn, zoals in de meeste technische opleidingen het geval is, in een zogenoemde tokenpositie; sommigen veronderstellen dat dit ertoe leidt dat zij zichtbaarheid gaan vermijden (Van Oost, 1986), weinig vragen stellen, en vooral geen opvallende prestaties leveren (Wolffensperger, 1992). Anderen wijzen erop dat vrouwelijke studenten ten opzichte van de docenten een uitzonderingspositie ervaren en dat dit zowel positieve als negatieve kanten heeft (Vermeulen, 1993). Voorts wordt gevonden dat het voor een goede academische integratie van belang is dat vrouwelijke studenten, als ze in de minderheid zijn, zowel met vrouwelijke als met mannelijke studiegenoten omgaan (Vermeulen, 1993). Naast het percentage vrouwelijke medestudenten blijkt ook het aandeel vrouwelijke docenten een relevante factor te zijn; vrouwelijke docenten kunnen - mits aan een aantal voorwaarden is voldaan - naar de studentes toe verschillende functies vervullen; zij laten zien dat vrouw zijn verenigbaar is met techniek, dat werken als technicus en gezinsverantwoordelijkheden combineerbaar zijn en dat ook vrouwen goede docenten kunnen zijn. Als voorwaarden worden genoemd dat de docenten qua leeftijd en bereikt niveau niet te ver van de vrouwelijke studenten af staan en aanspreken als identificatiemodel (Koballa, 1988). De mate waarin studenten zich kunnen identificeren met de normen en waarden van de studie is bepalend voor hun studiesucces. Wanneer mannelijke normen en waarden overheersen, kan dit de identificatie van vrouwelijke studenten met de studie belemmeren en zo een negatief effect hebben op hun studiesucces (Hermanussen en Overdijk, 1991).

4. DE ONDERZOEKSPOPULATIE

De onderzoekspopulatie voor het onderzoek bestaat uit studenten die instroomden in het studiejaar '88/89.¹ Van dit cohort zijn alle vrouwen in het survey-onderzoek betrokken. Dat komt

neer op 497 (oud)studentes met een minimum van 1 en een maximum van 200 per opleiding. Daarnaast is een groep mannelijke studenten geselecteerd.² In totaal zijn 1106 studenten c.q. oud-studenten/stoppers uit dit cohort bevestigd. Studenten van wie bekend was dat ze met de studie waren gestopt, kregen een beperkter vragenlijst.

In totaal stuurden 499 studenten, stoppers en afgestudeerden de toegezonden vragenlijst ingevuld retour. Dit komt neer op een responspercentage van 45%. Van de aangeschreven vrouwen reageerde 49%, van de mannen 42%. De respons onder studerende en afgestudeerden was aanzienlijk hoger dan die onder de stoppers (58% versus 22%). Een lage respons onder de stoppers was niet onverwacht; omdat relatief veel uitval in de beginfase van de studie plaatsvindt, was het voor het merendeel op het moment van bevestiging al vier of vijf jaar geleden dat zij aan de TU studeerden. Omdat het responsbestand scheef verdeeld was ten gevolge van het feit dat dit een gestratificeerde steekproef betrof en dat er sprake was van verschillen in responspercentages tussen opleidingen en tussen mannen en vrouwen, zijn ten behoeve van de analyses weegfactoren geconstrueerd, zodat uitspraken over de populatie (in dit geval het cohort als geheel) gedaan kunnen worden.³ De reconstructie van de studieloopbanen is uitgevoerd op basis van gegevens uit de studentenadministratie betreffende de studenten van cohort '88/89 en daarvan diegenen die hebben deelgenomen aan het survey.

5. DE ONDERZOEKSRESULTATEN

Bij de presentatie van de onderzoeksresultaten beantwoorden we eerst de vraag of de loopbanen van studenten aan de technische universiteit samenhangen met hun sekse. Vervolgens bekijken we welke factoren die loopbanen beïnvloeden en gaan we na of de invloed van deze factoren seksespecifiek is.

Tabel 1: Opleidingen naar percentage vrouwelijke instromers, TU Delft, cohort '88/89

percentage vrouwelijke studenten	opleidingen	
≤ 10%	Maritieme Techniek	4%
	Technische Natuurkunde	4%
	Elektrotechniek	5%
	Luchtvaart- en Ruimtevaarttechniek	5%
	Werktuigbouwkunde	6%
	Mijnbouwkunde en Petroleumwinning	9%
	Technische Informatica	9%
11 t/m 20%	Materiaalkunde	13%
	Civiele Techniek	18%
21 t/m 30%	Geodesie	21%
	Scheikundige Technologie	22%
	Technische Wiskunde	24%
	Industrieel Ontwerpen	29%
> 30%	Bouwkunde	37%

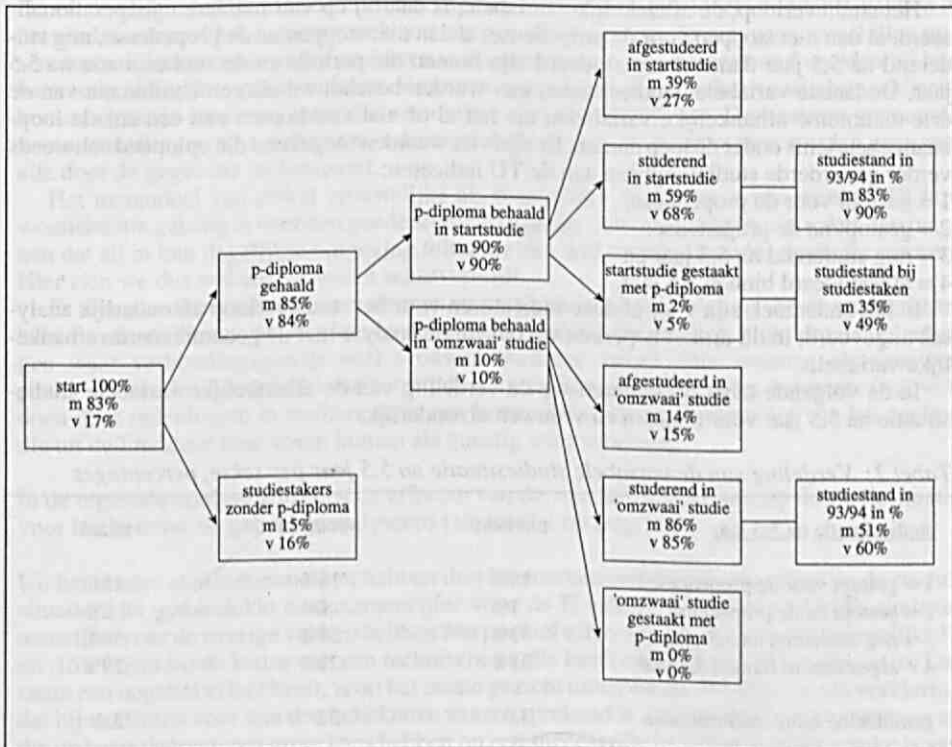
Studieloopbanen

In hoeverre zijn er verschillen tussen de studieloopbanen van mannelijke respectievelijk vrouwelijke studenten aan de TU Delft? Daarbij is gekeken naar studiekeuzen en naar het studieverloop.

Er zijn duidelijke verschillen in studiekeuzen tussen mannen en vrouwen. De percentages vrouwelijke studenten variëren sterk tussen de verschillende opleidingen. Opleidingen die in 1988/89 verhoudingsgewijs zeer weinig door vrouwen worden gekozen, zijn Technische Natuurkunde, Maritieme Techniek, Luchtvaart- en Ruimtevaarttechniek, Elektrotechniek, Werktuigbouwkunde, Mijnbouwkunde en Petroleumwinning en Technische Informatica. De meeste vrouwen zijn te vinden bij Bouwkunde en Industrieel Ontwerpen. Deze seksspecifieke keuzepatronen vertonen over de jaren heen een grote mate van consistentie (zie bijvoorbeeld Statistisch Jaarboek 1993/94, TU Delft, 1994).

De studieloopbanen zijn op basis van gegevens uit de studentenadministratie vervolgens gedetailleerder onderzocht. Als indicatoren voor het verloop van de studie is een drietal variabelen geconstrueerd: de studiestand (= het aantal behaalde studiepunten gedeeld door het aantal te behalen studiepunten x 100%), het gemiddeld aantal pogingen dat nodig was om te slagen voor de studie-onderdelen en het gemiddelde van de behaalde voldoende. Daarnaast is gebruikgemaakt van gegevens betreffende omzwaaien en voortijdige studiebeëindiging. Als studentvariabelen zijn, behalve sekse, opgenomen het aantal B-vakken in het vakkenpakket in het voortgezet onderwijs, het gemiddelde eindexamencijfer van de B-vakken, alsmede het gemiddelde eindexamencijfer voor de overige vakken.

Figuur 2: Het studieverloop van de onderzoeksgroep studenten '88/89 uitgesplitst naar sekse



Van de onderzochte groep studenten heeft 15% geen propedeutisch examen behaald aan deze TU. Van de 85% die dat wel hebben gedaan, is één op de tien voor de propedeuse omgezwaaid.⁴ Tot zover zijn er geen verschillen tussen de mannen en de vrouwen. Van de studenten die in de startstudie zijn doorgegaan, is een hoger percentage mannen dan vrouwen binnen 5,5 jaar afgestudeerd; onder de snelle studenten vinden we dus meer mannen. Onder diegenen die nog studeren zijn de vrouwen echter gemiddeld verder met hun studie dan de nog studerende mannen. Vrouwelijke studenten stoppen iets vaker dan mannen, en doen dat gemiddeld later in de studie.

Bij de omgezwaaiden zijn de percentages afgestudeerden binnen 5,5 jaar bij mannen en vrouwen even hoog (ongeveer 15%). Bij diegenen die nog studeren zien we ook bij de omzwaaiers dat de vrouwen gemiddeld verder met de studie zijn gevorderd dan de mannen.

Er zijn geen verschillen tussen mannen en vrouwen in het gemiddeld aantal pogingen dat zij nodig hebben om voor de studie-onderdelen te slagen en in het gemiddelde van de behaalde voldoende (zie voor uitgebreider bespreking Van Eck, Koopman en Robijns, 1995). We vinden dus op dit punt geen indicaties voor seksespecifieke verschillen in studieprestaties.

Determinanten van de studieloopbanen

Wat is de invloed van de hiervoor onderscheiden determinantencusters op het verloop van studieloopbanen? Bij de beantwoording van deze vraag dient het in Figuur 1 gepresenteerde onderzoeksmodel als basis. Dat impliceert dat kenmerken van de student en diens sociale omgeving, kenmerken van het studie-aanbod en kenmerken van de studie-omgeving van invloed zijn op het studieverloop. Voor de analyses is gebruikgemaakt van logistische en multipele regressie-analyses.

Het studieverloop, de afhankelijke variabele, is daarbij op vier manieren geoperationaliseerd: al dan niet stoppen voor de propedeuse, al dan niet stoppen na de propedeuse, nog studierend na 5,5 jaar dan wel afgestudeerd zijn binnen die periode en de studiesituatie na 5,5 jaar. De laatste variabele, studiesituatie, kan worden beschouwd als een combinatie van de drie dichotome afhankelijke variabelen die het al of niet voorkomen van een enkele loopbaangebeurtenis onder de loopnamen. Er zijn vier waarden toegekend die oplopend een steeds verder gevorderde studieloopbaan aan de TU indiceren:

- 1 = gestopt voor de propedeuse,
- 2 = gestopt na de propedeuse,
- 3 = nog studierend na 5,5 jaar en
- 4 = afgestudeerd binnen 5,5 jaar.

In het onderzoek zijn voor al deze indicatoren voor het studieverloop afzonderlijk analyses uitgevoerd; in dit artikel beperken we ons tot de analyse met de gecombineerde afhankelijke variabele.

In de volgende tabel presenteren we de verdeling van de afhankelijke variabele studiesituatie na 5,5 jaar voor mannen en vrouwen afzonderlijk.

Tabel 2: Verdeling van de variabele studiesituatie na 5,5 jaar per sekse, percentages

studiesituatie na 5,5 jaar	mannen	vrouwen	totaal
1 = gestopt voor de propedeuse	15%	16%	15%
2 = gestopt na de propedeuse	1%	4%	2%
3 = nog studierend na 5,5 jaar	53%	59%	54%
4 = afgestudeerd binnen 5,5 jaar	31%	22%	29%
gemiddelde score studiesituatie	3.0	2.9	3.0

Van deze studenten is ruim een kwart afgestudeerd binnen 5,5 jaar, ruim de helft studeert op dat moment nog, en een kleine 20% is gestopt.

Als onafhankelijke variabelen zijn - op basis van de bevindingen uit de literatuurstudie, de analyse van studie-aanbod en -omgeving en het survey onder studenten - uit elk van de categorieën enkele kernvariabelen geselecteerd (zie paragraaf 3).

Bij de *studentfactoren* zijn dat het aantal B-vakken in het eindexamenpakket in het voortgezet onderwijs, het gemiddelde eindexamencijfer voor die vakken en dat voor de overige vakken. Verder zijn als studentfactoren 'ondervonden steun bij de keuze' en 'voorlichting bijgewoond' opgenomen. Onder *sociale omgeving* zijn twee factoren in de loopbaananalyses opgenomen: ervaren steun in de studie-omgeving en de mate waarin de woonsituatie van de student gunstig wordt geacht voor de studievoortgang. Als kenmerken van het *studie-aanbod* zijn de factoren 'studeerbaarheid' en 'sfeer/omgang docent student binnen de opleiding' opgenomen. Daarnaast is de variabele 'gunstig voor vrouwen' in de analyse betrokken. Deze geeft aan in hoeverre de opleiding kenmerken vertoont waarvan op basis van de literatuur wordt verondersteld dat ze bevorderlijk zijn voor de loopbanen van vrouwen. Deze score is toegekend op basis van de analyses van de opleidingen (voor een uitgebreider onderbouwing met bronvermelding zie Van Eck, Koopman en Robijns, 1995). Onder het kopje *studie-omgeving*, ten slotte, zijn twee factoren ondergebracht; het percentage vrouwelijke docenten, respectievelijk studenten waar studenten mee te maken hebben.

Een overzicht van de geselecteerde variabelen staat in Tabel 3 (pag. 260). Daarin is verder te zien in hoeverre er sprake is van verschillen tussen mannen en vrouwen op de onafhankelijke variabelen.

De vrouwelijke studenten in de onderzoeksgroep hebben gemiddeld iets minder B-vakken in het eindexamenpakket en scoren ook wat lager op die B-vakken. Er zijn geen verschillen tussen mannelijke en vrouwelijke studenten in de mate waarin zij zich gesteund voelden bij hun beslissing in Delft te gaan studeren. Vrouwen woonden wel wat vaker voorlichtingsbijeenkomsten van de studie bij maar die verschillen zijn niet significant. De hiervoor genoemde veronderstellingen ten aanzien van sekseverschillen op de studentfactoren worden geen van alle door de gegevens ondersteund.

Het merendeel van zowel vrouwelijke als mannelijke studenten is van mening dat hun woonsituatie gunstig is voor een goede studievoortgang; vrouwen zeggen wel vaker dan mannen dat zij in hun dagelijkse omgeving feitelijkte dan wel morele steun bij de studie ervaren. Hier zien we dus wel een verwacht sekseverschil.

Er zijn geen verschillen tussen mannen en vrouwen in de beoordeling van de studeerbaarheid van de studie en van de sfeer en de omgang student-docent. Wel blijkt dat de opleidingen waar verhoudingsgewijs veel vrouwen studeren zowel door mannen als vrouwen positiever worden beoordeeld wat betreft de omgang tussen studenten en docenten. Ook voltooien deze opleidingen in sterkere mate aan kenmerken van de vormgeving van het aanbod die uit de literatuur naar voren komen als gunstig voor vrouwen.

In de regressie-analyse zijn eerst de effecten van de vier factorenclusters op de studiesituatie voor het bestand als geheel geanalyseerd (zie tabel 4 op pag. 261).

Uit het cluster studentkenmerken hebben drie kenmerken een significant effect op de studiesituatie. Het gemiddelde eindexamencijfer voor de B-vakken en het gemiddelde eindexamencijfer voor de overige vakken hebben een positief effect (de bèta's zijn respectievelijk .37 en .15). Steun bij de keuze van een technische studie heeft een negatief effect ($\beta = -.10$). Dat steun een negatief effect heeft, is op het eerste gezicht onverwacht. Wellicht is een verklaring dat bij studenten voor wie de studiekeuze vanzelfsprekend is qua capaciteiten en interesse en die op basis daarvan een grote kans hebben op een succesvolle loopbaan minder sprake is van

Tabel 3: De onafhankelijke variabelen; sekseverschillen¹

studentfactoren	Mannen		Vrouwen		sign. sekse verschillen
	gem.	(s.d)	gem.	(s.d)	
aantal B-vakken	3.8	(.74)	3.6	(.82)	*
gem. eindex. cijfer B-vakken	7.4	(.85)	6.9	(.87)	**
gem. eindex. cijfer overige vakken	6.9	(.70)	7.0	(.74)	n.s.
steun bij keuze	3.6	(.59)	3.7	(.56)	n.s.
voorlichting bijgewoond	.65	(.48)	.75	(.44)	n.s.
sociale omgeving					
ervaren steun in studie-omgeving	2.7	(1.2)	3.2	(1.1)	**
woonsituatie gunstig voor de studie	.66	(.48)	.69	(.47)	n.s.
studie-aanbod²					
studeerbaarheid ind.	2.2	(.87)	2.3	(.84)	n.s.
studeerbaarheid (gem. opleiding)	2.2	(.29)	2.2	(.21)	n.s.
sfeer/omgang student docent ind.	2.6	(.80)	2.6	(.85)	n.s.
sfeer/omgang student docent (gem. opleiding)	2.5	(.32)	2.7	(.32)	**
'gunstig voor vrouwen'	1.8	(.87)	2.5	(.77)	**
studie-omgeving					
perc. vrouwelijke docenten	9%	(6.1)	15%	(6.6)	**
perc. vrouwelijke studenten	14%	(10.2)	24%	(10.4)	**

* $p \leq .05$; ** $p \leq .01$; n.s. = niet significant

1. Nadere informatie over de samenstelling en wijze van scoring van de variabelen;

Steun bij keuze betreft een schaal van 6 items waarvoor geldt 1 = erg belemmerd - 5 = erg gesteund;

Bij Voorlichting bijgewoond is gescoord de fractie die deze vraag met ja beantwoordde;

Steun in de studie-omgeving heeft betrekking op de gemiddelde score op een schaal van 3 items 0 = geen - 5 = veel;

Bij Woonsituatie gunstig voor de studie betreft de score de fractie die deze vraag met ja beantwoordde

Studeerbaarheid betreft de gemiddelde score op een schaal van 21 items ($\alpha = .88$) waarbij 0 = positief - 5 = negatief oordeel;

Sfeer/omgang student docent betreft de gemiddelde score op een schaal met 17 items ($\alpha = .81$) waarbij 0 = negatief - 5 = positief oordeel;

'*Gunstig voor vrouwen*' 1 = minst gunstig - 3 = meest gunstig.

2. De variabelen studeerbaarheid en sfeer/omgang docent-student zijn op twee manieren in de analyses opgenomen, op geaggregeerd niveau als studierichtingkenmerk en op individueel niveau als studentbeoordeling. Ten slotte is in dit cluster een studierichtingscore 'gunstig voor vrouwen' opgenomen. Deze geeft aan in hoeverre de studierichting kenmerken vertonen waarvan op basis van de literatuur wordt verondersteld dat ze bevorderlijk zijn voor de loopbanen van vrouwen. Deze score is toegekend op basis van de analyses van de studierichtingen (voor een uitgebreider onderbouwing met bronvermelding zie Van Eck, Koopman en Robijns, 1995).

expliciete steun dan bij studenten voor wie de keuze minder vanzelfsprekend is. Daarnaast zijn de studenten bij drie opleidingen relatief ver gevorderd met de studie, terwijl bij één opleiding de gemiddelde studiesituatiescore relatief laag is. Van de factoren uit het cluster studie-aanbod die in de analyses zijn opgenomen, hebben er drie een significant effect. Studenten in opleidingen waar de sfeer en de omgang met docenten positiever worden beoordeeld, zijn verder gevorderd met de studie, dat wil zeggen scoren hoger op studiesituatie ($\beta = .22$). Dit speelt ook op individueel niveau, naarmate studenten tevredener zijn over hun opleiding op deze punten zijn zij verder gevorderd ($\beta = .34$). Het individuele oordeel over studeerbaarheid heeft eveneens een significant effect. Op het eerste gezicht lijkt dit effect 'verkeerd om'; stu-

Tabel 4: De resultaten van de regressie-analyses 'studiesituatie na 5,5 jaar', voor het totale bestand, en voor mannen en vrouwen afzonderlijk

Onafhankelijke variabelen	bèta-coëfficiënten		
	m + v	m	v
<u>studentfactoren</u>			
aantal B-vakken	-	-	-
gem. index. cijfer B-vakken	.37	.40	-
gem. index. cijfer overige vakken	.15	.13	-
steun bij keuze	-.10	-.17	-
voorlichting bijgewoond	-	-	-
sekse	-	x	x
opleiding ¹			
ET ²	.19	.15	-
TI	.20	-	-
Wb	.11	-	-
TW	-.09	-	-
IO	-	-.14	-
LR	-	-.18	-
<u>sociale omgeving</u>			
ervaren steun in omgeving	-	-	-
woonsituatie gunstig	-	-	-
<u>studie-aanbod</u>			
studeerbaarheid ind.	.20	.23	-
studeerbaarheid gem. opl.	-	.15	-
sfeer ind.	.34	.36	.30
sfeer gem. opl.	.22	-	-
'gunstig voor vrouwen'	-	-	-
<u>studie-omgeving</u>			
% vr. studenten	-	-	-
% vr. docenten	-	-	-
Verklaarde variantie	35%	39%	10%

1. De variabele opleiding is opgenomen in de vorm van 13 dummy-variabelen, de veertiende dummy geldt als referentie-studierichting. Hiervoor is steeds Civiele Techniek genomen omdat deze opleiding op de verschillende afhankelijke variabelen het meest gemiddeld scoort.

2. De namen van de opleidingen worden conform het gebruik op de TU Delft als volgt afgekort: Bouwkunde (Bk), Civiele Techniek (CT), Elektrotechniek (Et), Geodesie (Ge), Industrieel Ontwerpen (IO), Luchtvaart- en Ruimtevaarttechniek (LR), Materiaalkunde (Mk), Mijnbouwkunde en Petroleumwinning (MP), Maritieme Techniek (MT), Scheikundige Technologie (ST), Technische Informatica (TI), Technische Natuurkunde (TN), Technische Wiskunde (TW), Werktuigbouwkunde (Wb).

denten die een negatief oordeel hebben over de studeerbaarheid, scoren hoger op studiesituatie dan studenten die de studeerbaarheid positiever beoordelen ($\beta = .20$). Bij de analyse van de 'loopbaangebeurtenis' 'nog studierend' of 'al afgestudeerd' na 5,5 jaar bleek dat het effect van studeerbaarheid wel in de voor de hand liggende richting gaat: hoe studeerbaarder de studie, des te sneller men studeert. Een nadere analyse heeft geleerd dat het omgekeerde effect

bij deze analyse wordt veroorzaakt door de studenten die voor de propedeuse gestopt zijn. Deze studenten hebben een relatief gunstige beoordeling van de studeerbaarheid. Dit wijst erop dat studenten sneller kunnen beslissen of ze de goede studie hebben gekozen als de studie goed 'studeerbaar' is. Dat een student direct in het propedeutisch jaar duidelijkheid krijgt over de juistheid van zijn of haar studiekeuze, kan worden beschouwd als een kwaliteitscriterium. Het opleidingskenmerk 'gunstig voor vrouwen' heeft geen significant effect op de studiesituatie, evenmin als de ingevoerde kenmerken van de sociale omgeving en de studieomgeving. Ook sekse heeft geen significant effect.

Dezelfde analyse voor mannen afzonderlijk vertoont vergelijkbare resultaten. Ook hier significante positieve effecten van de beide categorieën eindexamencijfers (B-vakken; $\beta = .40$; overige vakken $\beta = .13$) en een negatief effect van steun bij de keuze ($\beta = -.17$). Drie opleidingen vertonen afwijkende resultaten, bij twee zijn de studenten minder ver gevorderd dan gemiddeld, bij één zijn ze verder dan gemiddeld. Bij studie-aanbod zijn drie factoren significant: het gemiddelde oordeel over de studeerbaarheid van alle studenten van een opleiding en het individuele oordeel van de student hebben een effect op de studiesituatie, in die zin dat een negatief oordeel over de studeerbaarheid samengaat met een gunstige studiesituatie (β 's zijn respectievelijk .15 en .23). Verder geldt ook voor mannelijke studenten dat diegenen die de sfeer positiever beoordelen ook verder zijn gekomen met de studie ($\beta = .36$). Kenmerken van de sociale omgeving en van de studie-omgeving hebben geen significant effect op de studiesituatie van de mannen.

De analyse voor vrouwelijke studenten levert maar één significant effect op: naarmate studentes een positiever oordeel hebben over de sfeer en de omgang met docenten hebben ze een gunstiger studiesituatie ($\beta = .30$).

6. CONCLUSIES EN DISCUSSIE

De studieloopbanen aan de TU Delft vertonen kenmerken die deels seksespecifiek, deels sekseneutraal zijn. Seksespecifiek zijn de studiekeuzen; vrouwelijke studenten in Delft hebben duidelijke voorkeuren voor bepaalde opleidingen. Het verloop van de studieloopbanen van de vrouwelijke respectievelijk mannelijke studenten vertoont heel weinig verschillen. De percentages stoppers en omzwaaiers zijn ongeveer even groot. Bij de mannen zijn er wat meer studenten die snel afstuderen, maar gemiddeld is er geen verschil in studietempo tussen vrouwelijke en mannelijke studenten.

Zowel kenmerken van de studenten als van de opleidingen blijken het verloop van de studieloopbanen te beïnvloeden. Belangrijke studentkenmerken in dit verband zijn het gemiddeld eindexamencijfer voor de exacte vakken, maar ook dat voor de overige vakken. Studenten met hoge eindexamencijfers doen het in het algemeen beter. Het aantal B-vakken in het pakket heeft geen effect op het verloop van studieloopbanen. Sekse blijkt in de analyses op de groep mannen en vrouwen tezamen geen effect te hebben op het loopbaanverloop. Van de kenmerken van het studie-aanbod blijkt de sfeer binnen de opleiding en de manier waarop docenten omgaan met studenten de belangrijkste determinant; daarnaast speelt de studeerbaarheid van de studie een belangrijke rol. Snel afgestudeerde studenten beoordelen de studeerbaarheid van de studie positiever dan studenten die na 5,5 jaar nog met de studie bezig zijn, terwijl studenten die hun propedeutisch examen hebben behaald een negatiever oordeel hebben over de studeerbaarheid dan studenten die voor de propedeuse zijn gestopt. De didactische kwaliteit en de mate waarin de opleiding kenmerken vertoont die gunstig worden geacht voor vrouwen als ook de twee omgevingskenmerken, percentage vrouwelijke docenten respectievelijk vrouwelijke studenten in de opleiding, blijken geen effect te hebben op het loopbaanverloop.

Bij de analyses op het bestand van mannelijke en vrouwelijke studenten tezamen blijkt

seks geen factor die het loopbaanverloop verklaart. Wanneer we de analyses voor mannelijke en vrouwelijke studenten apart uitvoeren, komen verschillende modellen voor mannen en vrouwen naar voren. De loopbanen van mannen blijken veel beter te verklaren te zijn op basis van de variabelen uit ons onderzoeksmodel dan die van de vrouwen. Voor de loopbanen van mannen zijn de gemiddelde eindexamencijfers voor exacte vakken en voor de overige vakken, sfeer en studeerbaarheid belangrijke determinanten van het loopbaanverloop. Het studieverloop van de vrouwelijke studenten blijkt met behulp van het door ons gehanteerde loopbaanmodel niet te verklaren; gemiddelde eindexamencijfers voor exacte vakken, gemiddelde eindexamencijfers voor de overige vakken, aantal B-vakken, steun bij de keuze van een technische studie en het bijwonen van de voorlichtingsbijeenkomsten van de gekozen opleiding en de mate waarin men de woonsituatie gunstig acht voor de studievoortgang blijken geen effect te hebben op de loopbanen van de vrouwelijke studenten van '88/89 aan de TU Delft. Ook kenmerken van het studie-aanbod, zoals studeerbaarheid, didactische kwaliteit, en de mate waarin de studie kenmerken vertoont die gunstig worden geacht voor vrouwen hebben geen aantoonbaar effect op de loopbanen van vrouwelijke studenten. Het percentage vrouwelijke studenten in de opleiding, een kenmerk dat in een landelijk onderzoek bleek samen te hangen met studiesucces (Engelfriet en Van der Veldt, 1993) blijkt hier evenmin van invloed op het loopbaanverloop, noch van mannen, noch van vrouwen. De enige factor die uit de analyse naar voren komt, is de beoordeling van de sfeer in de opleiding en van de wijze waarop docenten met de studenten omgaan. De loopbanen van vrouwen laten zich slecht voorspellen uit de door ons geselecteerde kenmerken van student en opleiding.

Hoe kunnen we deze bevinding interpreteren? In de eerste plaats is een technische verklaring mogelijk; verschillen in de spreiding op de onafhankelijke en afhankelijke variabelen zouden kunnen leiden tot een verschil in verklarend vermogen van het model voor de loopbanen van vrouwelijke respectievelijk mannelijke studenten. In het begin van deze paragraaf hebben we laten zien dat dit niet het geval was.

Het kan ook betekenen dat we een factor die essentieel is voor het verklaren van seksverschillen in de loopbanen niet in het model hebben opgenomen. Een factor die in dit verband relevant zou kunnen zijn is de waardecomponent uit het model van Parsons/Eccles. Hoe belangrijk is deze studiekeuze voor de student. Verondersteld kan worden dat meisjes die door naar de TU te gaan 'afwijken van de gebaande paden', zeer veel waarde hechten aan een succesvolle afronding van deze studie, meer dan de gemiddelde mannelijke student, voor wie techniek studeren een traditionele, meer voor de hand liggende keuze is. De waarde die de student hecht aan de studie bepaalt wellicht sterker dan de variabelen in ons onderzoeksmodel het studieverloop van vrouwelijke studenten in deze 'niet-traditionele vrouwelijke' opleidingen.

Dat de eindexamencijfers bij meisjes niet als significante determinant van studiesucces naar voren komen, is opmerkelijk. De statistische jaarboeken van de TU Delft laten zien dat voor alle opleidingen voor de opleidingspopulaties als geheel de kans op studiesucces sterk toeneemt met het stijgen van het gemiddeld eindexamencijfer voor exacte vakken. Overdijk (1995) komt echter in haar onderzoek naar studieloopbanen aan de Technische Universiteit Eindhoven ook tot de conclusie dat die relatie voor meisjes minder eenduidig is; hoewel de vrouwen in haar onderzoek gemiddeld lager scoren op eindexamencijfer exacte vakken, is hun studiesucces vergelijkbaar met dat van de jongens die gemiddeld betere eindexamenresultaten hebben. Zij geeft hiervoor twee mogelijke verklaringen; eindexamencijfers geven geen betrouwbaar beeld van de feitelijke capaciteiten van vrouwen en mannen voor exacte vakken (meisjes worden ondergewaardeerd, jongens overgewaardeerd) of vrouwen compenseren hun lagere cijfers met sociaal gedrag dat hun sociale integratie bevordert en op die manier leidt tot studiesucces. Dat de cijfers van meisjes en vrouwen voor exacte vakken geen goed beeld geven van hun capaciteiten wordt zowel in de tweede fase van het voortgezet onderwijs (zie bijvoorbeeld Deboer, 1986; Van Eck, 1996) als in het hoger onderwijs (zie bijvoorbeeld Wolffensperger, 1992) gevonden. Ook de veronderstelling dat de sociale integratie van

vrouwelijke studenten in de studie en de sociale omgeving soepel verloopt en dat zij daaraan steun ontfangen bij de studievoortgang vindt ondersteuning in ons onderzoek.

Om te kunnen nagaan in hoeverre de hiervoor genoemde veronderstellingen inderdaad een rol spelen zou in verder onderzoek naar seksespecifieke loopbanen in technische en exacte opleidingen uitgebreider aandacht besteed moeten worden aan de waarde die studenten toevoegen aan de studie en het succesvol afronden ervan, en zouden andere indicatoren van de capaciteiten op exact gebied gebruikt moeten worden dan eindexamencijfers.

LITERATUUR

- Alting, A., & Udo, S. (1990). De houding van MTS-docenten tegenover vrouwelijke leerlingen. *Tijdschrift voor Onderwijswetenschappen*, 20, 185-196.
- Brouwer, F., Groenendaal, W., Bosch, A., & Marinussen, L. (1990). *Docenten over vrouwen in het HTO. Een onderzoek naar houding en gedrag van HTO-docenten ten opzichte van studentes in het Hoger Technisch onderwijs*. Utrecht: VHTO.
- Bijleveld, R.J. (1993). *Numeriek rendement en studiestaking; een theoretische analyse van factoren die samenhangen met rendement en studiestaking in het wetenschappelijk onderwijs*. Enschede: CSHOB.
- Dam, G. ten, Eck, E. van & Volman, M. (1992). *Onderwijs en sekse; een reconstructie van researchprogramma's*. Trendrapport nummer 4. Den Haag: DCE/STEO.
- Deboer, G.E. (1986). Perceived science ability as a factor in the course selections of men and women in college. *Journal of Research in Science Teaching*, 23 (4).
- Dekkers, H. (1990). *Seksespecifieke studiekeuzen in het wetenschappelijk onderwijs*. Nijmegen: ITS.
- Dronkers, J. (1976). *Studenten en hun onderwijs. Een onderzoek naar de wisselwerking tussen universiteit en studenten*. Groningen: Tjeenk Willink.
- Eccles, J.S. (1987). Gender Roles and Women's Achievement-related Decisions. *Psychology of Women Quarterly*, 11, 135-172.
- Eck, E. van, & Volman, M. (red.). (1990). *Determinanten van seksespecifieke keuzen en prestaties. Verslag van symposium gehouden op de Onderwijsresearchdagen 1989*. Amsterdam: SPCP.
- Eck, E. van, Koopman, P. & Robijns, M. (1995). *Delfts blauw en rose. Studieloopbanen van mannelijke en vrouwelijke studenten aan de TU Delft*. Amsterdam: SCO-Kohnstamm Instituut.
- Eck, E. van (1996). *Apert en samen: een onderzoek naar implementatie en effecten van het project Coornhert Exact*. Amsterdam: SCO-Kohnstamm Instituut/ILO.
- Engelfriet, L., & Veldt, M.C. van der (1993). *Studentenemancipatiebeleid aan de Nederlandse universiteiten. Inventarisatie en beleidsaanbevelingen. Achtergrondstudies Hoger onderwijs en Wetenschappelijk onderzoek 13*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- Everts, S. & Oost, E. van (1985). *Vrouwelijke studenten aan de TH Twente*. Enschede: Technische Hogeschool Twente.
- Hermanussen, R. en Overdijk, K. (1991). *Technical Education: study choice and coping with the environment, a similarity hypothesis*. Eindhoven: TUE.
- Jong, U. de, Koopman, P., & Roeleveld, J. (1991). *Eindrapport Project 'Studieloopbanen in het Hoger Onderwijs'*. Amsterdam: SCO Kohnstamm-Instituut.
- Jong, U. de (1992). *De loopbaan doorlopen; keuze en selectie tijdens de loopbaan van basisonderwijs tot Open Universiteit*. Amsterdam: SCO-Kohnstamm Instituut.
- Jong, U. de, Webbink, H.D. & Roeleveld, J. (1996). *De subjectieve factor*. Den Haag, OC&W.
- Kempkens, L. (1987). *Deelname aan tentamens bij de Open Universiteit: een onderzoek onder langzame studenten en uitvallers*. Heerlen: Open Universiteit.
- Koballa, Th.R. (1988). Attitude Change and Gender-Equity in Science: Communicator Qualities that enhance Persuasion. In: *A Summary of Research in Science Education*. Ohio.
- Marinussen, L., Rinck, J., & Udo, S. (1987). *Vrouwen over het HTO. Verslag van een onderzoek naar studierendement en studie-ervaring van vrouwelijke studenten en studiestaaksters*. Wageningen: VHTO.
- Oost, E. van (1986). Hoe gaat het met Thea op de technische hogeschool? In: J. 't Hart, M. Hijman, E. Kas, A. Kemps, E. van Oost, M. den Ouden, N. Oudshoorn, & M. van den Wijngaard (red.), *Een barst in het bolwerk. Vrouwen, natuurwetenschappen en techniek* (p.61-74). Amsterdam: SUA.
- Oude Avenhuis, A.M. (1990). *Alledaags seksisme op de universiteit. Een onderzoek naar ongewenste intimiteiten*. Groningen: RION.

- Overdijk, K. (1995). *Onderweg op de technische universiteit*. Een onderzoek naar studielooptbanen van vrouwelijke en mannelijke studenten (Proefschrift TU Eindhoven). Eindhoven: z.u.
- Reerink, R. (1988). *Verschuivingen in deelname, rendement en beleving van studie. Onderzoek naar deelname en rendement van mannelijke en vrouwelijke studenten. Vergelijkend onderzoek naar achtergronden, kenmerken en studiegedrag van studenten in de studierichtingen Geneeskunde en Sociologie*. Oud-Annerveen: z.u.
- Saharso, S., & Westerbeek, J. (1983). De precare balans, een reconstructie van de ervaringen van vrouwelijke studenten met studeren, achterraken en uitvallen. *Tijdschrift voor Vrouwenstudies*, 4, (1), p. 96-110.
- Tinto, V. (1986). Theories of student departure revisited. *Higher Education: Handbook of theory and research*, Volume II. New York: Agathon Press.
- Tinto, V. (1987). *Leaving college. Rethinking the causes and cures of student attrition*. Chicago: The University of Chicago.
- TU Delft (1994). *Statistisch Jaarboek 1993/94. Studenten aan de TU Delft: Momentopnames, Studievoortgang, Raming aantallen*. Delft: TU Delft/Bureau van de Universiteit/Dienst Onderwijs en Onderzoek.
- Turkenburg, M., & Maenhout, J. (1990). *De positie van meisjes en jonge vrouwen in het informatica-onderwijs*. 's-Gravenhage: Ministerie van Sociale Zaken en Werkgelegenheid.
- Vermeulen, E. (1993). *Notitie ervaringen vrouwelijke studenten TU Delft*. Verslag van zes interviews met vrouwelijke studenten van de Technische Universiteit Delft (intern rapport). Delft: TUD Bureau Studentenzaken.
- Verwayen-Leijh, R.M. (1985). *Studie-uitval en vertraging van vrouwen in het hoger onderwijs: een studie naar de omvang van uitval en vertraging van vrouwen ten opzichte van mannen in het hoger onderwijs en naar in de literatuur genoemde belemmeringen die vrouwen in het hoger onderwijs ondervinden*. 's-Gravenhage: Studiecentrum OTO.
- Wolffensperger, (1992). *Gender Issues in Master Programs*. Contributions to the GASAT Conference 1992. Volume 1: Schooling.
- Wijnen, W.H.F.W. (1992). *Te doen of niet te doen*. Rapport van de Commissie Studeerbaarheid. Den Haag: Ministerie van O&W.
- Yin, R.K. (1984). *Case study research: Design and Methods*. Beverly Hills: Sage.

TREFWOORDEN T.B.V. SWIDOC

studielooptbanen - sekseverschillen - technisch wetenschappelijk onderwijs

1. Voor cohort '88/89 geldt dat ten tijde van de dataverzameling de nominale studieduur was verstreken.
2. Via een matchingsprocedure zijn daarnaast uit de studentenadministratie mannelijke instromers geselecteerd die met de vrouwelijke instromers overeenkomen wat betreft opleiding, vooropleiding, aantal exacte vakken in het pakket, gemiddeld eindexamencijfer voor de B-vakken en het gemiddeld eindexamencijfer op de overige vakken. Omdat bij matching de kans bestaat dat het aantal complete matchparen door non-respons te gering is, is bovendien een steekproef mannelijke studenten uit het bestand toegevoegd, dit ten behoeve van statistische controle.
De respons bleek inderdaad onvoldoende complete matchparen te bevatten om vergelijkingen tussen de veertien opleidingen mogelijk te maken. Daarom zijn de verdere analyses niet gedaan op het matchbestand maar op het bestand 'alle vrouwen en een steekproef mannen per opleiding'.
3. De responsbestanden die de surveys hebben opgeleverd zijn scheef verdeeld. Dit komt in de eerste plaats doordat sprake is van gestratificeerde steekproeftrekking. In de onderzoeksopzet is ervoor gekozen alle vrouwen te benaderen, terwijl van de mannelijke studenten een steekproef is getrokken van maximaal 30 per opleiding. In de tweede plaats is het responsbestand scheef verdeeld omdat er sprake is van selectieve non-respons. In de weegfactoren corrigeren wij voor opleiding en sekse, zodanig dat de verdeling in de responsgroep voor deze variabelen overeenkomt met die in de populatie.
Hoewel de respons ook scheef verdeeld bleek voor stoppers en doorgaande studenten kon daarvoor niet gecorrigeerd worden. Het geleverde administratieve bestand liet het niet toe een onderscheid te maken tussen vroege stoppers en bijvakkers.
4. Dit propedeuserendement is hoger dan dat in de Statistische Jaarboeken van de TU Delft omdat het is berekend op basis van een responsbestand waarin vroege stoppers ondervertegenwoordigd zijn.

Ontvangen 03-01-97

Definitief aanvaard 15-11-97

Centreren in multilevelanalyse: implicaties van twee centreringmethoden voor het bestuderen van schooleffectiviteit¹

Marie-Christine Opdenakker en Jan Van Damme²

Centrum voor Secundair en Hoger Onderwijs, Katholieke Universiteit Leuven

ABSTRACT

In educational effectiveness research, multilevel models are increasingly used because these models take the multilevel structure of the data into account. For computational ease in deriving parameter estimates, centering of the predictor variables is often suggested. With nested data, two different forms of centering are possible: centering around the grand mean and centering around the group mean. The international literature on multilevel modelling has paid little attention to the implications of both ways of centering for the parameter estimates and the research results. In this article, we study the effect of these two forms of centering for random intercept, and for random intercept and random slopes models by means of model equivalence formulas and the analysis of a concrete data set. The conclusion is that centering around the group mean amounts almost always to fitting a different model than that obtained with centering around the grand mean. Making a choice between the two forms of centering is therefore very important and can mostly only be made on the basis of the research questions and/or the theory under investigation.

INLEIDING

Bij het bestuderen van het effect van scholen op de leerprestaties en het welbevinden van leerlingen op school wordt gebruik gemaakt van multilevelmodellen (zie o.m. Bryk & Raudenbush, 1992; Goldstein, 1995). Deze modellen houden, in tegenstelling tot de klassieke regressiemodellen, rekening met het feit dat leerlingen genest zijn in scholen en laten toe dat parameters (intercept en hellingen) variëren over de groepen waarin de observaties genest zijn. Om de berekeningen voor het verkrijgen van de parameterschattingen te vergemakkelijken, wordt in de literatuur aangeraden om predictoren te centreren rond het gemiddelde (Longford, 1990). Bij gegroepeerde of geneste data (bv. leerlingen in scholen) kan op twee wijzen rond een gemiddelde gecentreerd worden: men kan centreren rond het algemeen gemiddelde of rond het gemiddelde van de groep waartoe de betreffende observatie behoort (het zogenaamde groepsgemiddelde). Tot op heden werd in de internationale literatuur weinig aandacht besteed aan de implicaties van beide wijzen van centreren voor de parameterschattingen en - hiermee samenhangend - voor de onderzoeksresultaten.

Uit een studie van Kreft, de Leeuw en Aiken (1995) blijkt dat centreren rond het groepsgemiddelde leidt tot het fitten van een ander model dan hetgeen verkregen wordt bij centreren rond het algemeen gemiddelde of bij het gebruik van ruwe scores. Het centreren van de scores van predictorvariabelen op leerlingniveau, dat gewoonlijk o.w.v. technische redenen aanbevolen wordt, effecteert in multilevelanalyse (in tegenstelling tot de traditionele regressieanalyse) meer dan alleen de intercept. Hierdoor heeft het centreren in multilevelanalyses

conceptuele implicaties die - aldus Kreft (1995) - tot nog toe onvoldoende bestudeerd en begrepen zijn. Het vaststellen van schijnbaar tegenstrijdige onderzoeksresultaten in schooleffectiviteitsonderzoek, (ook) bij het gebruik van multilevelanalyse, moet mede in deze context begrepen worden: in diverse studies wordt op verschillende wijzen gecentreerd en wordt soms geen of onvoldoende rekening gehouden met de implicaties van de gehanteerde centeringsmethode o.m. voor de interpretatie van de gecentreerde predictorvariabelen en voor de parameterschattingen op schoolniveau. Zo bespreekt men in sommige studies de modellen als ruwe-scoresmodellen terwijl men de predictorvariabele op leerlingniveau gecentreerd heeft rond het groepsgemiddelde en deze variabele aldus de afwijking ten aanzien van de groep waartoe de observaties behoren, weergeeft (Kreft, 1995).

Aan de hand van een concrete dataset ontleend aan het LOSO-project worden verschillen tussen scholen inzake leerprestaties van leerlingen in kaart gebracht. Hierbij worden op schoolniveau enkele schoolkenmerken (o.a. kenmerken m.b.t. het functioneren van de school als organisatie, onderwijskenmerken en een schoolpopulatiekenmerk) en op leerlingniveau een leerlingachtergrondkenmerk (intelligentie) ter verklaring van de leerprestaties gebruikt.

In een eerste deel van dit artikel wordt ingegaan op modelequivalentie bij hiërarchisch lineaire modellen. Aan de hand van formules en voorbeelden worden de verschillende effecten van de twee centeringsmethoden (centreren rond het algemeen gemiddelde en centreren rond het groepsgemiddelde) geïllustreerd. De resultaten van elf twee-niveau-modellen worden met elkaar vergeleken, met name het 'null'-model, dat informatie oplevert over de opsplitsing van de totale variantie in leerprestaties naar het leerling- en schoolniveau, vier random intercept-modellen en zes modellen met een random intercept en een random helling. Met betrekking tot de random interceptmodellen wordt het volgende onderscheid gemaakt: een model met de predictorvariabele op leerlingniveau gecentreerd rond het algemeen gemiddelde; hetzelfde model waarbij het schoolgemiddelde van de intelligentiescores eveneens als verklarende variabele opgenomen wordt; een model met de predictorvariabele op leerlingniveau gecentreerd rond het groepsgemiddelde en hetzelfde model waarbij het schoolgemiddelde van de intelligentiescores eveneens als verklarende variabele opgenomen wordt. Voor de modellen met een random intercept en random helling worden analoge modellen besproken, aangevuld met twee modellen (voor elke centeringsmethode één) waarbij het schoolgemiddelde van de intelligentiescores ook als variabele ter verklaring van de random helling gehanteerd wordt.

In een tweede deel worden de gevolgen van het kiezen voor een welbepaalde centeringsmethode geïllustreerd door de onderzoeksresultaten van de twee complexe 'fully random'-modellen inhoudelijk te bespreken.

1. MODELEQUIVALENTIE BIJ HIERARCHISCH LINEAIRE MODELLEN

Elk multilevelmodel bepaalt een formule voor de verwachte waarde van de onafhankelijke variabele Y_{ij} en de variantie/covariantie matrix van deze variabele. Twee verschillende modellen (met normale distributies) zijn equivalent als ze dezelfde set verwachte waarden en dispersies genereren. Dit wil echter niet zeggen dat de formules om de modellen te beschrijven dezelfde zijn; er kunnen immers andere parametrisaties gebruikt zijn. Voor modelequivalentie is wel vereist dat er een één op één transformatie tussen de beide parametrisaties is.

1.1. Random interceptmodellen

Bij de bespreking van random interceptmodellen onderscheiden we vier modellen m.n. modellen met en zonder opname van het groepsgemiddelde. We centreren rond het algemeen gemiddelde (we spreken in het vervolg over het CAG-model) en rond het groeps- of schoolgemiddelde (CSG-model).

In een eerste rubriek onderzoeken we de equivalentie van modellen zonder groepsgemiddelde. We besteden aandacht aan de algebraïsche relaties tussen de modellen en illustreren deze met voorbeelden uit eigen onderzoek.

1.1.1. Modellen zonder groepsgemiddelde

1.1.1.1 Algebraïsche relaties

Formules:

Model met centrering rond het algemeen gemiddelde	Model met centrering rond het groepsgemiddelde
$Y_{ij} = a_j + b_j(X_{ij} - X_{..}) + e_{ij}$	$Y_{ij} = a_j^* + b_j^*(X_{ij} - X_j) + e_{ij}^*$
$a_j = \gamma_{00} + u_j$	$a_j^* = \gamma_{00}^* + u_j^*$
$b_j = \gamma_{10}$	$b_j^* = \gamma_{10}^*$
met variantie van $e_{ij} = \sigma^2$	met variantie van $e_{ij}^* = \sigma^{2*}$
met variantie van $u_j = \varpi_{00}$	met variantie van $u_j^* = \varpi_{00}^*$

De *vaste gedeelten* van de modellen (vaste parameters) zijn equivalent als:

$$\begin{aligned} \gamma_{00} + \gamma_{10}(X_{ij} - X_{..}) &= \gamma_{00}^* + \gamma_{10}^*(X_{ij} - X_j) \\ \gamma_{00} + \gamma_{10}X_{ij} - \gamma_{10}X_{..} &= \gamma_{00}^* + \gamma_{10}^*X_{ij} - \gamma_{10}^*X_j \\ (\gamma_{00} - \gamma_{10}X_{..}) + \gamma_{10}X_{ij} &= \gamma_{00}^* + \gamma_{10}^*X_{ij} - \gamma_{10}^*X_j \end{aligned}$$

Uit het voorgaande volgt dat in het algemeen de vaste gedeelten van beide modellen niet equivalent zijn. Er is immers slechts een oplossing indien $\gamma_{10}^* = 0$ en $\gamma_{10} = 0$ omdat in het algemeen de constante vector, de vector met de elementen X_{ij} en de vector met de elementen X_j lineair onafhankelijk zijn.

De *random gedeelten* van de modellen zijn equivalent als de totale resterende variantie van het model met centrering rond het algemeen gemiddelde ($\sigma^2 + \varpi_{00}$) gelijk is aan de totale resterende variantie van het andere model ($\sigma^{2*} + \varpi_{00}^*$).

1.1.1.2. Voorbeelden

Onze dataset (een subgroep uit de LOSO-dataset) bestaat uit 4889 leerlingen (eerste leerjaar A) verdeeld over 52 secundaire scholen. We onderzoeken de relatie tussen het behaalde percentage punten op een schoolvorderingstoets wiskunde op het einde van het eerste leerjaar A (PCWSE1) en de score op een numerieke intellectuele vaardigheidsproef (NI) gemeten bij het begin van het schooljaar.

Voor de analyses werd gebruikgemaakt van het computerprogramma MLN (Woodhouse, 1995).

Tabel 1: 'Null'-model

		Schatting	SE
<i>Vaste parameters</i>			
Intercept		42.87	1.121
<i>Random parameters</i>			
Schoolniveau	ω_{00}	61.63	12.8
Leerlingniveau	σ^2	179.2	3.644
<i>Deviantie</i>		39408	

Uit de analyse van het zogenaamde 'null'-model (model zonder verklarende variabelen) (tabel 1) blijkt dat 26% van de variantie in het behaalde percentage wiskunde op het einde van het eerste leerjaar A kan toegeschreven worden aan verschillen tussen scholen. Dit percentage is vrij hoog in vergelijking met buitenlandse studies, waarin percentages tussen 8 en 20% gerapporteerd worden en met de metastudie van Bosker en Witziers (1996) waarin men, op basis van 103 studies, tot een percentage van 18% kwam. Het is evenwel in overeenstemming met een recente Vlaamse studie (Verhoeven et al., 1992) waarin 28,7 % van de variantie in wiskundescores op het einde van het eerste leerjaar A kon toegeschreven worden aan verschillen tussen scholen.

Wordt de numerieke intelligentie van de leerlingen als verklarende variabele (op leerlingniveau) opgenomen en respectievelijk gecentreerd rond het algemeen gemiddelde versus het schoolgemiddelde (tabel 2), dan zien we dat we twee verschillende modellen krijgen. Er zijn verschillen in de schattingen van de intercept, de helling en de variantie van de intercept. De randomgedeelten zijn equivalent.

In deze context willen we erop wijzen dat een verschillende vraagstelling aan de basis ligt van beide modellen. In het model met centrering rond het algemeen gemiddelde staat de vraag centraal of de relatieve positie op een verklarende variabele van invloed is op de positie op een afhankelijke of responsvariabele. In het model met centrering rond het groepsgemiddelde staat de vraag centraal of de relatieve positie in een groep (in ons voorbeeld de school) van invloed is op de afhankelijke of responsvariabele.

Uit het *CAG-model* concluderen we dat een leerling met een gemiddelde score op de numerieke intelligentieproef (d.i. 104.03) een verwachte waarde van 44.67% heeft op de wiskundetoets op het einde van het eerste leerjaar. Verder blijkt dat de numerieke intelligentie een belangrijke voorspeller is voor prestatie op de wiskundetoets: de coëfficiënt is zeer significant $t_{4889} = 42$ en $R_1^2 = 0.3238$ wat betekent dat ongeveer één derde van de variantie inzake wiskunde Prestaties op leerlingniveau kan verklaard worden door de numerieke intelligentie van de leerlingen. Niet alleen de variantie op leerlingniveau is afgenomen. Ook de interceptvariantie is lager omdat de scholen van elkaar verschillen m.b.t. de gemiddelde numerieke intelligentie van hun leerlingen, zodat deze leerlingvariabele ook een gedeelte van de verschillen tussen scholen verklaart ($R_2^2 = 0.4772$). Dit wijst er tevens op dat intellectueel sterke leerlingen meestal in scholen zitten met een intellectueel sterk publiek in het eerste leerjaar A en omgekeerd intellectueel zwakke leerlingen meestal in scholen zitten met een intellectueel zwakke bevolking.

Het *CSG-model* geeft aan dat leerlingen die qua numerieke intelligentie even sterk zijn als de gemiddelde leerling in hun school een verwachte waarde van 42.81 percent hebben op de wis-

kundetoets; anders geformuleerd voor de populatie van scholen verwachten we een gemiddeld schoolgemiddelde van 42.81 percent op de wiskundetoets. De relatieve positie inzake numerieke intelligentie van leerlingen binnen hun school is een belangrijke voorspeller voor hun presteren op wiskunde: de coëfficiënt is zeer significant $t_{4889} = 42$ en $R_1^2 = 0.3238$, één derde van de variantie inzake wiskunde-prestaties kan dus verklaard worden door de relatieve positie inzake numerieke intelligentie die de leerlingen in hun school innemen. In vergelijking met het vorige model zien we dat de variantie op het niveau van de scholen nagenoeg dezelfde is gebleven als bij het 'null'-model. Dit is logisch, vermits we door de waarden van de individuele leerlingvariabele te centreren rond hun groepsgemiddelde deze variabele 'uitzuiveren' voor het gemiddeld niveau van de groep en we aldus ook geen groepsvariantie kunnen verklaren met deze gecentreerde variabele.

Tabel 2: Random interceptmodellen zonder groepsgemiddelde

	CAG-model		CSG-model		
	Schatting	SE	Schatting	SE	
<i>Vaste parameters</i>					
Intercept	44.67	0.8136	42.81	1.127	
'NI'	0.5703	0.01338	0.567	0.01348	
<i>Random parameters</i>					
Schoolniveau	ϖ_{00}	31.64	6.727	63.26	12.94
Leerlingniveau	σ^2	131.2	2.668	131.2	2.667
Deviantie		37866.8		37899.2	

1.1.2. Modellen met groepsgemiddelde

1.1.2.1. Algebraïsche relaties

Formules:

Model met centrering rond het algemeen gemiddelde	Model met centrering rond het groepsgemiddelde
$Y_{ij} = a_j + b_j(X_{ij} - X_j) + e_{ij}$	$Y_{ij} = a_j^* + b_j^*(X_{ij} - X_j) + e_{ij}^*$
$a_j = \gamma_{00} + \gamma_{01}X_j + u_j$	$a_j^* = \gamma_{00}^* + \gamma_{01}^*X_j + u_j^*$
$b_j = \gamma_{10}$	$b_j^* = \gamma_{10}^*$
met variantie van $e_{ij} = \sigma^2$	met variantie van $e_{ij}^* = \sigma^{2*}$
met variantie van $u_j = \varpi_{00}$	met variantie van $u_j^* = \varpi_{00}^*$

De *vaste gedeelten* van de modellen (vaste parameters) zijn equivalent als:

$$\gamma_{00} + \gamma_{10}(X_{ij} - X_{..}) + \gamma_{01}X_{.j} = \gamma_{00}^* + \gamma_{10}^*(X_{ij} - X_{.j}) + \gamma_{01}^*X_{.j}$$

Na uitwerking verkrijgen we:

$$(\gamma_{00} - \gamma_{10}X_{..}) + \gamma_{10}X_{ij} + \gamma_{01}X_{.j} = \gamma_{00}^* + \gamma_{10}^*X_{ij} + (-\gamma_{10}^* + \gamma_{01}^*)X_{.j}$$

Deze vergelijking gaat op indien $\gamma_{00} = \gamma_{00}^* + \gamma_{10}X_{..}$; $\gamma_{10} = \gamma_{10}^*$ en $\gamma_{01} = -\gamma_{10}^* + \gamma_{01}^*$.

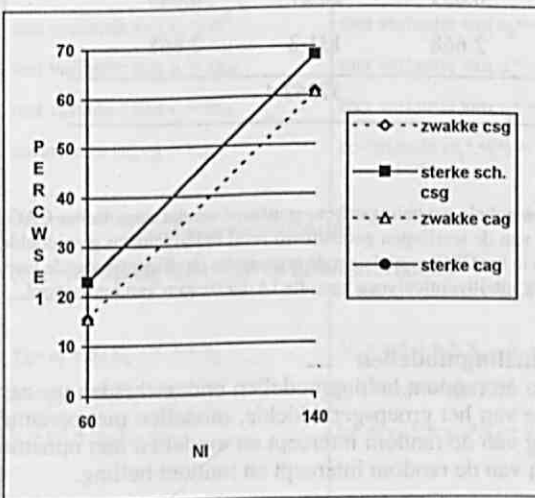
Uit het voorgaande volgt dat de vaste gedeelten van beide modellen equivalent zijn, want we kunnen voor een gegeven γ_{00} en γ_{01} altijd een γ_{00}^* en γ_{01}^* vinden voor alle 'i' en 'j' zodat hoger vermelde vergelijking opgaat.

De *random gedeelten* van de modellen zijn equivalent als de totale resterende variantie van het model met centrering rond het algemeen gemiddelde ($\sigma^2 + \omega_{00}$) gelijk is aan de totale resterende variantie van het andere model ($\sigma^{2*} + \omega_{00}^*$).

1.1.2.2. Voorbeelden

Ter illustratie van het voorgaande werken we opnieuw met dezelfde dataset. Uit tabel 3 blijkt de equivalentie van de modellen. Zowel het vaste als het random gedeelte van beide modellen zijn equivalent. De variabele 'gemiddelde numerieke intelligentie van de schoolpopulatie in het eerste leerjaar A' (SCHNI) heeft een verschillende coëfficiënt in beide modellen omdat in het model met centrering rond het algemeen gemiddelde deze coëfficiënt betrekking heeft op het effect van het intellectueel schoolpeil wanneer rekening gehouden wordt met het niveau dat de leerling zelf op de numerieke intelligentieproef behaalt (dus tussen- 'min' binnengroepseffect), terwijl in het andere model de coëfficiënt betrekking heeft op het tussen-groepseffect. De variabele 'SCHNI' heeft in beide modellen dus een andere betekenis³. Een grafische voorstelling wordt in figuur 1 weergegeven.

Figuur 1: Regressie van PERCWSEI op NI (voor SCHNI = hoog en laag) op basis van CAG- en CSG-model



Uit het CAG-model kan geconcludeerd worden dat naast de numeriek intellectuele vaardigheden van de leerling zelf ook het gemiddeld numeriek intellectueel peil van zijn/haar school bijdraagt tot het verklaren van zijn/haar wiskundeprestatie ($t_{52} = 2.3$). Op schoolniveau wordt 53% van de variantie in intercepten verklaard. Dit betekent dat 5% van de variantie in inter-

cepten kan verklaard worden door het gemiddeld intellectueel niveau van de leerlingen van de school, wanneer reeds rekening gehouden werd met het intellectueel niveau van de leerling zelf. Als leerling is het dus belangrijk terecht te komen in een school met een gemiddeld intellectueel sterk publiek. (Een mogelijke verklaring kan het stimulerend karakter van de omgeving zijn: beter onderwijs en/of hogere stimulering door school- en klasgenoten in scholen met een sterk intellectueel publiek.)

Uit het CSG-model blijkt dat het gemiddeld numeriek intelligentiepeil van de leerlingen van de school bijdraagt tot het verklaren van de wiskunde-prestatie van leerlingen ook wanneer rekening gehouden wordt met de relatieve positie die de leerling op intellectueel vlak inneemt in zijn/haar school. Op basis van deze schoolvariabele kan 53% van de interceptvariantie verklaard worden. Dit betekent dat het gemiddeld intellectueel niveau van de leerlingen van een school een belangrijke bijdrage levert ter verklaring van de wiskunde-prestaties.

Tabel 3: Random interceptmodellen met groepsgemiddelde

		CAG-model		CSG-model	
		Schatting	SE	Schatting	SE
<i>Vaste parameters</i>					
Intercept		20.97	11	-37.96	10.9
'NI' ¹		0.567	0.01348	0.567	0.01348
SCHNI		0.2345	0.1085	0.8009	0.1076
<i>Random parameters</i>					
Schoolniveau	ω_{00}	28.24	6.051	28.27	6.089
Leerlingniveau	σ^2	131.2	2.668	131.2	2.668
<i>Deviantie</i>		37862.4		37862.4	

1. De variabele 'NI' wordt in de tabellen tussen enkele aanhalingstekens geplaatst omdat deze in het CAG-model verwijst naar de numerieke intelligentie van de leerlingen gecentreerd rond het algemeen gemiddelde van alle IA-leerlingen in de onderzoeksgroep en in het CSG-model naar de numerieke intelligentie van de leerlingen gecentreerd rond het gemiddeld numeriek intelligentieniveau van alle IA-leerlingen van hun school.

1.2. Random intercept- en random hellingmodellen

Bij de bespreking van random intercept- en random hellingmodellen onderscheiden we zes modellen m.n. modellen zonder opname van het groepsgemiddelde, modellen met opname van het groepsgemiddelde ter verklaring van de random intercept en modellen met opname van het groepsgemiddelde ter verklaring van de random intercept en random helling.

1.2.1. Algebraïsche relaties

Formules:

Model met centrering rond het algemeen gemiddelde	Model met centrering rond het groepsgemiddelde
1. zonder groepsgemiddelde	
$Y_{ij} = a_j + b_j(X_{ij} - X_{.j}) + e_{ij}$ $a_j = \gamma_{00} + u_j$ $b_j = \gamma_{10} + v_j$ <p>met variantie van $e_{ij} = \sigma^2$ met variantie van $u_j = \omega_{00}$ met variantie van $v_j = \omega_{11}$ covariantie $(u_j, v_j) = \omega_{01}$</p>	$Y_{ij} = a_j^* + b_j^*(X_{ij} - X_{.j}) + e_{ij}^*$ $a_j^* = \gamma_{00}^* + u_j^*$ $b_j^* = \gamma_{10}^* + v_j^*$ <p>met variantie van $e_{ij}^* = \sigma^{2*}$ met variantie van $u_j^* = \omega_{00}^*$ met variantie van $v_j^* = \omega_{11}^*$ covariantie $(u_j^*, v_j^*) = \omega_{01}^*$</p>
2. met groepsgemiddelde (voor verklaring interceptvariantie)	
$Y_{ij} = a_j + b_j(X_{ij} - X_{.j}) + e_{ij}$ $a_j = \gamma_{00} + \gamma_{01}X_{.j} + u_j$ $b_j = \gamma_{10} + v_j$ <p>met variantie van $e_{ij} = \sigma^2$ met variantie van $u_j = \omega_{00}$ met variantie van $v_j = \omega_{11}$ covariantie $(u_j, v_j) = \omega_{01}$</p>	$Y_{ij} = a_j^* + b_j^*(X_{ij} - X_{.j}) + e_{ij}^*$ $a_j^* = \gamma_{00}^* + \gamma_{01}^*X_{.j} + u_j^*$ $b_j^* = \gamma_{10}^* + v_j^*$ <p>met variantie van $e_{ij}^* = \sigma^{2*}$ met variantie van $u_j^* = \omega_{00}^*$ met variantie van $v_j^* = \omega_{11}^*$ covariantie $(u_j^*, v_j^*) = \omega_{01}^*$</p>
3. met groepsgemiddelde (voor verklaring intercept- en hellingsvariantie)	
$Y_{ij} = a_j + b_j(X_{ij} - X_{.j}) + e_{ij}$ $a_j = \gamma_{00} + \gamma_{01}X_{.j} + u_j$ $b_j = \gamma_{10} + \gamma_{11}X_{.j} + v_j$ <p>met variantie van $e_{ij} = \sigma^2$ met variantie van $u_j = \omega_{00}$ met variantie van $v_j = \omega_{11}$ covariantie $(u_j, v_j) = \omega_{01}$</p>	$Y_{ij} = a_j^* + b_j^*(X_{ij} - X_{.j}) + e_{ij}^*$ $a_j^* = \gamma_{00}^* + \gamma_{01}^*X_{.j} + u_j^*$ $b_j^* = \gamma_{10}^* + \gamma_{11}^*X_{.j} + v_j^*$ <p>met variantie van $e_{ij}^* = \sigma^{2*}$ met variantie van $u_j^* = \omega_{00}^*$ met variantie van $v_j^* = \omega_{11}^*$ covariantie $(u_j^*, v_j^*) = \omega_{01}^*$</p>

De *vaste gedeelten* van de modellen *zonder groeps-gemiddelden* zijn equivalent indien:

$$\gamma_{00} + \gamma_{10}(X_{ij} - X_{..}) = \gamma_{00}^* + \gamma_{10}^*(X_{ij} - X_{.j}).$$

Omdat in het algemeen de constante vector, de vector met de elementen X_{ij} en de vector met de elementen $X_{.j}$ lineair onafhankelijk zijn, betekent dit dat de vergelijking slechts opgaat indien $\gamma_{10}^* = 0$ en $\gamma_{10} = 0$. Aangezien de vergelijking in het algemeen niet kan opgelost worden, zijn de vaste gedeelten van beide modellen niet equivalent.

De *random gedeelten* van de modellen *zonder groeps-gemiddelden* zijn equivalent indien:

$$\sigma^2 + \omega_{00} + \omega_{01}X_{ij} + \omega_{11}X_{ij}^2 = \sigma^{2*} + \omega_{00}^* + \omega_{01}^*(X_{ij} - X_{.j}) + \omega_{11}^*(X_{ij} - X_{.j})^2.$$

Wanneer dit uitgewerkt wordt, blijkt dat het rechtergedeelte van de vergelijking de termen $X_{.j}$, $X_{.j}X_{ij}$ en $X_{.j}^2$ bevat. De vergelijking kan slechts een oplossing hebben indien $\omega_{11}^* = 0$, $\omega_{11} = 0$, $\omega_{01}^* = 0$, $\omega_{00} = \omega_{00}^*$ en $\sigma^2 = \sigma^{2*}$. Dit betekent dat we geen random hellingen hebben.

De *vaste gedeelten* van de modellen *met groeps-gemiddelden ter verklaring van de intercept-variantie* zijn equivalent indien:

$$\gamma_{00} + \gamma_{01}X_{.j} + \gamma_{10}(X_{ij} - X_{..}) = \gamma_{00}^* + \gamma_{01}^*X_{.j} + \gamma_{10}^*(X_{ij} - X_{.j}).$$

Uitgewerkt betekent dit dat $\gamma_{00} = \gamma_{10}X_{..} + \gamma_{00}^*$, $\gamma_{01} = \gamma_{01}^* - \gamma_{10}^*$, $\gamma_{10} = \gamma_{10}^*$. De vaste gedeelten zijn dus equivalent.

De *random gedeelten* van de modellen *met groeps-gemiddelden ter verklaring van de intercept-variantie* zijn equivalent indien:

$$\sigma^2 + \omega_{00} + \omega_{01}X_{ij} + \omega_{11}X_{ij}^2 = \sigma^{2*} + \omega_{00}^* + \omega_{01}^*(X_{ij} - X_{.j}) + \omega_{11}^*(X_{ij} - X_{.j})^2.$$

Zoals bij de vorige twee modellen, zijn ook hier de random gedeelten niet equivalent.

De *vaste gedeelten* van de modellen *met groeps-gemiddelden ter verklaring van de intercept- en hellingvariantie* zijn equivalent indien:

$$\gamma_{00} + \gamma_{01}X_{.j} + (\gamma_{10} + \gamma_{11}X_{.j})(X_{ij} - X_{..}) = \gamma_{00}^* + \gamma_{01}^*X_{.j} + (\gamma_{10}^* + \gamma_{11}^*X_{.j})(X_{ij} - X_{.j}).$$

Dit betekent na uitwerking dat $-\gamma_{11}^* = 0$, $\gamma_{11}^* = \gamma_{11}$, $\gamma_{10} = \gamma_{10}^*$, $\gamma_{01} = \gamma_{01}^* - \gamma_{10}^*$, $\gamma_{00} = \gamma_{00}^* + \gamma_{10}X_{..}$.

Wanneer $\gamma_{11} = 0$ en $\gamma_{11}^* = 0$, betekent dit dat het groeps-gemiddelde geen verklaring kan bieden voor de variantie in de hellingen, m.a.w. we vallen terug naar de vorige twee modellen. Indien het groeps-gemiddelde wel relevant is (dus $X_{.j}X_{ij}$ een significante coëfficiënt heeft) dan zijn beide modellen niet equivalent.

De *random gedeelten* van de modellen *met groeps-gemiddelden ter verklaring van de intercept- en hellingvariantie* zijn equivalent indien:

$$\sigma^2 + \omega_{00} + \omega_{01}X_{ij} + \omega_{11}X_{ij}^2 = \sigma^{2*} + \omega_{00}^* + \omega_{01}^*(X_{ij} - X_{.j}) + \omega_{11}^*(X_{ij} - X_{.j})^2.$$

Zoals bij de vorige vier modellen, zijn ook hier de random gedeelten niet equivalent.

1.2.2. Voorbeelden

Ter illustratie van het voorgaande werken we opnieuw met dezelfde dataset. Uit tabel 4 blijkt dat het zinvol is om random hellingen te veronderstellen: in het geval van centrering rond het algemeen gemiddelde daalt de deviantie van het model met 40 eenheden (in vergelijking met de gegevens van tabel 2), wat met een verschil van 2 vrijheidsgraden een zeer significante

verbetering van het model is. Uit tabellen 4 tot en met 6 blijkt dat de modellen met centrering rond het algemeen gemiddelde en rond het groepsgemiddelde in geen van de drie 'condities' equivalent zijn (de devianties van de modellen zijn nooit gelijk). Dit komt omdat het random gedeelte van beide modellen nooit equivalent is. Het vaste gedeelte van beide modellen is gelijk indien het groepsgemiddelde (alleen) opgenomen wordt ter verklaring van de variatie in de intercept (tabel 5).

Tabel 4: Random intercept- en random hellingmodellen zonder groepsgemiddelde

	CAG-model		CSG-model	
	Schatting	SE	Schatting	SE
<i>Vaste parameters</i>				
Intercept	43.89	0.9231	42.73	1.136
'NI'	0.5206	0.02561	0.5065	0.02701
<i>Random parameters</i>				
Schoolniveau				
ω_{00}	40.51	8.627	64.38	13.15
ω_{01}	0.6195	0.1923	1.038	0.2647
		cor=.67		cor=.80
ω_{11}	0.02075	0.006327	0.02569	0.007249
Leerlingniveau				
σ^2	128.9	2.633	128.6	2.627
<i>Deviantie</i>	37826.8		37838.9	

De variabele 'SCHNI' heeft een verschillende coëfficiënt in beide modellen (tabel 5) omdat in het model met centrering rond het algemeen gemiddelde deze coëfficiënt betrekking heeft op het effect van het intellectueel schoolpeil wanneer rekening gehouden wordt met het niveau dat de leerling zelf op de numerieke intelligentieproef behaalt (dus tussen- 'min' binnengroepseffect), terwijl in het andere model de coëfficiënt betrekking heeft op het tussen-groepseffect.

Uit het *CAG-model* kan geconcludeerd worden dat het gemiddeld numeriek intellectueel peil van de school niet bijdraagt tot het verklaren van de wiskunde-prestatie van een leerling wanneer reeds rekening gehouden werd met de numerieke intelligentievaardigheden van de leerling en naast de random intercept ook een random helling m.b.t. de numerieke intelligentie verondersteld wordt. Deze bevinding staat in contrast met de resultaten van het random intercept model: hier vonden we dat het gemiddeld numeriek peil van de school wel bijdroeg tot het verklaren van de wiskunde-prestatie. (Uit het vervolg van deze bijdrage zal blijken dat er wat anders aan de hand is.)

Uit het *CSG-model* blijkt dat het gemiddeld numeriek intelligentiepeil van de leerlingen van de school bijdraagt tot het verklaren van de wiskunde-prestatie van leerlingen ook wanneer rekening gehouden wordt met de relatieve positie die de leerling op intellectueel vlak inneemt in zijn/haar school.

Tabel 5: Random intercept- en random hellingmodellen met groepsgemiddelde ter verklaring van de interceptvariantie

	CAG-model		CSG-model	
	Schatting	SE	Schatting	SE
<i>Vaste parameters</i>				
Intercept	36.3	10.56	-15.21	10.07
'NI'	0.5206	0.02561	0.5212	0.02638
SCHNI	0.07537	0.1043	0.5746	0.09925
<i>Random parameters</i>				
Schoolniveau				
ω_{00}	37.15	7.958	31.82	6.728
ω_{01}	0.5395	0.1811	0.5887	0.1735
		cor=.61		cor=.69
ω_{11}	0.02065	0.006344	0.02279	0.006715
Leerlingniveau				
σ^2	128.9	2.633	128.9	2.632
<i>Deviantie</i>	37826.6		37819.8	

De op het eerste zicht tegenstrijdige resultaten van beide modellen (wat het vaste gedeelte betreft) zijn evenwel tot elkaar te herleiden. De intercept van beide modellen is verschillend omdat de onafhankelijke variabele in het model met centering rond het algemeen gemiddelde gelijk is aan '0' indien 'NI' gelijk is aan het algemeen gemiddelde van deze variabele, terwijl de onafhankelijke variabele in het model met centering rond het groepsgemiddelde gelijk is aan '0' wanneer 'NI' gelijk is aan het groepsgemiddelde. Beide intercepten zijn tot elkaar te herleiden ($\gamma_{00} = \gamma_{10} \bar{X}_{..} - \gamma_{00}^*$): $36.3 = 0.5206 * 104.3 - 15.21$. De coëfficiënt van 'SCHNI' is in beide modellen verschillend omdat in het model met centering rond het algemeen gemiddelde deze slaat op het tussen- 'min' binnengroepseffect, terwijl in het andere model het slaat op het tussengroepseffect. Beide coëfficiënten zijn tot elkaar te herleiden ($\gamma_{01} = \gamma_{01}^* - \gamma_{10}^*$): $0.07537 = 0.5746 - 0.5212$.

Het random gedeelte van beide modellen is verschillend en niet herleidbaar omdat in het eerste model de variantie op groepsniveau functie is van 'NI-NI_j', terwijl in het tweede model deze functie is van 'NI-NI_j'.

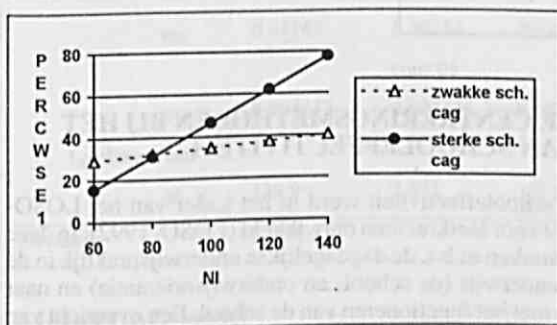
Wanneer het groepsgemiddelde ook opgenomen wordt ter verklaring van de hellingvariantie (tabel 6), krijgen we twee modellen waarvan noch het random gedeelte, noch de vaste coëfficiënten tot elkaar te herleiden zijn.

Uit de resultaten van het CAG-model kan geconcludeerd worden dat de opname van het interactie-effect 'NI*SCHNI' leidt tot een significante verbetering van het model. De deviantie van het model daalt met 32.3 eenheden, wat met een verschil van 1 vrijheidsgraad zeer significant is. Tevens blijkt dat bij opname van de interactieterm, de coëfficiënt behorend bij 'SCHNI' (opnieuw) significant wordt. Om de interpretatie van het interactie-effect en de hoofdeffecten te vergemakkelijken, werden de coëfficiënten van het vast gedeelte van het model grafisch voorgesteld in figuur 2.

Tabel 6: Random intercept- en random hellingmodellen met groepsgemiddelde ter verklaring van de intercept- en hellingvariantie

	CAG-model		CSG-model	
	Schatting	SE	Schatting	SE
<i>Vaste parameters</i>				
Intercept	-2.206	11.3	-38.48	10.98
'NI'	-1.52	0.3331	-1.283	0.3483
SCHNI	0.4515	0.1111	0.8038	0.1083
'NI'*SCHNI	0.01999	0.003227	0.01769	0.003381
<i>Random parameters</i>				
Schoolniveau				
ω_{00}	26.81	5.851	28.73	6.131
ω_{01}	0.1821	0.1093	0.27	0.1215
		cor=.38		cor=.49
ω_{11}	0.008415	0.003602	0.01074	0.004119
Leerlingniveau				
σ^2	128.8	2.63	128.7	2.628
Deviantie	37794.3		37797.6	

Figuur 2: Regressie van PERCWSE1 op NI (voor SCHNI = hoog en laag) op basis van CAG-model

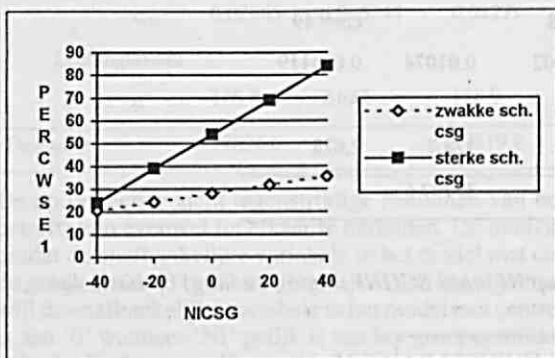


De numerieke intelligentie (NI) van de leerlingen vertoont een positief verband met het percentage wiskunde dat de leerlingen behalen op het einde van het eerste leerjaar A (in tegenstelling tot wat de negatieve coëfficiënt in het model laat vermoeden)⁴. Ook het gemiddeld numeriek intelligentieniveau van de leerlingen van een school draagt in positieve zin bij tot het verklaren van de wiskundeprestatie ($t_{52} = 4.06$) nadat reeds rekening gehouden werd met de numerieke intelligentie van de leerlingen. (Voor een leerling met een gemiddeld numeriek intelligentieniveau bedraagt de coëfficiënt van 'SCHNI' 0.4515.) Bovendien blijkt de samenhang tussen 'NI' en de wiskundeprestatie sterker te zijn naarmate de scholen een gemiddeld

sterkere leerlingpopulatie hebben in het eerste leerjaar A. Hoe hoger de numerieke intelligentie van de leerling hoe belangrijker het voor hem/haar is om naar een school te gaan met een gemiddeld sterk publiek in het eerste leerjaar.

Uit de resultaten van het *CSG-model* kan geconcludeerd worden dat het model significant verbeterd door de opname van het interactie-effect 'NI*SCHNI'. De deviantie van het model daalt met 22.2 eenheden, wat met een verschil van 1 vrijheidsgraad een zeer significant verschil is. In figuur 3 worden de hoofdeffecten en het interactie-effect grafisch weergegeven. Uit deze figuur blijkt dat de relatieve positie van de leerlingen inzake numerieke intelligentie een positief verband vertoont met het presteren op de wiskundetoets op het einde van het eerste leerjaar A. Het gemiddeld numeriek intelligentieniveau van de leerlingen van de school waartoe men behoort, draagt in positieve zin bij tot het verklaren van deze prestatie, ook wanneer rekening gehouden wordt met de relatieve positie van de leerlingen in hun school inzake numerieke intelligentie. Bovendien kan uit de figuur geconcludeerd worden dat hoe hoger het numeriek niveau van de gemiddelde leerling in een school is, hoe sterker de samenhang tussen de relatieve positie van de leerlingen in de school en hun presteren op de wiskundetoets.

Figuur 3: Regressie van PERCWSE1 op NICSIG (voor SCHNI = hoog en laag) op basis van CSG-model



2. TOEPASSING VAN TWEE CENTRERINGSMETHODEN BIJ HET BESTUDEREN VAN SCHOOLEFFECTIVITEIT

Ten behoeve van het onderzoek naar schooleffectiviteit werd in het kader van het LOSO-project een schoolkenmerkenvragenlijst voor leerkrachten ontwikkeld (LOSO, 1992). In deze vragenlijst werd o.m. gepeild naar kenmerken m.b.t. de dagdagelijkse onderwijspraktijk in de klas, de visie van de leerkrachten op onderwijs (de school- en onderwijsoriëntatie) en naar schoolkenmerken die te maken hebben met het functioneren van de school. Een overzicht van de schalen, die in het kader van onderstaande analyses gebruikt werden, worden in een bijlage gegeven.

Na een aantal verkennende analyses, waarbij naast de in de bijlage vermelde schalen ook de omvang van de schoolpopulatie in het eerste leerjaar A (GROOTTE) en het net waartoe de school behoort (NET) opgenomen werd, kwamen we tot de modellen voorgesteld in tabel 7. Alvorens de modellen te bespreken, rest ons nog één voorafgaande opmerking. De variabele 'SCHNI' (gemiddeld numeriek intellectueel niveau van de leerlingen van de school) werd gecentreerd rond het algemeen gemiddelde van de schoolgemiddelden (SCHNIC). Hierdoor

krijgen we voor de leerlingvariabele 'NI' een coëfficiënt die slaat op een conditie met een gemiddelde school d.w.z. een school met een gemiddeld intelligente leerlingpopulatie⁵.

Tabel 7: *Complexe random intercept- en random hellingmodellen*

	CAG-model		CSG-model	
	Schatting	SE	Schatting	SE
<i>Vaste parameters</i>				
Intercept	59.66	31.35	55.3	32.48
'NI'	2.518	0.5755	2.744	0.6039
LLNRICH	-7.698	4.482	-5.171	4.652
CONTR	-8.109	3.242	-7.825	3.424
LLNBEHOUD	5.19	1.626	5.714	1.718
ORDEKLAS	6.203	4.176	4.449	4.346
GROOTTE	0.03308	0.0118	0.03419	0.01251
SCHNIC	0.2662	0.1294	0.6841	0.1341
NET	-4.618	1.645	-5.238	1.741
'NI'*LLNRICH	-0.3762	0.1298	-0.4292	0.1357
'NI'*CONTR	-0.2239	0.08382	-0.2368	0.08789
'NI'*SCHNIC	0.01416	0.003171	0.01244	0.003375
<i>Random parameters</i>				
Schoolniveau				
ω_{00}	13.31	3.097	15.47	3.494
ω_{01}	0.01149	0.06254	0.03045	0.07069
		cor=.05		cor=.1
ω_{11}	0.003612	0.002426	0.004386	0.002647
Leerlingniveau				
σ^2	128.9	2.631	128.8	2.63
<i>Deviantie</i>	37756.8		37762.1	

2.1. Het CAG-model

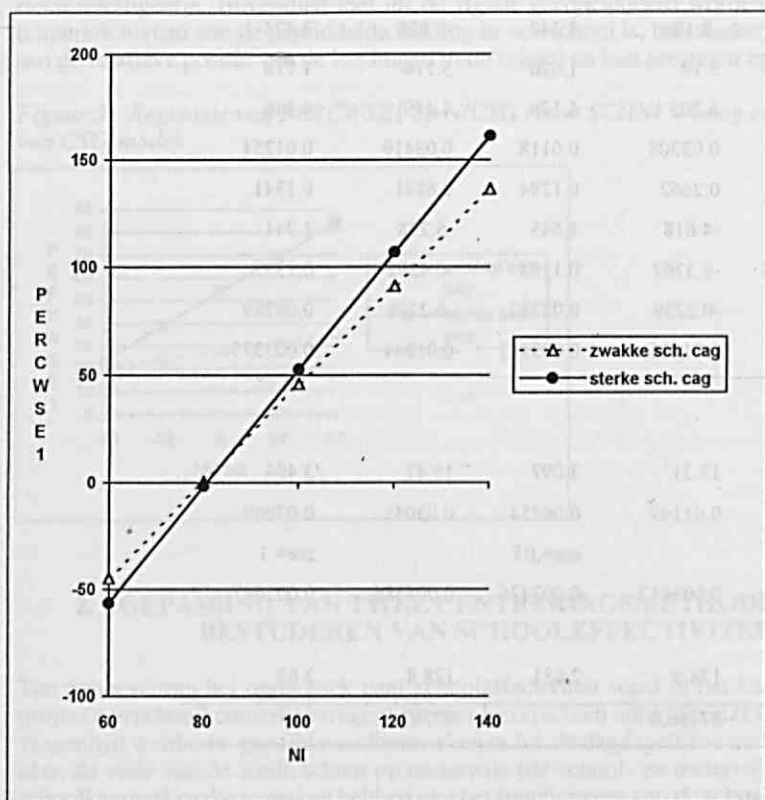
Zoals in voorgaande analyses reeds vastgesteld, blijkt de numerieke intelligentie van de leerlingen een belangrijke predictor voor het latere schoolsucces i.c. de wiskundeprestaties. Wat de schoolkenmerken betreft: zowel kenmerken die het onderwijsgebeuren in de klas beschrijven, kenmerken van de leerlingpopulatie evenals algemene schoolkenmerken staan in voor het verklaren van verschillen tussen scholen.

Niet alleen de eigen numerieke intelligentie van de leerling speelt een rol bij zijn/haar later succes op de wiskundetoets. Ook het gemiddeld niveau van de schoolgenoten in het eerste leerjaar A. Voor de gemiddelde leerling heeft het niveau van zijn/haar leerjaargenoten een

positieve 'invloed' op zijn/haar wiskunde-prestatie: hoe hoger het niveau van de leerjaar-noten, hoe meer de leerling behaalt. Wellicht heeft het vertoeven in een sterk intelligent milieu een gunstige invloed o.w.v. de rijkere leeromgeving. (Dit kan zowel een gevolg zijn van de leerlingen, van de onderwijsactiviteiten van de leerkrachten als van beide samen!)

Uit figuur 4 blijkt evenwel dat voor vrij zwakke 1A-leerlingen ('NI' < 86) het niet opportuun is om onderwijs te volgen in een 'sterke' school. Bij hen heeft dit eerder een negatieve invloed op hun prestaties. Voor de overige leerlingen geldt: hoe sterker hun eigen intellectuele capaciteiten, hoe belangrijker het is om ook naar een school te gaan met een sterk publiek in het eerste leerjaar. (Voor de sterkste leerlingen maakt het een verschil uit van ongeveer 25%.)

Figuur 4: Regressie van PERCWSE1 op NI (voor SCHNI = hoog en laag) op basis van CAG-model

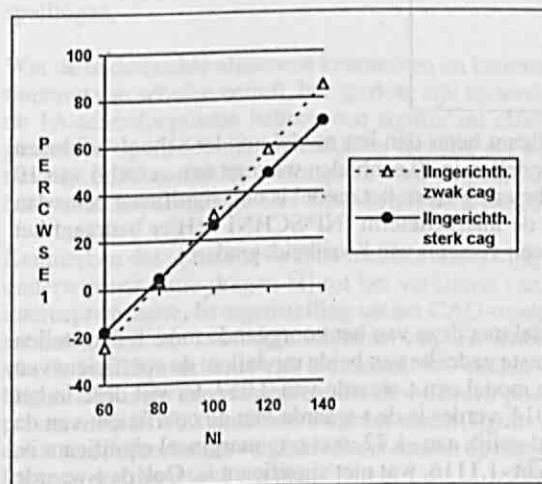


Van de onderzochte algemene schoolkenmerken en kenmerken die te maken hebben met het functioneren van scholen, hebben 'het gericht zijn op leerlingenbehoud in de advisering' (LLNBEHOUD), het net waartoe de school behoort (NET) en de omvang of grootte van de 1A-leerlingpopulatie (GROOTTE) een significant effect. Zij verklaren interceptvariantie. De leerlingen van een school gericht op leerlingenbehoud, of van een school met een grote 1A-populatie of van een school behorend tot het gemeenschapsonderwijs zullen meer behalen op de wiskunde-proef dan verwacht kon worden op basis van hun numerieke intelligentie. Voor

de gemiddelde leerling uit het gemeenschapsonderwijs bedraagt de meerwaarde respectievelijk 5.19% en 0.03% per eenheid van deze variabelen. Voor de gemiddelde leerling uit het vrij onderwijs moet het verwacht percentage met 4.618% verminderd worden.⁶ Een verklaring voor het effect van de eerstgenoemde variabele hebben we vooralsnog niet.⁷ De variabele 'GROOTTE' zou kunnen beschouwd worden als een indicator voor door de ouders gepercipieerde onderwijskwaliteit: een school trekt veel 1A-leerlingen aan omdat veel ouders de school beschouwen als een 'goede school'. Wat de variabele 'NET' betreft: deze variabele incorporeert een vorm van 'opportunity to learn'. Immers, in het gemeenschapsonderwijs worden in het eerste leerjaar A standaard 5 uur wiskunde gegeven, terwijl in het vrij onderwijs (hier: katholiek onderwijs) standaard 4 uur wiskunde wordt gegeven, waarbij de scholen of de leerlingen over de mogelijkheid beschikken om dit uit te breiden met 1 tot maximum 2 uur.

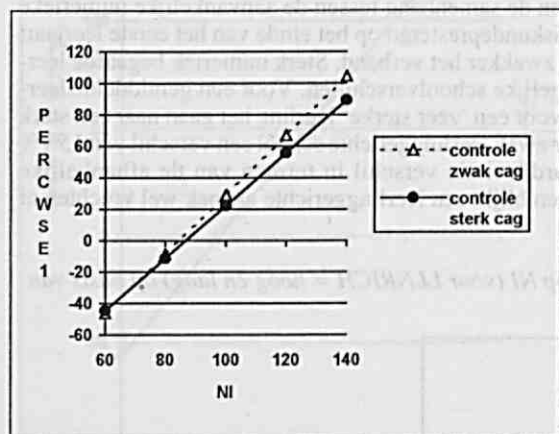
Kenmerken die te maken hebben met het onderwijsgebeuren in de klas of met de school- en onderwijsoriëntatie dragen bij tot de variantie in intercept en meestal ook in de hellingsvariantie. Wanneer in een school de lessen gemiddeld genomen ordelijk en rustig verlopen, de leerkrachten de klas goed kunnen managen (ORDEKLAS), heeft dit een positieve 'invloed' op het presteren van de leerlingen, zelfs wanneer rekening gehouden wordt met de numerieke intelligentie van die leerlingen. De invloed van de leerlinggerichtheid van de leerkrachten van een school (LLNRICH) verloopt via het reduceren van de intercept- en vooral van de hellingsvariantie. Uit figuur 5 blijkt dat wanneer de doorsnee leerkracht van een school een brede persoonsontwikkeling nastreeft, dit samengaat met een iets minder goed presteren van de leerlingen op de schoolvorderingstoets wiskunde. Bovendien is er een significant verband tussen de mate van leerlinggerichtheid en de samenhang tussen de aanvankelijke numerieke intelligentie van de leerlingen en hun wiskundepresteren op het einde van het eerste leerjaar: hoe meer leerlinggericht de school, hoe zwakker het verband. Sterk numeriek begaafde leerlingen zijn het meest gevoelig voor dergelijke schoolverschillen. Voor een gemiddelde leerling bedraagt het verschil 5.5%, terwijl voor een 'zeer sterke' leerling het gaat naar een sterk leerlinggerichte school versus een eerder zwak leerlinggerichte school een verschil van 15.7% betekent. Dit is meer dan één standaarddeviatie verschil in termen van de afhankelijke variabele. Voor de zwakste 1A-leerlingen blijkt een leerlinggerichte aanpak wel vruchten af te werpen.

Figuur 5: Regressie van PERCWE1 op NI (voor LLNRICH = hoog en laag) op basis van CAG-model



Deze - misschien op het eerste zicht verwonderlijke - vaststelling blijkt in overeenstemming te zijn met wat de leerkrachten zelf aangeven. Hoog scoren op leerlinggerichtheid houdt o.a. in dat men "het belangrijker vindt dat de leerstof aansluit bij de belangstellingssfeer van de leerlingen dan veel leerstof zien" (item 147 uit de vragenlijst); dat sociale vorming en persoonlijkheidsvorming minstens zo belangrijk is, zonet belangrijker dan een beroepsopleiding en/of intellectuele vorming (item 141). Wellicht impliceert het aandacht besteden aan de brede persoonlijkheidsontwikkeling van de leerling dat er m.b.t. de intellectuele vorming van de leerling en in het bijzonder het behandelen van leerstof keuzes moeten gemaakt worden.⁷ Met betrekking tot de variabele 'strikt uitvoeren van controle' (CONTR) komen we tot gelijkaardige bevindingen (figuur 6). In scholen waarbij de doorsnee leerkracht veel controle uitvoert, dit houdt o.m. in een strak werkritme aanhouden om de leerlingen in de hand te houden, presteren de leerlingen minder goed dan in scholen waar dit niet zo sterk gebeurt. Enige vergelijking kan hier gemaakt worden met het onzekere agressieve type uit de typologie van het interpersoonlijk leraarsgedrag (Brekelmans, et al., 1993). Ook in de studies van de vernoemde auteurs presterden de leerlingen van leerkrachten die dergelijk gedrag vertoonden minder goed. Verder stellen we vast dat de samenhang tussen de numerieke intelligentie van de leerlingen en hun presteren minder groot is naarmate de leerkrachten van de scholen meer controle uitvoerend gedrag vertonen. Ook hier geldt weer: hoe intelligenter de leerling, hoe meer nadeel hij ondervindt van een dergelijke aanpak.

Figuur 6: Regressie van PERCWSE1 op NI (voor CONTR = hoog en laag) op basis van CAG-model



Tot slot: ons uiteindelijk model is significant beter dan het model zonder schoolvariabelen (het verschil in deviantie tussen beide modellen is 70 eenheden wat met een verschil van 10 vrijheidsgraden een zeer significante verbetering is) en het model is ook significant beter dan het model met alleen maar 'SCHNI' en de interactieterm 'NI*SCHNI'. (Hier bedraagt het verschil in deviantie 37.5 eenheden met een verschil van 8 vrijheidsgraden.)

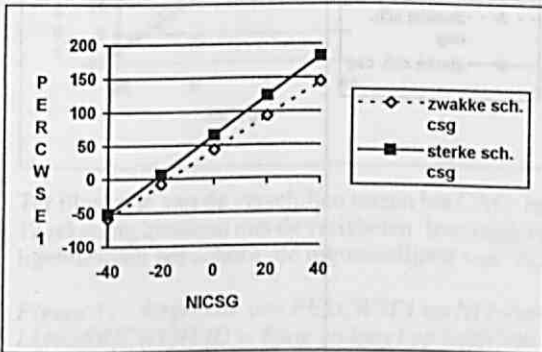
2.2. Het CSG-model

Vergelijken we de resultaten van dit model met deze van het voorgaande model, dan stellen we vast dat er verschillen zijn m.b.t. het vaste gedeelte van beide modellen: de coëfficiënt van de variabele 'SCHNI' heeft in het eerste model een t-waarde van 2.057, terwijl deze in het tweede model een waarde heeft van 5.1014, verder is de t-waarde van de coëfficiënt van de variabele 'LLNRICH' in het CAG-model gelijk aan -1.72, wat nog marginaal significant is, terwijl deze in het CSG-model gelijk is aan -1.1116, wat niet significant is. Ook de t-waarde

van de variabele 'ORDEKLAS' is in het CAG-model gelijk aan 1.4854, wat nog marginaal significant is, terwijl de coëfficiënt in het CSG-model niet significant is (t-waarde is gelijk aan 1,0237). Ook de coëfficiënt van de interactieterm 'NI*SCHNIC' wordt minder significant (t-waarde respectievelijk 4.4654 versus 3.6859). Qua interpretatie van de resultaten betekent dit dat we zoals in voorgaande analyses van het CSG-model vaststellen dat de relatieve positie die een leerling inneemt in zijn/haar school qua numerieke intelligentie een belangrijke predictor is voor zijn/haar latere wiskundeprestatie. Het algemeen niveau van de school qua numerieke intelligentie van de leerlingenpopulatie draagt ook in positieve zin bij tot dit succes. Bovendien is het verband tussen de relatieve positie van de leerlingen in de school en de wiskundeprestatie functie van het algemeen niveau van de school: hoe hoger het niveau van de school, hoe hoger de samenhang.

Uit figuur 7 blijkt bovendien dat naarmate de numerieke intelligentie van de leerlingen in de buurt ligt van het algemeen niveau van de 1A-leerlingen van hun school of hoger is, ze gevoeliger zijn voor het algemeen niveau van de 1A-leerlingen van hun school.

Figuur 7: Regressie van PERCWSE1 op NICSG (voor SCHNI = hoog en laag) op basis van CSG-model

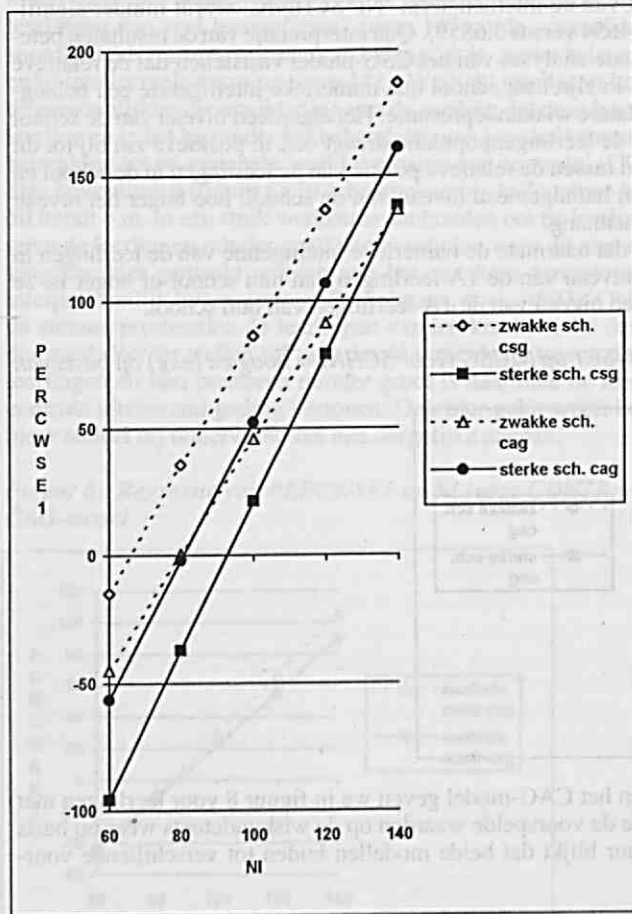


Ter vergelijking van het CSG- en het CAG-model geven we in figuur 8 voor leerlingen met eenzelfde numerieke intelligentie de voorspelde waarden op de wiskundetoets weer op basis van beide modellen. Uit de figuur blijkt dat beide modellen leiden tot verschillende voorspellingen.

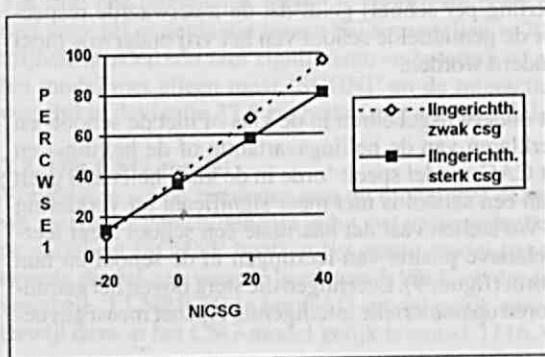
Wat de onderzochte algemene kenmerken en kenmerken die te maken hebben met het functioneren van scholen betreft, het 'gericht zijn op leerlingenbehoud', het net en de grootte van de 1A-schoolpopulatie hebben een significant effect. Voor de gemiddelde school van het gemeenschapsonderwijs (gemiddelde leerling per school) geldt dat de meerwaarde respectievelijk 5.7% en 0.03% bedraagt en voor de gemiddelde school van het vrij onderwijs moet het verwacht percentage met 5:24 verminderd worden.

Kenmerken die te maken hebben met het onderwijsgebeuren in de klas of met de school- en onderwijsoriëntatie dragen bij tot het verklaren van de hellingsvariantie of de hellings- en interceptvariantie. In tegenstelling tot het CAG-model speelt 'orde in de klas' helemaal geen rol meer. Ook de 'leerlinggerichtheid' van een school is niet meer significant ter verklaring van de interceptvariantie van het model. We stellen vast dat naarmate een school meer leerlinggericht is de samenhang tussen de relatieve positie van leerlingen in de school en hun presteren op de wiskundetoets zwakker wordt (figuur 9). Leerlingen die sterk boven het gemiddelde van de leerlingen van hun school scoren op numerieke intelligentie zijn het meest gevoelig voor een dergelijke aanpak.

Figuur 8: Regressie van PERCWSE1 op NI (SCHNI = hoog en laag) op basis van CAG- en CSG-model

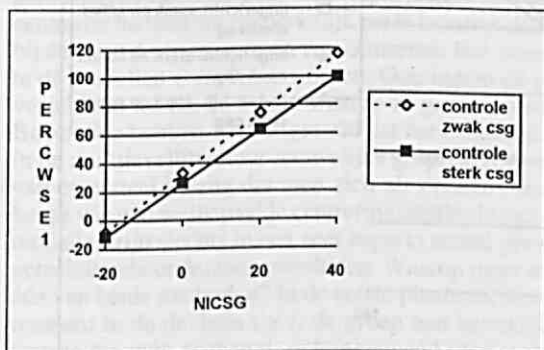


Figuur 9: Regressie van PERCWSE1 op NICSG (voor LLNRICH = hoog en laag) op basis van CSG-model



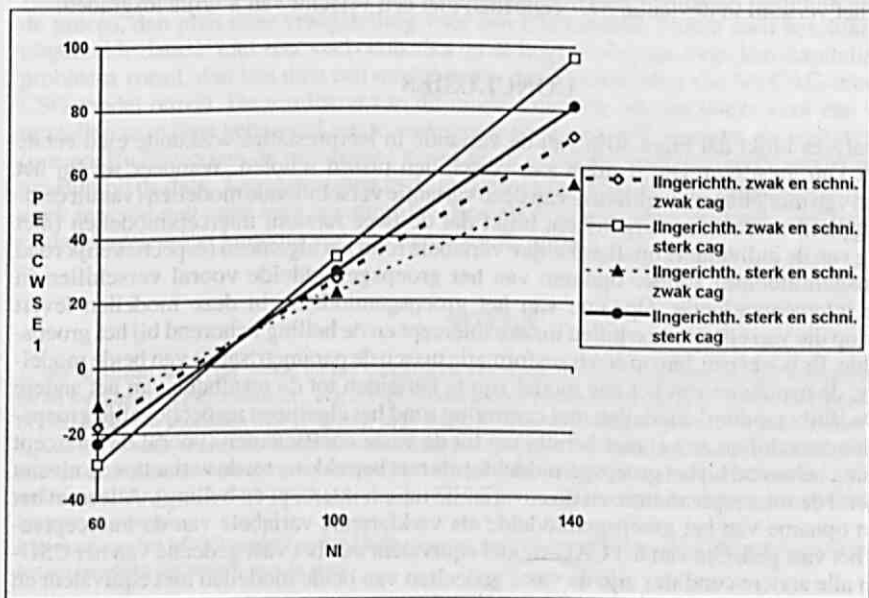
Met betrekking tot de variabele 'strikt controle uitvoeren' vinden we zowel een significant hoofdeffect van deze variabele op het behaald percentage wiskunde eind eerste leerjaar als een significant interactie-effect: in scholen waar de doorsnee leerkracht veel controle uitvoert presteren de leerlingen minder goed dan verwacht kon worden op basis van hun relatieve positie qua numerieke intelligentie in hun school. Bovendien is de samenhang tussen de relatieve positie in de school en het behaald percentage op de wiskundetoets hoger naarmate de doorsnee leerkracht van de school meer controle uitvoert (figuur 10).

Figuur 10: Regressie van PERCWSE1 op NISG (voor CONTR = hoog en laag) op basis van CSG-model

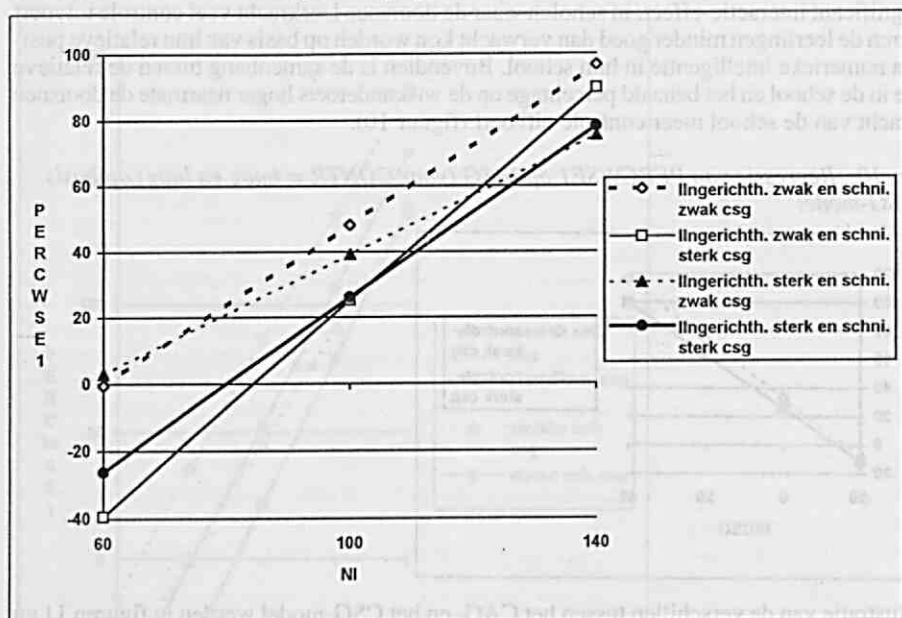


Ter illustratie van de verschillen tussen het CAG- en het CSG-model werden in figuren 11 en 12 rekening houdend met de variabelen 'leerlinggerichtheid' en 'gemiddelde numerieke intelligentiescore per school' de regressielijnen van 'PERCWSE1' op 'NI' weergegeven.

Figuur 11: Regressie van PERCWSE1 op NI (voor SCHNI = hoog en laag en LLNGERICHTHEID = hoog en laag) op basis van CAG-model



Figuur 12: Regressie van PERCWSE1 op NI (voor SCHNI = hoog en laag en LLNGERICHTHEID = hoog en laag) op basis van CSG-model



Tot slot: het uiteindelijke CSG-model is significant beter dan het CSG-model zonder schoolvariabelen (tussen beide modellen is het verschil in deviantie 76.8 eenheden wat met een verschil van 10 vrijheidsgraden een zeer significante verbetering is). Het model is ook significant beter dan het model met alleen maar 'SCHNI' en de interactieterm 'NICSMSCHNI'. Hier bedraagt het verschil in deviantie 57.7 eenheden met een verschil van 8 vrijheidsgraden.

CONCLUSIES

Uit de analyses blijkt dat bijna 30% van de variantie in leerprestaties wiskunde eind eerste leerjaar A kan toegeschreven worden aan verschillen tussen scholen. Wanneer we bij het opbouwen van modellen ter verklaring van deze variantie verschillende modellen (vanuit centeringsoogpunt) met elkaar vergelijken, blijkt dat de twee random interceptmodellen (met centering van de individuele onafhankelijke variabele rond het algemeen respectievelijk rond het groepsgemiddelde) zonder opname van het groepsgemiddelde vooral verschillen in residuele interceptvariantie. Opname van het groepsgemiddelde in deze modellen levert modellen op die van elkaar verschillen inzake intercept en de helling behorend bij het groepsgemiddelde. Er is wel een één op één transformatie tussen de parametrisaties van beide modellen m.a.w. de resultaten van het ene model zijn te herleiden tot de resultaten van het andere model. De 'fully random'-modellen met centering rond het algemeen respectievelijk groepsgemiddelde verschillen zowel met betrekking tot de vaste coëfficiënten (vooral de intercept en de helling behorend bij het groepsgemiddelde) als met betrekking tot de varianties op niveau twee (vooral de interceptvariantie en de covariantie tussen intercept en helling). Alleen in het geval van opname van het groepsgemiddelde als verklarende variabele van de interceptvariantie is het vast gedeelte van het CAG-model equivalent aan het vast gedeelte van het CSG-model. In alle andere condities zijn de vaste gedeeltes van beide modellen niet equivalent en

verkrijgen we vaste coëfficiënten voor beide modellen die niet tot elkaar te herleiden zijn. Het kiezen voor een welbepaalde centreringmethode heeft dus gevolgen voor de onderzoeksresultaten, hetgeen geïllustreerd werd aan de hand van een inhoudelijke bespreking van de resultaten van twee complexe 'fully random' modellen. We vinden verschillen tussen beide modellen met betrekking tot het belang van zowel het groepsgemiddelde als met betrekking tot het belang van schoolkenmerkenvariabelen ter verklaring van de intercept- en hellingsvariantie.

Uit de studie kan geconcludeerd worden dat de centreringproblematiek meer aandacht verdient dan tot nog toe gebeurd is. Dat centreren belangrijk is, ervaren we reeds bij een vroegere analyse van onze gegevens (Opdenakker & Van Damme, 1996). Zonder enige vorm van centreren hadden we opmerkelijk meer iteraties nodig opdat het model zou convergeren dan bij één van de twee vormen van centreren. Het verschil in aantal iteraties neemt toe naarmate de modellen complexer worden. Ook indien de predictorvariabelen zeer sterk van elkaar verschillen inzake de schaal waarop ze gescoord worden, is het gunstig om ze te centreren. Bovendien hebben we vastgesteld dat het centreren van hogere niveau variabelen, waarover in de multilevelliteratuur nauwelijks gesproken wordt, soms ook niet onbelangrijk is. Het is van essentieel belang dat men zich als onderzoeker bewust is van de consequenties die de keuze van een welbepaalde centreringmethode met zich meebrengt. Immers, CAG- en CSG-modellen zijn slechts in een zeer beperkt aantal gevallen equivalent en leiden dus meestal tot verschillende onderzoeksresultaten. Waarop moet men zich aldus baseren om te kiezen voor één van beide methoden? In de eerste plaats zal de concrete vraagstelling (bv. is men geïnteresseerd in de deviatie t.o.v. de groep met betrekking tot een welbepaalde variabele) of de theorie die men aanhangt, richtinggevend (moeten) zijn voor de keuze (Bosker & Snijders, 1991; Kreft et al., 1995). Indien men aldus wenst na te gaan of de absolute waarde die individuen op een welbepaalde onafhankelijke variabele hebben predictief is voor hun score op de onafhankelijke variabele en men daarnaast ook geïnteresseerd is in het effect van de context waarin deze individuen zich bevinden (bv. behoort het individu tot een omgeving die hoog of laag scoort op de onafhankelijke variabele), dan vraagt dit om een CAG-model. Is men echter geïnteresseerd in de relatieve positie van individuen ten opzichte van de groep waartoe ze behoren (bv. scoort een individu hoog ten opzichte van het 'gemiddelde' lid van de groep), dan pleit deze vraagstelling voor een CSG-model. Indien men beschikt over een uitgebreide dataset met zeer veel eenheden op de hogere niveaus zodat kanskapitalisatie geen probleem vormt, dan kan men een model testen dat de variabelen van het CAG-model en het CSG-model omvat. De resultaten van dit model laten toe om een keuze voor één van beide modellen te maken gebaseerd op de onderzoeksgegevens zelf. Immers, de analyseresultaten (o.a. coëfficiënten van de variabelen) geven in dit geval aan welk van beide modellen het best aansluit bij de data. Aangezien men in de onderzoekspraktijk van het onderwijskundig onderzoek meestal niet over zeer omvangrijke datasets beschikt, heeft de laatst geschetste mogelijkheid veeleer het karakter van een theoretische optie.

NOTEN

1. We beperken ons in dit artikel tot het vergelijken van modellen met centreren rond het algemeen gemiddelde en rond het groepsgemiddelde. Ten eerste, omdat uit een studie van Kreft et al. (1995) en Opdenakker et al. (1996) blijkt dat modellen met centreren rond het algemeen gemiddelde en ruwe scoresmodellen equivalent zijn en ten tweede, omdat de vraag van een onderzoeker bij het kiezen voor een model in het licht van het voorgaande en gezien de voordelen van centreren, niet zozeer zal zijn of hij/zij al dan niet moet centreren maar wel welke centreringmethode hij/zij het best gebruikt bij het analyseren van de data ter beantwoording van de onderzoeksvragen.

Met dank aan het LOSO-project en het Ministerie van de Vlaamse gemeenschap, Departement Onderwijs voor het ter beschikking stellen van de data.

2. Adres: Vesaliusstraat 2, B-3000 Leuven, e-mail: Marie-Christine. Opendakker@ped.kuleuven.ac.be.
3. Dit geldt overigens ook voor het toevoegen van verklarende variabelen op het leerlingniveau: indien in het model met centering rond het algemeen gemiddelde een bijkomende leerlingvariabele opgenomen wordt die gecorreleerd is met de variabele die gecentreerd werd, dan kan deze bijkomende variabele, in tegenstelling tot dezelfde variabele in het model met centering rond het groepsgemiddelde, geen tussengroepsvariantie binden.
4. Bij de opname van een interactieterm en hoofdeffecten verwijst de coëfficiënt bij 'NI' naar de situatie waarbij 'SCHNI' = 0; in ons geval is dit een onbestaande situatie. De coëfficiënt wordt positief in reële (d.w.z. in de dataset voorkomende) situaties.
5. Werken met de niet gecentreerde variabele 'SCHNI' leverde een niet significante coëfficiënt van 'NI' op. Omdat de coëfficiënt in dit geval slaat op een onbestaande situatie m.n. een situatie waarbij het schoolgemiddelde op 'NI' gelijk is aan 0, heeft het geen zin om parameters te laten schatten voor een dergelijke situatie.
6. Het 'NET' werd als volgt gecodeerd: 1 = vrij onderwijs, 0 = gemeenschapsonderwijs.
7. Het betreft een schaal bestaande uit slechts twee items. Enig voorbehoud met betrekking tot het gevonden resultaat is dus wellicht op zijn plaats.
8. Deze vaststelling impliceert niet dat leerlinggerichte scholen geen goede scholen zouden zijn. We stellen alleen vast dat m.b.t. de kennis en kunde van wiskunde op het einde van het eerste leerjaar de leerlingen van deze scholen het minder goed doen dan verwacht. Wellicht beschikken de leerlingen van deze scholen over andere troeven o.a. sociale en persoonlijkheidsvorming en voelen zij zich misschien beter thuis op de school, wat in het vervolg van hun studieloopbaan misschien ook niet onbelangrijk is.

REFERENTIES

- Bosker, R.J. & Snijders, T. (1991). *Center forward*. Handout bij de voordracht voor de MULOLOG-groep op 21 maart 1991.
- Bosker, R.J. & Witziers, B. (1996, April). *The magnitude of school effects. or: Does it really matter which school a student attends?* Paper presented at the annual meeting of the AERA. New York.
- Brekelmans, M., Levy, J. & Rodriguez, R. (1993). A typology of teacher communication style. In T. Wubbels & J. Levy (Eds.). *Do you know what you look like. Interpersonal relationships in education* (p. 56-63). London: The Falmer Press.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models. Applications and data analysis methods*. Newbury Park: Sage.
- Damme, J. Van, Opendakker, M.-C. & Damme, G. Van (1994). *Klaspraktijk, school- en onderwijsoriëntatie en het functioneren van scholen: individuele schoolgegevens verkregen via de schoolkenmerkenvragenlijst voor de leerkrachten*. Bundel voor de meewerkende scholen.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Eduard Arnold.
- Kreft, I.G.G., Leeuw, J. de & Aiken, L.S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30 (1), 1-21.
- Kreft, I.G.G. (1995). *The effects of centering in multilevel analysis: is the public school the loser or the winner? A new analysis of an old question*. Paper presented at the annual meeting of AERA. San Francisco.
- Longford, N.T. (1990). *VARCL. Software for variance component analysis of data with nested random effects (maximum likelihood)*. Princeton, New York: Educational Testing Service.
- LOSO. (1992). *Vorderingsverslag van het F.K.F.O.-project M.J. nr. 89.10 "De Studieloopbaan van Vlaamse jongeren. Het secundair-onderwijssysteem doorgelicht op basis van longitudinaal onderzoek"*. Leuven: K.U.Leuven, Centrum voor Secundair en Hoger Onderwijs.
- Opendakker, M.-C. & Damme, J. Van (1996). *Centreren in multilevelonderzoek. Een verkenning op basis van LOSO-gegevens*. Leuven, K.U.Leuven, Onderzoekscentrum voor Secundair en Hoger Onderwijs. Intern document, LIVO-seminarie.
- Verhoeven, J., Vandenberghe, R., Damme, J. Van, Clement, M., Maetens, D. & Vergauwen, G. (1992). *Schoolmanagement en kwaliteitsverbetering van het onderwijs. Een empirisch onderzoek in secundaire scholen*. Leuven: K.U.Leuven, Departement Sociologie, Sociologisch Onderzoeksinstituut.
- Woodhouse, G. (Ed.) (1995). *A guide to MLn for New users*. University of London: Institute of Education.

BIJLAGE

Beschrijving van de schalen uit de schoolkenmerkvragenlijst voor de leerkrachten

De vragenlijst 'Schoolkenmerken voor de leerkrachten' bevatte een uitgebreide reeks uitspraken over o.a. de beleving van het werk als leerkracht, de dagdagelijkse onderwijspraktijk, de school- en de onderwijsoriëntatie, het functioneren van de school als organisatie, de besluitvorming op school en de oriëntatie van de school op studiepakketten.

Aan de hand van een aantal analysetechnieken (o.a. factoranalyse met varimaxrotatie) hebben we deze uitspraken in groepen ondergebracht. Belangrijk hierbij was zo zuiver mogelijke schalen verkrijgen die inhoudelijk goed kunnen worden geïnterpreteerd.

Per schaal wordt de naam van de schaal, de afkorting die we in de tabellen voor deze schaal gebruiken, de betrouwbaarheid van de schaal (alfa-coëfficiënt als maat van interne consistentie tussen de opgenomen uitspraken), een verduidelijking van de betekenis van de schaal en enkele voorbeelden van representatieve uitspraken of items m.n. de drie hoogstladende items.

Bij het benoemen van de schalen wordt de afkorting 'LLN' gebruikt voor 'leerlingen'.

Een (-) teken voor een item betekent dat dit item in tegenovergestelde zin bijdraagt tot de betekenis van de schaal: hoe minder men akkoord gaat met dit item, des te hoger scoort men op de betreffende schaal.

We onderscheiden 16 inhoudelijk betekenisvolle schalen die elk een facet van de dagdagelijkse onderwijspraktijk, de school- en onderwijsoriëntatie of het functioneren van de scholen meten. Sommige schalen behandelen een eerder specifiek aspect van het 'schoolse leven' (bv. schaal 4: 'een persoonlijke relatie met leerlingen hebben'), terwijl andere schalen een ruim geheel behandelen (bv. schaal 11: 'goed functioneren van de school als organisatie').

SCHAAL 1: ARBEIDSTEVREDENHEID (ARBTEVR) (n = 14, alfa = .87)

Deze schaal geeft aan in welke mate de leerkracht tevreden is over zijn/haar werk als leerkracht in het secundair onderwijs.

items: Ik sta graag voor de klas.

Ik vind van mezelf dat ik goed geschikt ben voor het werk als leerkracht in het secundair onderwijs.

Ik heb over het algemeen het gevoel dat ik voor de leerlingen waaraan ik les geef veel kan betekenen.

SCHAAL 2: LLN ACTIEF BETREKKEN IN LESGEBEUREN (LLNACT)

(n = 7, alfa = .77)

Deze schaal geeft aan in welke mate de leerkracht de leerling op een actieve manier laat deelnemen aan het lesgebeuren.

items: Ik onderhandel met de leerlingen over wat er tijdens de les besproken of vervaardigd zal worden.

Ik laat de leerlingen zelf allerhande toepassingen zoeken van hetgeen ze in de klas of in het atelier geleerd hebben.

Het gebeurt dat ik er met de leerlingen opuit trek om de werkelijkheid zelf te verkennen.

SCHAAL 3: ORDELIJK EN RUSTIG VERLOOP IN DE KLAS (ORDEKLAS)

(n = 6, alfa = .74)

Deze schaal is een indicator voor een ordelijk en rustig verloop van de activiteiten in de klas.

items: (-) Het komt voor dat er tijdens mijn lessen onnodige en storende onderbrekingen zijn.

(-) Het duurt dikwijls lang vooraleer het rustig genoeg is om met de les te beginnen.

De overgangen tussen de lessen verlopen bij mij rustig en vlot.

SCHAAL 4: PERSOONLIJKE RELATIE MET LLN HEBBEN (PERSLLN) (n = 6, alfa = .67)

Deze schaal geeft aan of de leerkracht een persoonlijke relatie aangaat met zijn/haar leerlingen.

- items:* Wanneer een leerling mij in vertrouwen wil nemen dan zal ik daar op ingaan.
Ik zal (zou) me t.a.v. de leerlingen openlijk excuseren wanneer ik hen foutief of unfair behandeld heb (zou hebben).
Tussen mezelf en de leerlingen is er een relatie die gebaseerd is op wederzijds vertrouwen.

SCHAAL 5: AAN DIFFERENTIATIE DOEN (DIFF) (n = 6, alfa = .67)

Deze schaal geeft aan of de leerkracht een onderscheid maakt tussen zijn/haar leerlingen en verschillende groepen leerlingen ook anders benadert (bv. andere oefeningen geeft, extra uitleg, ...).

- items:* Ik heb speciaal materiaal (handboeken, oefeningen ...) voor het remediëren van tekorten bij zwakkere leerlingen.
Ik voorzie extra-oefeningen of andere taken voor leerlingen die vlugger van begrip zijn (uitbreidings- of verdiepingsstof).
Ik geef zelf extra-hulp buiten de lesuren voor leerlingen die wat trager of zwakker zijn en dreigen te mislukken.

SCHAAL 6: CONTROLE UITVOEREN (CONTR) (n = 3, alfa = .50)

Via deze schaal wordt nagegaan in welke mate de leerkracht het werkritme en eventuele sancties hanteert om de leerlingen onder controle te houden.

- items:* Om de leerlingen in de hand te houden, hou ik een strak werkritme aan.
De sancties die voor een overtreding gelden, worden door mij strikt toegepast.
Bij ongeregelheden in de klas reageer ik met straf.

SCHAAL 7: LEERLINGGERICHTHEID (LLNRICH) (n = 14, alfa = .77)

Deze schaal geeft aan in welke mate de leerkracht in zijn/haar visie op het onderwijs en de school gericht is op de leerlingen als totale persoon.

- items:* De ontwikkeling van de creativiteit van leerlingen is één van de belangrijkste taken van de school.
Leerlingen moeten op school vooral zichzelf kunnen zijn en zichzelf kunnen ontdekken.
Aansluiting van de leerstof bij de belangstellings sfeer van de leerlingen vind ik belangrijker dan veel leerstof zien.

SCHAAL 8: GERICHTHEID OP COGNITIEVE ONTWIKKELING EN OP LEERSTOF (COGNRICH) (n = 9, alfa = .70)

Deze schaal geeft aan in welke mate de leerkracht in zijn/haar visie op het onderwijs en de school gericht is op kennisoverdracht, op vakdeskundigheid en op het verwerven van leerstof.

- items:* Ik ben er op de eerste plaats voor verantwoordelijk dat de leerlingen voldoende vak-kennis wordt bijgebracht. Al de rest is eerder bijzaak.
De voornaamste taak van de school is de leerling voor te bereiden op verdere studies.
Kennisoverdracht en intellectuele ontwikkeling blijven in het onderwijs belangrijker dan algemene vorming.

SCHAAL 9: POSITIEVE DIFFERENTIATIEATTITUDE (POSDIFF) (n = 7, alfa = .68)

Deze schaal inventariseert of de leerkracht een positieve houding heeft ten aanzien van differentiëren.

items: (-) De goede leerling is de dupe van de differentiatie in het onderwijs.

Gedifferentieerd onderwijs komt alle leerlingen ten goede.

(-) In onderwijs waarbij verschillende leerlingen aan verschillende taken werken, verwerven de leerlingen minder kennis.

SCHAAL 10: DISCIPLINE (DISCIPL) (n = 9, alfa = .70)

Deze schaal meet in welke mate de leerkracht belang hecht aan gehoorzaamheid en discipline.

items: Gehoorzaamheid is goed voor de vorming van de leerling.

Voor goed onderwijs is het belangrijk dat de leerlingen in de eerste plaats gehoorzamen en zich schikken naar de regels van de school.

(-) Discipline en orde verhinderen de vrije ontplooiing van de leerlingen en zouden tot een minimum beperkt moeten worden.

SCHAAL 11: GOED FUNCTIONEREN VAN DE SCHOOL ALS (INTERNE) ORGANISATIE (FUNCTIORG) (n = 34, alfa = .94)

Deze schaal peilt naar het goed functioneren van de school qua management, relatie tussen directie en leerkrachten en leerkrachten onderling m.a.w. naar het intern functioneren van de school.

items: De directie op onze school heeft een duidelijke visie op wat een goede school is.

Als organisatie draait onze school goed.

De directie stimuleert de leerkrachten regelmatig tot gezamenlijk nadenken over wat we nastreven met het onderwijs in onze school.

SCHAAL 12: POSITIEVE ATTITUDE T.O.V. P.M.S. (POSPMS) (n = 8, alfa = .89)

Deze schaal gaat na of de leerkracht positief staat t.a.v. de werking van het P.M.S.-centrum.

items: Over de huidige werking van het P.M.S.-centrum ben ik erg tevreden.

Het P.M.S.-centrum doet voor onze school veel nuttig werk.

Het P.M.S.-centrum brengt vaak interessante aanvullende gegevens aan bij de bespreking over een leerling met leerproblemen (bijv. op de klassenraad).

SCHAAL 13: POSITIEVE ATTITUDE T.O.V. OUDERS (POSOD) (n = 10, alfa = .68)

Deze schaal geeft aan dat de leerkracht inspraak van ouders in schoolse aangelegenheden wenst en goede contacten met ouders heeft.

items: Ik vind het positief dat ouders deelnemen aan het schoolse gebeuren.

(-) De oudervereniging of ouderraad behoort geen invloed te hebben op het schoolbeleid.

Ik vind dat ouders invloed moeten hebben op de organisatorische vormgeving van het onderwijs bij ons op school.

SCHAAL 14: MATE VAN STUDIEKEUZE BEGELEIDING (STKBGL) (n = 7, alfa = .72)

Via deze schaal wordt nagegaan in welke mate de school aan studiekeuzebegeleiding doet bij de leerlingen.

items: Op onze school wordt er met het P.M.S.-centrum samengewerkt rond de studiekeuzebegeleiding van de leerlingen.

De begeleiding van de leerlingen op onze school in verband met hun studiekeuzes verloopt volgens een welbepaalde strategie.

Op het oudercontact op het einde van het schooljaar bespreken we met de ouders van de leerling welke keuze deze leerling het best zou maken.

SCHAAL 15: HIERARCHIE TUSSEN STUDIERICHTINGEN EN VAKKEN (HIESTR) (n = 5, alfa = .71)

Deze schaal meet of de studierichtingen die in de school aangeboden worden, door de leerkrachten als evenwaardig beschouwd worden.

items: Sommige studierichtingen hebben bij ons op school meer aanzien dan andere.

Niet alle studierichtingen worden bij ons op school als evenwaardig beschouwd.

Bij het lesgeven en het beoordelen van de leerlingen worden duidelijk andere normen gehanteerd in de verschillende studierichtingen.

SCHAAL 16: GERICHTHEID OP LLNBEHOUD IN ADVISERING (LLNBEHOUD) (n = 2, alfa = .70)

Deze schaal meet of de school bij het adviseren van leerlingen probeert om zo veel mogelijk leerlingen binnen de eigen school te houden.

items: De kans dat een bepaalde richting onvoldoende leerlingen telt, bepaalt wel eens mede welk advies aan bepaalde leerlingen gegeven wordt.

Op onze school krijgen leerlingen die wellicht beter naar een andere school zouden gaan, toch soms het advies een richting te kiezen die bij ons bestaat.

Ontvangen 01-09-97

Definitief aanvaard 15-11-97

Notities en Commentaren

IS VERKLEINING VAN KLASSEN EEN GOEDE INVESTERING?

Hessel Oosterbeek & Dinand Webbink¹

Enkele weken geleden bood de voorzitter van de Commissie Kwalitatieve Aspecten van Groeps-grootte in het Basisonderwijs (Van Eijndhoven) het eindrapport 'Klassenverkleining' aan, aan staatssecretaris Netelenbos.² Het advies van de commissie, de reactie van de staatssecretaris en beschouwingen in de pers waren vrijwel unaniem. Eindelijk was het 'wettig en overtuigend' bewijs geleverd dat verkleining van klassen in het basisonderwijs, met name in de onderbouw, een noodzakelijke voorwaarde is voor beter basisonderwijs. En politiek ligt de zaak kennelijk ook goed, want minister Zalm zou - die indruk werd althans gewekt - wel bereid zijn de extra 923 miljoen gulden die daarvoor jaarlijks (!) nodig zijn, beschikbaar te stellen.

De unanimititeit ten aanzien van deze majeure investering was niet verwacht. Weliswaar was er al een eerste aankondiging toen uitkomsten van experimenten in de Amerikaanse staat Tennessee bekend werden. Toen hield de staatssecretaris echter nog een flinke slag om de arm. Maar vooral omdat er nu ook specifiek aan de Nederlandse situatie ontleende onderzoeksuitkomsten beschikbaar zijn, lijkt zij overtuigd.

Naar onze mening zou de staatssecretaris het rapport dat haar is aangeboden eens goed moeten lezen, en het daarin aangeboden bewijsmateriaal kritisch tegen het licht moeten houden. Zij zal dan moeten concluderen dat de bewijzen veel minder hard zijn dan de commissie suggereert; in feite zijn ze boterzacht.

De stellingname van de commissie is dat klassenverkleining noodzakelijk is voor betere prestaties van leerlingen in het basisonderwijs; maar klassenverkleining is niet voldoende. Er is flankerend beleid nodig, daarbij noemt de commissie ruimtelijke voorzieningen, het afleggen van rekenschap door de scholen en professionalisering van leerkrachten. Zo geformuleerd is het een voorzichtige stellingname. Immers als empirisch wordt gevonden dat klassengrootte geen invloed heeft op de prestaties, dan kan beweerd worden dat het flankerend beleid onvoldoende is geweest om de gecreëerde randvoorwaarden te benutten. Pas als leerlingen in grote klassen beter presteren dan leerlingen in kleine klassen, is de stellingname niet houdbaar meer.

Toch meent de commissie dat het voor de onderbouwing van haar advies belangrijk is om materiaal aan te dragen dat laat zien dat leerlingen in grote klassen minder goed presteren dan leerlingen in kleine klassen. Voor die onderbouwing worden brokstukken bewijs geleverd die zijn ontleend aan: (i) internationaal onderzoek, (ii) de mening van door de commissie gehoorde deskundigen, (iii) onderzoek op basis van Nederlandse gegevens, en (iv) internationaal vergelijkende gegevens over klassengrootte, uitgaven per leerling en leerprestaties.

In het vervolg van dit commentaar bespreken we elk van deze onderdelen van het bewijs.

Buitenlands onderzoek

Veel van het onderzoek dat in andere landen is verricht geeft weinig aanleiding voor de stellingname dat kleine klassen gepaard gaan met betere leerprestaties. De commissie gaat in op methodologische tekortkomingen van de betreffende studies. Hoewel daarmee op sommige plaatsen terecht vraagtekens bij de uitkomsten worden gezet, geven deze onderzoeken geen bewijs voor een negatief verband tussen klassengrootte en leerprestaties van kinderen.

Er zijn echter enkele onderzoeken die wel op zo'n relatie wijzen. Het bekendst daarvan is het recente STAR-experiment dat is uitgevoerd in de Amerikaanse staat Tennessee. In dit onderzoek zijn in 79 scholen klassen gevormd die aan drie verschillende condities zijn blootgesteld. Deze condities zijn: (i) klassen van 13-17 leerlingen, (ii) klassen van 22-25 leerlingen,

en (iii) klassen van 22-25 leerlingen met een fulltime onderwijsassistent. Gevonden wordt dat zowel voor taal als voor rekenen leerlingen in de kleine klassen belangrijk beter scoren dan in de grote klassen. De aanwezigheid van een onderwijsassistent heeft geen invloed op de prestaties in deze domeinen. Dit lijkt sterk bewijs, maar er zijn wel wat kanttekeningen te maken. Ten eerste, scholen deden op basis van vrijwilligheid mee aan het experiment, en niet - zoals zou moeten - op basis van random toewijzing. Ten tweede, de range van klassengrootte waar de betere resultaten worden vastgesteld is 13-17 leerlingen. Dit is aanzienlijk buiten de range waarover de discussie in Nederland momenteel plaatsvindt; en niets zegt dat vergelijkbare effecten bij andere schalen ook worden gerealiseerd. Ten derde, de onderwijzers van de grote klassen hadden er belang bij om het experiment te laten slagen, omdat zij dan in de toekomst ook met kleinere klassen kunnen werken. Evenzo hadden leerkrachten in kleine klassen reden er een schepje bovenop te doen om het werken in kleine klassen ook voor de toekomst veilig te stellen. Tot slot, er wordt in Tennessee gevonden dat kleine klassen tot betere prestaties leiden, maar volgens de commissie zijn kleine klassen wel noodzakelijk maar niet voldoende voor betere prestaties. Dat maakt de lezer van het rapport er natuurlijk nieuwsgierig naar hoe men in Tennessee de geboden condities heeft benut; of in de termen van de commissie, wat het flankerend beleid is geweest. Helaas, daar wordt met geen woord over gerept.

Deskundigen

De commissie heeft ten behoeve van haar standpuntbepaling een aantal deskundigen geraadpleegd. De waarde van de mening van deze deskundigen is moeilijk te schatten. Waarom juist de gekozen deskundigen zijn geselecteerd en geen anderen wordt niet verantwoord. Wat opvalt is dat hun aantal beperkt is; slechts 17. Verder dat het aantal direct betrokkenen gering is; vier van de 17 zijn verbonden aan een school, waarvan er slechts een (waarschijnlijk) onderwijzer is. En hoewel sommige deskundigen wellicht ook ouder van kinderen in de basisschoolleeftijd zijn, is niemand op die titel als deskundige geraadpleegd. Kinderen zijn helemaal niet gehoord.³

Met de mening van de deskundigen wordt nogal opportunistisch omgegaan: als die mening voor het betoog niet uitkomt, wordt die aan de kant gezet ten gunste van uitkomsten uit buitenlands onderzoek (p.125). Dat kan natuurlijk best als dat buitenlands onderzoek overtuigend is, maar waarom zouden we de deskundigen dan op andere plaatsen wel moeten geloven. Kennelijk valt hun deskundigheid tegen. Bovendien wordt nooit aangegeven hoeveel deskundigen een bepaalde mening aanhangen. Verhullende formuleringen als: "opvatting van deskundigen ..", "Deskundigen vinden ...", worden gebruikt. Hoewel dat er niet expliciet bij staat suggereren die formuleringen dat kennelijk niet alle deskundigen dezelfde mening zijn toegedaan. Ook dat kan natuurlijk best, maar zonder uitgebreide verslaglegging van wat alle geraadpleegde deskundigen ergens van vinden, zijn zulke opmerkingen weinig waard. En zelfs als de geraadpleegde deskundigen wel unaniem zouden zijn, dan is hun aantal te klein, en is bovendien niet vermeld op basis waarvan zij wel en anderen niet geraadpleegd zijn.

Nederlands onderzoek

De derde bron van bewijs voor de stelling dat klassenverkleining gunstig is voor leerprestaties wordt gevormd door een speciaal voor de commissie uitgevoerd onderzoek van Roel Bosker en Joop Hox. Dit onderzoek is als bijlage IV bij het rapport gevoegd. De onderzoekers zijn op basis van het zogenoemde PRIMA cohort nagegaan wat de relaties zijn tussen klassengrootte enerzijds en variabelen als rekentaalvaardigheid, taalvaardigheid, welbevinden van leerlingen, en het functioneren van leerkrachten (werkbeleving en attitude) anderzijds. Wie de hoofdstuktekst leest en de aankondiging van het rapport heeft gevolgd, verwacht dat in deze bijlage het onomstotelijk bewijs wordt geleverd dat kleinere klassen zorgen voor betere scores op deze variabelen. Maar dat is geenszins het geval. Laten we de resultaten maar stuk voor stuk samenvatten.

De effecten van klassengrootte op reken- en taalvaardigheid zijn afzonderlijk berekend voor jaargroep 2 enerzijds en jaargroepen 4, 6, 8 anderzijds. Voor rekenvaardigheid in jaargroep 2 wordt gevonden dat leerlingen in klassen van 5-9, 10-14, 15-19 leerlingen het even goed doen als leerlingen in klassen van 20-24 leerlingen: geen effect van klassengrootte derhalve.⁴ Tweede groepers in groepen van 25-29, 30-34, 35-39 leerlingen scoren slechter dan leerlingen in klassen van 20-24 leerlingen. Maar, het patroon is bepaald niet monotoon: in klassen van 30-34 doen leerlingen het beter dan in klassen van 25-29; en 35-39 scoort even goed als 25-29. De rekenvaardigheid van 2de groepers gaat er dus op vooruit als leerlingen uit een klas van 25-29 naar een kleinere klas van 20-24 gaan. Een deel van de vooruitgang wordt echter ook geboekt als diezelfde leerlingen in een grotere klas van 30-34 worden ondergebracht. Belangrijk verschil is natuurlijk dat klassenverkleining geld kost en klassenvergroting geld bespaart.

De samenhang tussen klassengrootte en rekenvaardigheid bij 4de, 6de en 8ste groepers is nog opmerkelijker. Er zijn geen significante verschillen in rekenprestaties tussen leerlingen in klassen van 5-9, 10-14, 15-19, 20-24 en 25-29 leerlingen. Leerlingen in klassen van 30-34 scoren significant beter (!), en alleen als de klassen meer dan 35 leerlingen bevatten zijn de prestaties slechter. Stuitend (een ander woord is er echt niet) is hoe selectief deze bevindingen in de hoofdttekst zijn opgenomen. "De effecten die gerapporteerd zijn voor jaargroep-2 leerlingen in Nederland zijn behoorlijk groot ... Voor de hogere groepen zijn de bevindingen minder geprononceerd ... Desalniettemin zijn er indicaties dat leerlingen in klassen van 35 of meer enigszins achterblijven in hun leerprestaties vergeleken met leerlingen in kleinere klassen" (p.78). Dat leerlingen in grote (30-34) klassen het beter doen dan leerlingen in kleine (5-29) klassen, wordt verzwegen.

Op het domein van de taalvaardigheid is het beeld niet veel anders. In jaargroep 2 scoren leerlingen uit klassen van 25-29 en 35-39 leerlingen significant lager dan leerlingen in klassen van elke andere omvang. Dus: leerlingen in klassen van 30-34 leerlingen scoren beter(!) dan leerlingen in klassen van 25-29 of 35-39 leerlingen en even goed als leerlingen in klassen met tussen de 5 en 24 leerlingen. Voor de jaargroepen 4, 6 en 8 is alleen de score bij 35-39 leerlingen lager dan in kleinere klassen. Over het hele segment van 5 tot 34 leerlingen is in die jaargroepen geen verschil te constateren. Kortom, ook hier is er feitelijk geen bewijs dat verkleining van klassen met betere scores voor taalvaardigheid samen gaan.

De samenhang tussen klassengrootte en het welbevinden van leerlingen, de werkbeleving van leerkrachten en de attitudes van leerkrachten zijn snel samengevat: er is geen samenhang (p.186-188).⁵

Twee kanttekeningen moeten naar aanleiding van de bijlage van Bosker en Hox nog worden gemaakt. Ten eerste, de analyses zijn uitgevoerd op niet-OVB gebiedsscholen.⁶ Dat betekent dat juist de scholen met de leerlingpopulaties waar we ons het meest zorgen over maken niet in de analyse zijn betrokken. Dit is op zich geen onbelangrijke vaststelling omdat juist op de OVB gebiedsscholen via de gewichtenregeling al veel extra financiële middelen worden ingezet waarmee klassen worden verkleind. Een van de belangrijke conclusies uit de evaluaties van het OVB-beleid is dat deze inzet van middelen weinig effect heeft gesorteerd.⁷

Niet alleen worden de uitkomsten gerapporteerd van de samenhang van klassengrootte en prestaties, maar ook zijn enkele covariaten opgenomen. Vermeldenswaardig is dat voor leerlingen in de jaargroepen 4, 6 en 8 de rekenprestaties significant hoger zijn als de leerkracht meer ervaring heeft.⁸ Dit heeft als belangrijke implicatie dat als klassenverkleining gepaard gaat met inzet van onervaren krachten (als die al te vinden zijn) de resultaten wel eens tegen kunnen vallen; een deel van de (veronderstelde) winst door klassenverkleining wordt - in elk geval voorlopig - teniet gedaan door de inzet van onervaren leerkrachten.

Internationale vergelijking

De laatste bron van bewijs wordt tot slot gevormd door tabellen met internationaal vergelijkende cijfers. In tabel 2.1 van het rapport worden voor 15 landen gegevens gepresenteerd over

de gemiddelde klassengrootte in het basisonderwijs en indices voor de uitgaven per leerling. Alleen Spanje, Nieuw-Zeeland en Ierland hebben gemiddeld grotere klassen dan de 25,7 van Nederland. Qua uitgaven per leerling zijn alleen Hongarije, Spanje, Nieuw-Zeeland en Ierland zuiniger.

1. Hessel Oosterbeek is universitair docent bij de vakgroep algemene economie van de Universiteit van Amsterdam. Dinand Webbink is senior-onderzoeker bij de Stichting voor Economisch Onderzoek (SEO). Thans is hij als onderwijs econoom werkzaam bij de Onderwijsraad.

2. In het vervolg van dit commentaar zullen we de commissie volgen en - zij het met lichte aarzeling - ook spreken over klassen in plaats van groepen.

3. Anekdotisch: een van de auteurs vroeg (zo neutraal mogelijk) aan zijn kind van 6 (die in een groep van 31 leerlingen zit) of hij liever in een kleinere klas zou willen zitten. Hij antwoordde dat hij juist in een grotere klas zou willen zitten omdat hij nu 'niet genoeg ruimte voor zijn kleedje heeft'. Niet het aantal kinderen is het probleem, maar de fysieke infrastructuur.

4. Waarom klassengrootte is gemeten in intervallen van 5, en niet in een tweede- of derdegraads polynoom, is onduidelijk.

5. Alleen leerkrachten in klassen van 5-9 leerlingen hebben een aanzienlijk minder positieve attitude dan andere leerkrachten.

6. OVB staat voor Onderwijsvoorangsbeleid.

7. Sociaal Cultureel Planbureau, Sociaal en Cultureel Rapport 1994, Rijswijk, 1994, blz.310.

8. Voor leerlingen in jaargroep 2 wordt deze samenhang niet gevonden en ook in het domein van de taalvaardigheid is zo'n samenhang er niet.

⁹ Roel Bosker, Klein is fijn, NRC Handelsblad, 30 november 1996, p. 49.

10. Zie: Levin, H.M., Cost-effectiveness and educational policy, Educational Evaluation and Policy Analysis 10 (1988), 51-69.

DE EMPIRISCHE BEWIJSLAST VOOR EXTRA UITGAVEN IN HET PRIMAIR ONDERWIJS

R.J. Bosker^a

Cecil Graham: 'What is a cynic?'

Lord Darlington: 'A man who knows the price of everything and the value of nothing'
(Uit: 'Lady Windermere's fan' van Oscar Wilde)

In hun stuk 'Is verkleining van klassen een goede investering?' raden de Amsterdamse economen Oosterbeek en Webbink de staatssecretaris van onderwijs aan het rapport Klassenverkleining goed te lezen. Dat is een verstandig advies, want als je dat niet doet, zoals zij, dan ontstaan er alleen maar misverstanden.

"Hoewel we er werkelijk van overtuigd zijn dat de argumenten en bewijzen voor verkleining van klassen te zwak zijn om een majeure investering van jaarlijks 923 miljoen gulden te rechtvaardigen, hebben we niet de illusie dat onze stellingname veel invloed zal hebben op de besluitvorming." Dat is de conclusie die Oosterbeek en Webbink verbinden aan hun beoordeling van het materiaal dat door de commissie is aangedragen. De kern van hun kritiek richt zich op de empirische onderbouwing van het advies. Daar zal ik in deze reactie derhalve met name bij stilstaan. Maar vooraf wil ik eerst iets kwijt over het flankerend beleid, waar beide economen wel een heel erg merkwaardige opvatting over hebben, en ik zal aan het einde nog iets zeggen over de goedkopere alternatieven die de critici te berde brengen en de rol van experimenteel onderzoek. Slecht voorbeeld doet goed volgen: de anekdotische bewijsvoering, misleidende bewijsvoering, aparte onjuistheden waarvan mijn weerlegging de lezer maar van de hoofdttekst zou afleiden, stel ik aan de kaak in de voetnoten.

De beide onderwijseconomen vinden het kinderachtig dat de commissie bij haar advies tot klassenverkleining ook aangedrongen heeft op flankerend beleid: "Immers als empirisch wordt gevonden dat klassengrootte geen invloed heeft op de prestaties, dan kan beweerd worden dat het flankerend beleid onvoldoende is geweest om de gecreëerde randvoorwaarden te benutten" (Oosterbeek en Webbink, p. ...). In de sociale wetenschappen, waartoe ook de economie behoort, bestuderen we mensen. Bij klassenverkleining draait het om de leerkrachten, want over klassenverkleining in klassen zonder leerkrachten is niets verstandigs te zeggen. Zien we de klassengrootte als een werkomstandigheid van de leerkracht, dan leidt klassenverkleining tot verbetering van de werkomstandigheden. Zien we de klas als een verzameling leerlingen waarmee de leerkracht onder meer individueel moet interacteren, dan leidt klassenverkleining tot vergroting van de voor individueel-gerichte interactie ter beschikking staande tijd. Ik ken geen *universele* wet die mij iets zegt over de effecten van verbetering van werkomstandigheden van mensen noch over het beschikbaar stellen van meer tijd voor een bepaalde handeling. "Het creëren van een kans leidt niet automatisch tot een doelpunt", is de enige wijsheid die we bezitten. Het STAR-onderzoek laat overigens zien, dat zelfs met weinig flankerend beleid (extra klaslokalen, extra leerkrachten en de dreiging dat klassenverkleining niet door zou kunnen gaan waren kennelijk genoeg), klassenverkleining helpt.

Gaan we nu in op de empirische bewijslast. In het commissierapport (hoofdstuk 4 om precies te zijn) wordt een overzicht gegeven van het onderzoek naar klassengrootte. Het gaat in hoofdzaak om vijf belangrijke bronnen, die alle hebben meegewogen in het uiteindelijke advies:

1. een door Glass c.s. uitgevoerde statistische meta-analyse van al het tot 1979 verrichte klassengrootte-onderzoek en een kritiek daarop;
2. het Amerikaanse Prime Time onderzoek naar klassenverkleining in Indiana;
3. een samenvattend overzicht van de onderwijseconoom Hanushek en een kritiek daarop met een hernieuwde (statistische) meta-analyse van het materiaal;
4. het Amerikaanse experimentele onderzoek naar klassenverkleining in Tennessee;
5. het onderzoek naar klassengrootte en het functioneren van leerlingen en leerkrachten in Nederland van Joop Hox en ondergetekende op basis van de PRIMA-data.

Voor het gemak slaan Oosterbeek en Webbink de eerste drie bronnen maar over in hun overzicht. Dan laten ze kritische geluiden over het STAR-onderzoek horen, welke kritiek niets nieuws bevat: alle punten van kritiek staan in het commissierapport genoemd. Zo staat op p. 35, voordat het relevante onderzoek besproken gaat worden, dat ook een experiment nooit het doorslaggevende bewijs kan leveren (daar is de sociale werkelijkheid toch iets te weerbarstig voor), waarna bij een tussenbalans (p. 43) de kritiek herhaald wordt en ook nog eens op p. 47, als de conclusies geformuleerd worden.¹

Overigens zijn de scholen in het STAR-experiment inderdaad niet aselekt toegewezen aan de experimentele en controleconditie, maar wel degelijk zijn binnen de geselecteerde scholen (die in Amerika heel wat groter zijn dan in Nederland) leerlingen en leerkrachten aselekt toegewezen aan kleine en grote klassen.

Bij de bespreking van de PRIMA-resultaten² vinden Oosterbeek en Webbink dat als de zaak wordt samengevat en de hoofdlijnen worden weergegeven alles herhaald moet worden. Dat is natuurlijk onzin. De voornaamste bevindingen die in de technische bijlage bij het rapport staan³, staan ook in de hoofdtekst van het rapport. Zo staat op p. 42 van de hoofdtekst waar de rekenresultaten van de leerlingen van 8 jaar en ouder besproken worden: "de leerlingen in klassen met 30-34 leerlingen komen zelfs tot de hoogste scores." En wat is er mis met een (uitermate voorzichtig geformuleerde) samenvatting waarin niet alles wordt uitgesplitst naar vakken, maar waar een trend gepresenteerd wordt voor de bovenbouwklassen die zowel voor rekenen als taal geldt: 'desalniettemin zijn er indicaties dat leerlingen in klassen van 35 of

meer leerlingen enigszins achterblijven in hun leerprestaties'. In dit verband is het natuurlijk ook belangrijk om te vermelden dat de commissie voor de bovenbouwklassen geen verkleining heeft voorgesteld.

Gaan we naar de internationale gegevens, die trouwens in het rapport in het geheel niet gebruikt zijn om een causaal verband tussen klassengrootte en leerprestaties aan te tonen - wat Oosterbeek en Webbink de lezer echter wel willen laten geloven. In het commissierapport wordt opgemerkt dat Nederland voor alle sectoren van het onderwijs (basisonderwijs, voortgezet onderwijs en hoger onderwijs) vergeleken met andere OESO-landen relatief weinig geld uitgeeft. Dat is natuurlijk prachtig, want waarom geld over de balk smijten? En het zou natuurlijk nog mooier zijn als die zuinigheid met doeltreffendheid gepaard zou gaan. Waarom immers kritiek leveren op die zuinigheid als de Nederlandse leerlingen goed presteren? Maar als dat laatste niet het geval is, zoals blijkt uit het internationaal vergelijkende IEA-onderzoek naar het begrijpend lezen van 9-jarige leerlingen en uit de in 1993 uitgevoerde evaluatie van het basisonderwijs, dan moeten we ons toch eens achter de oren krabben over die zuinigheid. Aangezien zuinigheid zich vertaalt in grote klassen, vormden de internationale gegevens dus alleen maar een aanleiding om eens goed naar die grote klassen te kijken. Meer niet, want ook deze passages zijn omgeven met waarschuwingen voor al te lichtzinnige interpretaties.

Oosterbeek en Webbink beweren op grond van die internationale gegevens over de samenhang tussen klassengrootte en leerprestaties: 'eerder lijken er aanwijzingen te zijn voor een positieve samenhang.' Grotere klassen, betere prestaties, zou dan de interpretatie van die samenhang zijn. Maar dat is misleidend: het verband tussen leerprestaties en klassengrootte wordt door beide economen geschat *nadat* het effect van de onderwijsuitgaven per leerling op leesprestaties is verdisconteerd. Dat dit laatste effect positief is (hoe hoger de uitgaven per leerling des te beter de leerprestaties in een land) wordt verder niet uitgewerkt. Problematisch is ook, dat klassengrootte en uitgaven per leerling sterk samenhangen: hoe groter de uitgaven per leerling hoe kleiner de klassen. Maar het is sowieso niet verstandig om te pretenderen, dat op basis van gegevens uit slechts vijftien onderling zeer verschillende landen, iets beweerd kan worden over het (oorzakelijk) verband tussen klassengrootte, uitgaven en leerprestaties, en daarom heeft de commissie daarvan dus ook afgezien.⁴

Al het onderzoek samenvattend komt de commissie tot de conclusie dat er voldoende, weliswaar niet 'wettig', maar toch overtuigend bewijs is voor de stelling: klassengrootte doet ertoe. En om tot die conclusie te geraken worden nergens de door de commissie op verzoek van de staatssecretaris gehoorde deskundigen misbruikt om de gaten in de bewijsvoering te dichten. Die zijn voornamelijk geraadpleegd over kwesties als de pedagogisch-didactische aanpak. Het 'wettig bewijs' kan overigens niet geleverd worden als het onderwerp van studie de sociale werkelijkheid is. We kunnen op zijn best de empirische bewijslast op zijn waarde proberen te wegen.⁵

Het is overigens een raadsel waar Overbeek en Webbink de empirische onderbouwing vinden voor de - overigens sympathieke - stelling dat de opleiding tot onderwijzer verlengd en verzaamd moet worden. En dat Amerikaans onderzoek zou laten zien dat klassenverkleining minder kosten-effectief is dan computers is niet waar. Levin (1988, p. 57) schrijft zelf: "The examples are presented to illustrate the issues rather than to provide firm conclusions on cost-effectiveness for each topic". Dat de auteur zelf in computers geloven⁶ (voor 4, 5, 6, 7-jarigen? Probeer je dat eens voor te stellen!), blijkt onder meer uit het feit dat ze verzwijgen dat in Levin's voorbeeldanalyse 'cross-age tutoring' als het meest kosten-effectief uit de bus komt.

Tenslotte is er een *deus ex machina*. Er is een cruciaal experiment denkbaar, volgens beide economen. Dat is interessant, want eerder onderschreven ze ruimhartig de kritische kanttekeningen die de commissie bij het Amerikaanse STAR-experiment plaatste, wat hen mede

ertoe bracht de onderbouwing van het advies 'Klassenverkleining' zwak te noemen. Hun plan: verdeel Nederland in regio's, voer in sommige regio's een klassenverkleining door en in andere niet. Een geniaal plan, want in Twente komen ouders en leerkrachten er vast niet achter dat de klassen in de randstad kleiner zijn. En in de randstad zal niemand zich afvragen waarom alle scholen plotseling zo maar extra leerkrachten krijgen. Kortom: voor dit experiment gelden natuurlijk dezelfde bezwaren als voor het STAR-experiment. Maar nu blijkt ook dat Oosterbeek en Webbink het rapport slecht gelezen hebben: zij willen de *maximale* klassengrootte en niet de *gemiddelde* klassengrootte per regio laten variëren. Dat maximum is echter maar een ondergeschikt element uit het advies. Waar het primair om draait is dat er volgens het advies van de commissie meer leerkrachten moeten komen voor de 4- tot en met 7-jarige leerlingen, en wel zo dat klassen in de onderbouw nooit veel groter dan 20 leerlingen hoeven te zijn. De gemiddelde klassengrootte voor de onderbouw kan dan dalen van 25.7 nu, naar 18 leerlingen in 2001. Vervolgens wordt een wettelijke maximering voorgesteld: de onderbouwklassen mogen niet meer dan 25 en de bovenbouwklassen mogen niet meer dan 34 leerlingen bevatten. Beide maatregelen geven een signaal af aan scholen, dat klassenverkleining het meest succesvol kan zijn als het wordt ingezet voor de allerjongste leerlingen. Als scholen met de beschikbare formatie verstandig omspringen, is dus een enorme reductie van de grootte van de onderbouwklassen te realiseren, een zodanige reductie dat de beoogde effecten realiseerbaar moeten zijn.

"Een onzalig idee" hebben Oosterbeek & Webbink het voorstel tot klassenverkleining elders genoemd (NRC, 1996). Het is niet door de commissie gezegd dat het een zaligmakend idee is: het biedt een kans die vervolgens door het onderwijs verzilverd moet worden. En er moet voor gewaakt worden dat dat ook inderdaad gebeurt, en dat het gemeenschapsgeld niet besteed wordt om leerkrachten in kleine klassen een vakantiegevoel te geven. Daartoe worden voorstellen gedaan voor flankerend beleid, zoals het aan scholen opleggen van de verplichting rekenschap af te leggen over de opbrengsten van het onderwijs. Het is voorstelbaar dat scholen niet precies weten hoe ze de kans die geboden wordt het beste kunnen verzilveren: de kleinste klassen voor de kleinste kinderen; of de kleinste klassen voor groep 2 en 3 als het rekenen en lezen goed op gang moet komen; of een leerkracht vrijroosteren die waar nodig individueel hulp kan bieden; of gewoon toch maar alle klassen in alle jaargroepen in onder- en bovenbouw op ongeveer 22 leerlingen? Zo'n experiment (zie: Bosker & Meijnen, 1996) zou zijn waarde kunnen bewijzen in het zo goed mogelijk bepalen van de mate waarin de verschillende varianten effectief zijn.

a. Roel Bosker is als universitair hoofddocent verbonden aan de vakgroep Onderwijsorganisatie en -management van de Faculteit der Toegepaste Onderwijskunde (Universiteit Twente) en was lid van de commissie Groeps-grootte (beter bekend als de commissie-Van Eijndhoven).

1. Bij STAR was er ook nog een extra steekproef van scholen (zoals overigens ook bij Prime Time) om na te kunnen gaan of de grotere klassen in de onderzoeksscholen het slechter gingen doen (i.p.v. dat de kleinere klassen het beter gingen doen). Het antwoord luidt ontkennend.

2. In het door Bosker & Hox (1996) uitgevoerde onderzoek voor de samenhang tussen klassengrootte en het functioneren van leerlingen en leerkrachten is gebruikgemaakt van de referentiesteekproef uit het PRIMA-onderzoek. Dit is een voor de gehele populatie van Nederlandse basisscholen representatieve steekproef. De gebiedsscholen van het Onderwijs Voorrang Beleid zijn daarin naar rato vertegenwoordigd en niet zoals Oosterbeek en Webbink beweren door ons verwijderd.

3. Waarom Oosterbeek en Webbink het effect van klassengrootte in tweede- of derdegraads polynomen gemiddeld zouden willen zien (zie hun noot 3) is onduidelijk.

4. De gebruikte gegevens zijn afkomstig van de IEA en van de OECD. De oorspronkelijke IEA-gegevens hebben betrekking op 27 landen, maar voor slechts 15 landen zijn via de OECD gegevens over onderwijsuitgaven bekend. Voor die 25 landen rapporteren Ross & Postethwaite (1994, p.130) een negatieve correlatie tussen de

leerling/leerkracht-ratio en het landgemiddelde leesniveau van -0.61: Hoe gunstiger (kleiner) de leerling/leerkracht-ratio hoe hoger de prestaties.

5. Het was natuurlijk wel makkelijker geweest om bewijs te voeren op basis van anekdotes, zoals Oosterbeek en Webbink doen ten aanzien van het belang van de fysieke infrastructuur.

6. Dat delen zich overigens met een andere criticus van het advies, namelijk De Klerk (zie: Burgers, 1996).

Literatuur

Bosker, R.J. 'Klein is Fijn'. *NRC-Handelsblad*, 30-11-1996, p. 49.

Bosker, R.J. & Meijnen, G.W. (1996). *Organisatie van klassen. Het nationale OK-experiment*. Enschede/Amsterdam: UT/UvA.

Bosker, R.J. (1997). Het einde van de klassenstrijd? Achtergronden bij de discussie over klassengrootte in het basisonderwijs. *Pedagogische Studiën*, 74, 3, 210-227.

Burgers, L. (1996). 'Minder leerlingen niet altijd beste oplossing.' *Nieuwsblad van het Noorden*, 18-11, 1996, p. 6.

Commissie Kwalitatieve Aspecten van Groeps grootte in het Basisonderwijs (1996). *Klassenverkleining. Advies over de betekenis van klassenverkleining voor beter basisonderwijs*. Den Haag: SDU.

Levin, H.M. (1988). Cost-effectiveness and educational policy. *Educational Evaluation and Policy Analysis*, 10, 1, 51-69.

Oosterbeek, H. & Webbink, D. 'Hoe fijn is klein', *NRC-Handelsblad*, 23-11-1996, p. 45.

Ross, K.N. & Postelthwaite, T.N. (1994). Differences among countries in school resources and achievement. In: W.B. Elley (Ed.) *The IEA study of reading literacy: Achievement and Instruction in thirty-two school systems*. Oxford: Pergamon. (pp. 123-148).

GEWOGEN LOTING IN COMBINATIE MET RECHTSTREEKSE TOELATING VANAF EXAMENCIJFER 8

Henk A.L. Kiers en Willem K.B. Hofstee
Rijksuniversiteit Groningen

In de beleidsnotitie Toelating tot opleidingen met een numerus fixus van 12 mei 1997 stelt de Minister van Onderwijs, Cultuur en Wetenschappen de Tweede Kamer het volgende voor: gegadigden voor numerus-fixus-studierichtingen met gemiddeld eindexamencijfer 8 en hoger worden rechtstreeks geplaatst; voor de overige gegadigden wordt gewogen loting toegepast met als gewichten:

Eindexamencijfer:	<6.5	≥6.5	≥7.0	≥7.5
Lotingsgewicht:	0.75	1.00	1.50	2.25

Deze nieuwe gewichten lopen sneller op dan de oude, die er op een vergelijkbare schaal als volgt uitzien:

0.83	1.00	1.25	1.56
------	------	------	------

Argument voor het kiezen van de nieuwe gewichten is het 'mitigeren' van het verschil in toelatingskans tussen degenen met cijfer 7.5 en degenen met 8 of hoger.

Een bezwaar tegen vaste gewichten is dat hun feitelijke uitwerking afhangt van de toelatingsratio (het aantal plaatsen in verhouding tot het aantal gegadigden). Enerzijds kan het bij gebruik van de nieuwe gewichten, in geval van een hoge toelatingsratio, gemakkelijk voorkomen dat gegadigden met cijfer <8 toch rechtstreeks worden geplaatst, hetgeen niet de bedoeling was: de beleidsnotitie bevat een argumentatie om de grens voor rechtstreekse toelating niet beneden de 8 te leggen. Anderzijds kan, ook bij de sneller oplopende nieuwe gewichten, de toelatingskans voor een studie met een zeer lage toelatingsratio (bijv. diergeneeskunde) voor een gegadigde met examencijfer 7.9 onevenredig laag zijn, hetgeen wederom niet strookt

met de bedoelingen. Meer in het algemeen is de gekozen combinatie van rechtstreekse toelating en vaste lotingsgewichten niet coherent.

Een oplossing voor dit probleem ligt besloten in het algemene model dat door Hofstee en Kiers (1997) werd gepresenteerd. In dat model zijn de lotingsgewichten eveneens een versnelde (nl. exponentiële) functie van het examencijfer; de snelheid waarmee de gewichten oplopen kan echter worden gevarieerd door aanpassing van de centrale parameter in het model. Wanneer vervolgens aan dit model de randvoorwaarde wordt opgelegd dat de toelatingkans de waarde 1 bereikt bij examencijfer 8, resulteert een unieke en verrassend eenvoudige oplossing:

Eindexamencijfer:	<6.5	≥6.5	≥7.0	≥7.5
Inlotingskans:	p^4	p^3	p^2	p

Het bewijs wordt gegeven in de Appendix bij dit artikel. De waarde van p kan via standaardtechnieken worden gevonden uit:

$$n_6 p^4 + n_{612} p^3 + n_7 p^2 + n_{712} p = N - n_{8+},$$

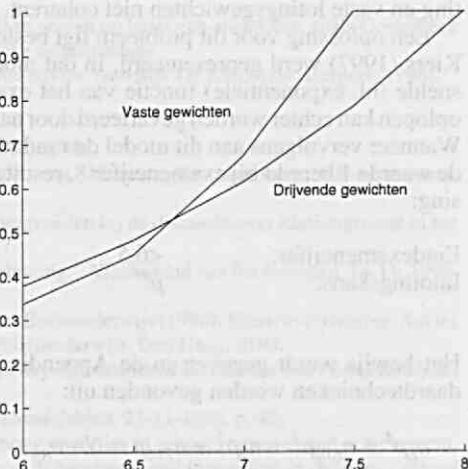
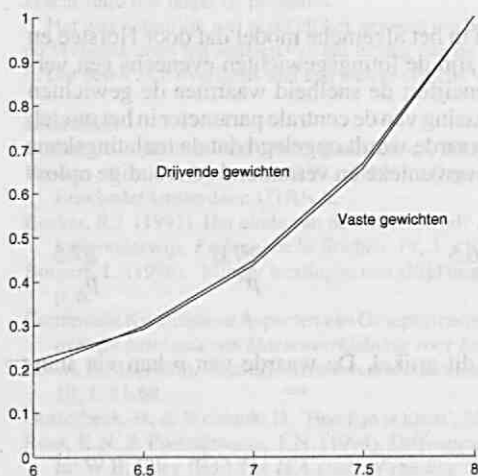
waarbij n staat voor het aantal gegadigden in de betreffende lotingsklasse, en N voor het totaal aantal beschikbare plaatsen.

Hiermee zijn gewogen loting en rechtstreekse toelating op één noemer gebracht, met als gevolg dat het niet meer 'wringt' op de grens tussen beide. Als argument tegen de gepresenteerde oplossing zou kunnen worden aangevoerd dat de lotingsgewichten niet meer vooraf kunnen worden gepubliceerd en dat het systeem daardoor minder doorzichtig wordt voor de betrokkenen. Dat argument is ongeldig. Voor de betrokkenen zijn de lotingsgewichten als zodanig oninteressant. Van belang zijn hun toelatingkansen (gegeven hun verwacht eind-examencijfer); die kansen zijn in alle gevallen mede afhankelijk van de toelatingsratio, en kunnen alleen worden ingeschat op basis van gegevens uit vorige cohorten.

Simulatie

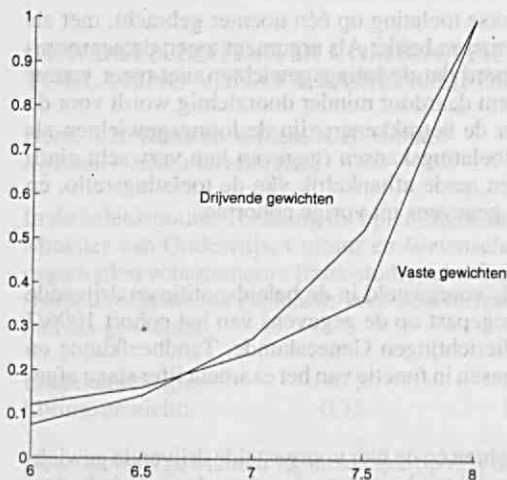
Beide modellen - vaste lotingsgewichten zoals voorgesteld in de beleidsnotitie en drijvende gewichten zoals hier voorgesteld - werden toegepast op de gegevens van het cohort 1996/7 (gegadigden per 25 september) voor de studierichtingen Geneeskunde, Tandheelkunde en Diergeneeskunde. De resulterende inlotingskansen in functie van het examencijfer staan afgebeeld in figuur 1 t/m 3.

Bij Geneeskunde 1996/7 komen de vaste gewichten en de hier voorgestelde drijvende gewichten vrijwel op hetzelfde neer; ook de vaste gewichten beantwoorden daar dus aan de bedoelingen. Bij verschuiving echter van de toelatingsratio - bijvoorbeeld voortvloeiend uit maatregelen zoals beperking van het aantal herkansingen - zal dat niet meer het geval zijn. Bij Tandheelkunde, met een hogere toelatingsratio, zouden de vaste gewichten hebben geresulteerd in rechtstreekse toelating vanaf cijfer 7.5; de kanskromme vertoont een scherpe knik. Bij Diergeneeskunde treedt bij de vaste gewichten het omgekeerde effect op: daar keldert de toelatingkans van iemand die net geen 8 heeft gehaald naar 0.36. Drijvende gewichten trekken, binnen de beperkingen van de lage toelatingsratio, die kans op tot 0.52. De voorbeelden illustreren dat de vaste gewichten, die blijkbaar zijn gebaseerd op de situatie bij Geneeskunde in de afgelopen jaren, te rigide zijn om andere studierichtingen en toekomstige ontwikkelingen aan te kunnen.



Figuur 1 (linksboven): Toelatingskans Geneeskunde in functie van eindexamen-cijfer

Figuur 2 (rechtsboven): Toelatingskans Tandheelkunde in functie van eind-examencijfer



Figuur 3 (links): Toelatingskans Dier-geneeskunde in functie van eindexamen-cijfer

Conclusie

Politieke overwegingen die ten grondslag liggen aan de keuze voor een combinatie van rechtstreekse toelating en gewogen loting bleven hier buiten beschouwing. We hebben ons beperkt tot het presenteren van een model dat, beter dan de beleidsvoornemens van de minister, de bedoelingen incorporeert die eraan ten grondslag lagen. De eerste stap daarbij (zie Hofstee & Kiers, 1997) bestond uit het vangen van de 'Vermaat-gewichten' in een algemener, exponentieel model. De tweede, hier gepresenteerde stap bestaat uit het toespitsen van dat model op de situatie waarin rechtstreekse toelating vanaf een bepaald examencijfer (in casu 8) wordt beoogd. Dat blijkt te resulteren in een coherente en transparante oplossing.

Verwijzing

Hofstee, W.K.B., & Kiers, H.A.L. (1997). Een algemeen model voor loting en selectie bij numerus clausus. *Tijdschrift voor Onderwijsresearch*, 22, 81-85.

Appendix: Afleiding van het drijvende-gewichtenmodel

Het algemeen model voor de bepaling van lotingsgewichten (Hofstee & Kiers, 1997) luidde:

$$w_X = e^{h(X-X_g)+C} \quad (1)$$

met w_X het gewicht dat hoort bij klasse X (hier: $X = 6, 6\frac{1}{2}, 7, 7\frac{1}{2}$) en h de hellingsparameter van het model; de constanten X_g en C zullen zo dadelijk blijken weg te vallen en blijven hier onbesproken.

Lotingsgewichten fungeren als tellers en noemers in kansverhoudingen, dat wil zeggen: als bijvoorbeeld een gegadigde met cijfer 8 een bepaalde toelatingskans heeft, is bij cijfer 7 $\frac{1}{2}$ de toelatingskans $w_{7\frac{1}{2}}/w_8$ maal die kans. Voor het geval van rechtstreekse toelating vanaf cijfer 8 (kans 1) is de toelatingskans bij cijfer 7 $\frac{1}{2}$ dus gelijk aan $p = w_{7\frac{1}{2}}/w_8$. Substitutie van uitdrukking (1) levert op:

$$p = e^{-1/2h} \quad (2)$$

Op dezelfde manier vinden we de inlotingskansen voor de overige lotingsklassen:

cijfer:	<6.5	≥6.5	≥7.0	≥7.5	8+
kans:	e^{-2h}	$e^{-1/2h}$	e^{-h}	$e^{-1/2h}$	e^0
ofte wel:	p^4	p^3	p^2	p	$p^0=1$

Boekbesprekingen

J.B. Kuhlemeier (1996)

Taalvaardigheid, taalactiviteiten en taalattituden. Proefschrift UvA. Arnhem: CITO

Inleiding

Er kan gezegd worden van peilingsonderzoek wat men wil, maar niet dat het tot weinig studies leidt. Voor het vak Nederlands legde Wesdorp (1987 †) de basis voor een groot aantal onderzoeken waarin het meten van taalvaardigheid (proefschriften: (Wesdorp, 1974; Blok, 1987; De Gloppe, 1987, Van den Bergh (1987), Oostdam (1991), Van Gelderen (1991), Schoonen (1993)) en later ook attituden en onderwijsaanbod (proefschrift Otter, 1993, studies van Overmaat 1992, Wesdorp, Van den Bergh, Bos, Hoeksma., Oostdam, Scheerens, & Triesscheijn, (1986)). Het proefschrift van Hans Kuhlemeier (HK), behoort tot deze Wesdorptraditie, al is hij niet via directe participatie opgeleid in deze Amsterdamse onderzoeksgroep; zijn co-promotor Van den Bergh is dat wel, en de promotor Mellenbergh was van menig bovengenoemde onderzoeker promotor.

HK's bewijs van 'academisch middenstandsdiplooma' (zie HK's Voorwoord) bevat een vier-tal validatiestudies, gebaseerd op data, die zijn verzameld in een studie naar de haalbaarheid van nationale peilingen in het voortgezet onderwijs (derde leerjaar). De proefpeiling vond eind jaren tachtig plaats, en in de eindrapportage werd geconcludeerd dat peilingen in het voortgezet onderwijs haalbaar zijn. Opzettelijk koos men destijds voor het vak Nederlands: als peilingen bij dit weerbarstig te meten vak mogelijk zouden zijn, dan zijn ze voor elk schoolvak mogelijk, was de redenering.

Peilingsonderzoek richt zich op een beschrijving van de output en de input van het onderwijs. Bij de beschrijving van de output wordt een onderscheid gemaakt tussen onderwijsleerresultaten (veelal gelijkgesteld met cognitieve onderwijsleerresultaten of prestaties) en affectieve doelen (veelal niet benoemd als onderwijsleerresultaten). Van de input wordt vooral het onderwijsaanbod beschreven, naar inhoud, werkwijzen en toetsvormen. Enkele algemene school-, docent- en leerlingkarakteristieken worden eveneens verzameld. In Nederland worden onderwijspeilingen in opdracht van de Minister van Onderwijs, Cultuur en Wetenschappen uitgevoerd door het Instituut voor Toetsontwikkeling. HK was betrokken als medewerker van het CITO bij de proefpeiling Nederlands in het voortgezet onderwijs.

In deze proefpeiling werden de taalvaardigheden van leerlingen gemeten (schrijven, lezen, luisteren en spreken) met veelal functionele taaltaken: taken die leerlingen in hun dagelijkse taallevens (zouden) kunnen moeten verrichten. Daarnaast werd aandacht besteed aan niet-cognitieve doelen van het taalonderwijs, zoals schrijfatteitude, schrijfangst en spreekangst, enerzijds omdat affectieve doelen op zichzelf een belang vormen, anderzijds omdat er een relatie verondersteld wordt tussen attituden en prestaties. Om tot een goede beschrijving van het aanbod te komen werden vragenlijsten voorgelegd aan docenten Nederlands, maar ook aan leerlingen (lijsten van taalactiviteiten). Omdat leerlingen niet alleen bij Nederlands oefenen in de taalvaardigheid Nederlands, maar ook bij andere vakken, en ook buiten de school, werden ook docenten Engels en Geschiedenis bevraagd naar de frequenties van taalactiviteiten, en werd aan leerlingen een lijst voorgelegd van activiteiten buiten de school. Enkele van deze instrumenten, maar in feite de hele exercitie van peilingsonderzoek bij het vak Nederlands in het VO, wordt in dit proefschrift door een ingewijde onder de loep genomen. Ik vrees na bestudering van dit degelijke proefschrift dat de conclusie uit het haalbaarheidsonderzoek te positief was. Peilingsonderzoek zoals in het haalbaarheidsonderzoek uitgevoerd is letterlijk invalide onderzoek: de vergelijking tussen schooltypen in het voortgezet onderwijs gaat mank,

omdat het instrumentarium onvoldoende tegemoet komt aan de verschillen tussen de opleidingstypen, en de heterogeniteit van de leerlingenpopulatie.

Korte beschrijving van het proefschrift

Het proefschrift behandelt twee belangrijke vragen. De eerste vraag betreft de efficiëntie van de omvangrijke en gedetailleerde meting en rapportage van de taalprestaties, de taalactiviteiten op school en in de vrije tijd, en de taalattitudes. Een belangrijke basisbeslissing bij de opzet van peilingsonderzoek is immers de rapportage-eenheid: hoe gedetailleerd of globaal moet/kan gerapporteerd worden? In de peiling zijn veel kostbare instrumenten betrokken (veel testresultaten zoals door leerlingen geschreven teksten moesten door getrainde beoordelaars beoordeeld worden), en de gedetailleerdheid van de rapportage (per schrijfo opdracht, per beoordelingsaspect) doet de auteur de welhaast retorische vraag stellen; "Of dit wel zo efficiënt is, valt niet zonder meer te zeggen." (p. 18). De tweede vraag betreft een verkenning van 'enkele mogelijkheden en problemen van het gebruik van peilingsdata in het kader van onderzoek naar schooleffectiviteit.' (p. 18). Deze exploratie moet aanbevelingen opleveren over de wijze waarop peilingsonderzoek bij kan dragen aan verschillen tussen scholen in onderwijseffectiviteit.

Het boek bevat vier empirische hoofdstukken. In hoofdstuk 2 wordt de validiteit van de meting van taalvaardigheid onderzocht, in het derde hoofdstuk de validiteit van de meting van taalactiviteiten op school en in de vrije tijd en in hoofdstuk 4 de validiteit van de meting van attitude jegens schrijven en schrijf- en spreekangst.

In het laatste empirische hoofdstuk wordt onderzocht of peilingsonderzoek een valide bijdrage kan leveren aan schooleffectiviteitsonderzoek. Ondanks een niet voor dit doeleinde optimaal geordende dataverzameling, - de auteur is zich hier goed van bewust -, weet de auteur vrijwel het maximale uit de dataset te halen. Hierdoor levert het proefschrift een bijdrage aan het inzicht in de meting van taalvaardigheid, en biedt de auteur discussiebijdragen die uitdagen tot nader empirisch onderzoek en theoretische verdieping.

Hoofdstuk 2 (ca. 60 pagina's) is geheel gewijd aan een secundaire analyse van de vaardigheidsmetingen. Het systematische patroon van dit hoofdstuk staat model voor de twee volgende hoofdstukken. Onderzocht worden (1) de interne structuur van de meting, (2) de relaties tussen de vier vaardigheden, (3) de bijdrage van opdrachtfactoren aan de totale variantie en (4) de generaliseerbaarheid over opleidingstypen (Vwo/HAVO, MAVO, LBO).

De LISREL-analyses voor luisteren, lezen, spreken en schrijven tonen aan dat in geen van de gevallen er sprake is van een één-factor model. Telkens paste het model waarin gecorrleerde factoren waren opgenomen het beste: voor lezen, luisteren en spreken waren dat opdrachtfactoren (leerlingen kregen taken voorgelegd die uit verscheidene, meestal functionele, communicatieve opdrachten bestonden); voor schrijven zowel een opdrachtfactor als een beoordelingsaspectfactor (geschreven teksten leveren zes scores op, die verschillende kwaliteiten van de tekst weerspiegelden). Een interessante uitstap maakt HK dan om binnen leesvaardigheid te onderzoeken in hoeverre het onderscheid 'functionele leesvaardigheid' en 'conventionele leesvaardigheid' aanwezig is: empirisch blijkt dat onderscheid houdbaar, al is de correlatie wel erg hoog (.87). Een andere nevenvraag over de beoordelingsaspecten van schrijfvaardigheid, leverde het antwoord op dat de onderzoekers er niet in geslaagd waren empirisch onderscheidbaar te meten: de aspecten globale kwaliteit, inhoud, stijl, en organisatie van een tekst, gemeten aan een en dezelfde tekst, door (deels) verschillende beoordelaars, correleren perfect (na correctie voor onbetrouwbaarheid en na correctie voor onbedoeld meegemeten taakeffecten).

Uit de analyses van de verschillende modellen die de samenhang tussen de vier vaardigheden beschrijven, blijkt het model waarin zowel de factor receptief-productief en mondeling-schriftelijk zijn opgenomen het beste te passen. De correlaties tussen de vaardigheden

zijn echter vrij hoog (range .64-.93). In de derde stap van de validiteitsanalyse gaat HK na in hoeverre er naast een algemene taalvaardigheidsfactor ook een taakfactor gepostuleerd dient te worden. Die taakeffecten zijn er wel degelijk: bijna 30 tot meer dan 40% van de variantie bij lezen, luisteren en spreken wordt verklaard door taakeffecten: het doet er dus nogal wat toe uit welke taken de peiling samengesteld wordt. Waarop schrijfproducten worden beoordeeld, lijkt er minder toe te doen (de beoordelingsaspecten verklaren 15% van de variantie). Uit een nadere analyse blijkt de grootte van deze taakeffecten nog eens te verschillen per steekproef en per instrument. De laatste stap betreft de generaliseerbaarheid van de structuur van taalvaardigheid over opleidingstypen. De samenhang tussen de vier vaardigheden blijkt in de drie onderscheiden opleidingstypen (VWO/HAVO, MAVO, LBO) niet te verschillen; wel verschillen de betrouwbaarheden van de geobserveerde scores, de taakeffecten en de mate waarin opdrachten een beroep doen op taalvaardigheid. Het is dus niet toegestaan om de scores van de drie opleidingstypen met elkaar in verband te brengen.

Uit het tweede empirische hoofdstuk, de validatiestudie van de twee leerlingenvragenlijsten naar taalactiviteiten op school (bij Nederlands en bij andere vakken) en buiten school (55 pagina's) blijkt min of meer hetzelfde patroon als bij de vaardigheden: met de interne structuur is niets mis, maar overige validiteitsaanwijzingen zijn zwak of afwezig. Zo wordt onderscheid tussen activiteiten bij Nederlands en bij andere schoolvakken niet aangetoond, en zijn de gegevens van de vragenlijsten niet generaliseerbaar over de opleidingstypen; bij VWO/HAVO wordt iets anders gemeten met deze vragenlijsten dan bij LBO. Als validiteitsaanwijzing acht HK het bevestigen van een relatie tussen activiteiten en vaardigheden van belang. Inderdaad blijken (gerapporteerde frequenties) van activiteiten samen te hangen met vaardigheden, maar de verklaarde varianties zijn klein, en diffuus: zo worden er nogal wat relaties geobserveerd tussen niet overeenkomstige paren, zoals de activiteit spreken met de vaardigheid lezen.

Het derde empirische hoofdstuk gaat over de meting van schrijffattitude, schrijfangst en spreekangst. Deze metingen dienden de niet-cognitieve onderwijsopbrengstbeschrijving. Ook dit onderzoek vertoont het patroon van de resultaten in de twee eerdere hoofdstukken: de interne structuur van de instrumenten voldoet, maar andere indices voor validiteit ontbreken: de generalisatie over opleidingstypen kan niet gemaakt worden, en van de drie vragenlijsten vertoont er slechts een zwakke relatie met prestaties (schrijfangst voorspelt 9% van de variantie in schrijfprestaties).

Hoofdstuk 5 heeft als titel 'De factorstructuur van schrijfvaardigheid: een toepassing van multilevel analyse'. Het behandelt de onderzoeksvraag in hoeverre peilingsonderzoek een bijdrage kan leveren aan schooleffectiviteitsonderzoek. HK demonstreert de mogelijkheden voor de schrijfvaardigheid. Vaststellen van prestatieniveau is één ding, vaststellen van onderwijsaanbod een tweede, maar vaststellen van de relatie (lieftst nog: causale relatie) is waar het beleid naar vraagt en de onderwijskundige theorie naar hunkert. HK neemt vooraf al stelling: de mogelijkheid tot dieptestudies naar in peilingsstudies geconstateerde verschillen tussen scholen is gering wegens de onbetrouwbaarheid van schoolscores, gezien het afnameontwerp van peilingstudies (een klein aantal leerlingen per school, een gemakkelijk te verhelpen probleem), en de statische aannames in peilingsonderzoek (unilevel, met overschatting van het aantal vrijheidsgraden door het clustereffect; invariantie van de factorstructuur over niveaus). De hiërarchische structuur van de data in peilingsonderzoek (leerlingen binnen klassen, klassen binnen scholen) moet in de data-analyse gereflecteerd worden om overschatting van vrijheidsgraden wegens intraclasscorrelaties (clustereffecten) tegen te gaan en tot zuiverder schattingen te komen. De belangrijkste uitkomst van dit hoofdstuk is wel dat de schoolgemiddelden voor schrijfvaardigheid onvoldoende betrouwbaar zijn gemeten (.60 voor stelvaardigheid, .22-.37 voor schrijftechniek) en dat scholen sterk verschillen van elkaar in het niveau van stelvaardigheid (20% van de variantie wordt

verklaard door schoollidmaatschap); de verschillen tussen leerlingen binnen scholen zijn groter (ca. .35).

Evaluatie

Het proefschrift van Hans Kuhlemeier op DAP-niveau (Degelijk Arnheems Peil): zorgvuldige dataverzameling, zorgvuldige analyses, helder gestructureerd en gerapporteerd, adequate en bondige theoretische inbeddingen, en, heel belangrijk, zelfkritisch. Van tijd tot tijd plaats- te ik een vraag of opmerking in de kantlijn, en dan bleek de auteur op die vraag terug te komen in de volgende paragraaf 'discussie'. Ook al is het dataontwerp niet optimaal voor een validatiestudie, als secundaire analyse voldoet de studie prima. Laten ik zeggen, in de termen van dit proefschrift: interne validiteit is o.k. Maar soms blijft de auteur voor mijn gevoel wat te dicht op de gegevens, is HK wat te voorzichtig of wat te beknopt in de aanbevelingen. De prangende vraag die de lezer stelt na dit proefschrift is dezelfde als de vraag die ten grondslag lag aan het onderzoek waaruit de dataverzameling voortkwam: is peilingsonderzoek haalbaar? Moet het oorspronkelijke antwoord ('ja!') na alles wat dit proefschrift oplevert herzien worden? HK gaat deze vraag als medewerker bij een peilingsonderzoeksinstituut wijselijk uit de weg. Ik ben sterk geneigd, op basis van deze studie, te zeggen: nee, nog niet. We weten niet wat we meten, de validiteit van taalvaardigheidsmetingen, aanbodsmetingen, attitude-metingen ligt nog te zeer onder vuur. Ik loop deze punten hieronder langs.

Het meten van taalvaardigheid

In de discussie van hoofdstuk 2 gaat HK vooral in op de taakeffecten. Hij geeft, m.i. terecht, in overweging het oude standpunt af te wijzen dat taakeffecten beschouwd zouden moeten worden als ruis in de meting van de vaardigheid. Te lang wordt de gangbare indeling in de vier vaardigheden spreken, schrijven, lezen en luisteren gebezigd zonder een theoretische onderbouwing en analyse van overeenstemming tussen en verschillen in deelvaardigheden, communicatieve functies (zoals argumenteren, rapporteren, beschouwen, beschrijven), die andere onderscheidingen als receptief-productief en mondeling-schriftelijk weleens in de schaduw zouden kunnen stellen. In de moderne didactiek gaat het steeds minder om de vier hoofdvaardigheden dan om belangrijk geachte taken. Lezing van de kerndoelen voor het basisonderwijs en het voortgezet onderwijs doet vermoeden dat de commissies voor Nederlands hun best hebben gedaan meer vanuit taken te denken en te ordenen dan vanuit de aloude vierdeling. Een theoretisch gefundeerde typologie van taaltaken kan een belangrijke bijdrage leveren aan het taalonderwijs, het taalonderwijsbeleid en het peilingsonderzoek. HK wijst er in ander verband op dat het valideren van zo'n theorie erg begroetelijk is. We zullen er toch aan moeten geloven, vrees ik.

Op de nogal miserabele kwaliteit van de meting van schrijfvaardigheid (de beoordelingsaspecten correleerden perfect) had HK dieper kunnen ingaan. Het is voor mij onbegrijpelijk dat telkens maar weer bij het meten van vaardigheden als schrijven en spreken zoveel mogelijk verschillende aspecten van schrijf- en spreekproducten beoordeeld moeten worden: telkens blijkt dat dat niet kan. Het maakt een beetje een onbeholpen indruk: we weten niet wat schrijfvaardigheid is, maar als we allerlei aspecten mee laten tellen zullen we wel in de buurt komen. Is het gek dat kenmerken en oordelen hoog correleren als ze gemeten worden aan dezelfde tekst? In plaats van veel verschillende aspecten per opdracht te meten, zouden veel opdrachten voorgelegd moeten worden, die ieder op een voor die opdracht karakteristiek kenmerk beoordeeld wordt (bijvoorbeeld afhankelijk van de communicatieve functie die de tekst moet vervullen: een verslag, een argumentatie, een beschouwing) of op een globaal, overkoepelend kenmerk als de communicatieve effectiviteit van de tekst.

Het meten van taalactiviteiten door leerlingen als aanbodmeting

In de discussie van hoofdstuk 3 wijst HK er terecht op dat in feite geen taalactiviteiten zijn gemeten, maar gepercipieerde taalactiviteiten, en argumenteert op basis van cijfermateriaal

(p. 138) dat leerlingen niet goed in staat zijn om verschillen in frequenties van taalactiviteiten te onderkennen. Fundamenteeler vind ik de vraag, waarom je leerlingen dergelijke lijsten zou voorleggen in peilingsonderzoek. Vragenlijsten zijn sowieso al ondingen voor het meten van onderwijsaanbod (zie Otter, 1993). Bovendien, het gaat er toch niet om vast te leggen wat het door leerlingen gepercipieerde curriculum is? Men zou dat hooguit in zeker opzicht als een onderwijsopbrengst kunnen bestempelen, maar toch niet als een onderwijsaanbodindex? En mocht het dat wel zijn, dan rijst de vraag hoe men zulke onderwijsopbrengstcores moet waarderen. Ook al is de vragenlijst een intern valide instrument, de relevantie ervan in peilingsonderzoek is geheel andere koek, ook gezien de haken en ogen die de auteur zelf aanvoert in de discussie.

Het meten van attitudes in peilingsonderzoek

Het lijkt lovenswaardig dat peilingsonderzoek zich niet beperkt tot louter een prestatie-meting. Maar aan het meten, maar vooral het waarderen van bijvoorbeeld schrijffattitudescores of spreekangstcores, zijn nogal wat problemen verbonden. Wat te doen met een uitslag dat 5% van de leerlingen spreekangst vertoont bij het houden van een spreekbeurt, of dat 20% van de leerlingen een uitgesproken hekel heeft aan het schrijven van opstellen of functioneel communicatieve teksten? Wat vinden we daarvan? Teveel spreekangstige kinderen? Valt het wel mee? Het wordt (pas?) interessant als er een relatie gelegd kan worden tussen aanbodenmerken en angst en als die relatie van te voren gespecificeerd kan worden: levert een bepaald complex van lesinhouden en werkvormen, zoals het in het kader van het vergroten van schrijfplezier veelgeprezen 'vrije schrijven' in het schrijfonderwijs, andere, positievere attitudes op? Een ander probleem is dat er steeds maar weer onderzoekers zijn die een relatie willen vaststellen tussen attitudes en prestaties; veelal zijn ze uitsluitend op zoek naar eenvoudige lineaire verbanden. Ook in deze validatiestudie wordt dat verband onderzocht, en wederom moet de onderzoeker besluiten dat er geen verband vastgesteld kan worden, en dat daarom een validiteitsaanwijzing voor de meting van attitudes in deze zin ontbreekt. (Overigens heeft HK naast de gerapporteerde verbanden ook naar non-lineaire verbanden gezocht, en evenmin gevonden; mondelinge mededeling HK.)

In de discussie van hoofdstuk 4 roert HK een paar interessante punten aan, waaronder de veronderstelde relatie tussen attitudes en prestaties. Dat HK geen verband aantreft tussen schrijffattitudes en spreekangst met schrijf- en spreekprestaties, doet hem concluderen dat deze onderzoeksresultaten in ernstige mate afbreuk doen aan de validiteit van de vragenlijsten en de theoretische veronderstellingen die eraan ten grondslag liggen. Die conclusie gaat mij iets te ver: we weten nog steeds niet of de instrumenten niet deugen, of dat de relatie niet bestaat. En zo langzamerhand is er toch nog al wat onderzoek waarin de aanwijzingen voor zo'n relatie erg zwak zijn. En waarom zou je geïnteresseerd zijn in zo'n zwak verband? HK blijft in zijn afrondende conclusie voorzichtig als hij stelt dat de bevindingen misschien een juiste afspiegeling van de werkelijkheid zouden kunnen zijn: misschien is er bij vijftienjarigen inderdaad geen sprake van een duidelijk relatie tussen taalattitudes en taalprestaties. Hij neemt dan als empiricus wetenschappelijk gas terug: "Gezien de hiervoor besproken relativerende kanttekeningen en alternatieve verklaringen zijn wij vooralsnog niet geneigd deze conclusie definitief te trekken." (Pag. 212).

Over peilingsonderzoek

Over het interessante thema van de mogelijke koppeling tussen peilingsonderzoek en school-effectiviteitsonderzoek, is de auteur echt te karig. In tegenstelling tot de discussies in de andere hoofdstukken, blijft de discussie van deze bevindingen mager; in het laatste hoofdstuk, Conclusies en aanbevelingen, wordt dit niet goedgemaakt. HK doet twee aanbevelingen voor het geval men peilingsonderzoek mede ten dienste wil stellen aan schooleffectiviteitsonderzoek: het steekproefontwerp moet erop worden afgestemd en de analyse moet de hiërarchische structuur van de data (leerlingen binnen scholen) weerspiegelen. Ik zou daaraan willen toevoegen, als het om peilingen in het voortgezet onderwijs gaat: leerlingen binnen klassen

binnen scholen. Didactische keuzen worden in Nederland immers nauwelijks op school gemaakt in het voortgezet onderwijs, maar binnen de klas. Er is niet alleen een clustereffect wegens schoollidmaatschap, maar eveneens wegens klaslidmaatschap. Dat vergt een niet aan de school, maar aan de klas aangepast steekproefontwerp. Ik had graag gezien tot welke prijs een peilingsonderzoek als het haalbaarheidsonderzoek waarover HK rapporteert uitgebreid had kunnen worden t.b.v. de schooleffectiviteitsvraag. HK raakt hier overigens een probleem dat veel verder reikt dan peilingsonderzoek en schooleffectiviteitsonderzoek: elk onderzoek waarin een didactiek niet aan een individu maar aan een groep wordt aangeboden, zou rekening moeten houden met het clustereffect om statistische overschatting van effecten te voorkomen. Mij is gebleken dat HK deze aanbevelingen reeds in praktijk bracht in het onderzoeksontwerp en de instrumentatie van peilingsonderzoek voor aardrijkskunde in het voortgezet onderwijs (Kuhlemeier e.a., 1994). Dit onderzoek kende een longitudinaal onderzoeksontwerp waarbij per school alle derde klassen in het onderzoek werden betrokken. Daardoor kon in de statistische analyse een expliciet onderscheid worden gemaakt tussen de leerwinst op school-, klas- en leerlingniveau (Van den Bergh & Kuhlemeier, 1997). Daarnaast werd de reguliere aanbodmeting met vragenlijsten aangevuld met klasobservaties, interviews, check-lists en logboeken (Heideberg, Van den Bergh, Kuhlemeier & Hoeks, 1993; Kuhlemeier & Van den Bergh, 1995). Het is kenmerkend voor de bescheidenheid van de auteur van het proefschrift dat niet al in het proefschrift erop gewezen wordt dat de auteur lessen trok uit het dissertatie-onderzoek. Wat mij betreft valse bescheidenheid en koudwaterrees: de aardrijkskundepeiling toont de progressie van peilingsonderzoek, en een mededeling daarover had de lezer van dit proefschrift wat zorgen weg kunnen nemen over de kwaliteit van peilingsonderzoek.

En daar begon het allemaal mee: de haalbaarheid van peilingsonderzoek. Naast de gebrekkige validiteit van het instrumentarium, roert HK twee hoofdproblemen van peilingsonderzoek aan: de 'didactisch relevante eenheid' als uitgangspunt voor rapportage, en de problematiek van de onvergelykbaarheid van gegevens uit verschillende opleidingstypen.

HK toont overduidelijk taakeffecten aan, die inhoudelijk goed te interpreteren zijn, en niet tot ruis in de meting bestempeld kunnen worden. Taken dienen dus de eenheid van rapportage te zijn. Dat betekent dat er een domeinbeschrijving van taken aan de peiling ten grondslag zou moeten liggen. Deze empirische studie roept om een bezinning op zo'n domeinbeschrijving, niet alleen als output van het onderwijs, maar ook als onderliggende theorie voor de beschrijving van het aanbod. Zo'n exercitie is zeker nuttig, niet alleen voor de onderbouwing van het peilingsonderzoekinstrumentarium.

HK toont eveneens duidelijk aan dat vergelijken tussen schooltypen in VO-I niet kan op basis van het huidige peilingsinstrumentarium: in veel gevallen meten de instrumenten niet dezelfde zaken, en zijn vergelijkingen tussen opleidingstypen daardoor uitgesloten. Mijn conclusie zou zijn: peilingsonderzoek is (vooralsnog) niet mogelijk, in elk geval niet zolang een brede populatie niet beter bediend wordt dan met één toetsbatterij. De vraag is trouwens of zo'n vergelijking tussen schooltypen zinvol is; telkens vaststellen dat het VBO slechter scoort dan MAVO, en dat MAVO slechter dan HAVO/VWO is een weinig interessante vaststelling. Zorgvuldige adaptie van het instrumentarium aan de populatie van het opleidingstype is geboden: de verschillen tussen opleidingstypen moeten niet alleen in de scores tot uitdrukking komen, maar ook in het instrumentarium. De vraag is zelfs, of de heterogeniteit in sommige van deze populaties niet zo groot is, dat ook binnen sommige opleidingstypen een gedifferentieerd instrumentarium aangeboden zou moeten worden. Wie bedacht toch ook al weer dat de basisvorming voor alle leerlingen met één en dezelfde toets moet worden afgesloten?

Literatuur

Bergh, H. van den (1988). *Examens geëxamineerd*. Diss. UvA. 's-Gravenhage: SVO.

Blok, H. (1987). *Taal voor alledag: Feiten en meningen over het taalgebruik van LBO-en MAVO-leerlingen in alledaagse situaties*. Diss. UvA. 's-Gravenhage: SVO.

- Gelderen, A. van (1992). *Fonetisch en linguïstische validatie van oordelen over spreekprestaties*. Diss. UvA. Amsterdam: SCO.
- Glopper, K. De (1988). *Schrijven beschreven. Inhoud, opbrengsten en achtergronden van het schrijfonderwijs in de eerste vier leerjaren van het voortgezet onderwijs*. Diss. UvA. 's-Gravenhage: SVO.
- Bergh, H. van den & Kuhlemeier, H. (1997). Multiniveaue modellen voor de analyse van leerwinst vergeleken. *Tijdschrift voor Onderwijsresearch*, 22, 2.
- Heidelberg, J., Bergh, H. van den, Kuhlemeier, H. & Hoeks, J. (1993). Tijdsbesteding van docenten in het voortgezet onderwijs op verschillende manieren gemeten: een validatiestudie. *Tijdschrift voor Onderwijsresearch*, 18, 306-316.
- Kuhlemeier, J.B., Bergh, H. van den, Notté, H., Wagenaar, H., Verstralen, H., & Cappers, R. (1994). *Balans van het aardrijkskunde-onderwijs in het derde leerjaar van het voortgezet onderwijs*. Arnhem: Instituut voor Toetsontwikkeling.
- Kuhlemeier, J.B. & Bergh, H. van den (1995). *Meting van het onderwijsaanbod met behulp van twee checklists, een logboek en een vragenlijst*. PPO-VO Werkdocument Nr 7. Arnhem: Instituut voor Toetsontwikkeling.
- Oostdam, R.J. (1991). *Argumentatie in de peiling. Aanbod en prestatiepeiling van argumentatievaardigheden in het voortgezet onderwijs*. Diss. UvA. Amsterdam: SCO.
- Otter, M.E. (1993). *Leesvaardigheid, leesonderwijs en buitenschools lezen. Instrumentatie en effecten*. Diss. UvA. Amsterdam: SCO.
- Otter, M.E., Oostdam, R.J. & Vergeer, M.M. (1998). *Voorstudie periodieke peiling van het onderwijsniveau in het special onderwijs*. Deel 4: Het onderwijsaanbod voor taal op LOM-scholen. Amsterdam: SCO.
- Overmaat, M. (1992). *De kwaliteit van meetinstrumenten voor het aanbod van taal en lezen*. Amsterdam: SCO.
- Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen. Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basis-onderwijs*. Diss UvA. Amsterdam: SCO.
- Wesdorp, H. (1974). *Het meten van productief-schriftelijke taalvaardigheid*. Purmerend: Muusses.
- Wesdorp, H., Bergh, H. van den, Bos, D.J., Hoeksma, J.B., Oostdam, R.J., Scheerens, J. & Triesscheijn, B. (1986). *De haalbaarheid van peilingsonderzoek; een voorstudie op het gebied van het taalonderwijs in de lagere school*. Lisse: Swets & Zeitlinger.

Gert Rijlaarsdam
Instituut voor de Lerarenopleiding UvA
Wibautstraat 2-4
1091 GM Amsterdam
(rijlaars@ilo.uva.nl)

Verslag VOR-ledenvergadering 21 mei 1997

18.00 - 18.45 uur Katholieke Universiteit Leuven (tijdens ORD 1997)

Aanwezig

Leden:

N. de Bekker-Ketelaars, M. Berkhout, Th. C.M. Bergen, F. van Bussel, S. Dijkstra, A.J. van Essen, B. van der Heijden, H. van Hoeven, F. de Jong, T. de Jong, G. Kanselaar, W. Kuiper, J.L. van der Linden, D.D. Longayroux, A. Pilot, J.G.L. Thijssen, K.J. Westerhof

Bestuursleden:

J.J.H. van den Akker (namens Divisie Curriculum), T. Imbos (penningmeester), M. Mulder (secretaris), J.M. Pieters (namens Divisie Leren en Instructie), N. Verloop (voorzitter)

1. Opening en agendavaststelling

De voorzitter opent de vergadering en heet alle aanwezigen hartelijk welkom. De agenda wordt in ongewijzigde vorm vastgesteld.

2. Verslag

Wat betreft de tekst van het verslag van de VOR-ledenvergadering van 5 juni 1996 wordt meegedeeld dat onder het punt 'Rondvraag' de naam van De Jong moet zijn J.L. van der Linden. Verder wordt het verslag ongewijzigd goedgekeurd met dank aan de secretaris. Naar aanleiding van het verslag worden verder geen opmerkingen gemaakt.

3. Mededelingen

De voorzitter deelt mee dat mogelijkheden voor samenwerking met het Vlaams Forum voor Onderwijsonderzoek (VFO) worden geëxploreerd. Het VOR-bestuur staat positief tegenover het verzoek van het VFO om VFO-leden reductie te geven voor deelname aan ORD's in Nederland. Gesteld is dat dit principe wederzijds moet gelden: VOR-leden zouden ook tegen VFO-tarieven moeten kunnen deelnemen aan VFO-activiteiten. Het VFO heeft hiermee ingestemd en derhalve heeft het VOR-bestuur het VFO-verzoek welwillend toegestaan.

Wat betreft de ECER in 1997 in Frankfurt stelt de voorzitter dat afgesproken is de papervoorstellen strenger te beoordelen. Het proces van beoordeling loopt nu ook beter. De ECER-conferentie in 1998 is in Ljubljana en in 1999 wellicht in Göteborg, zo mogelijk samen met de EARLI-conferentie. Als het niet lukt om de ECER- en EARLI-conferentie gezamenlijk te organiseren dan zal de ECER-conferentie in 1999 in Finland worden georganiseerd. De ECER-conferentie in 2000 is waarschijnlijk in Parijs.

Ten slotte deelt de voorzitter mee dat het functioneren van de VOR-Divisies momenteel wordt doorgelicht. Sinds het oprichten van de Divisies functioneren deze in hoge mate autonoom. De mate van activiteit is per Divisie echter sterk verschillend. Uit een inventarisatie is gebleken dat de meeste Divisies redelijk tot zeer actief zijn. Met de besturen van de Divisies die minder of niet actief zijn, wordt afzonderlijk gesproken. Het VOR-bestuur is van mening dat minder actieve Divisies moeten worden aangespoord meer activiteiten te ontplooiën, dan wel moeten worden opgeheven.

De secretaris deelt mee dat de Faculteit der Toegepaste Onderwijskunde van de Universiteit Twente is benaderd met de vraag of zij de 25e ORD in 1998 wil organiseren. Op dat verzoek is positief gereageerd. Voorlopig zijn de data 13 tot en met 15 mei 1998. Begin juni wordt vastgesteld of er tijdens die data ruimte beschikbaar is. Indien dat het geval is, kunnen de data definitief worden vastgesteld.

4. Jaarverslag

Naar aanleiding van het jaarverslag merkt de voorzitter op dat het aantal leden is afgenomen. Dat is verontrustend, maar het heeft vanaf heden de aandacht van het bestuur. Omdat er in de begroting voor het komend jaar meer ruimte zit dan in de afgelopen jaren, zijn er meer financiële middelen beschikbaar. Het voornemen is daarvan een deel aan te wenden voor ledenwerving. Er is dit jaar al een nieuwe folder uitgekomen die kan worden gebruikt voor ledenwerving.

Het jaarverslag wordt verder ongewijzigd vastgesteld en zal worden gepubliceerd in het VOR-Bulletin.

5. Tariefstructuur/TOR/Tijdschriften Divisies

Alvorens over te gaan tot het bespreken van het voorstel van de nieuwe tariefstructuur vraagt de voorzitter de heer Westerhof (redactiesecretaris van het TOR) verslag te doen over het gevoerde TOR-redactiebeleid gedurende het afgelopen jaar.

Westerhof deelt mee dat in het verleden werd gevonden dat er in het TOR veel artikelen op het gebied van de onderzoeksmethodologie werden geplaatst die door grote groepen binnen de VOR als minder interessant werden ervaren. Voorts speelde de inhoud van het TOR weinig in op actuele ontwikkelingen. Op basis van deze constatering is de TOR-redactie door het VOR-bestuur een aantal aandachtspunten voor het redactiebeleid meegegeven. De redactie heeft vervolgens gesteld dat de kwaliteit van het TOR op hetzelfde peil diende te blijven, maar dat er inderdaad niet steeds meer van hetzelfde in TOR gepubliceerd zou moeten worden. Het streven was vervolgens gericht op variatie. Ter versterking van de band tussen de redactie en de Divisies is besloten een Divisievertegenwoordiger aan te wijzen als contactpersoon voor het TOR binnen de Divisie. Nog niet alle Divisies hebben een kandidaat voorgesteld.

Gedurende de jaargang 1996 zijn vier themanummers gepubliceerd waaruit een groeiende variatie blijkt: het jonge kind, organisatie en management, rendement in het beroepsonderwijs en leerstijlen in het onderwijs. In de reeks van manuscripten die nu in de pijplijn zitten, is ook een grote variatie aan onderwerpen te onderkennen. Verder is de rubriek 'Notities en commentaren' weer nieuw leven ingeblazen. Er is ook afgesproken meer overzichtsartikelen te publiceren. Daar moet nog flink aan gewerkt worden. De overgang van het TOR van Swets naar Ten Brink heeft veel energie gekost, zodat er weinig tijd is overgebleven voor dit onderwerp. De redactie is in grote lijnen tevreden met de voortgang in het redactiebeleid maar stelt dat nog niet alle doelen zijn bereikt. Wat betreft de aard van de kopij die de laatste tijd wordt ingestuurd, wordt opgemerkt dat deze steeds meer afkomstig is van jonge indieners en dat er veel onaf werk wordt ingestuurd. Dat laatste is wel een punt van zorg.

Kanselaar vraagt of het kiezen voor veel themanummers in 1998 een noodgreep was. Westerhof antwoordt dat dat niet het geval was. Hij verwacht dat het aantal themanummers niet zal toenemen. De verwachting is dat het aantal themanummers zeker zal afnemen. Kanselaar merkt op dat het publiceren van themanummers de inhoud van die nummers niet gevarieerder maakt.

De voorzitter constateert dat het redactiebeleid zich gunstig ontwikkelt, maar dat er nog verdere aandacht nodig is voor de genoemde beleidsontwikkelingen.

De heer Imbos merkt op dat er geen mededeling is gedaan over het beleidsplan ten behoeve van de subsidie van het MOCenW. De heer Westerhof deelt mee dat dit punt de aandacht heeft van de TOR-redactie en dat hij dit verder zal afhandelen met de secretaris van het VOR-bestuur.

Vervolgens stelt de voorzitter voor over te gaan naar de voorgestelde tariefstructuur voor het VOR-lidmaatschap. Het voorstel is een basislidmaatschapsbedrag vast te stellen van f 110,-. Vervolgens ondersteunt de VOR het abonnement op het TOR door daarvoor een prijs van f 25,- in rekening te brengen. Dit bedrag voor het TOR zal voor alle groepen leden gelden.

Verder is het mogelijk zich te abonneren op een aantal andere tijdschriften. Onderstaand het bestuursvoorstel voor de tariefstructuur.

Basislidmaatschap gewone leden	f 110,-
Basislidmaatschap studenten	30,-
Basislidmaatschap promovendi (AIO-OIO-bursaal)	60,-
Basislidmaatschap combinatieleden NVvO/Opleiders	75,-

Voor elke *extra* Divisie betalen de leden f 10,-.

Prijzen van tijdschriftabonnementen	
Tijdschrift voor Onderwijs Research	f 25,-
International Journal of Educational Research	55,-
Learning and Instruction	45,-
Educational Research and Evaluation	55,-

De voorzitter stelt voor de voorgestelde tariefstructuur voor 1998 vast te stellen. De vergadering stelt deze tariefstructuur vervolgens inderdaad vast.

Ten slotte merkt de voorzitter op dat de Divisies reeds enkele keren is gevraagd of zij belangstelling hebben voor het met korting aanbieden van andere wetenschappelijke internationale tijdschriften op het gebied van de Divisie. Op dit moment is nog niet van alle Divisies bekend of zij van deze optie gebruik willen maken. De Divisies is gevraagd vóór 1 september 1997 aan te geven welk tijdschrift zij eventueel extra zouden willen aanbieden aan de leden. Als Divisies niet op dit herhaalde verzoek reageren, neemt het VOR-Bestuur aan dat deze Divisies geen belangstelling hebben voor deze mogelijkheid. In het najaar zal een inventarisatie volgen bij de leden over de vraag onder welke condities zij voor welke tijdschriften opteren. Per 1 januari 1998 dient deze lijst definitief te zijn vastgesteld.

6. Financiën

De penningmeester deelt de volgende stukken uit:

- winst- en verliesrekening VOR 1996;
- begroting 1998;
- balans 1996 VOR.

Vervolgens licht de penningmeester de stukken toe. Uit de winst- en verliesrekening VOR 1996 blijkt dat er een nadelig saldo is van f 6613,-. Dit heeft verschillende redenen. Allereerst was de begroting 1996 aan de krappe kant. De kosten van de secretaris en penningmeester zijn gestegen. En er was bij het opmaken van de cijfers een bedrag van f 12.000,- aan contributie niet geïnd. Het bestuur is bezig de niet-betalende leden te bellen. Van de achterstallige betalingen is inmiddels al een bedrag van circa f 6000,- geïnd. Verder wordt opgemerkt dat er aan het probleem van niet-betalers weinig meer valt te doen dan nu wordt gedaan. Wat betreft de balans wordt opgemerkt dat de financiële positie van de VOR niet ongunstig is. De heer Dijkstra vraagt waar het bedrag aan EERA-kosten aan is besteed en welke rechten de VOR-leden ontleenen aan het EERA-lidmaatschap van de VOR. Het bestuur antwoordt dat de kosten zijn gemaakt in verband met de contributie die de VOR jaarlijks aan de EERA betaalt en aan de Executive Committee meeting die wordt bezocht door de voorzitter. Aan het lidmaatschap is korting op de jaarlijkse ECER-conferentie verbonden. Verder kunnen leden het EERA-Bulletin ontvangen. Alleen krijgt de VOR slechts 50 exemplaren toegestuurd. Het individuele lidmaatschap is een heikel punt. EERA is opgezet als een vereniging van verenigingen, maar er zijn inmiddels ook individuele leden. Zij hebben echter geen stemrecht. Een ander probleem is dat er de laatste tijd in het wilde weg netwerken worden opgericht.

Daar is een betere structuur voor nodig, bijvoorbeeld in de vorm van AERA-Divisies. Daar wil het Executive Committee echter nog niet aan.

Alvorens over te gaan tot de behandeling van de begroting, stelt de voorzitter voor het verslag van de kascontrolecommissie aan de orde te stellen. Bij ontstentenis van de leden van de kascontrolecommissie heeft de heer Stijnen mede namens mevrouw Peters-Sips en de heer Van Wieringen de volgende tekst aangeboden aan de ledenvergadering:

'De financiële overzichten van de VOR voor 1996 zijn door de leden van de commissie gecontroleerd. De winst- en verliesrekening en de balans voor 1996 zijn in orde bevonden. De penningmeester heeft uitstekend werk gedaan en kan - wat de commissie betreft - gedechargeerd worden.

Naar aanleiding van de opmerking van de kascontrolecommissie van vorig jaar, brengt de commissie nog het volgende onder uw aandacht:

- de commissie is uitermate gelukkig met de kostenreductie die ten aanzien van het TOR is bereikt;
- de penningmeester heeft de suggestie voor een beter cash-flow management onderzocht. Helaas bleek hier nauwelijks winst te behalen. De commissie is de penningmeester erkentelijk voor het uitgevoerde onderzoek;
- de declaraties van de Divisies zijn nu duidelijk traceerbaar.

Verder deel ik u mee dat de commissie de voorgestelde tariefstructuur onder agendapunt 5 van de vergadering ondersteunt. De commissie beschouwt de financiële situatie van de vereniging voor de toekomst als gezond, met andere woorden: de commissie vindt de begroting voor 1998 realistisch.

Tot slot kan de commissie u meedelen dat zij de pogingen van het bestuur de achterstallige contributie alsnog te innen zeer waardeert. Een mogelijke suggestie van de commissie is in uiterste omstandigheden publicatie van de lijst van wanbetalers te overwegen. Wellicht zijn er uit de vergadering nog andere - minder vergaande - suggesties om deze problematiek creatief te lijf te gaan.'

De voorzitter stelt voor het verslag van de commissie vast te stellen en de penningmeester met dank voor de verrichte werkzaamheden decharge te verlenen, hetgeen wordt besloten. Vervolgens deelt de voorzitter mee dat de leden van de kascontrolecommissie bereid zijn de door hen verrichte werkzaamheden volgend jaar opnieuw uit te voeren en stelt voor dat aanbod te accepteren, hetgeen door de vergadering met instemming wordt ondersteund

Vervolgens wordt overgegaan tot het behandelen van de begroting 1998. De penningmeester stelt dat een aantal zaken nog onzeker zijn, zoals het jaarlijkse bedrag aan royalties dat werd verkregen van Swets. Hierover moeten met Ten Brink afspraken worden gemaakt. De secretaris merkt op dat de niet-leden abonnees geen rekening krijgen van Ten Brink maar van de penningmeester. Van ieder abonnement voor niet-leden ontvangt de VOR dan het volle bedrag van circa f 195,-. Daar gaan de kosten van productie af. Deze bedragen circa f 60,- per jaargang. Dat betekent dat er met deze verkoop meer geld wordt gegenereerd dan het bedrag aan royalties dat van Swets werd ontvangen.

Mevrouw De Bekker-Ketelaars vraagt toelichting op de categorieën leden. De penningmeester verduidelijkt dit punt. De categorieën zijn:

1. Gewone leden;
2. AIO's;
3. Studenten;
4. Ereleden;
5. Combinatielieden.

De voorzitter stelt voor de begroting 1998 goed te keuren, hetgeen vervolgens gebeurt.

7. Jaarboek/ledenlijst

De voorzitter memoreert dat verschillende onderzoeksverenigingen zoals bijvoorbeeld de AERA een ledenlijst hebben die gebruikt kan worden door de leden voor de communicatie binnen de vereniging. Het bestuur stelt voor een jaarboek van de VOR te maken met daarin de ledenlijst. Vooruitlopend op de besluitvorming is met de vergaderstukken een vraag verstuurd of de leden bezwaar zouden hebben tegen dit voorstel. Indien de vergadering kan instemmen met het voorstel zal er in het najaar een mailing verstuurd worden naar de leden met hun gegevens. Zij kunnen die dan aanvullen en corrigeren. De vergadering stemt in met het voorstel.

8. Elektronische communicatie binnen VOR

Tijdens de vorige algemene ledenvergadering van de VOR is de elektronische communicatie aan de orde geweest. Binnen de Divisie Hoger Onderwijs is gewerkt aan dit onderwerp en de voorzitter vraagt de heer Pilot de plannen van de Divisie toe te lichten. De heer Pilot deelt mee dat de Divisie denkt aan een web-site en een elektronische nieuwsbrief, waarin bijvoorbeeld literatuurbronnen kunnen worden opgenomen en conferenties kunnen worden aangekondigd. De kosten zijn redelijk beperkt te houden en vloeien voornamelijk voort uit het huren van een plaats op het web. Verder zijn de kosten vooral aan de betrokkenen die participeren in de ontwikkeling van deze vormen van communicatie. De voorzitter merkt op dat in het VOR-bestuur de verwachting leefde dat het goed opzetten en bijhouden van elektronische communicatiemiddelen arbeidsintensief is. Verder stelt hij dat het bestuur zich vooralsnog terughoudend opstelt ten aanzien van het elektronisch uitgeven van het TOR. Hij stelt voor de ontwikkelingen in de Divisie Hoger Onderwijs een jaar te volgen om te bezien of de voorzieningen die zijn getroffen voor deze Divisie op het niveau van de gehele VOR kunnen worden getild. Zo zouden bijvoorbeeld het VOR-Bulletin en de adreslijst op een VOR-website geplaatst kunnen worden. Van Bussel merkt op dat het bestuur terughoudend moet zijn met het publiceren van adressenlijsten met e-mailnummers op een web-site omdat diverse leden vervolgens mogelijk bestookt gaan worden met junk-mail. Verloop sluit zich hierbij aan. De vergadering stemt in met het plan om de ontwikkelingen binnen de Divisie Hoger Onderwijs af te wachten.

9. Rondvraag

De heer Thijssen vraagt of het bestuur wel eens gedacht heeft aan het inschakelen van een incasso-bureau om niet-betalers tot betaling te manen. Het bestuur antwoordt dat daar inderdaad aan is gedacht, maar dat de kosten die daarmee gemoeid zijn niet opwegen tegen de opbrengsten. Vandaar dat is besloten tot het bellen van de niet-betalers door bestuursleden, waarvan de tijd niet wordt doorberekend.

Mevrouw De Bekker-Ketelaars vraagt of de NVvO en de VELON de enige verenigingen zijn waarbij combinatielidmaatschappen mogelijk zijn. De voorzitter antwoordt dat dat momenteel inderdaad feitelijk het geval is. Mevrouw De Bekker-Ketelaars suggereert ook combinatie-lidmaatschappen met andere verenigingen mogelijk te maken, bijvoorbeeld met de NVO. De voorzitter antwoordt dat daar verder over zal worden nagedacht binnen het bestuur.

Tot slot wordt gevraagd of de Divisiebudgetten al zijn vastgesteld. Het bestuur antwoordt dat dat inderdaad het geval is en dat de Divisies dit jaar f 10,- per lid van de Divisie krijgen. Dat is ook verwerkt in de financiële stukken die eerder tijdens de vergadering zijn vastgesteld.

10. Sluiting

De voorzitter sluit de vergadering en bedankt de aanwezigen.

Jaarverslag VOR 1996 - 1997

Samenstelling bestuur VOR 1996-1997:

Prof. dr. N. Verloop, voorzitter

Prof. dr. J.F.M.J. van Hout, vice-voorzitter

Dr. M. Mulder, secretaris

Dr. Tj. Imbos, penningmeester

Dr. J. Ax

Prof. dr. J.J.H. van den Akker

Prof. dr. J. Dronkers

Prof. dr. J.M. Pieters

Prof. dr. P.F. Sanders

Dr. A.M. Versloot

Prof. dr. Th. Wubbels

1. Algemene gegevens

De VOR telt ultimo 1996 549 leden. Ultimo 1995 waren dat er nog 596. Dat houdt in dat het ledenaantal in een jaar met bijna 10% is gedaald. Deze daling is niet meteen alarmerend, maar wel zorgelijk. Vandaar dat het bestuur heeft besloten meer aandacht te gaan schenken aan de ledenwerving. De daling is redelijk gelijkelijk verdeeld over de divisies. Onderstaand is de verdeling van de leden per categorie lid en per divisie weergegeven:

Ledenaantal	1996	1997	
Categorieën leden:			
gewone leden	451	417	
promovendi	48	43	
studenten	35	29	
ereleden	7	10	
combinatie-leden	55	50	
Totaal	596	549	
Ledenaantal	1996	1997	verschil
Divisies:			
B&BO	170	162	-8
LO&LG	95	86	-9
Bel&Org	91	88	-3
Curr	74	72	-2
M&E	85	77	-8
L&I	139	128	-11
HO	110	102	-8
O&S	60	56	-4

Op een totaal van 549 leden zijn er 771 registraties voor divisies.

Het voorzitterschap van de VOR is overgedragen op prof. dr. N. Verloop. Prof. dr. Th. Wubbels is toegetreden tot het algemeen bestuur als vertegenwoordiger van de divisie Lerarenopleiding & Leraarsgedrag.

Boekjaar 1996 is met een verlies van circa f 7.000,- afgesloten. De divisiebudgetten zijn der-

halve voor 50% toegekend aan de divisies. De andere helft van deze budgetten is in het jaar 1996 gereserveerd als algemene VOR-middelen. De begroting voor het jaar 1997 ziet er beter uit dan in de afgelopen jaren, met name door de bezuinigingen op de uitgave van het TOR.

2. TOR

De overgang van het TOR naar uitgeverij Ten Brink is in de loop van 1996 geëffectueerd. Deze overgang heeft bij uitgeverij Swets & Zeitlinger geen problemen opgeleverd. Het TOR paste niet in het strategisch uitgeefbeleid van Swets en daarom werd de overgang door deze uitgeverij in het geheel niet als problematisch ervaren.

Een belangrijk voordeel van het uitgeven van het TOR bij Ten Brink is dat het veel minder kosten met zich meebrengt. Daardoor staat de begroting voor 1997 minder onder druk dan die van 1996.

3 ORD

In 1997 wordt de ORD georganiseerd in Leuven. Na een aanvullende oproep van de zijde van de organisatiecommissie te Leuven en een verlengde inzendtermijn is er een voldoende aantal voorstellen binnengekomen.

Na een consultatieronde bij verschillende universiteiten is de Universiteit Twente bereid gevonden de ORD 1998 te organiseren. Dat zal de 25e ORD zijn.

4. EERA/ECER

Er was de nodige kritiek op de ECER conferentie in Sevilla. Vooral het programma liet te wensen over. Een aantal sprekers heeft verstek laten gaan. De verwachting is dat de ECER-conferentie in 1997 in Frankfurt beter zal verlopen. De netwerken en leden van de organisatiecommissie zijn gevraagd voorstellen te beoordelen. Netwerken mogen ook symposia-voorstellen opstellen en indienen. Er is ruim tijd gereserveerd voor netwerk-symposia.

De ontwikkelingen binnen EERA verlopen verder goed. Er is nog wel een discussie gaande over wie stemrecht heeft binnen de algemene ledenvergadering. De discussie spitst zich toe op de vraag in welke mate individuele leden stemrecht hebben. Volgens sommigen moet de EERA een vereniging van individuele leden worden (vergelijkbaar met de AERA). De EERA is echter opgezet als vereniging van verenigingen.

De ECER-conferentie in 1998 (17-20 september) zal gehouden worden te Ljubljana.

5. VOR-congres

Besloten is een VOR-congres te organiseren over het studiehuis. Deze conferentie zal in februari 1998 worden gehouden. Prof.dr. J. van den Akker en Prof.dr. J. Pieters hebben hierin het voortouw.

6. Doorlichten divisies

Het bestuur van de VOR heeft geconstateerd dat de divisies in zeer verschillende mate actief zijn. Sommige zijn zeer actief, andere eigenlijk in-actief. Binnen het VOR-bestuur is van gedachten gewisseld over hun activiteiten. Er is gesproken over mogelijkheden om in-actieve divisies meer actief te doen worden. Het standpunt van het bestuur is dat in-actieve divisies niet binnen de vereniging thuishoren. Ten aanzien van de constatering dat er sprake is van te veel verschillende organisaties waaraan onderwijsonderzoekers hun bijdrage moeten leveren, stelt het bestuur dat waar mogelijk geprobeerd kan worden meer met andere organisaties samen te werken. In verschillende divisies wordt dit in ruime mate gedaan en de ervaringen daarmee zijn positief.

7. Promovendi

Er is overleg geweest met het promovendi-overleg. Daarin is afgesproken de band tussen het VOR-bestuur, de vereniging en de promovendi aan te halen. Voorgesteld is een promovendus te benoemen in het algemeen bestuur van de VOR en een in elk divisiebestuur. Het voorstel is door het algemeen bestuur van de VOR goedgekeurd en tevens als aanbeveling doorgegeven aan de divisiebesturen.

Publikaties jaargang 22

Nr. 1

Artikelen

Themanummer Meta-analyse

- Haanstra, Folkert, Effecten van beeldende vakken op
visueel-ruimtelijk vermogen en esthetische waarneming: twee meta-analyses 42-63
- Hox, J.J. en E.D. de Leeuw, Meta-analyse via Multi-level modellen 16-29
- Kamp, van der L.J.Th. en T.A.B. Snijders, Voorwoord 2
- Severiens, S. en G. ten Dam, Een multiniveau meta-analyse naar sekse-verschillen in
leren 29-42
- Snijders, Tom A.B., De populaties van de meta-analyse 3-16
- Otter, M.E., R. Schoonen en K. de Gopper, De relatie tussen leesprestaties en buiten-
schoolse lezen in leerjaar 5-8 63-81

Notities en Commentaren

- Hofstee, Willem K.B. en Henk A.L. Kiers, Een algemeen model voor loting en
selectie bij numerus clausus 81-86

Boekbesprekingen

- Lowyck, J., en N. Verloop, red. Onderwijskunde: een kennisbasis voor professionals.
Door N.A.J. Lagerweij. 86-88
- Schellings, Gonny. An educational approach to selecting main points in texts.
Motives, methods, tasks and individual differences. Door Jan Eling. 88-91

Nr. 2

Artikelen

- Bergh, H. van den en H. Kuhlemeier, Multi-level modellen voor de analyse van leer-

winst vergeleken	54-76
Boekaerts, M., R. Otten en R. Simons, Leerstijl in de onderbouw van het voortgezet onderwijs. Een onderzoek naar de bruikbaarheid van de ILS.	15-37
Hacquebord, H., Constructie en evaluatie van een tekstbegrip-toets voor de brugklas. Een instrument voor signaleren en typeren van leesproblemen.	37-54
Jansen, E., Curriculumorganisatie en studievoortgang	3-15
Otter, M., R. Schoonen en K. de Klopper, Televisie kijken en lezen. Het effect van tv-kijken op de leesfrequentie en de leerprestaties in leerjaar 5 tot en met 8 van het basisonderwijs.	76-96

Boekbesprekingen

Bokhove, Joop, Frank van der Schoot, Theo Eggen, Balans van het rekenonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling rekenen/wiskunde einde basisonderwijs. Door E.G. Harskamp en C. Suhre.	99-103
Polder, K.J., Implementatie van wetgeving in onderwijs: Beleidsevaluatie-onderzoek in basisonderwijs en volwasseneneducatie. Academisch proefschrift Universiteit van Amsterdam. Door J. van Gennip.	96-99

Nr. 3

Artikelen

Grift, W.J.C.M. van de, Ontwikkelingen in onderwijskundig leiderschap in het basisonderwijs tussen 1989 en 1993	109-123
Taconis, R., M. Ferguson-Hessler en G. Verkerk, Physics Problem Solving and the Transition from General Secondary Education to Higher Education	123-145
Van de redactie, Publiceren in Nederlandstalige wetenschappelijke tijdschriften?	107-109
Verschaffel, L., E. De Corte, C. Lamote en N. Dhert, Ontwikkeling van een adaptieve aanpakstrategie voor het schatten van hoeveelheden	161-186

Notities en Commentaren

Claessen, J.F.M., Commissie kwalitatieve aspecten van groeps grootte in het basisonderwijs	189-197
Meijnen, G. Wim, Indicatiestelling onderwijsachterstanden: een commissie rapporteert	186-189

Boekbesprekingen

Ganzeboom, H.B.G. & W.C. Ultee, De sociale segmentatie van Nederland in 2015. Door J. Dronkers.	204-206
---	---------

Hajer, Maaïke, *Leren in een tweede taal. Interactie invakonderwijs aan een meertalige MAVO-klas.* Door Hilde Hacquebord. 202-204

Rijmenans, R., V. Geudens, H. Coucke, H. van den Bergh & F. Daems, *Effectiviteit van Vlaamse secundaire scholen.* Door Gerry Reezigt. 198-200

Vermeulen, Marc, *Human capital in the hinterland. An analysis of causes and consequences of regional variation in educational participation.* Door Bart Bakker. 200-202

Nr. 4

Artikelen

E. van Eck, P. Koopman, M. Robijns, *Delfts blauw en rose; studieloopbanen aan een technische universiteit* 249-263

E.P.W.A. Jansen, *Invloed van het vakkenpakket VWO op het studieresultaat in de propedeuse* 238-248

M.C. Opdenakker, J. van Damme, *Centreren in multilevelanalyses: implicaties van twee centreringmethoden voor het bestuderen van schooleffectiviteit* 264-290

H.D. Webbink, *Op welk niveau verder studeren na VWO of HAVO; veranderingen tussen 1982 en 1991* 228-237

W. de Wit, H. Dekkers, *Drop-out of startkwalificatie* 211-227

Notities & Commentaren

R.J. Bosker, *De empirische bewijslast voor de extra uitgaven in het primaire onderwijs* 294-298

H.A.L. Kiers, W.K.B. Hofstee, *Gewogen loting in combinatie met rechtstreekse toelating vanaf examencijfer 8* 298-301

H. Oosterbeek & D. Webbink, *Is verkleining van klassen een goede investering?* 291-294

Boekbesprekingen

J.B. Kuhlemeier, *Taalvaardigheid, taalactiviteiten en taalattituden.* Door G. Rijlaarsdam 302-308

VOR-bulletin 309-316

Publikaties in jaargangen 1997 316-318

Lijst van redactionele medewerkers in 1997 319

Redactionele medewerkers

In 1997 werd redactionele medewerking verleend door:

W. Admiraal	IVLOS Rijksuniversiteit Utrecht
D. Beijaard	ICLON Rijksuniversiteit Leiden
J.J. Beishuizen	Rijksuniversiteit Leiden
J.M.F. ten Berg	Rijksuniversiteit Groningen
H.J.M. van Berkel	Katholieke Universiteit Brabant
R.J. Bosker	Universiteit Twente Enschede
H.P. Brandsma	OCTO Universiteit Twente
J.A. van Buuren	Open Universiteit Heerlen
C.E. Clason	Rijksuniversiteit Groningen
A.H. Corporaal	Inspectie van het Onderwijs De Meern
Th. van Dellen	Rijksuniversiteit Groningen
P.J.D. Drenth	Vrije Universiteit Amsterdam
P. van Eeden	Vrije Universiteit Amsterdam
D. de Haan	OPTIC Open universiteit
C. Hamaker	Universiteit van Amsterdam
W.K.B. Hofstee	Rijksuniversiteit Groningen
B.H.A.M. van Hout-Wolters	IVLOS, Universiteit van Amsterdam
E. Jansen	GION/COWOG Groningen
R. de Jong	GION Groningen
P. Jungbluth	ITS Nijmegen
E. de Kat	SCO-Kohnstamm Instituut Amsterdam
F. Korthagen	IVLOS Rijksuniversiteit Leiden
W.J. van der Linden	Universiteit Twente Enschede
K. Meerum Terwogt-Kouwenhoven	Universiteit van Amsterdam
W. Molenaar	Rijksuniversiteit Groningen
M. Mulder	Universiteit Twente Enschede
L. Mulder	ITS Nijmegen
Y.J. Pijl	GION Groningen
Tj. Plomp	Universiteit Twente
G. Reezigt	GION Groningen
P. Reitsma	Paedologisch Instituut Duivendrecht
J. Roeleveld	SCO-Kohnstamm Instituut Amsterdam
W.F. Siero	Rijksuniversiteit Groningen
T.A.B. Sniijders	Rijksuniversiteit Groningen
H. Starren	Rijksuniversiteit Groningen
A. van Streun	Rijksuniversiteit Groningen
C. Sühre	GION Groningen
P. Tesser	SCP Rijswijk
G.J. Tillekens	Rijksuniversiteit Groningen
L. Verhoeven	Katholieke Universiteit Nijmegen
M.J.M. Voeten	Katholieke Universiteit Nijmegen
T.H.A. van der Voort	Rijksuniversiteit Leiden
M.G. Weide	NIVEL Utrecht
B. Witziers	Universiteit Twente Enschede