



**Tijdschrift voor onderwijsresearch 1998 : een twee-
maandelijks tijdschrift voor onderzoekers van onderwijs,
gewijd aan theoretische, methodologische en professionele
problemen in de onderwijsresearch**

<https://hdl.handle.net/1874/214718>

Inhoud

Universiteit Utrecht
BIBLIOTHEEK CENTRUM UITHOF

Gezin en schoolloopbaan: een onderzoek naar de invloed van cultureel en sociaal kapitaal in het gezin op de schoolloopbaan van kinderen / *pag. 3* SW
K. van Veen, E. Denessen, P. van der Kley en J. Gerris

Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen? / *pag. 17*

J. Dronkers, M. van Erp, M. Robijn en J. Roeleveld

Hoe verhoog ik de betrouwbaarheid van mijn toetsen? / *pag. 31*

H. Kuhlemeier, F. Kleintjes en P. Goldebeld

De actuele en structurele arbeidsmarktpositie van het MBO / *pag. 46*

L. Borghans, A. de Grip en A.G.M. Matheeuwsen

Notities en Commentaren

M.H.J. Wolbers - Langetermijnontwikkelingen in de opbrengsten van diploma's op de Nederlandse arbeidsmarkt; doet sociale herkomst er wat toe? / *pag. 58*

B.P.M. Creemers - Waarzeggen en luchtfietsen, commentaar op 'Toekomst voor het funderend onderwijsbeleid' / *pag. 62*

A.C. Glebbeek - Commentaar bij artikel Gezin en schoolloopbaan / *pag. 67*

K. van Veen - Antwoord op reactie Glebbeek / *pag. 69*

Boekbesprekingen

J.J.C.M. Visser - Lezen op AVI-niveau. De scholing, normering en validering van de nieuwe AVI-toetskaarten / *pag. 72*

A.B. Dijkstra, J. Dronkers, R. Hofman - Verzuiling in het onderwijs. Actuele verklaringen en analyse / *pag. 76*

M.M.H. Frederiks - Beslissen over kwaliteit: verklaring van besluitvorming over het gebruik van onderwijsvisitatierapporten in Nederlandse universiteiten / *pag. 78*

BIBLIOTHEEK UNIVERSITEIT UTRECHT



3085 823 1

Inhoud

Evaluatie van de mogelijkheden voor doorstroming binnen het MBO / *pag. 95*
A.W. Lazonder en F.P.C.M. de Jong

Leerkrachtwisselingen en verschillen in toetsscores tussen leerling-
cohorten / *pag. 108*
H. Luyten en S. Doolaard

Pedagogisch klimaat in het basisonderwijs / *pag. 121*
W.J.C.M. van de Grift en H.P. Brandsma

Vijfentwintig jaar onderzoek naar "Leren en instructie": een persoonlijk
perspectief / *pag. 143*
E. de Corte

Notities en Commentaren

J. Dronkers - Het Trouw-rapportcijfer van scholen in het voortgezet onderwijs;
een analyse van de inspectiegegevens over de schooljaren 1995/96 en
1996/97 / *pag. 159*

W. Pijs - De berekeningsmethode in "Schoolprestaties" / *pag. 176*

J. Dronkers - De berekening van het totaal intern rendement van scholen; een
antwoord op Pijs' bezwaren / *pag. 179*

Boekbesprekingen

L. van Wessum - De sectie als eenheid: Samenwerking en professionaliteits-
opvattingen van docenten in het voortgezet onderwijs / *pag. 182*

J.J.C. Blom - Use-oriented courseware development for agricultural education:
An ecological approach / *pag. 185*

L. Sontag - Vormgeving en effecten van onderwijs aan vier- tot zevenjarige
leerlingen / *pag. 187*

Inhoud

Grafische calculators in de wiskundeles: een onderzoek naar de invoering van de grafische calculator in het VWO / *pag. 195*

E. Harskamp, A. van Streun en C. Suhre

Doeloriëntaties, waargenomen Eigen Competentie en Studieresultaten / *pag. 210*

A. Vught en J. Hoogstraten

Het didactisch handelen in het basisonderwijs / *pag. 224*

J.F. Lam en W.J.C.M. van de Grift

Vaardig oplossen van contextgebonden wiskunde problemen in de bovenbouw van de basisschool / *pag. 242*

L. Verschaffen e.a.

Notities en Commentaren

G.G.H. Jansen en K.J. Westerhof - Interbeoordelaarsovereenstemming bij het meten van de kwaliteit van het onderwijs met behulp van een beoordelingsschaal / *pag. 262*

A. Dirkwager - Inzichtelijkheid van scoringsregels en tolerantie voor fouten / *pag. 268*

Boekbesprekingen

H.J.A. Biemans - Fostering activation of prior knowledge and conceptual change, door K.S. Ali / *pag. 273*

M. Kral - Instructie en leren in combinatieklassen, door D.J. Bosker / *pag. 276*

J. Prins - Studieuitval in het W.O., door C. Terlouw / *pag. 280*

VOR-bulleting

Jaarverslag en verslag ledenvergadering / *pag. 285*

Mededeling

ORD'99 / *pag. 293*

Inhoud

Vakkenpakketkeuze en de aansluiting tussen het voortgezet en hoger onderwijs / **pag. 299**

R.K.W. van der Velden en M.H.J. Wolbers

Verlenging van de expositietijd van ondertitels op televisie:
Effecten op leestijd en de herinneringen van kinderen / **pag. 316**

C.M. Koolstra, T.H.A. van der Voort en G. d'Ydewalle

De kwaliteit van leerlingenoordelen over leraarsgedrag / **pag. 329**

R. de Jong en K.J. Westerhof

Mogelijkheden en Beperkingen van Statistische Uitzuivering binnen
Schooleffectiviteitsonderzoek / **pag. 356**

E. Maris

Boekbesprekingen

J. Onstenk - Brede vakbekwaamheid en de integratie van leren, werken en innoveren / **pag. 364**

M. Vermeulen - De school als arbeidsorganisatie: Personeelsbeleid, organisatiekenmerken en arbeidsbeleving in scholen voor voortgezet onderwijs / **pag. 368**

M.Th. Glaudé - 'Werkplekopleiden als innovatie' / **pag. 371**

VOR-bulletin

In memoriam: Frans van Bussel / **pag. 375**

Gezin en schoolloopbaan: een onderzoek naar de invloed van cultureel en sociaal kapitaal in het gezin op de schoolloopbaan van kinderen

Klaas van Veen, Eddie Denessen, Peter van der Kley & Jan Gerris*
Subfaculteit Pedagogische Wetenschappen & Onderwijskunde / KUN

Family and educational attainment; a study on the influence of cultural and social capital in the family on educational attainment of children.

ABSTRACT

The influence of cultural and social capital within families on children's educational attainment is analysed on the basis of the social capital theory of JS Coleman. It is hypothesized that social capital in the relationship between parents and children will be a precondition for the relation between the cultural capital of the parents and the school career of the children. Data of a representative sample of Dutch families (N = 235) were used. Children of these families are attending secondary education. Several interactions between social and cultural variables were tested. Results did not confirm the hypothesis of social capital as a precondition for the transfer of cultural capital. On the contrary, children of highly educated parents who have a strong cultural orientation and who spend less time with their child, have better school careers.

Het belang van het gezin voor zowel het persoonlijk als het maatschappelijk functioneren van het kind wordt door velen in toenemende mate benadrukt (Swam, 1994; Ryan, Adams, Gullotta, Weisberg & Hampton, 1995). Het kind zou in het gezin een startkapitaal meekrijgen voor de rest van zijn leven. In deze discussie wordt onder andere verondersteld dat het gezin van grote invloed is op het schoolsucces van kinderen. In het kader van deze discussie over het gezin en de schoolloopbaan van kinderen is dit artikel te plaatsen. Aan de hand van onderwijssociologische theorieën over cultureel kapitaal en vooral sociaal kapitaal in het gezin wordt in dit artikel nader ingegaan op de samenhang tussen gezinsrelaties en het schoolsucces van kinderen.

In het onderwijssociologisch onderzoek naar de ongelijkheidsproblematiek in het onderwijs kan onderscheid worden gemaakt tussen een structurele en een culturele benadering. In deze laatste benadering is met name veel onderzoek gedaan naar de invloed van bepaalde vormen van hulpbronnen oftewel 'kapitaal' in het gezin (financieel, cultureel en sociaal) op de schoolloopbaan. Het merendeel van dit onderzoek heeft zich gericht op het cultureel kapitaal van het gezin aan de hand van theorieën van Bourdieu, Bernstein en Kohn (zie Peschar & Wesselingh, 1995). Er zijn aanwijzingen dat onderzoek naar de verdeling van sociaal kapitaal in het gezin een nieuw licht zou kunnen werpen op de discussie over ongelijkheid van kansen in het onderwijs (Coleman, 1988; Hüttner & Kleijer, 1990).

In de discussie over sociaal kapitaal zijn globaal twee benaderingen te onderscheiden. In de neo-Weberiaanse benadering ligt de nadruk op het feit of men deel uitmaakt van bepaalde sociale netwerken die sterk gekoppeld zijn aan sociale milieus. Deze netwerken geven toe-

gang tot sociale posities omdat het 'elk van zijn leden de ruggesteun van het collectieve kapitaalbezit geeft, een 'geloofsbrief' die hen in de ruime zin des woords kredietwaardig maakt' (Bourdieu, 1989, p. 132). De beschikking over dit soort sociaal kapitaal hangt sterk samen met de beschikking over cultureel kapitaal en is sterk afhankelijk van het sociaal milieu waar iemand zich in bevindt (Bourdieu, o.c.).

In de andere benadering ligt de nadruk meer op de kwaliteit van de relaties tussen personen. Coleman (1990, p. 302) definieert sociaal kapitaal in het algemeen als middelen die zich bevinden in de relaties tussen personen of eigen zijn aan een bepaald sociaal netwerk, waarmee bepaalde handelingen vergemakkelijkt en bepaalde doelen efficiënter bereikt worden. Vormen hiervan zijn onder andere gedeelde normen tussen personen en verwachtingen en verplichtingen ten opzichte van elkaar. Als Coleman het heeft over de relatie tussen sociaal kapitaal en het functioneren in het onderwijs, vertaalt hij dit begrip in de mate van ondersteuning ('attention, personal interest and intensity of involvement') in gezinsrelaties en in de sociale organisatie van een gemeenschap. De mate waarin sociaal kapitaal beschikbaar is, is volgens Coleman afhankelijk van de hechtheid ('closure') van deze sociale netwerken. Hoe hechter het sociale netwerk, hoe meer sociaal kapitaal aanwezig is (Coleman, 1987; 1990; Coleman & Hoffer, 1987).

Sociaal kapitaal kan betrekking hebben op verschillende niveaus in de relatie tussen personen. In dit artikel willen we de relatie tussen sociaal kapitaal in gezinsrelaties en de schoolloopbaan onderzoeken aan de hand van de benadering van Coleman. We beperken ons in dit artikel tot het sociaal kapitaal in gezinsrelaties.

Het belang van sociaal kapitaal in het gezin lijkt toe te nemen onder invloed van moderniseringsprocessen (zoals onder andere toenemende technologisering en individualisering). ...Eén van de aannames in dit verband is dat de kwaliteit en de structuur van de relatie tussen ouders en kinderen zou worden aangetast door ontwikkelingen als een toenemende scheiding van beroepsarbeid en gezinsleven, de toenemende beroepsarbeid van moeders, het stijgend aantal eenoudergezinnen, een toename van de invloed van de media en een toenemende generatiespecifieke vrijetijdsbesteding. Door een sterker accent op de individuele belangen en behoeftebevrediging zouden ouders minder bereid zijn om tijd en energie te steken in hun opvoedingstaken en verantwoordelijkheden dan vroeger. Verder wordt verondersteld dat deze veranderingen in het gezin en daarmee ook in de socialisatie van de kinderen, onder andere gevolgen zouden hebben voor het functioneren van kinderen in het onderwijs. De oorzaak voor eventuele tegenvallende schoolprestaties zou kunnen liggen in de afname van sociaal kapitaal in het gezin (Coleman, 1990; Meijnen, 1990).

In de theorie van Coleman over sociaal kapitaal in gezinsrelaties heeft het sleutelbegrip ondersteuning een sterk pedagogische connotatie. Coleman hanteert voor het begrip ondersteuning equivalenten als persoonlijke aandacht, betrokkenheid, een hechte emotionele relatie tussen ouders en kinderen, gekenmerkt door een bepaalde mate van intimiteit.

In onderwijssociologisch onderzoek wordt algemeen aangenomen dat de opleiding van de ouders van invloed is op het functioneren van kinderen in het onderwijs¹. Kinderen waarvan de ouders hoog opgeleid zijn, hebben over het algemeen succesvollere schoolloopbanen dan kinderen waarvan de ouders laag opgeleid zijn. Ter verklaring van deze samenhang wordt gewezen op de theorie over culturele hulpbronnen. Culturele hulpbronnen hebben betrekking op bepaalde houdingen, gedragingen, gewoontes en competenties; taalgebruik, taalcodes, esthetische voorkeuren, consumptiepatronen en opvattingen over opvoeding en onderwijs. Kinderen uit gezinnen waar de ouders een hoog opleidingsniveau hebben, functioneren beter in het onderwijs, omdat ze beschikken over meer culturele hulpbronnen vergeleken met kinderen uit gezinnen waar de ouders een lager opleidingsniveau hebben. De afname van sociaal kapitaal in de gezinsrelaties zou als gevolg hebben dat kinderen steeds minder profijt hebben van het cultureel kapitaal van de ouders. Coleman (1988) veronderstelt namelijk dat voor de 'overdracht' van dit ouderlijk kapitaal de intensieve en hechte band tussen de ouders en het kind een belangrijke voorwaarde is. Hoe hechter de relatie tussen de ouders en het kind, des

te sterker zal de 'overdracht' plaatsvinden en des te meer profijt zal het kind hebben van het in de opleiding van de ouders belichaamd cultureel kapitaal. Zoals we gezien hebben verwijst Coleman (1987; 1990) voor een meer nauwkeurige afbakening van de inhoud van het begrip sociaal kapitaal naar een tweetal niveaus van sociale relaties: de mate van ondersteuning in gezinsrelaties en de hechtheid ('closure') van meer omvattende sociale netwerken. Zoals gezegd beperken we ons in dit onderzoek tot gezinsrelaties en met name tot aspecten van gezinsopvoeding (cf. Gerris, 1989). Zoals we zullen zien in de methodesectie is er a priori geen precieze operationalisering te geven van het begrip 'sociaal kapitaal'.

In dit artikel richten we ons op de vraag naar de invloed van sociaal kapitaal op de schoolloopbaan van kinderen in de leeftijd van 12 tot 17 jaar. Er wordt verondersteld dat sociaal kapitaal een belangrijke voorwaarde is voor de invloed van het cultureel kapitaal van de ouders².

Uit het bovenstaande hebben we de volgende onderzoeksvragen afgeleid:

1. *Wat is de invloed van de culturele en sociale hulpbronnen in het gezin op de schoolloopbaan van het kind?*
2. *Wat is de invloed van de sociale hulpbronnen in het gezin op de samenhang tussen de culturele hulpbronnen in het gezin en de schoolloopbaan van het kind? Oftewel, is sociaal kapitaal een belangrijke voorwaarde voor deze samenhang?*

In onderwijssociologisch onderzoek is veelvuldig aangetoond dat een hoog opleidingsniveau en het beschikken over veel culturele hulpbronnen een positiever effect heeft op de schoolloopbaan van kinderen dan een laag opleidingsniveau en het beschikken over weinig culturele hulpbronnen. Daarbij komt dat er een sterke samenhang bestaat tussen opleidingsniveau en culturele hulpbronnen: Naarmate het opleidingsniveau hoger is, beschikken ouders over meer culturele hulpbronnen (Peschar & Wesselingh, 1995). Colemans centrale stelling is dat sociaal kapitaal een belangrijke voorwaarde is voor de invloed van het opleidingsniveau en de culturele hulpbronnen in het gezin op de schoolloopbaan van het kind. Hieruit kan de volgende hypothese worden geformuleerd: *De invloed van het cultureel kapitaal van de ouders op de schoolloopbaan van het kind is meer positief naarmate er meer sociaal kapitaal aanwezig is in de relatie tussen de ouders en het kind. Deze invloed is minder positief naarmate er minder sociaal kapitaal in deze relatie aanwezig is. Met andere woorden, er is sprake van een interactie-effect tussen sociaal en cultureel kapitaal op de schoolloopbaan van het kind.*

METHODE

Onderzoekssubjecten

Voor het onderzoek wordt gebruik gemaakt van gegevens uit het OGGIN-project van de vakgroep Algemene Pedagogiek van de Katholieke Universiteit Nijmegen. OGGIN (Opvoeding en Gezin In Nederland) is een grootschalig veldonderzoek onder de Nederlandse bevolking met als doel de samenhang tussen gezinscultuur, -structuur en opvoedingspraktijken te beschrijven en te onderzoeken (Gerris, Van Boxtel, Vermulst, Janssens, Van Zutphen & Felling, 1993). Uitgangspunt en voornaamste aandachtspunt in OGGIN is de opvoeding in het gezin: het opvoedingsproces, de wijze waarop ouders opvoeden en hoe hun kind dit ervaart. Daarnaast worden de directe sociale omgeving (familie-, vrienden- en werkkring), maatschappelijke krachten (religie, politiek en economie) en ontwikkelingen die van invloed zouden zijn op het interne functioneren van het gezin en de gezinscontext (demografische, wetenschappelijke, technologische en culturele ontwikkelingen) beschreven en onderzocht.

De gegevens van OGGIN betreffen een representatieve steekproef van 788 ouderparen met tenminste één kind. Voor dit onderzoek zijn er 235 ouderparen beschikbaar, waarbij het kind in 1990 het voortgezet onderwijs bezocht³, en waar voor elke respondent de scores op de in het model gebruikte variabelen bekend zijn.

Op drie verschillende manieren zijn de gegevens voor OGGIN verzameld. Ten eerste zijn

beide ouders afzonderlijk geïnterviewd met behulp van vragenlijsten over hun opvoedingsgedrag, -oriëntaties en ervaringen in het grootbrengen van het desbetreffende kind dat onderdeel uitmaakt van het onderzoek. Ten tweede is het betreffende kind gevraagd naar de wijze waarop het de opvoeding van de ouders ervaart. Ten derde zijn de vader en de moeder gevraagd naar hun eigen opvoeding zoals zij dat hebben ervaren van hun eigen vader en moeder.

Naast deze vragenlijsten over opvoeding, is er een aantal algemene demografische gegevens verzameld (Gerris et al., 1993).

Variabelen

Sleutelbegrip in Colemans theorie over sociaal kapitaal is het begrip ondersteuning dat, zoals gezegd, een sterk pedagogische connotatie kent. Coleman hanteert voor het begrip ondersteuning equivalenten als persoonlijke aandacht, betrokkenheid, een hechte emotionele relatie tussen ouders en kinderen, gekenmerkt door een bepaalde mate van intimiteit. *'The essential qualities (of social capital in the family) have been attention, personal interest and intensity of involvement, some persistence and continuity over time, and a certain degree of intimacy'* (Coleman, 1987, p. 38). De mate waarin deze ondersteuning voorhanden is tussen ouders en kinderen, hangt samen met de hechtheid ('closure') van het gezin in de zin van de fysieke aanwezigheid van de ouders. De mate waarin sociaal kapitaal in gezinsrelaties beschikbaar is, kan dan geoperationaliseerd worden als de mate van fysieke aanwezigheid en de mate van ondersteuning. *'Social capital (...) depends both on the physical presence of adults in the family and on the attention given by adults to the child'* (Coleman, 1988, p. 111).

In dit onderzoek worden de *sociale hulpbronnen* geconcretiseerd met gegevens over de aanwezigheid en ondersteuning van de ouders. Voor aanwezigheid zijn voor vaders en moeders afzonderlijk clusteranalyses uitgevoerd, met behulp van de hiërarchische methode van Ward (1963), om te komen tot een aantal typen die zich onderscheiden naar de mate van aanwezigheid⁴.

Voor de clusteranalyse voor de *aanwezigheid* van vaders en moeders is uitgegaan van drie variabelen die betrekking hebben op de hoeveelheid tijd die de ouder besteedt aan het werk, de hoeveelheid tijd die de ouder samen met het kind besteedt en het aantal dagen dat de ouder gemiddeld thuis is wanneer het kind uit school komt⁵ (Gerris, Van Boxtel, Vermulst, Janssens, Van Zutphen, & Felling, 1992).

Er zijn drie clusters gevormd voor wat betreft de *aanwezigheid van de vader*, die als volgt getypeerd kunnen worden:

- Cluster 1: De vader besteedt *veel* tijd aan zijn werk, doet *weinig* samen met zijn kind en is *weinig* thuis als zijn kind uit school komt (n = 135).
- Cluster 2: De vader besteedt *veel* tijd aan zijn werk, doet *veel* dingen samen met zijn kind en is *weinig* thuis als zijn kind uit school komt (n = 66).
- Cluster 3: De vader besteedt erg *weinig* tijd aan zijn werk, doet *weinig* dingen samen met zijn kind, maar is zeer *vaak* thuis als zijn kind uit school komt (n = 16).

In Cluster 1 bevinden zich de vaders die het minst aanwezig zijn. In Cluster 2 zijn de vaders meer aanwezig in de zin van dat ze veel samen doen met hun kind en in Cluster 3 zijn de vaders wel veel aanwezig maar doen ze weinig samen met hun kind.

Wat betreft de *aanwezigheid van de moeder* zijn er ook drie clusters gevormd, die getypeerd kunnen worden als:

- Cluster 1: De moeder besteedt zeer *veel* tijd aan haar werk, doet *niet veel*, maar ook *niet weinig* samen met haar kind en is zeer *weinig* thuis als het kind uit school komt (n = 32).
- Cluster 2: De moeder besteedt *gemiddeld* tijd aan haar werk, doet *weinig* samen met haar kind, maar is wel *vaak* thuis als haar kind uit school komt (n = 107).
- Cluster 3: De moeder besteedt *weinig* tijd aan haar werk, doet *veel* dingen samen met haar kind en is *vaak* thuis als haar kind uit school komt (n = 96).

In Cluster 1 bevinden zich de moeders die het minst aanwezig zijn. In Cluster 2 zijn de

moeders meer aanwezig maar doen ze weinig samen met hun kind en in Cluster 3 zijn de moeders veel aanwezig en ook veel samen met hun kind.

Opvoedingsondersteuning door de ouders van de kinderen is gemeten met een drietal variabelen, bestaande uit drie schalen (responsiviteit, vertrouwensrelatie en affectie).

De schaal van responsiviteit heeft betrekking op de mate waarin het kind aangeeft in de ouderlijke opvoeding responsiviteit te ervaren voor behoeften, signalen en gesteldheid van zichzelf (Vaders $\bar{x} = 462.26$, $sd = 122.51$, Cronbachs $\alpha = .91$ / Moeders $\bar{x} = 552.48$, $sd = 109.18$, Cronbachs $\alpha = .85$). In deze schaal zijn items opgenomen als 'Ik kan goed met mijn ouders praten over van alles', 'Mijn ouders weten precies wanneer ik het ergens moeilijk mee heb' en 'Als het niet zo goed met mij gaat, lukt het mijn ouders om mij te troosten'. Hier is dus het kind informant.

De schaal vertrouwensrelatie geeft de mate weer waarin ouders aangeven een goede, persoonlijke vertrouwensrelatie met hun kind te ervaren, waarin ze hun kind begrijpen en er vertrouwelijk mee omgaan (Vaders $\bar{x} = 598.77$, $sd = 79.76$, Cronbachs $\alpha = .79$ / Moeders $\bar{x} = 604.38$, $sd = 85.34$, Cronbachs $\alpha = .81$). In deze schaal zijn items⁶ opgenomen als 'Soms denk ik dat ik niet zo'n hechte band heb met mijn kind', 'Ik heb minder hechte en warme gevoelens voor mijn kind dan ik verwacht had' en 'Mijn kind gaat te weinig vertrouwelijk met mij om'. Hier is dus de ouder informant.

De schaal affectie heeft betrekking op de mate waarin het kind aangeeft in de opvoeding van de ouders voelbare en waarneembare uitingen van positieve affectie en genegenheid te ervaren. (Vaders $\bar{x} = 301.47$, $sd = 143.69$, Cronbachs $\alpha = .83$ / Moeders $\bar{x} = 346.72$, $sd = 137.13$, Cronbachs $\alpha = .78$). In deze schaal zijn items opgenomen als 'Mijn ouders knuffelen, kussen en omarmen mij vaak', 'Mijn ouders noemen mij vaak bij mijn troetelnaam' en 'Mijn ouders zeggen vaak: Jij bent mijn lieve schat' (Gerris, et al., 1992). Hier is dus het kind informant.

Het *opleidingsniveau* van de ouders, gemeten met het hoogst behaalde diploma, hebben we tot drie groepen (laag, midden en hoog⁷) gehercodeerd.

Opleidingsniveau kan als indicator voor het sociaal milieu worden gezien, maar aan het opleidingsniveau kan nog een andere betekenis worden toegekend, namelijk dat bepaalde soorten culturele hulpbronnen inherent zijn aan een bepaald opleidingsniveau. Eigen aan een bepaald opleidingsniveau zijn de vertrouwtheid en bekendheid met bepaald taalgebruik, bepaalde gewoontes en opvattingen (Dronkers & De Graaf, 1995). In dit onderzoek wordt, naast het opleidingsniveau, een meer specifieke vorm van *culturele hulpbronnen* onderscheiden: de culturele oriëntatie van vaders en moeders. Deze culturele oriëntatie heeft betrekking op de interesse voor en gerichtheid op gebeurtenissen en activiteiten van culturele, intellectuele en politieke aard (Vader $\bar{x} = 371.51$, $sd = 95.35$, Cronbachs $\alpha = .75$ / Moeders $\bar{x} = 376.38$, $sd = 111.89$, Cronbachs $\alpha = .77$). In deze schaal, die eigenlijk een subschaal is in de gezinsklimaatsschaal, zijn items opgenomen als 'De leden van ons gezin houden van kunst', 'De leden van ons gezin lezen veel' en 'We houden ervan een museum te bezoeken' (Gerris, et al., 1993).

Uit Tabel 1 blijkt dat de culturele oriëntatie van vaders sterk positief samenhangt met die van moeders ($r = .67$). Verder zien we dat culturele oriëntatie geen noemenswaardige samenhang vertoont met de opvoedingsondersteuningsvariabelen. Voor dit onderzoek hebben we de verwachting uitgesproken dat de relatie tussen culturele oriëntatie en schoolloopbanen afhankelijk is van de aanwezige sociale hulpbronnen. Een samenhang tussen sociale en culturele hulpbronnen is daarbij niet van belang en wordt ook niet verondersteld. De opvoedingsondersteuningsvariabelen onderling vertonen een sterke samenhang, zowel voor vaders als moeders apart als tussen vaders en moeders. Opvallend is dat responsiviteit van de vader (volgens het kind) en responsiviteit van de moeder (volgens het kind) een zeer sterke samenhang vertoont ($r = .61$), evenals affectie van de vader (volgens het kind) en affectie van de moeder (vol-

gens het kind) ($r=.63$). Deze laatste bevindingen wijzen erop dat het kind blijkbaar een consistent beeld heeft van de belangrijkste opvoedingsondersteuningsvariabelen.

Tabel 1: Correlatiematrix van culturele oriëntaties en opvoedingsondersteuningsvariabelen

		Vader				Moeder			
		Cult.	Resp.	Vertr.	Aff.	Cult.	Resp.	Vertr.	Aff.
Vader	Culturele oriëntatie	1.00							
	Responsiviteit	.05	1.00						
	Vertrouwen	.03	.30**	1.00					
	Affectie	.07	.38**	.23**	1.00				
Moeder	Culturele oriëntatie	.67**	.11	-.05	.15*	1.00			
	Responsiviteit	.10	.61**	.19**	.30**	.19**	1.00		
	Vertrouwen	.07	.09	.26**	.17**	.03	.20**	1.00	
	Affectie	.03	.27**	.14*	.63**	.09	.39**	.15*	1.00

* $p<.05$ ** $p<.01$

In Tabel 2 zijn kruistabellen weergegeven van het opleidingsniveau van de vader en de aanwezigheidsclusters van de vader en daarnaast kruistabellen van het opleidingsniveau van de moeder en de aanwezigheidsclusters van de moeder. De samenhang tussen de opleiding van de vader en de aanwezigheidsclusters van de vader is significant (Pearson chi-kwadraat=13.40; $df=4$; $p=.009$) maar is klein (Cramer's $V=.17$). De kruistabel van opleidingsniveau met aanwezigheid (Tabel 3) laat zien dat hoog opgeleide vaders relatief weinig aanwezig zijn (Cluster 1) en dat laag opgeleide vaders meer aanwezig zijn (Cluster 2).

Tabel 2: Kruistabel van opleiding vader en aanwezigheid vader en daarnaast van opleiding moeder en aanwezigheid moeder (Cursief = verwachte celfrequentie)

Opleiding vader	Aanwezigheid vader				Opleiding moeder	Aanwezigheid moeder			
	clu 1	clu 2	clu 3	Totaal		clu 1	clu 2	clu 3	Totaal
Laag	53 <i>61,9</i>	30 <i>26,7</i>	12 <i>6,5</i>	95	Laag	11 <i>15,4</i>	46 <i>51,5</i>	56 <i>46,2</i>	113
Midden	53 <i>52,1</i>	23 <i>22,5</i>	4 <i>5,4</i>	80	Midden	12 <i>11,6</i>	42 <i>38,7</i>	31 <i>34,7</i>	85
Hoog	47 <i>39,1</i>	13 <i>16,9</i>	0 <i>4,1</i>	60	Hoog	9 <i>5,0</i>	19 <i>16,8</i>	9 <i>15,1</i>	37
Totaal	153	66	16	235	Totaal	32	107	96	235

Ook de samenhang tussen de opleiding van de moeder en de aanwezigheidsclusters van de moeder is significant (Pearson chi-kwadraat=10.49; $df=4$; $p=.033$) en klein (Cramer's $V=.15$). Hoog opgeleide moeders blijken minder aanwezig te zijn (Cluster 1) en laag opgeleide moeders blijken meer aanwezig te zijn (Cluster 2 en 3).

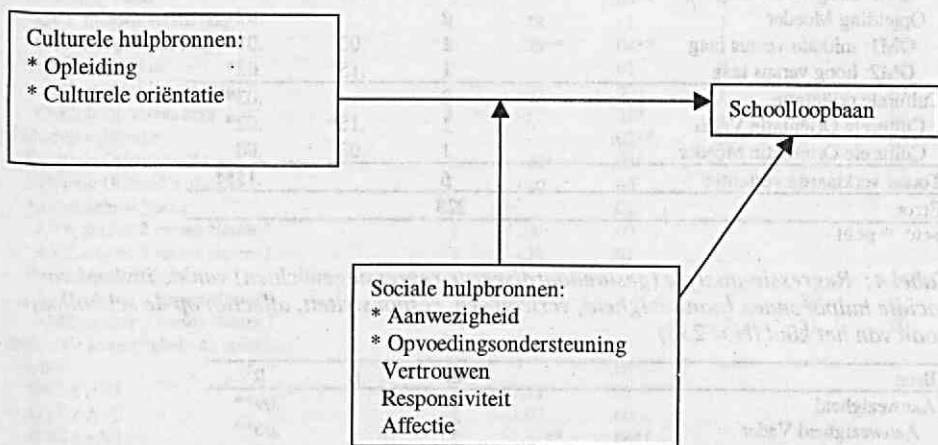
Wat betreft de relatie tussen aanwezigheid en ondersteuning blijkt dat er geen verband tussen deze variabelen bestaat, noch bij de vaders ($F=1.34$, $p=.24$), noch bij de moeders ($F=1.06$, $p=.39$).

De *schoolloopbaangegevens*⁸ zijn met behulp van het GIVON-model (Jansen Heijtmajer, 1992) vertaald tot een leerjarenladder, die we verder beschouwen als een variabele op interval niveau. Het GIVON-model heeft als principe dat aan elk leerjaar een vast aantal punten wordt toegekend en dat elk schooltype ook een aantal punten krijgt, al naar gelang de maatschappelijke waardering. Voor elk basisschooljaar worden twee punten toegekend, voor elk LBO-jaar twee en een half punt, voor elk MAVO-jaar drie punten, voor elk HAVO-jaar vier punten, voor elk VWO-jaar 5 punten en voor elk MBO-jaar vier punten. De uitkomst wordt dan nog met twee vermenigvuldigd (zie Jansen Heijtmajer, 1992).

De leerjarenladder bestaat dan uit het schooltype en het aantal leerjaren dat het kind in het basis- en voortgezet onderwijs heeft doorgebracht. Kinderen die tijdens hun schoolloopbaan ooit zijn gedoubleerd hebben dus een lagere score op de leerjarenladder dan als zij niet zouden zijn gedoubleerd⁹. Het aantal leerjaren hangt sterk samen met de leeftijd van het kind. Vanwege deze sterke samenhang hebben we de leerjarenladder per leeftijdsklasse gestandaardiseerd. De steekproef bestaat uit 6 leeftijdsklassen (12 tot en met 17 jaar). De gestandaardiseerde scores op de leerjarenladder variëren van -2.26 tot 1.97.

In schema ziet het onderzoeksmodel er als volgt uit:

Figuur 1: Onderzoeksmodel voor de samenhang tussen cultureel en sociaal kapitaal in het gezin enerzijds en de schoolloopbaan van het kind anderzijds



Met behulp van regressie-analyses worden de relaties, zoals beschreven in het bovenstaande model, onderzocht. In eerste instantie zullen regressiemodellen voor de invloed van culturele hulpbronnen op de schoolloopbaan en de invloed van sociale hulpbronnen op de schoolloopbaan afzonderlijk worden onderzocht. In tweede instantie zullen interacties tussen sociale en culturele hulpbronnen worden getoetst. Zowel de opleidingsniveaus als de aanwezigheidsclusters zullen ten behoeve van deze regressie-analyses worden gedummificeerd. Bij de opleidingsvariabelen zijn de laagste opleidingsgroepen als referentiegroepen gekozen, bij de aanwezigheidsclusters die clusters, waarvan de leden het minst aanwezig zijn (zowel voor vaders als moeders betreft dit Cluster 1).

RESULTATEN

Onderzoeksvraag 1: *Wat is de invloed van de culturele en sociale hulpbronnen in het gezin op de schoolloopbaan van het kind?*

Allereerst is de invloed van opleiding en culturele oriëntatie van vaders en moeders op de ladderscore van de kinderen getoetst. In Tabel 3 zijn de resultaten van de regressie-analyse weergegeven. Dit model verklaart 15% van de totale variantie, waarbij er met name sprake is van een hoofdeffect van de culturele oriëntatie van de vader en verschillen tussen de hoogste en de laagste opleidingsgroepen. Alle regressiecoëfficiënten zijn positief: hoe hoger de culturele oriëntatie en hoe hoger het opleidingsniveau, des te hoger is de ladderscore.

Tabel 3: *Regressie-analyse (gestandaardiseerde regressiegewichten) van de invloed van culturele hulpbronnen (opleiding en culturele oriëntatie) op de schoolloopbaan van het kind (N = 235)*

Bron	df	β	R ²
Opleiding	4		.06**
Opleiding Vader	2		.02
OV1: midden versus laag	1	.07	.00
OV2: hoog versus laag	1	.18*	.02*
Opleiding Moeder	2		.02
OM1: midden versus laag	1	.08	.01
OM2: hoog versus laag	1	.15*	.02*
Culturele oriëntatie	2		.03*
Culturele Oriëntatie Vader	1	.18*	.02*
Culturele Oriëntatie Moeder	1	.03	.00
Totaal verklaarde variantie	6		.15**
Error	228		

*p<.05 ** p<.01

Tabel 4: *Regressie-analyse (gestandaardiseerde regressiegewichten) van de invloed van sociale hulpbronnen (aanwezigheid, vertrouwen, responsiviteit, affectie) op de schoolloopbaan van het kind (N = 235)*

Bron	df	β	R ²
Aanwezigheid	4		.09**
Aanwezigheid Vader	2		.06**
AV1: cluster 2 versus cluster 1	1	-.15	.02
AV2: cluster 3 versus cluster 1	1	-.24**	.05**
Aanwezigheid Moeder	2		.03*
AM1: cluster 2 versus cluster 1	1	-.24*	.02*
AM2: cluster 3 versus cluster 1	1	-.26*	.03*
Vertrouwen	2		.01
Vertrouwen Vader	1	.07	.00
Vertrouwen Moeder	1	.08	.01
Responsiviteit	2		.00
Responsiviteit Vader	1	.01	.00
Responsiviteit Moeder	1	.01	.00
Affectie	2		.00
Affectie Vader	1	.04	.00
Affectie Moeder	1	-.06	.00
Totaal verklaarde variantie	10		.15**
Error	224		

*p<.05 ** p<.01

Als tweede is het model met de effecten van sociale hulpbronnen getoetst (zie Tabel 4). Dit model verklaart 12% van de totale variantie. Deze verklaring komt vrijwel volledig voor rekening van de aanwezigheid van de ouders. De variabelen die de ondersteuning in het gezin meten hebben een verwaarloosbare invloed op de schoolloopbaan van het kind. De regressiecoëfficiënten van de aanwezigheidsvariabelen zijn allen negatief. Omdat voor zowel vaders als moeders Cluster 1 de referentiegroep is en omdat dit cluster de vaders en moeders bevat die het minst aanwezig zijn, betekent dit dat aanwezigheid een negatieve invloed heeft op de schoolloopbaan: hoe meer aanwezig, des te lager de ladderscore.

Onderzoeksvraag 2: *Wat is de invloed van de sociale hulpbronnen in het gezin op de samenhang tussen de culturele hulpbronnen in het gezin en de schoolloopbaan van het kind? Oftewel, is sociaal kapitaal een belangrijke voorwaarde voor deze samenhang?*

Tabel 5: *Regressie-analyse (gestandaardiseerde regressiegewichten en bijdrage aan verklaarde variantie) van de invloed van Opleiding, Culturele Oriëntatie en Aanwezigheid op de schoolloopbaan van het kind (N = 235)*

Bron	Vaders			Moeders		
	df	β	R ²	df	β	R ²
Opleiding	4		.07**	4		.03
Opleiding Vader	2		.04**	2		.02
OV1: midden versus laag	1	.12	.01	1	.07	.00
OV2: hoog versus laag	1	.29**	.04**	1	.18*	.02*
Opleiding Moeder	2		.01	2		.01
OM1: midden versus laag	1	.06	.00	1	-.34	.01
OM2: hoog versus laag	1	.11*	.01*	1	-.11	.00
Culturele oriëntatie	2		.03**	2		.03*
Culturele Oriëntatie Vader	1	.20*	.02*	1	.19*	.02*
Culturele Oriëntatie Moeder	1	-.00	.00	1	.01	.00
Aanwezigheid Vader	2		.02			
AV1: cluster 2 versus cluster 1	1	.09	.00			
AV2: cluster 3 versus cluster 1	1	-.12	.01			
Aanwezigheid Moeder				2		.04**
AM1: cluster 2 versus cluster 1				1	-.38	.02
AM2: cluster 3 versus cluster 1				1	-.51	.00
Interactie aanwezigheid en opleiding*						
Vader	3		.04*			
OV1 x AV1	1	-.11	.01			
OV1 x AV2	1	-.03	.00			
OV2 x AV1	1	-.27**	.04**			
Moeder				4		.03
OM1 x AM1				1	.29	.01
OM1 x AM2				1	.43**	.03**
OM2 x AM1				1	.17	.01
OM2 x AM2				1	.19	.01
Totaal verklaarde variantie	11		.21**	12		.19**
Error	223			222		

*p<.05 **p<.01

noot: omdat de combinatie hoog opleidingsniveau en aanwezigheidscluster 3 bij vaders niet voorkomt, zijn er slechts drie toetsbare interactie-effecten

In de beantwoording van deze vraag zijn we op zoek naar de invloed van sociale hulpbronnen op de relatie tussen culturele hulpbronnen en de schoolloopbaan. De toetsing van deze vraag betreft dus een interactie-effect van sociale en culturele hulpbronnen. De betreffende interacties zijn in een aantal stappen getoetst.

Uitgangsmodel voor deze toetsing is het model met de invloeden van de culturele hulp-

bronnen (zie Tabel 3). Vervolgens is dit basismodel uitgebreid met de effecten van de sociale hulpbronnen.

Beurtelings zijn sociale hulpbronnen en interacties tussen sociale en culturele hulpbronnen aan het basismodel toegevoegd¹⁰. In eerste instantie is het effect van aanwezigheid van de vaders en de interactie tussen diens aanwezigheid en opleiding getoetst, gevolgd door een model met het effect van aanwezigheid en het interactie-effect met culturele oriëntatie. Vervolgens is dezelfde procedure voor de moeders herhaald. Op deze manier zijn de effecten van de vier sociale hulpbronnen successief getoetst. In Tabel 5 zijn alle modellen, waarbij sprake is van significante (hoofd- of interactie-)effecten van de sociale hulpbronnen weergegeven.

Uit de analyses blijkt dat aanwezigheid de enige sociale hulpbron is, waarbij sprake is van significante hoofd- of interactie-effecten. Effecten van responsiviteit, vertrouwen en affectie zijn geen van alle significant.

De interacties tussen aanwezigheid van vaders en culturele hulpbronnen levert een bijdrage van 4% verklaarde variantie op; het toevoegen van interactie-effecten van moeders een bijdrage van 3%. Uit de regressie-analyses blijken interactie-effecten tussen de opleiding en de mate van aanwezigheid. Dit betekent dat per aanwezigheidscluster de invloed van het opleidingsniveau op de ladderscore van het kind verschillend is. Dit geldt zowel voor vaders als voor moeders.

Uit de gemiddelde scores van de kinderen (zie Tabel 6) blijkt dat binnen de clusters waarbij de vader relatief weinig aanwezig is (Cluster 1) de invloed van het opleidingsniveau vrij sterk is. Van de hoger opgeleiden zijn het juist de kinderen, van wie de vader weinig aanwezig is, degenen met de hoogste scores. Binnen Cluster 3, de vaders die vaak thuis zijn als het kind uit school komt, zijn de gemiddelden relatief erg laag. Binnen Cluster 2, waar de vaders veel met de kinderen samen doen, maar weinig thuis zijn als het kind uit school komt, zijn het juist de kinderen van laag opgeleide vaders die de hoogste ladderscores hebben.

Tabel 6: De gemiddelden op de leerjarenladder van de combinaties van opleidingsniveaus en aanwezigheidsclusters van vaders en moeders

Opleiding vader	Aanwezigheid vader			Opleiding moeder	Aanwezigheid moeder		
	Cluster 1	Cluster 2	Cluster 3		Cluster 1	Cluster 2	Cluster 3
Laag	.11	.13	-.72	Laag	.58	-.38	-.50
Midden	.37	-.36	-.69	Midden	.00	-.36	.27
Hoog	.71	-.23	-	Hoog	.44	.45	.54

Voor de combinaties van opleiding en aanwezigheid van de moeders blijkt dat binnen Cluster 3, de moeders die het meest aanwezig zijn, de kinderen van lager opgeleide moeders de laagste scores halen. In Cluster 1, de moeders die het minst aanwezig zijn, hebben de kinderen van de laag en hoog opgeleide moeders de hoogste scores op de leerjarenladder.

Als we de resultaten van de vaders en moeders met elkaar vergelijken dan zien we dat een combinatie van geringe aanwezigheid en een hoge opleiding samengaan met hoge scores, en dat een combinatie van veel aanwezig zijn en een lage opleiding juist samengaan met lage scores van kinderen.

CONCLUSIE EN DISCUSSIE

In dit onderzoek staat de volgende, aan Coleman ontleende vraagstelling centraal: Wat is de invloed van sociale hulpbronnen in het gezin op de samenhang tussen het cultureel kapitaal in het gezin en de schoolloopbaan van het kind? Hierbij hebben we de volgende hypothese geformuleerd: *De invloed van het cultureel kapitaal van de ouders op de schoolloopbaan van het kind is meer positief naarmate er meer sociaal kapitaal aanwezig is in de relatie tussen de ouders en het kind. Deze invloed is minder positief naarmate er minder sociaal kapitaal in deze relatie aanwezig is. Met andere woorden, er is sprake van een interactie-effect tussen sociaal en cultureel kapitaal op de schoolloopbaan van het kind.*

Uit de resultaten komt het volgende beeld naar voren. De mate van aanwezigheid blijkt in die zin van invloed te zijn op de schoolloopbaan dat juist ouders die weinig aanwezig zijn een positievere invloed hebben op de schoolloopbaan van hun kinderen dan ouders die veel aanwezig zijn. De mate van ondersteuning blijkt niet eenduidig verband te houden met schoolsucces. De ouders die niet veel aanwezig zijn, zijn veelal hoog opgeleid en sterk gericht op activiteiten van culturele, intellectuele en politieke aard. Colemans theorie over sociaal kapitaal in het gezin, meer specifiek dat sociaal kapitaal een belangrijke voorwaarde is voor de invloed van cultureel kapitaal, gaat dus niet op voor deze Nederlandse steekproef. Ons onderzoek bevestigt de in de onderwijssociologie algemene aannames, dat vooral het opleidingsniveau van de ouders en de culturele oriëntatie van het gezin een positieve invloed hebben op het schoolsucces van het kind. De veronderstelde positieve invloed van de sociale hulpbronnen aanwezigheid en ondersteuning kon niet worden bevestigd. Dit lijkt een enigszins onverwachte conclusie, want op grond van Colemans these hadden we verwacht dat de mate van aanwezigheid en ondersteuning een positieve invloed zouden hebben op de schoolloopbaan.

Wil dit zeggen dat daarmee ook is aangetoond dat sociaal kapitaal in de zin van Coleman van minder belang is? We menen van niet, en wel op grond van de volgende redenering. Er is sprake van een overduidelijke samenhangende invloed van een relatief hoog opleidingsniveau van ouders met een relatief geringe aanwezigheid op het schoolsucces van adolescente kinderen. Men zou kunnen veronderstellen dat hoger opgeleide ouders veel vertrouwen hebben in de eigen verantwoordelijkheid van het kind en dat er in hoge mate sprake kan zijn van een gedeeld norm- en taakbesef tussen ouders en kinderen. Uit onderzoekingen van de theorie van Kohn blijkt ondermeer dat ouders met een hoog opleidingsniveau meer gericht zijn op de autonomie van het kind in opvoedingssituaties en daardoor eerder een appel doen op zelfbeheersing en norminternalisatie dan ouders met een relatief laag opleidingsniveau (cf. Gerris & Janssens, 1987; Janssens & Gerris, 1988; Van der Slik, Gerris & Felling, 1996). Het ligt voor de hand dat een op autonomie gerichte opvoeding bij het kind ook een zelfstandig taakbesef teweeg brengt. Een zelfstandig taakbesef dat van belang is om op school met succes de leertaken te kunnen uitvoeren. Binnen de context van het gezin kan dit zelfstandig taakbesef of de toeschrijving van een zelfstandig taakverantwoordelijkheid mede verklaren waarom ouders volstaan met minder aanwezigheid en daarmee met minder toezicht. In dit verband willen we bovendien wijzen op de mogelijkheid dat meer affectieve en responsieve uitingen en 'closure' werkzaam zijn als sociaal kapitaal op jongere leeftijden, terwijl in de vroeg-adolescentie en adolescentie ouders meer gebruik maken van een geïnternaliseerd norm- en plichtsbesef bij hun kinderen met name wanneer het gaat om schoolse prestaties. Dit zou betekenen dat de hypothese van sociaal kapitaal in samenhang met cultureel kapitaal een leeftijdsspecifieke inhoud zou kunnen krijgen in de pedagogische relatie tussen ouder en kind. Dit wordt bevestigd in het onderzoek van Amato (1989) die laat zien dat ouderlijke ondersteuning van meer belang blijkt te zijn voor het schoolsucces van kinderen in de basisschoolleeftijd dan voor het schoolsucces van adolescenten.

Een ander kenmerk naast opleidingsniveau en culturele oriëntatie van de ouders uit deze

steekproef is dat ze veelal ook een hoog beroepsniveau hebben (Van Veen, 1995). Vaders en moeders met een hoog beroepsniveau kunnen een sterke voorbeeldfunctie vervullen. Presteren en het werken voor je toekomst wordt als vanzelfsprekend beschouwd; onderwijs als middel om iets in je leven te bereiken, enz. Een verklaring, die hier mee samenhangt, is dat juist het veel afwezig zijn van ouders kinderen eerder zelfstandig zou kunnen maken dan wanneer de ouders veel aanwezig zijn. Aanwezigheid van de ouders zou het kind bepaalde verantwoordelijkheden kunnen ontnemen.

In het verlengde hiervan ligt dan ook een tweede kanttekening. Dit betreft een alternatieve operationalisering van sociaal kapitaal in termen van het begrip opvoedingsstijl. Dit begrip verwijst naar een constellatie van houdingen, waarden- en rolopvattingen van ouders ten opzichte van het kind die te zamen voor het kind een emotioneel klimaat creëren waarin ouderlijke gedragingen worden uitgedrukt (Baumrind, 1971; Steinberg, Elmen & Mounts, 1989; Steinberg, Mounts, Lamborn & Dornbusch, 1991; Van As & Janssens, 1994). In een aantal onderzoeken wordt dit begrip gebruikt om schoolsucces te verklaren. Uit de onderzoeken van Steinberg, et al. blijkt dat er een positieve samenhang is tussen een autoritatieve opvoedingsstijl en het schoolsucces van adolescenten. Deze opvoedingsstijl kenmerkt zich door een hoge mate van responsief gedrag en 'demanding control'. Dit laatste is te omschrijven als het uitoefenen van invloed door sterk de nadruk te leggen op de eigen verantwoordelijkheid van het kind (Van As & Janssens, o.c.). Deze jongeren hebben veelal meer zelfvertrouwen en een meer positieve houding ten opzichte van werk, waardoor zij het beter zouden doen op school dan jongeren waarvan de ouders een autoritaire of permissieve opvoedingsstijl hebben. Dit effect bleek onafhankelijk te zijn van etnische, sociaal-economische achtergrond en gezinsstructuur (Steinberg, et al., 1989; Steinberg, et al., 1991).

De autoritatieve opvoedingsstijl komt enigszins overeen met de in dit onderzoek gehanteerde operationalisering van opvoedingsondersteuning, al omvat de eerste meer aspecten van de ouder-kind relatie dan de mate van ondersteuning alleen. Wellicht verklaart dit waarom Steinberg, et al. (1989; 1991) een positieve relatie vinden tussen een autoritatieve opvoedingsstijl en schoolsucces en in ons onderzoek de mate van ondersteuning geen effect blijkt te hebben.

Het concept opvoedingsstijl of -gedrag brengt wellicht de relatie tussen ouders en kinderen ruimer in kaart dan de mate van aanwezigheid en ondersteuning. In vervolgonderzoek zou sociaal kapitaal in gezinsrelaties kunnen worden uitgebreid met het concept opvoedingsstijl.

Een derde kanttekening betreft de afhankelijke variabele van ons onderzoek. In het onderwijssociologisch onderzoek naar de invloed van cultureel en sociaal kapitaal vormt de schoolloopbaan veelal de afhankelijke variabele. In de meer pedagogische onderzoeken, zoals dat van Steinberg, et al., worden als afhankelijke variabele, naast schoolsucces, ook variabelen genomen als psycho-sociale competentie, gedragsproblemen, zelfstandigheid, normbesef, zelfbeeld, etc. Dit type onderzoek laat onder meer zien dat bepaalde gezinskenmerken niet alleen van invloed zijn op schoolsucces maar ook op bepaalde persoonlijkheidskenmerken, die bovendien sterk samenhangen met en wellicht medebepalend zijn voor het schoolsucces. In het onderwijssociologisch onderzoek naar de samenhang tussen 'gezinskapitaal' en schoolloopbanen zou dit soort persoonlijkheidskenmerken kunnen worden toegevoegd.

* Adres: Vakgroep Onderwijskunde, Montessorilaan 3, postbus 9104, 6500 HE Nijmegen.
E-mail: K.vanVeen@ped.kun.nl

1. Het beroepsniveau van de ouders laten we buiten beschouwing, vanwege de in de onderwijssociologie algemene aanname dat opleidingsniveau als voorspeller van de schoolloopbaan van meer belang is dan beroepsniveau (Peschar & Wesselingh, 1995).

2. In dit onderzoek ligt dus niet zozeer de nadruk op het onderzoeken van een verklaringsmodel van schoolloopbanen, maar op het onderzoeken van bepaalde factoren (sociaal kapitaal) die van invloed kunnen zijn op de schoolloopbanen van kinderen.
3. Onder deze 235 ouderparen bevinden zich 21 eenoudergezinnen (5.6%). Landelijk is het percentage eenoudergezinnen 15.9% (Centraal Bureau voor de Statistiek, 1995). Het betreft 3 alleenstaande vaders en 18 alleenstaande moeders. Vanwege deze kleine aantallen hebben we in dit onderzoek geen onderscheid gemaakt tussen een- en tweeoudergezinnen.
4. We volstaan hier met het weergeven van de typering van de clusters. Voor de clusteranalyses verwijzen we naar Van Veen, Denessen, Van der Kley & Gerris (1995) en Van Veen (1995).
5. De resultaten van de analyses veranderen niet wezenlijk als de variabelen, waaruit de aanwezigheidsclusters bestaan, afzonderlijk worden meegenomen. In ons onderzoek hebben we gekozen om te werken met aanwezigheidsclusters omdat we geïnteresseerd zijn in de invloed van aanwezigheid in het algemeen en niet zozeer in de verschillende soorten aanwezigheid, zoals gemeten wordt door de afzonderlijke variabelen.
6. Deze negatief geformuleerde items zijn in positieve richting gehercodeerd.
7. Laag = lagere school tot en met LBO, midden = MAVO, MBO, hoog = HAVO of hoger (zie Gerris et al., 1992).
8. In OGGIN zijn de volgende schoolloopbaangegevens aanwezig: Het schooltype, het leerjaar, de doubleerstatus en het wel of niet met een diploma verlaten van het voortgezet onderwijs.
9. Van de 235 kinderen zijn er 61 kinderen (16.2%) één keer en 8 kinderen (2.1%) twee keer blijven zitten.
10. Op deze manier krijgen alle interactie-effecten een kans en worden ze niet onderdrukt door effecten van de andere sociale hulpbronnen.

LITERATUUR

- Amato, P.R. (1989). Family Processes and the competence of adolescents and primary school children. *Journal of youth and adolescence*, 18, 39-53.
- As, N.M.C. van & Janssens, J.M.A.M. (1994). Kwaliteit ouder-kind relatie: samenhang tussen indicatoren. *Tijdschrift voor Orthopedagogiek*, 33, 206-224.
- Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology Monograph*, 4 (1), 2.
- Bourdieu, P. (1989). *Opstellen over smaak, habitus en het veldbegrip*. Amsterdam: Van Gennep.
- Centraal Bureau voor de Statistiek (1995). *Statistisch Jaarboek 1995*. 's-Gravenhage: SDU
- Coleman, J.S. (1987). Families and schools. *Educational Researcher*, 16 (6), 32-38.
- Coleman, J.S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94 (supplement), 95-120.
- Coleman, J.S. (1990). *Foundations of social theory*. Harvard: The Belknap Press.
- Coleman, J.S. & Hoffer, Th. (1987). *Public and private high schools: The impact of communities*. New York: Basic Books.
- Dronkers, J. & De Graaf, P.M. (1995). Ouders en het onderwijs van hun kinderen. In J. Dronkers & W.C. Ultee (Eds.). *Verschuivende ongelijkheid in Nederland. Sociale gelaagdheid en mobiliteit (pp. 46-66)*. Assen: Van Gorcum.
- Gerris, J.R.M. & Janssens, J.M.A.M. (1987). Opvoedingsgedrag van ouders in overtredingssituaties: Een onderzoek naar beïnvloedende factoren. *Pedagogische Studietoelichting*, 64, 295-307.
- Gerris, J.R.M. (1989). Gezinsonderzoek: een multidisciplinair werkteerrein op weg naar een interdisciplinaire benadering? *Gezin*, 1 (1), 5-31.
- Gerris, J., Van Bostel, D.A.A.M., Vermulst, A.A., Janssens, J.M.A.M., Van Zutphen, R. A. H & Felling, A.J.A. (1992). *Child-rearing and family in the Netherlands. Documentation of a national representative survey of child-rearing, family relations and family processes in 1990*. Nijmegen: University of Nijmegen.
- Gerris, J., Vermulst, A.A., Van Bostel, D.A.A.M. Janssens, J.M.A.M., Van Zutphen, R. A. H. & Felling, A.J.A.

- (1993). *Parenting in Dutch families. A representative description of Dutch family life in terms of validated concepts representing characteristics of parents, children, the family as a system and parental socio-cultural value orientations*. Nijmegen: Institute of Family Studies.
- Hüttner, H.J.M. & Kleijer, H. (1990). Marginaal: Cultureel en sociaal kapitaal: een waardevol bezit? *Sociologische Gids*, 37 (5), 270-274.
- Janssens, J.M.A.M. & Gerris, J.R.M. (1988). Sociaal milieu en reacties van ouders op disciplineringsituaties: Een empirisch verklaringsmodel. *Pedagogische Studietoën*, 65, 185-197.
- Jansen Heijtmajer, W. (1992). De GIVON, de getrapte indeling naar voltooid onderwijsniveau. *Tijdschrift voor Onderwijsresearch*, 17 (4), 217-237.
- Meijnen, G.W. (1997). *Maatschappelijke achtergronden van intellectuele ontwikkeling*. Groningen: Wolters-Noordhoff.
- Meijnen, G.W. (1990). Hulpbronnen voor schoolcarrières. Een poging tot synthese. *Sociologische Gids*, 37 (5), 288-303.
- Peschar, J.L. & Wesselingh, A.A. (1995). *Onderwijssociologie*. Groningen: Wolters-Noordhoff.
- Ryan, B.A., Adams, G.R., Gullotta, T.P., Weissberg, R.P. & Hampton, R.L. (Eds.). *The Family-School Connection. Theory, Research, and Practice*. London: SAGE.
- Slik, F.W.P. van der, Gerris, J.R.M. & Felling, A.J.A. (1996). Opleiding van ouders en de opvoeding van hun kinderen. Een onderzoek in de Kohn-traditie. *Pedagogische Studietoën*, 73 (5), 338-356.
- Steinberg, L., Elmen, J.D. & Mounts, N.S. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Development*, 60, 1424-1436.
- Steinberg, L., Mounts, N.S., Lamborn, S.D. & Dornbusch, S.M. (1991). Authoritative parenting and adolescent adjustment across varied ecological niches. *Journal of research on adolescence*, 1 (1), 19-36.
- Swan, J. van der (1994). Leve het gezin! *Elseviers Weekblad*. 3 december.
- Veen, K. van (1995). *In Loco Parentis. De invloed van sociaal kapitaal in het gezin op de schoolloopbaan van kinderen*. Doctoraalscriptie Onderwijskunde. Nijmegen: Vakgroep Onderwijskunde.
- Veen, K. van, Denessen, E., Kley, P. van der & Gerris, J. (1995). Het effect van sociaal kapitaal in het gezin op het functioneren in het onderwijs. In: J. Dronkers, H. Kleijer, H. Riddersma & G.C. de Vries (Red.). *Bundel voor de Elfde Onderwijssociologische Conferentie*. Amsterdam: SISWO.
- Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.

Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen?

De relaties tussen taal- en rekencores en advies binnen en buiten de Randstad onderzocht.

J. Dronkers, M. van Erp, M. Robijns & J. Roeleveld*

INLEIDING

In een plan voor het onderwijs in Amsterdam 1994-1998 (Gemeente Amsterdam, 1994: 8) wordt door het gemeentebestuur van die stad geconstateerd dat er in Amsterdam veel meer leerlingen naar het VWO gaan dan in de rest van het land en veel minder naar het VBO, ondanks de gemiddelde achterstand in de schoolprestaties van Amsterdamse basisschoolleerlingen. Daaruit wordt, ook door de schrijvers van dit plan, vaak maar al te snel de conclusie getrokken dat er sprake is van overadvisering en dat deze overadvisering negatieve gevolgen heeft voor de verdere loopbaan in het voortgezet onderwijs (zie echter: Rozendaal, de Jong & van der Eeden, 1987; Koeslag & Dronkers, 1995; Ledoux, 1996).

In deze bijdrage concentreren wij ons op de analyse van de achtergronden van deze veronderstelde overadvisering in Amsterdam. Speciale aandacht gaat hierbij uit naar een mogelijke speciale positie van Amsterdam in vergelijking met de drie andere grote steden, de andere steden en het platteland in de Randstad en de steden en het platteland buiten de Randstad. Met overadvisering bedoelen wij dat een bepaalde categorie leerlingen (in ons geval leerlingen woonachtig in Amsterdam) bij gelijke prestaties aan het eind van de basisschool een hoger advies krijgen dan andere categorieën leerlingen (in ons geval leerlingen woonachtig buiten Amsterdam). Deze omschrijving van overadvisering sluit aan bij die van Tesser, van Dugteren & Merens (1996: 188) en die van Bosma & Cremers (1996: 262-263). Uit verschillende studies is duidelijk geworden dat verschillende categorieën leerlingen overgeadviseerd kunnen worden: allochtone leerlingen (Tesser, e.a. 1996; Bosma & Cremers, 1996), maar ook leerlingen met hooggeschoolde autochtone ouders (Koeslag & Dronkers, 1995). Dit artikel handelt echter niet over de mogelijke overadvisering van deze overige categorieën leerlingen. Wij concentreren ons hier op het verschil tussen leerlingen binnen en buiten de grote steden. Wel zullen wij rekening houden met mogelijke verschillen in spreiding van deze andere categorieën leerlingen binnen en buiten de grote steden, omdat deze verschillen kunnen leiden tot vertekening in de mate van overadvisering van leerlingen, woonachtig in de grote steden.

In de inleiding behandelen wij het Sociaal en Cultureel Rapport van het SCP (1996) dat expliciet gewijd is aan de problemen van de grote steden. In de volgende paragraaf behandelen wij mogelijke verklaringen van overadvisering van leerlingen uit de grote steden. Na een paragraaf over de Prima-data komt de paragraaf met de analyses, en wij sluiten af met een concluderende paragraaf.

De belangstelling voor de afwijkende positie van het onderwijs in de grote steden is het laatste decennium sterk toegenomen. Dit hangt samen met een accentverschuiving tussen twee gezichten van de grote stad: de grote stad met zijn problemen als werkloosheid, kansarmoede, verloedering, onveiligheid en criminaliteit versus de grote stad als centrum van economische, politieke en culturele activiteit en vernieuwing (SCP, 1996: 316). Deze tegenstelling tussen de twee gezichten van de grote stad is reeds oud en heeft een lange geschiedenis in de sociale wetenschappen

maar ook in de literatuur. Wat betreft het onderwijs was het positieve gezicht van de grote stad tot ver na de tweede wereldoorlog dominant. Het grotere aanbod van onderwijsvoorzieningen in de grote stad en de politiek en cultureel actievere bevolking in de grote steden maakte de onderwijsresultaten van leerlingen woonachtig in de grote steden beter dan die van leerlingen in andere steden (Vrooman & Dronkers, 1986). In de loop van de zestiger jaren verdween deze voorsprong langzaam, om de laatste decennia te veranderen in een achterstand (SCP, 1996).

Het Sociaal en Cultureel Rapport 1996 is expliciet gewijd aan de problemen van de grote steden. Ook het onderwijs in de grote steden komt daarin aan de orde (SCP, 1996: 316-349). De allereerste vraag die in dit rapport gesteld wordt is in hoeverre de situatie in de vier grote steden afwijkt van die in de rest van het land. Als criterium neemt het SCP de verschillen in onderwijsresultaten. Vier factoren worden in dit rapport genoemd die verantwoordelijk kunnen zijn voor de geconstateerde verschillen.

De eerste factor is de bevolkingssamenstelling van de grote steden. Kenmerken van gezinnen, zoals het opleidings- en beroepsniveau en de etnische achtergrond van ouders zijn van invloed op de onderwijsresultaten van de kinderen. In het SCP rapport (1996: 319-320) komt de afwijkende bevolkingssamenstelling aan de orde. Het dubbele gezicht van de grote steden blijkt daaruit duidelijk. Aan de ene kant zijn er in de steden relatief veel mensen met niet meer dan lager onderwijs, aan de andere kant zijn er relatief ook veel hoog opgeleiden: de opleidingsverdeling is in de grote steden meer gepolariseerd dan elders. De grote steden verschillen overigens onderling aanzienlijk. Rotterdam onderscheidt zich voornamelijk van de rest van het land door een groot aantal laag opgeleiden. In Amsterdam en Utrecht wonen daarnaast ook veel meer hoog opgeleiden dan elders in het land. Dit weerspiegelt zich tot op zekere hoogte in de sociaal-culturele achtergrond van de leerlingen in het basisonderwijs. Het totale percentage autochtone én allochtone leerlingen met laag opgeleide ouders is in de vier grote steden ongeveer anderhalf maal zo hoog als daarbuiten. Rotterdam spant de kroon: meer dan zeven op de tien basisschoolleerlingen vallen onder de gangbare definitie van OVB-doelgroep leerlingen. Daarnaast is er ook verschil tussen de bevolkingssamenstelling van de vier grote steden en de sociaal-culturele achtergrond van basisschoolleerlingen in diezelfde steden, veroorzaakt door de relatieve kinderloosheid van de hoger opgeleide grote stadsbewoner en door zijn vertrek uit de stad nadat er kinderen geboren zijn.

De tweede factor is de concentratie van achterstandsgroepen in de grote steden, en binnen de grote steden in bepaalde wijken en scholen. Onderzoek naar leerresultaten in het basisonderwijs levert volgens het SCP aanwijzingen op dat leerlingen met eenzelfde sociale herkomst slechter presteren naarmate de concentratie allochtone leerlingen op een school hoger is. Door de grootschaligheid van de steden kunnen daar sneller verschillen tussen scholen ontstaan in sociale en/of etnische samenstelling dan in kleinere steden. Het SCP rapport (1996: 320-324) concludeert dat de segregatie tussen allochtone en autochtone leerlingen het grootst is in Rotterdam en Den Haag, gevolgd door Amsterdam en Utrecht. In de overige steden zijn dergelijke concentraties van allochtone leerlingen (zwarte en witte scholen) een zeldzaamheid.

De derde factor zoekt het SCP (1996: 317) in het door steden gevoerde onderwijsbeleid. De vier grote steden onderscheiden zich door een relatief actief onderwijsbeleid. Het achterstandsbeleid krijgt van ouds veel aandacht. Naar verhouding beschikken de grote steden over een omvangrijk en geprofessionaliseerd ambtelijk apparaat. Een actieve rol van de gemeentelijke overheid kan ondersteunend zijn voor de scholen, maar kan volgens het SCP ook leiden tot een overmaat aan projecten en activiteiten, waardoor de schoolleidingen worden afgehouden van hun onderwijskundige taken. De lokale overheid vormt tevens het bestuur van het openbaar onderwijs. Het aandeel openbare scholen in de grote steden is hoog en er is dus sprake van een aanzienlijke bestuurlijke schaal van in ieder geval de openbare scholen. Deze grootschalige besturen bieden volgens het SCP wellicht minder goede voorwaarden voor de effectiviteit van het onderwijs. Amsterdam vormt overigens een uitzondering: hier is de

bestuurlijke schaal van het openbaar onderwijs verkleind door de vorming van deelraden.

Als vierde factor noemt het SCP de mogelijke invloed van de grootstedelijke leefomgeving. De grootstedelijke cultuur brengt aan de ene kant een zekere vrijheid en ruimte voor vernieuwing met zich mee; aan de andere kant zijn er ook negatieve aspecten. Ook voor het onderwijs zouden er zowel positieve als negatieve gevolgen zijn. De verscheidenheid aan voorzieningen, evenementen en leefstijlen vormt een educatief potentieel. Tegelijkertijd zouden de grote stadscultuur en de daarin besloten liggende alternatieve tijdsbestedingsmogelijkheden ook leiden tot meer concentratieproblemen, meer schoolverzuim, meer gezagsproblemen en meer tegenstellingen tussen leerlingen en docenten. Ook zou van het culturele klimaat een zekere vrijblijvendheid uitgaan. Het SCP (1996: 331-332) geeft enige evidentie voor deze negatieve aspecten aan de hand van de hoeveelheid spijbelen, geweld en pesten in de vier grote steden en elders in Nederland.

Deze vier factoren kunnen verschillen in schoolresultaten tussen leerlingen uit de vier grote steden enerzijds en de rest van Nederland anderzijds wellicht verklaren. Het SCP rapport behandelt deze verschillen.

Allereerst analyseert het SCP (1996: 326-327, 350-352) aan de hand van de CITO-scores van het VOCL'93 cohort (basisschoolverlaters uit 1993) het effect van het leven in grote steden. De conclusie is dat het wonen in de vier grote steden geen invloed heeft op het prestatieniveau van basisschoolleerlingen als men rekening houdt met de afwijkende samenstelling van de leerlingpopulatie in deze grote steden: meer eenoudergezinnen, meer allochtone leerlingen, een lager gemiddeld sociaal milieu. De achterblijvende toetsprestaties in de grote steden zijn dus te herleiden tot de bevolkingssamenstelling. Bij nadere beschouwing blijkt dat sommige groepen (eenoudergezin, Surinaamse of Antilliaanse achtergrond) in de grote steden slechter presteren dan buiten die steden, maar andere groepen (Turkse of Marokkaanse achtergrond) niet. De meest kansrijke groepen (kinderen van hoog opgeleide ouders) doen het zelfs iets beter in de grote steden dan daarbuiten. Dit laatste is een duidelijk verschil met de bevindingen uit de Prima-rapportages over basisscholen met hoge concentraties allochtone leerlingen in Den Haag en Rotterdam (Roeleveld, 1995; ITS, 1995), die het SCP ook bespreekt (1996: 325)¹. De uit de Prima-data getrokken conclusie dat kinderen van hoog opgeleide ouders het in Den Haag en Rotterdam slecht doen, geldt kennelijk alleen indien deze kinderen basisscholen met veel kansarme leerlingen bezoeken. In dit artikel zullen wij ons niet bezighouden met de verklaring van verschillen in schoolprestaties aan het eind van de basisschool, maar de conclusie van het SCP hierover als uitgangspunt voor de verdere analyse gebruiken.

Ten tweede analyseert het SCP (1996: 327-329; 352-353) de overgang naar het voortgezet onderwijs, opnieuw met het VOCL'93 cohort (basisschoolverlaters uit 1993). Uit deze analyse blijkt dat grosso modo bij gelijk prestatieniveau, in de vier grote steden hogere adviezen worden verstrekt en voor hogere vormen van vervolgonderwijs wordt gekozen dan elders in het land. De grotere aanwezigheid van allochtone leerlingen in de grote steden biedt geen verklaring voor dit verschijnsel: ook indien rekening wordt gehouden met de etnische achtergrond van de leerlingen blijven de gevonden verschillen bestaan. Het SCP geeft voor deze 'overadvisering' en 'te hoge keuze' bij gelijke schoolprestaties van leerlingen in de vier grote steden geen verklaring². Het verschijnsel van overadvisering in een grote stad is eerder gevonden uit analyse van ouder materiaal, waarbij schoolloopbanen in Amsterdam werden vergeleken met die in de rest van het land (Rozendaal, de Jong & van der Eeden, 1987). Ook Saris & Blok (1982) vonden reeds dat de relatie tussen scores op prestatietoetsen aan het eind van de basisschool en adviezen varieert met de urbanisatiegraad van steden. Hoe urbaner de gemeente, hoe lager de gemiddelde prestatiescore nodig voor een bepaald advies. Zij gaven een plausibele verklaring voor de 'onderadvisering' van plattelandsgemeenten (zie onze verklaring 5 in de volgende paragraaf).

De 'overadvisering' in de vier grote steden in vergelijking met andere steden en het platteland binnen en buiten de Randstad is het onderwerp van dit artikel. Wij maken gebruik van de eerder genoemde Prima-data.

VERKLARINGEN VOOR OVERADVISERING IN DE GROTE STEDEN EN IN AMSTERDAM

Er kunnen zes mogelijke verklaringen voor het bestaan van overadvisering in de grote steden op grond van de genoemde literatuur opgesteld worden:

1. De gepolariseerde verdeling van de opleidingsniveaus van de bevolking in de grote steden (SCP, 1996: 316-349) kan tot hogere adviezen leiden dan in de rest van het land. Koeslag & Dronkers (1994) laten zien dat autochtone leerlingen met hooggeschoolde ouders vaker overgeadviseerd worden dan autochtone leerlingen met lager geschoolde ouders. Ook vinden zij dat autochtone leerlingen met laaggeschoolde ouders vaker overgeadviseerd worden dan leerlingen met hoger geschoolde ouders. Derhalve zullen wij bij de vaststelling van de mate van overadvisering in de grote steden controleren voor het opleidingsniveau van de ouders.

2. Kinderen van migranten, die veel in grote steden wonen, worden vaker overgeadviseerd. De oorzaken daarvan zijn volgens Tesser e.a. (1996: 188) nog niet geheel duidelijk. Een van de factoren die zij noemen is dat leerkrachten bij de vaststelling van hun advies naast prestaties ook op andere factoren letten als de houding van de leerling, de ondersteuning die thuis wordt gegeven en de wensen van ouders en kinderen. Daardoor kunnen leerkrachten aan toetsuitslagen van allochtone leerlingen een andere interpretatie geven dan aan die van autochtone leerlingen. Derhalve zullen wij bij de vaststelling van de mate van overadvisering in de grote steden controleren voor de allochtone achtergrond van de ouders, dit in combinatie met hun opleidingsniveau.

3. In de grote steden bestaat er een scherper onderscheid tussen scholen met relatief veel en met relatief weinig slecht presterende (allochtone) leerlingen. Bovendien zijn er in de grote steden meer scholen met relatief veel slecht presterende (allochtone) leerlingen dan buiten de grote steden (SCP, 1997: 320-324). Tesser e.a. (1996:188) wijzen er in hun bespreking van de mogelijke verklaring van de overadvisering van allochtone leerlingen op dat leerkrachten bij de vaststelling van de adviezen zouden worden beïnvloed door het gemiddelde prestatieniveau van hun school. Een leerling die het op een school met een laag prestatieniveau goed doet, zou een hoger advies krijgen dan wanneer dezelfde prestatie zou worden geleverd op een school met een hoog prestatiegemiddelde. De betekenis van prestatiegemiddelden van scholen voor de mate van overadvisering hangt nauw samen met de sociale compositie van scholen (naar sociale klasse en naar etniciteit). Derhalve zullen wij bij de vaststelling van de mate van overadvisering in de grote steden ook controleren voor de sociale compositie van de basisscholen³.

4. De effectiviteit van openbare scholen in de grote steden zou o.a. als gevolg van de grotere bestuurlijke schaal van dat openbaar onderwijs geringer zijn, waardoor de schoolprestaties lager liggen dan in het bijzonder onderwijs (SCP, 1996: 317; Hofman, 1997). Met het oog op de concurrentie-verhoudingen met het bijzonder onderwijs zouden de adviezen in het grootstedelijk openbaar onderwijs hoger kunnen zijn (Roeleveld, 1997). Omdat er in de grote steden relatief veel openbaar onderwijs is, controleren wij bij de vaststelling van de mate van overadvisering in de grote steden in onze analyses voor de richting van de bezochte basisschool.

5. De hogere onderwijstypen (VWO, HAVO) zijn in steden dichterbij voorhanden dan daarbuiten. Daardoor kunnen leerkrachten in steden bij gelijke schoolprestaties hogere adviezen geven (Saris & Blok, 1982). In hun HAVO- of VWO-advies behoeven zij immers geen rekening te houden met de moeite van het afleggen van de grotere afstand tot de dichtstbijzijnde HAVO of VWO, in tegenstelling tot leerkrachten buiten de steden. Indien deze verklaring juist is moet het verschijnsel van overadvisering in alle steden binnen en buiten de Randstad in min of meer gelijke mate voorkomen, omdat daar altijd wel een HAVO en een VWO staat. Daarom onderscheiden wij in onze analyses de vier grote steden (Amsterdam, Den Haag, Rotterdam, Utrecht), steden in de Randstad, steden buiten de Randstad, platteland binnen de Randstad en platteland buiten de Randstad.

6. Kinderen in de vier grote steden groeien in een assertiever cultureel klimaat op (SCP, 1996: 331-332). Binnen een dergelijk groot-stedelijk cultureel klimaat zouden leerkrachten minder gezag hebben en zouden ouders en kinderen meer en sneller opkomen voor hun (vermeende) rechten, zoals een hoger advies. Leerkrachten in de grote steden zouden daardoor sneller en meer onder druk komen te staan hogere adviezen te geven. Indien deze verklaring juist zou zijn, moet overadvisering vooral voorkomen in de vier grote steden en moeten de overige steden binnen en buiten de Randstad een min of meer dezelfde maar geringere mate van overadvisering kennen.

De bovenstaande zes verklaringen zijn geen reden voor een mogelijk hoger overadviseren in Amsterdam in vergelijking van de drie andere grote steden. Als blijkt dat er na controle in Amsterdam nog steeds meer wordt overgeadviseerd dan in de drie andere grote steden, dan zouden wij kunnen veronderstellen dat de processen die in de vier grote steden tot overadvisering leiden in Amsterdam heftiger verlopen dan in Rotterdam, Den Haag en Utrecht als gevolg van het specifiek culturele klimaat in Amsterdam. Deze verklaring zullen wij aan de hand van interacties tussen de afhankelijke variabelen en het al dan niet in Amsterdam wonen toetsen.

Voor de goede orde wijzen wij er nogmaals op dat een grotere mate van overadvisering niet een grotere uitval uit het voortgezet onderwijs verklaart, hoe plausibel dat ook lijkt. De uitval uit het Amsterdamse voortgezet onderwijs trof althans in het verleden overgeadviseerde en ondergeadviseerde leerlingen gelijk (Rozendaal, de Jong & van der Eeden, 1987).

HET PRIMA-COHORT EN DE VARIABELEN

Het Prima-cohortonderzoek (Jungbluth, Peetsma & Roeleveld, 1996) is een voortzetting en uitbreiding van de cohortonderzoeken voor de Landelijke Evaluatie Onderwijsvoorrang. In het schooljaar 1994/95 zijn de eerste metingen verricht op bijna 700 basisscholen (en bijna 100 scholen voor speciaal onderwijs). Ruim 400 van de basisscholen vormen samen de landelijk representatieve referentiesteekproef; daarnaast is een aanvullende steekproef getrokken van OVB-gebiedsscholen (Driessen, Jungbluth, van Langen & Vierke, 1996). Om het beeld van de grote steden zo breed mogelijk te houden, worden de analyses in dit artikel uitgevoerd op de volledige steekproef⁴.

We gebruiken gegevens van ruim 10.000 leerlingen uit groep 8 van de basisscholen, die deel hebben genomen aan de Prima taal- en rekentoetsen (Vierke, 1995) en waarvan het advies voor voortgezet onderwijs bekend is. Het advies is gecodeerd als⁵: 1 = ivbo; 2 = ivbo/vbo; 3 = vbo; 4 = vbo/mavo; 5 = mavo; 6 = mavo/havo; 7 = havo; 8 = havo/vwo; en 9 = vwo.

Voor het ouderlijk milieu van deze leerlingen gebruiken we de in het Prima-onderzoek gehanteerde zesdeling naar sociaal-etnische herkomst (Driessen e.a., 1996):

- beide ouders maximaal LBO-opleiding en van Turkse of Marokkaanse herkomst;

- beide ouders maximaal LBO-opleiding en afkomstig uit andere landen dan Nederland, Turkije of Marokko (overige allochtonen);
- beide ouders maximaal LBO-opleiding en van Nederlandse herkomst;
- de hoogst opgeleide ouder heeft meer dan een LBO-opleiding, maar geen hoger onderwijs;
- de hoogst opgeleide ouder heeft een opleiding op HBO- of WO-niveau;
- en een restcategorie onbekend.

Voor de (sociaal-etnische) compositie van de school gebruiken we de zesdeling, die binnen het Prima-onderzoek is ontwikkeld: elitescholen, burgerscholen, representatieve scholen, arbeidersscholen, gemixt allochtone scholen en Turks/Marokkaanse scholen (zie Jungbluth e.a., 1996: 9-12 voor precieze definities).

Bij de steden onderscheiden we, naast de vier grote steden, verder nog kleinere steden binnen en buiten de Randstad (Utrecht, Noord- en Zuid-Holland) en (evt. verstedelijkte) plattelandsgemeenten binnen en buiten de Randstad.

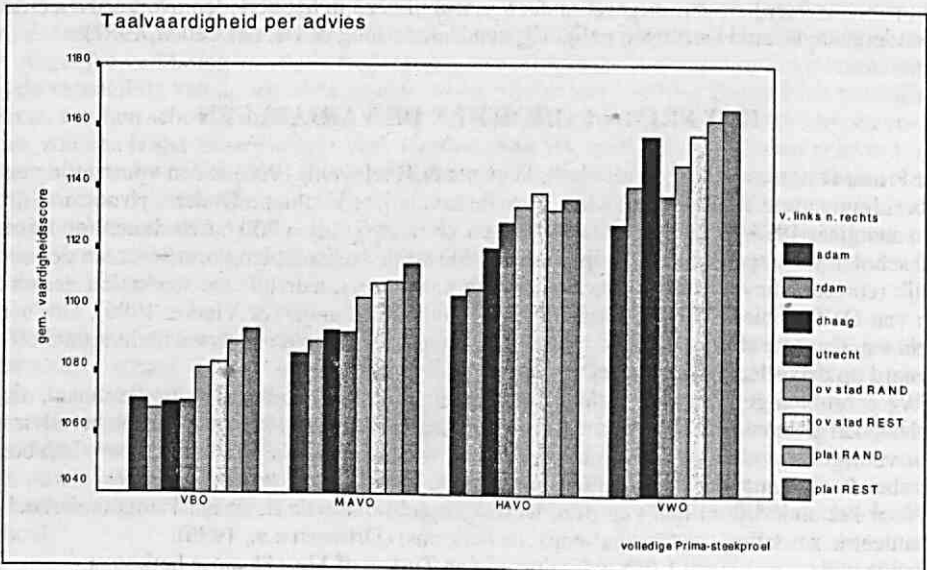
In bijlage 1 geven wij de verdeling van de Prima-leerlingen over deze kenmerken, uitgesplitst per grote stad, etc.

DE OVERADVISING IN DE VIER GROTE STEDEN

In afbeelding 1 zijn de gemiddelde taalscores weergegeven die leerlingen met een bepaald advies uit Amsterdam, Rotterdam, Den Haag, Utrecht, de overige steden in de Randstad, plattelandsgemeenten in de Randstad, steden buiten de Randstad en plattelandsgemeenten buiten de Randstad behalen.

Uit deze afbeelding blijkt dat leerlingen uit de grote steden met een bepaald advies steeds lagere taalscores hebben dan leerlingen buiten de grote steden (alleen de VWO-geadvisee-

Afbeelding 1: De gemiddelde taalvaardigheidsscore per advies voor Amsterdam, Rotterdam, Den Haag, Utrecht, overige steden in de Randstad, platteland in de Randstad, steden buiten de Randstad en platteland buiten de Randstad



Tabel 1: De mate van overadvisering in de vier grote steden en overige steden en gemeenten binnen en buiten de Randstad, gecontroleerd voor verschillen in ouderlijk milieu, schoolcompositie en richting van het onderwijs (beta's)

	Model			
	1	2	3	4
<i>Schoolprestatie</i>				
Rekenscore	.44**	.43**	.39**	.27**
Taalscore	.41**	.39**	.34**	.21**
<i>Gemeente</i>				
Amsterdam	.00	.00	.00	.00
Rotterdam	-.02**	.03*	-.02	-.10**
Den Haag	-.04**	.02	-.04*	-.12**
Utrecht	-.02**	.01	-.03**	-.09**
steden Randstad	-.12**	-.05**	-.11**	-.19**
plattel. Randstad	-.11**	-.03	-.13**	-.25**
overige steden	-.22**	-.11**	-.20**	-.29**
overig platteland	-.19**	-.08**	-.17**	-.32**
<i>Ouderlijk milieu</i>				
Turk/Marokkaan <=LBO		-.03**	-.03*	-.02
Overige alloctonen <=LBO		-.01	-.01	.02
Autochtoon <=LBO		-.12**	-.07**	-.04*
MAVO/HAVO/VWO/MBO		.00	.00	.00
HBO/Universiteit		.07**	.09**	.08**
Milieu onbekend		-.05**	-.05**	-.04
<i>Compositie school</i>				
Eliteschool		.05**	.03*	-.03
Burgerschool		.00	-.03*	-.05*
Arbeiderschool		.02	.01	-.02
Gemengde alloctone school		.00	.00	-.01
Turks/Marokkaanse school		.01	.02	-.02
Overige school		.00	.00	.00
<i>Richting school</i>				
Openbaar onderwijs		.04**	.04**	.04*
<i>Interactie met A'dam</i>				
Gemengde alloctone school		.04**		
Turks/Marokkaanse school		.05**		
HBO/Universiteit				-.04*
Openbaar onderwijs				-.05*
adjusted R ²	.55	.58	.40	.16

Legenda: ** p<.01; * .01<p<.05

den uit Den Haag hebben relatief hoge scores). Leerlingen uit plattelandsgemeenten binnen en buiten de Randstad hebben vrijwel steeds de hoogste scores bij elk advies. De HAVO-geadviseerden buiten de vier grote steden hebben gemiddeld een hogere taalscore dan de VWO-geadviseerden in Amsterdam.

Ook blijkt uit deze afbeelding dat de verschillen tussen de grote steden in de gemiddelde taalscores bij een bepaald advies het geringst zijn bij het VBO en het MAVO advies. Maar bij het HAVO en VWO advies zijn ook de verschillen tussen de grote steden aanzienlijk. Een Amsterdamse leerling met een VWO advies heeft gemiddeld een gelijke taalvaardigheidscore als een Utrechtse leerling met een HAVO advies. Dit kan erop wijzen dat relatief lagere taalscores in Amsterdam vooral voorkomt bij de hogere adviezen. Daarom zullen wij ook

afzonderlijke analyses uitvoeren over de leerlingen met een AVO advies en over leerlingen met een HAVO/VWO advies.

Deze afbeelding 1 geeft echter om twee reden een misleidend beeld van de mate van mogelijke overadvisering. Allereerst omdat nog niet is gecorrigeerd voor de verschillen in bevolkingssamenstelling, de concentratie van achterstandsgroepen in scholen en de sterkere positie van het openbaar onderwijs in de grote steden. Ten tweede heeft Pijl (1994) er terecht op gewezen dat lagere prestatiescores bij gelijke adviezen geen voldoende grond zijn om tot overadvisering te besluiten. Wat moet worden aangetoond is dat bij gelijke prestaties een bepaalde categorie leerlingen hogere adviezen krijgen dan de andere categorieën leerlingen. Daarom worden in tabel 1 de resultaten van multivariate regressie analyses weergegeven met het advies van de basisschool als afhankelijke variabele. Dit advies wordt voorspeld door de twee prestatievariabelen (reken- en taalscore), zodat overeenkomstig de procedure van Bosma & Cremer (1996) aan het bezwaar van Pijl (1994) wordt tegemoet gekomen. De overige onafhankelijke variabelen kunnen ook bijdragen aan de voorspelling van het advies en aan hun parameters kan men aflezen of en in welke mate zij bijdragen aan een verhoging van dat advies (positieve parameter) of aan de verlaging daarvan (negatieve parameter). Alle onafhankelijke variabelen zijn ongeacht de significantie van hun bijdrage toegevoegd aan de vergelijkingen. Hierop is een uitzondering: de interacties van de onafhankelijke variabelen met de variabele 'Amsterdam' zijn aan het slot stapsgewijs op grond van hun significantie toegevoegd.

De parameters (gestandaardiseerde regressie-coëfficiënten) van model 1 uit tabel 1 geeft de mate van overadvisering van leerlingen uit Amsterdam, als nog niet gecontroleerd is voor ouderlijk milieu, compositie van scholen en richting van de school. De resultaten van model 1 kunnen als volgt gelezen worden. Taal- en rekenscore dragen bijna in gelijke mate bij aan de voorspelling van de hoogte van het advies, maar de rekenscore iets meer dan de taalscore. Hoe hoger deze scores, des te hoger het advies. Bij de gemeentevariabelen vergelijken wij steeds het advies aan leerlingen uit Rotterdam, Den Haag, Utrecht, de overige randstedelijke steden, het randstedelijk platteland, de steden in de rest van Nederland en het platteland in de rest van Nederland, met het advies aan leerlingen uit Amsterdam. Naarmate de parameter van een van deze gemeenten negatiever is, krijgen leerlingen daar een lager advies in vergelijking met Amsterdamse leerlingen die vergelijkbare taal- en rekenscores hebben.

De resultaten van model 1 laten duidelijk zien dat Amsterdamse leerlingen een hoger advies krijgen dan gelijk presterende leerlingen in de andere steden en gemeenten binnen en buiten de Randstad. Gegeven de hoogte van de ongestandaardiseerde parameters van de gemeente variabelen is het verschil tussen het advies aan Amsterdamse leerlingen en aan leerlingen uit de steden buiten de Randstad het grootst (.95, d.w.z. één niveau op onze advies variabele), gevolgd door het platteland buiten de Randstad (.87), de steden binnen de Randstad (.82) en het platteland binnen de Randstad (.70). De mate van overadvisering is dus niet triviaal of onbeduidend.

Ten tweede valt op dat de overadvisering in de vier grote steden hoger ligt dan in de rest van de Randstad, maar dat de overadvisering binnen de randstedelijke gemeenten (steden en platteland) weer hoger is dan buiten de Randstad. Het is blijkbaar onjuist om de vier grote steden rechtstreeks te vergelijken met de rest van Nederland: de vier grote steden vertonen in verhevigde mate een verschijnsel dat kenmerkend voor de gehele Randstad is: leerlingen uit de Randstad (en met name die uit de vier grote steden) ontvangen bij een bepaalde taal- en rekenscore een hoger advies dan leerlingen buiten de Randstad. Voor alle duidelijkheid wijzen wij erop dat dit resultaat ook gelezen kan worden als een onderadvisering van leerlingen buiten de vier grote steden c.q. de leerlingen buiten de Randstad.

Model 2 geeft de resultaten opnieuw, maar nu ook gecontroleerd voor het ouderlijk milieu, de schoolcompositie, de richting van de school en de afwijkende en significante interacties

van deze variabelen met 'Amsterdam'. Allereerst valt nu op dat Amsterdamse leerlingen ver-geleken met gelijk presterende leerlingen uit de andere grote steden geen hogere adviezen krijgen, eerder het tegenovergestelde (Rotterdamse leerlingen krijgen de hoogste adviezen). Dit verschil tussen model 1 en model 2 ontstaat doordat nu ook rekening is gehouden met het ouderlijk milieu van de leerlingen, de compositie van de scholen en de richtingen in het onder-wijs. Blijkbaar is de verdeling van ouderlijke milieus, schoolcomposities en onderwijsrich-tingen over de vier grote steden zodanig dat het hogere advies aan Amsterdamse leerlingen volledig daarop terug te voeren is. Het gaat daarbij bijvoorbeeld om het percentage HBO/uni-versitair opgeleide ouders (hun kinderen krijgen een hoger advies dan leerlingen met lager opgeleide ouders bij gelijke taal- en rekenscores), om het percentage openbaar onderwijs (leer-lingen op openbare scholen krijgen hogere adviezen dan op niet-openbare bij gelijke taal- en rekenscores) en om het percentage laag geschoolde autochtone ouders (hun kinderen krijgen een lager advies dan andere leerlingen bij gelijke taal- en rekenscores). HBO/universitair opge-leide ouders (5.9% tegen 3.0%) en openbaar onderwijs (49.7% tegen 35.3%) komen in Amster-dam meer voor dan in Rotterdam. In Utrecht ligt het percentage laag opgeleide autochtone ouders veel hoger (25.0% tegen 9.3%) dan in Amsterdam. (zie bijlage 1)

Er blijft een onderscheid tussen de vier grote steden en de omringende randstedelijke gemeen-ten, hoewel veel minder scherp dan vóór de controle. Gegeven de hoogte van de ongestan-daardiseerde parameters van de gemeente variabelen is het verschil tussen het advies aan Amsterdamse leerlingen en aan leerlingen uit de steden buiten de Randstad nog steeds het grootst (.47, d.w.z. een half niveau op onze advies variabele), gevolgd door het platteland bui-ten de Randstad (.34), de steden binnen de Randstad (.31) en het platteland binnen de Ran-dstad (.22). De mate van overadvisering is dus ook na controle nog steeds niet triviaal of onbeduidend. Dit wijst op steun aan de zesde hypothese, die in de vier grote steden een asser-tiever cultureel klimaat veronderstelde waardoor de adviezen daar hoger uitvallen. Ook het verschil tussen de adviezen in de vier grote steden en de rest van de Randstad enerzijds en de adviezen in overig Nederland anderzijds blijft bestaan, waarbij het verschil tussen steden en het platteland buiten de Randstad overigens niet groot is. Dit is in strijd met de vijfde hypo-these, die veronderstelde dat de geringere afstand tot de dichtstbijgelegen HAVO en VWO een belangrijke verklaring van overadvisering in alle steden zou zijn.

Onze resultaten weerleggen de tweede hypothese die een oorzaak van de overadvisering in de grote steden legde bij het hoge percentage kinderen van migranten. Uit de negatieve para-meters blijkt dat allochtone leerlingen met laag geschoolde ouders bij gelijke taal- en reken-scores eerder een lager advies krijgen dan een hoger advies.

Onze resultaten weerleggen én bevestigen de eerste hypothese die een oorzaak van de over-advisering legde bij de verdeling van de opleidingsniveaus in de grote steden en daar buiten. Leerlingen met ouders, die een HBO/universitaire opleiding hebben, worden overgeadviseerd, maar ze zijn schaarser in de grote steden (zie bijlage 1): dit lage percentage hooggeschoolde ouders weerlegt de eerste hypothese. Leerlingen met laaggeschoolde autochtone ouders wor-den ondergeadviseerd, maar ze zijn ook schaarser in de grote steden (zie bijlage 1): dit per-centage laaggeschoolde ouders bevestigt de eerste hypothese.

De derde hypothese gaat alleen op voor 'zwarte' scholen in Amsterdam. Maar de derde hypothese gaat niet op voor 'zwarte' scholen buiten Amsterdam of voor arbeiderscholen bin-nen of buiten Amsterdam. Alleen Amsterdamse scholen met veel (gemengde) allochtone leer-lingen én Amsterdamse scholen met veel Turks/Marokkaanse leerlingen geven aan hun leerlingen een hoger advies (twee enige significante interactie-termen in model 2). Niet-Amsterdamse scholen met veel allochtone leerlingen geven hun leerlingen geen hoger advies (de parameters van Turks/Marokkaanse school en gemengde allochtone school zijn niet sig-nificant in model 2). In het kader van de derde hypothese is het van belang erop te wijzen dat overadvisering niet alleen voorkomt in Amsterdamse scholen met veel allochtone leerlingen, maar ook in alle scholen met veel kinderen met hooggeschoolde ouders (elite scholen). Het advies ligt in deze scholen hoger dan in andere scholen met een meer gemengde populatie.

Blijkbaar is een overadvies niet automatisch een indicatie van slecht functionerend onderwijs, maar is het een bepaalde beoordeling van de prestaties van een leerling in zijn maatschappelijke context.

De vierde hypothese (openbare scholen overadviseren vaker) wordt door onze resultaten bevestigd. Dat doen die openbare scholen in Amsterdam even veel als in alle andere steden (geen significante interactie tussen Amsterdam en openbaar onderwijs in model 2). Amsterdam heeft evenwel veel openbare scholen, waardoor de schijn ontstaat dat door specifieke Amsterdamse omstandigheden veel overgeadviseerd wordt.

De resultaten van model 2 laten dus zien dat in Amsterdam geen specifieke processen plaatsvinden die leiden tot extra overadvisering in vergelijking met de andere grote steden. We zien wel dat in de Amsterdamse 'zwarte' scholen meer overadvisering plaatsvindt dan in de 'zwarte scholen' in de andere steden.

Model 3 geeft dezelfde analyse, maar alleen voorzover er een AVO advies werd gegeven. De uitkomst van deze analyse over alleen die leerlingen die een AVO advies gekregen hebben staat in model 3 van tabel 1. De adviezen in Amsterdam zijn significant hoger dan die in Den Haag en Utrecht, en zeker dan in de rest van de Randstad en overig Nederland, maar wijken niet significant af van de adviezen in Rotterdam. De effecten van de overige kenmerken van de leerlingen en hun scholen verschillen niet erg tussen model 2 en 3, wat erop wijst dat de advisering in Amsterdam en Rotterdam binnen de AVO adviezen niet een geheel afwijkend proces is. De interacties tussen Amsterdam en 'zwarte' scholen zijn nu niet meer significant. Dit wijst erop dat de overadvisering in Amsterdamse 'zwarte' scholen vooral betrekking heeft op het verschil tussen VBO en AVO en niet op verschillen binnen het AVO.

Model 4 beperkt dezelfde analyse tot die leerlingen aan wie een HAVO- of VWO advies werd gegeven. De uitkomsten van de analyse met alleen leerlingen met een HAVO of VWO advies staan in model 4 van tabel 1.

Het percentage VWO adviezen ligt, bij gelijke taal- en rekenscores en andere leerling- en schoolkenmerken, in Amsterdam beduidend hoger dan in Rotterdam, Den Haag, en Utrecht. Dit percentage VWO adviezen ligt in deze steden weer hoger dan in de rest van de Randstad, zowel steden als platteland. Het percentage VWO adviezen ligt binnen de Randstad (exclusief de vier grote steden) weer hoger dan in de rest van Nederland, zowel in steden als op het platteland. Opnieuw blijkt het dus onjuist de vier grote steden met de rest van Nederland te vergelijken zoals het Sociaal en Cultureel Rapport deed. De vier grote steden zijn in het beste of ergste geval een extreme variant binnen de Randstad waarbinnen zij eerder knooppunten zijn dan afwijkende eenheden. Tussen de Randstad en de rest van Nederland bestaat een minstens even interessant verschil.

In Amsterdam wordt evenwel niet aan iedereen gemakkelijker een VWO advies gegeven, zoals de twee significante interacties met Amsterdam laten zien. Leerlingen met HBO/universitair geschoolde ouders krijgen in Amsterdam minder snel een VWO advies dan vergelijkbare leerlingen elders in het land, die bij gelijke taal- en rekenscores sneller een VWO advies krijgen. Zijn HBO/universitair geschoolde ouders die met kinderen in Amsterdam blijven wonen een afwijkende categorie die minder hecht aan een hoog advies en daaruit voortvloeiende schoolloopbaan? Of zijn deze adviezen het resultaat van een (democratiserende) neiging van Amsterdamse leerkrachten om kinderen van hooggeschoolde ouders minder het voordeel van de twijfel te gunnen? Ook in een tweede opzicht wijkt het percentage VWO adviezen af van dat elders. Leerlingen op Amsterdamse openbare basisscholen krijgen minder snel een VWO advies dan leerlingen op andere openbare scholen buiten Amsterdam. Met andere woorden: openbare en bijzondere scholen in Amsterdam geven even snel een VWO advies. Elders in de land wordt juist op de openbare scholen bij gelijke taal- en rekenscores sneller een VWO advies gegeven dan op bijzondere scholen. Is er in Amsterdam geen concurrentie tussen openbaar en bijzonder onderwijs?

CONCLUSIE

In de vier grote steden krijgen leerlingen hogere adviezen dan in de rest van de Randstad en dan in de rest van Nederland. Dit blijft waar als wij rekening houden met de verschillen in leerling- en schoolkenmerken. Met name het verschil tussen de vier steden en de steden buiten de Randstad in de hoogte van de uitgebrachte adviezen is duidelijk en niet triviaal. Het verschil tussen de adviezen in de vier grote steden en de rest van de Randstad is minder duidelijk. De vier grote steden lijken in dit opzicht geen uitzonderingen, liggend binnen een geheel andere omgeving, maar lijken eerder knooppunten binnen een Randstad die wat de hoogte van de adviezen betreft afwijkt van de rest van Nederland. De Randstad als conglomeraat van steden zou in dat geval een betere omschrijving zijn dan de tegenstelling tussen de vier grote steden en hun omgeving. Het assertiever cultureel klimaat, waarvan sprake is in hypothese 6, lijkt dus eerder een randstedelijk kenmerk dan een exclusief groot-stedelijk kenmerk. Een deel van de verschillen tussen de grote steden en de rest van Nederland is terug te voeren op de lagere percentages laaggeschoolde autochtone ouders in de grote steden (hypothese 1), de grote percentages kinderen van migranten (hypothese 2) en door het overwicht van het openbaar onderwijs in de vier grote steden (hypothese 4). Het groter aantal scholen met veel slecht presterende (allochtone) leerlingen (hypothese 3) en de nabijheid van VWO en HAVO in de grote steden (hypothese 5) bieden over het algemeen geen verklaring voor overadvisering in de grote steden.

Binnen de vier grote steden en de Randstad zijn de Amsterdamse adviezen in het algemeen niet uitzonderlijk, als men rekening houdt met ouderlijk milieu, schoolcompositie en de richting van de school. Wel zien wij dat op Amsterdamse scholen met veel allochtone leerlingen (hetzij gemengd, hetzij Turks/Marokkaans) eerder dan elders een AVO- in plaats van een VBO-advies wordt gegeven. Verder zien wij dat er in Amsterdam gemakkelijker een VWO-advies in plaats van een HAVO-advies wordt gegeven, overigens niet in de eerste plaats aan kinderen van hoogopgeleide ouders en niet in de eerste plaats aan leerlingen in het openbaar onderwijs. De eerste conclusie van deze analyse voor het Amsterdamse onderwijsbeleid zou daarom kunnen zijn dat het VBO in Amsterdam voor leerlingen van zwarte basisscholen blijkbaar weinig aantrekkelijk is en dus opwaardering zou behoeven. Maar men kan deze uitkomst ook interpreteren als een weigering van leerkrachten op Amsterdamse scholen met veel allochtone leerlingen zich automatisch neer te leggen bij een VBO loopbaan voor hun leerlingen. Een tweede conclusie voor het Amsterdamse onderwijsbeleid zou kunnen zijn dat er in het Amsterdamse onderwijs aan de bovenkant (VWO) wat voorzichtiger geadviseerd kan worden. Hoe het ook zij: de algemene stelling dat er speciaal in Amsterdam over de gehele linie sprake is van overadvisering behoeft nodig nuancering en dus ook het op die stelling te voeren onderwijsbeleid.

* Dit artikel is eerst gepresenteerd op het SCO-KI congres "Opvoeding, Onderwijs en Sociale Integratie", 17-18 oktober 1996, Amsterdam. Alle auteurs zijn werkzaam op het SCO-Kohnstamm Instituut. Correspondentie richten aan M. van Erp, SCO/Kohnstamm Instituut, Universiteit van Amsterdam, Wibautstraat 4, 1091 GM Amsterdam.

1. Het Sociaal en Cultureel Rapport zegt op pagina 325 dat in het Prima-cohort vergelijkbare gegevens voor Amsterdam en Utrecht niet of beperkt beschikbaar zijn. Zoals uit dit artikel waarin Prima-gegevens gebruikt worden, blijkt is deze passage misleidend. Feit is dat alleen in opdracht van Rotterdam en Den Haag een Prima-rapportage op stedelijk niveau heeft plaatsgevonden (ITS, 1995; Roeleveld, 1995). In Amsterdam is een dergelijke opdracht wel verstrekt door enkele stadsdeelraden.

2. Deze terughoudendheid is een merkwaardig verschijnsel omdat elders in het Sociaal en Cultureel Rapport het SCP niet terugschrikt voor voorzichtige speculaties. Lag dit onderwerp soms te gevoelig?

3. Wij geven de voorkeur aan de sociale compositie van de school als controle variabele boven het gemiddelde prestatieniveau van de school om mogelijke afhankelijkheden tussen de meting van de variabelen te voorkomen.
4. Deze volledige steekproef kent een oververtegenwoordiging van OVB-scholen, maar in de analyses zal gecontroleerd worden voor zowel de sociaal-etnische herkomst van de leerlingen als voor de compositie van de school.
5. In de regressie-analyses vatten we de afhankelijke variabele advies op als interval-variabele. De lineaire relatie met de taal en rekentoetsen (Pearson's r .635 resp. .652) is nauwelijks lager dan de totale relatie (η .636 resp. .654).

LITERATUUR

- Bosma, H. & Cremers, P. (1996). Schooladviezen van allochtone leerlingen. *Tijdschrift voor Onderwijsresearch*, 21, 262-271.
- Driessen, G., Jungbluth, P., Langen, A. van & Vierke, H. (1996). *PRIMA-cohortonderzoek. Technische rapportage*. Nijmegen/Amsterdam: ITS/SCO-KI.
- Gemeente Amsterdam (1994). *Een plan voor het onderwijs in Amsterdam 1994-1998*. Amsterdam: Gemeente Amsterdam.
- Hofman, R. H. (1997). Effectieve besturen? Verzuiling en bestuursvorm. In A. B. Dijkstra, J. Dronkers & R. H. Hofman (eds.), *Verzuiling in het onderwijs. Actuele verklaringen en analyse* (p. 271-288) Groningen: Wolters-Noordhoff.
- ITS (1995). *Landelijk vergelijkend gebiedsrapport FAO-scholen Rotterdam, schooljaar 1994/95*. Nijmegen/Amsterdam: ITS/SCO-KI.
- Jungbluth, P., Peetsma, T. & Roeleveld, J. (1996). *Leerlingprestaties en leerlinggedrag in het primair onderwijs*. Nijmegen/Amsterdam: ITS/SCO-KI.
- Koeslag, M. & Dronkers, J. (1994). Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs. *Tijdschrift voor Onderwijsresearch*, 19, 240-258.
- Ledoux, G. (1996). Onderwijs en etnische herkomst. In *Onderwijskundig Lexicon* (p. 61-84) Alphen aan de Rijn: Samson.
- Pijl, Y. L. (1994). De overschatte overadvisering van allochtonen. *Tijdschrift voor Onderwijswetenschappen*, 24, 240-246.
- Roeleveld, J. (1995). *Gebiedsrapport Den Haag in aanvulling op het Landelijk vergelijkend gebiedsrapport, schooljaar 1994/95*. Nijmegen/Amsterdam: ITS/SCO-KI.
- Roeleveld, J. (1997). Concurrentie tussen scholen? Verzuiling en leerlingmarkten. In A. B. Dijkstra, J. Dronkers & R. H. Hofman (eds.), *Verzuiling in het onderwijs. Actuele verklaringen en analyse* (p. 257-269) Groningen: Wolters-Noordhoff.
- Rozendaal, A., Jong, U. de & Eeden, P. van den (1987). *Vergelijking van de schoolloopbanen van leerlingen in het voortgezet onderwijs in Amsterdam en in de rest van Nederland*. Amsterdam: Amsterdams samenwerkingsverband/vakgroep Methoden en technieken VU.
- Saris, W. E. & Blok, H. (1982). Het onderwijzersadvies nader bekeken. *Tijdschrift voor Onderwijsresearch*, 7, 49-59.
- Sociaal en Cultureel Planbureau (1996). *Sociaal en Cultureel Rapport 1996*. Rijswijk: Sociaal en Cultureel Planbureau.
- Vierke, H. (1995). *De PRIMA toetsen gecalibreerd*. Nijmegen: ITS.
- Vrooman, J. C. & Dronkers, J. (1986). Changing educational attainment processes: some evidence from the Netherlands. *Sociology of Education*, 59, 69-78.

Bijlage 1: Kenmerken van Prima-leerlingen in acht typen gemeenten; leerlingen die hebben deelgenomen aan de Taal- en Rekentoetsen en waarvan het advies bekend is

	grote steden				overige steden		platteland	
	a'dam	r'dam	dhaag	utrecht	randst	rest	randst	rest
aantal scholen	30	22	28	10	56	191	62	178
aantal leerlingen	590	462	712	128	996	3431	1135	2739
percentage	5.8	4.5	7.0	1.3	9.8	33.7	11.1	26.9
Soc-etn. herkomst (kolom%)								
max.Lbo; TuMa	40.0	29.9	20.1	23.4	15.4	12.4	3.2	1.1
max.Lbo; ov.all	9.2	26.8	16.6	7.0	4.5	5.0	3.3	1.5
max.Lbo; autocht	9.3	9.7	10.8	25.0	24.0	28.7	25.9	34.0
Mavo/Havo/Vwo/Mbo	11.7	12.8	19.9	21.1	20.1	22.5	25.7	32.1
Hbo/universiteit	5.9	3.0	4.8	7.8	10.7	14.3	18.3	19.2
onbekend	23.9	17.7	27.8	15.6	25.3	17.2	23.6	12.1
schoolcompositie (kolom%)								
eliteschool	-	-	-	13.3	7.4	16.6	19.1	13.6
burgerschool	2.7	7.4	37.5	15.6	22.2	18.8	26.4	42.7
representatieve school	13.1	9.5	3.4	9.4	20.9	12.0	31.7	29.9
arbeidersschool	7.6	-	10.7	23.4	27.5	39.1	14.0	11.5
gemixt allocht. school	19.5	54.1	18.0	21.9	4.0	5.5	7.6	2.3
Turks/Marok. school	57.1	29.0	30.5	16.4	18.0	8.0	1.1	-
openbaar onderwijs (%)	49.7	35.3	35.3	31.3	44.9	31.0	33.6	25.4
Taalscore								
gem.	1088	1090	1095	1099	1107	1110	1118	1121
st.dev.	31	34	37	38	38	39	36	37
Rekenscore								
gem.	1165	1175	1168	1188	1189	1192	1196	1199
st.dev.	43	41	45	47	48	48	45	48
advies (kolom%)								
ivbo	2.4	.9	3.7	.8	2.7	3.3	1.5	1.9
ivbo/vbo	.8	-	3.1	3.1	.6	.9	.7	1.3
vbo	22.9	26.4	22.2	17.2	22.5	25.7	18.7	19.1
vbo/mavo	15.3	16.0	13.1	14.1	16.5	12.1	7.6	10.4
mavo	21.2	21.9	19.7	23.4	20.1	18.8	22.8	17.6
mavo/havo	14.7	10.6	15.6	10.2	11.6	13.2	14.2	14.9
havo	8.6	10.6	11.9	14.8	11.7	10.0	16.5	13.3
havo/vwo	9.2	10.0	6.0	12.5	8.3	9.9	10.1	14.3
vwo	4.9	3.7	4.8	3.9	5.9	6.1	7.9	7.3

DO PUPILS IN THE BIG DUTCH CITIES ESPECIALLY IN AMSTERDAM GET TOO HIGH ADVICES ON THE MOST SUITED SECONDARY SCHOOL?

The relations between language and mathematics scores and teachers' advices within and outside the 'Randstad'.

An important phase in the school career of Dutch pupils is the transfer from primary education to secondary. Next to the scores on national tests for language and mathematics the advice on the teachers of their primary schools on the most suited type of secondary education is an important predictor on the level the pupils will start and end their secondary education. The relations between the language and mathematics scores and the teachers' advices is not equal within the Netherlands: pupils in the four largest cities (Amsterdam, Den Haag, Rotterdam, Utrecht) get higher advices from their teachers than pupils from the rest of the Netherlands, even if the pupils have equal scores on the national tests. In this article we test five possible explanations of this phenomenon with recent data from the PRIMA-cohort. The results show that the main difference lays not between the four largest cities and the rest of the Netherlands, but between the 'Randstad' (Holland and Utrecht) and the other provinces of the Netherlands. The cultural climate in the former is more assertive than in the latter. Other explanations, supported by the analyses, are the lower percentages of autochthon low-education parents in the four largest cities and the dominance of public education in these cities. Other explanations, like the larger percentages of allochthonous pupils, the concentration of certain groups of pupils in different schools and the proximity of schools, are not supported by the analyses.

Hoe verhoog ik de betrouwbaarheid van mijn toetsen?

Effect van moeilijkheid en spreiding van itemmoeilijkheden op de betrouwbaarheid van de afsluitingstoetsen basisvorming

Hans Kuhlemeier, Frans Kleintjes & Peter Goldebeld
Instituut voor Toetsontwikkeling (Cito), Nieuwe Oeverstraat 68, 6801 MG Arnhem,
tel: 026-3521536, Email: hans.kuhlemeier@cito.nl

ABSTRACT

In this article the empirical relationship between test reliability and test difficulty was studied. The analysis has been conducted on the data of the nation-wide administration of 51 tests for basic education in the eight and ninth grade of Dutch secondary education (students aged 14 and 15). Tests were equal for students in each of the five major tracks. The effect of test difficulty and the within-test heterogeneity of item difficulties on test reliability were studied in a multiple regression analysis, taking into account the effect of test length, the percentage of multiple choice items, and the observed test score variation. It is shown that the difficulty of the test and the dispersion of the item difficulties are relatively strong predictors of test reliability. It is demonstrated that the poor test reliability in the 'higher' tracks of secondary education may be ascribed to the fact that the tests were relatively easy for the students in those tracks. Recommendations are given for improving test reliability by altering the procedures of test construction and item selection.

INLEIDING

De basisvorming is een van de meest ingrijpende vernieuwingen in het voortgezet onderwijs sinds de Mammoetwet. Ter ondersteuning van het vernieuwingsproces krijgen de scholen de beschikking over afsluitingstoetsen basisvorming. De eerste generatie afsluitingstoetsen is in december 1994 aan alle scholen voor voortgezet onderwijs ter beschikking gesteld. Dit toetspakket omvatte voor 21 vakken van de basisvorming in totaal 70 toetsen. Voor de meeste vakken werd ten minste één schriftelijke toets en ten minste één zogeheten praktijktoets ter beschikking gesteld (Kremers, 1994). Elke toets is bedoeld voor de volle breedte van het voortgezet onderwijs (Procesmanagement Basisvorming, 1995). De afsluitingstoetsen zijn met andere woorden gelijk voor het ivbo, vbo, mavo, havo en vwo.

De toetsen hebben een schoolinterne functie. Ze laten zien wat leerlingen van de basisvorming hebben opgestoken. In de onderwijspraktijk fungeren de toetsen veelal als 'gewone' proefwerken waarvoor cijfers worden gegeven die meetellen voor het rapport, de overgang en de keuze van het type vervolgonderwijs. De toetsen dienen dan ook zo betrouwbaar mogelijk te zijn. Anno 1997 is de eerste generatie afsluitingstoetsen massaal afgenomen bij tweede- en derdeklassers van het eerste basisvormingscohort en bij tweedeklassers van het tweede basisvormingscohort.¹ Een unieke gelegenheid om de betrouwbaarheid van deze toetsen te evalueren in relatie tot de gevolgde procedure van toetsconstructie. Dit artikel is gericht op

aanbevelingen voor toetsconstructie teneinde de betrouwbaarheid - gegeven de onveranderbare randvoorwaarden - te maximaliseren.

Hoe zijn de afsluitingstoetsen geconstrueerd? De commissie die de opgaven vaststelt, de Commissie Opgaven Basisvorming (COB), heeft het Cito gevraagd de toetsen zo samen te stellen dat leerlingen van het voorbereidend beroepsonderwijs gemiddeld ten minste de helft van het aantal te behalen punten scoren. In concreto: de gemiddelde moeilijkheidsgraad van een afsluitingstoets moest in het voorbereidend beroepsonderwijs minimaal 50% zijn ($\bar{p} > 50\%$ goed). Tegelijkertijd moest het Cito vermijden dat de items door grote aantallen vwo-leerlingen allemaal goed gemaakt zouden worden (Commissie Opgaven Basisvorming, 1994).

Het Cito heeft getracht de opdracht van de COB zo goed mogelijk uit te voeren. Ten eerste is tijdens de constructie van de items doorgaans 'gemikt' op hetgeen een gemiddelde (mavo-)leerling in het voortgezet onderwijs na meestal drie jaar onderwijs zou moeten kennen en kunnen. De concept-items zijn vervolgens bij een omvangrijke groep scholen uit het (i)vbo, mavo, havo en vwo gepretest. Uit het arsenaal van gepreteste opgaven zijn vooral die opgaven in de uiteindelijke afsluitingstoetsen opgenomen die noch te moeilijk zijn voor het (i)vbo, noch te gemakkelijk voor het vwo. Waar mogelijk en zinvol zijn die opgaven gekozen waarbij de 'kloof' tussen (i)vbo en vwo het kleinst is, zonder overigens concessies te doen aan onder meer de vakinhoudelijke dekking van de kerndoelen (Sluiter, 1994).

Ten tweede heeft het Cito getracht een grote spreiding naar moeilijkheidsgraad van de opgaven in de toetsen aan te brengen. De afsluitingstoetsen moesten immers geschikt zijn voor de 'volle breedte' van het voortgezet onderwijs. Die ene uniforme toets voor allen moest met andere woorden voor 'elk wat wils' bevatten. Dus zowel gemakkelijker opgaven die ook voor het (i)vbo te doen zouden zijn, als ook moeilijker opgaven die voor het vwo nog voldoende uitdaging zouden bieden.

Tegen deze procedure van item- en toetsconstructie zijn bedenkingen aan te voeren van toetstechnische aard. De eerste betreft de relatie tussen de moeilijkheid en de betrouwbaarheid van de toets. Theoretisch bestaat er een curvilineaire relatie tussen beide (Guilford & Fruchter, 1978; Feldt, 1993). Ceteris paribus is de betrouwbaarheid van toetsen met dichotoom gescoorde opgaven lager naarmate de gemiddelde moeilijkheid van een toets sterker van 50% goed afwijkt (als we raden voor het gemak even buiten beschouwing laten). Zeer gemakkelijke en zeer moeilijke toetsen zijn in theorie minder betrouwbaar dan toetsen met een gemiddelde moeilijkheidsgraad van 50%. De resultaten van de afname van de eerste generatie afsluitingstoetsen bij tweede- en derdeklassers in de schooljaren 1994/95 en 1995/96 laten zien dat het percentage goed gemaakte opgaven zeer sterk verschilt van opleidingstype tot opleidingstype.² In het ivbo/vbo, vbo en vbo/avo ligt de gemiddelde moeilijkheidsgraad met respectievelijk 47%, 52% en 53% dicht in de buurt van de toetstechnisch optimale 50%. In het ivbo maken de leerlingen daarentegen slechts 37% van de opgaven goed, in het mavo+ gaat het om 63% van de opgaven en in het havo+ en vwo+ betreft het respectievelijk 69% en 75%. De afstand tot de 50% goed doet vermoeden dat de betrouwbaarheid in het ivbo/vbo, vbo en vbo/avo het hoogst zal zijn en in het havo+ en het vwo+ het laagst (Kuhlemeier, Kremers & Kleintjes, 1996; 1997).

De tweede bedenking betreft de welbewust aangebrachte spreiding van de itemmoeilikheden in de toetsen. Alhoewel vanuit de toenmalige situatie begrijpelijk, is een grote spreiding om toetstechnische redenen niet optimaal. Theoretisch wordt namelijk een lagere betrouwbaarheid verwacht naarmate de moeilijkheden van (dichotoom gescoorde) opgaven sterker spreiden rond de 50% goed (Horst, 1953; Guilford & Fruchter, 1978; Feldt, 1993). Men mag derhalve verwachten dat de afsluitingstoetsen met een sterkere spreiding van de itemmoeilikheden wat minder betrouwbaar zijn dan die met een minder grote spreiding.

Uiteraard zijn de moeilijkheid en de spreiding van de itemmoeilikheden niet de enige toetskenmerken die een toetsontwikkelaar kan manipuleren om de betrouwbaarheid te verhogen. Een derde factor is de toetslengte. In het algemeen geldt: hoe meer items, hoe betrouwbaarder de toets. Dit geldt echter alleen bij homogene testverlenging. Dit wil zeggen: de toege-

voegde items moeten qua inter-itemcorrelaties en moeilijkheidsgraad vergelijkbaar zijn met de items waaraan ze worden toegevoegd (Guilford & Fruchter, 1978). Overigens is de toetslengte in de praktijk slechts tot op zekere hoogte te beïnvloeden. Een van de randvoorwaarden is dat voor afname van een afsluitingstoets doorgaans slechts één lesuur beschikbaar is.

Een vierde toetskenmerk is de verhouding tussen open en gesloten vragen. Open vragen leveren vaak een wat grotere bijdrage aan de toetsbetrouwbaarheid dan gesloten vragen (Van den Bergh, 1988; Frary, 1985), en dit blijkt ook bij de eerste generatie afsluitingstoetsen het geval (Kuhlemeier, Kremers & Kleintjes, 1997). Dit wil overigens niet zeggen dat open vragen te allen tijde de voorkeur verdienen boven gesloten vragen. Omdat het beantwoorden van een open vraag relatief veel afnametijd kost, zijn gesloten vragen meestal efficiënter dan open vragen (Goldeveld, 1995, 1996). Daarnaast kunnen het type toets en de scoringswijze een rol spelen. De meeste schriftelijke afsluitingstoetsen zijn (semi-) objectief scorebaar, maar bij de praktijktoetsen wordt meestal een sterk beroep gedaan op het menselijk oordeelsvermogen (vgl. Meuffels, 1994). Als de overeenstemming tussen beoordelaars minder dan honderd procent is, wordt de betrouwbaarheid van de scores overschat (Goldeveld, 1992; 1995).

De moeilijkheid, spreiding van itemmoeilijkheden, toetslengte en vraagtype kan de toetsontwikkelaar tot op zekere hoogte manipuleren. Dit geldt echter niet voor de populatie bij wie de toets is afgenomen en de afnamesituatie. Verschillen tussen leerlingen in bijvoorbeeld motivatie, testwijsheid en spiekgedrag kunnen bijdragen tot de toevallige meetfouten, waardoor de betrouwbaarheid daalt (Ebel & Frisbie, 1991). Een andere factor die buiten de invloedssfeer van de toetsontwikkelaar valt, is de variabiliteit van de vaardigheid in de populatie. Voor vakken als wiskunde, economie, Engels en geschiedenis lijken de verschillen tussen leerlingen bijvoorbeeld veel groter dan voor vakken als handenarbeid, techniek, verzorging en muziek (Kuhlemeier, Kremers & Kleintjes, 1997). Naarmate de vaardigheidsverschillen tussen leerlingen groter zijn, is een hoge betrouwbaarheid makkelijker te realiseren (Guilford & Fruchter, 1978, p. 419 en 420). Ook de samenstelling van de groep leerlingen aan wie de toets is voorgelegd kan een rol spelen. Zo is de geobserveerde scorevariantie in de totale responsgroep van leerlingen uit het voortgezet onderwijs vrijwel altijd groter dan in elk van de afzonderlijke opleidingstypen ivbo, ivbo/vbo, vbo, vbo/mavo, mavo+, havo+ en vwo+.² Evenzo is de heterogeniteit in de 'lagere' opleidingstypen van het voortgezet onderwijs doorgaans groter dan in de 'hogere' opleidingstypen. Een gerapporteerde toetsbetrouwbaarheid moet derhalve altijd worden geïnterpreteerd in het licht van de samenstelling van de groep aan wie de toets is voorgelegd.

Naast de (schaalonafhankelijke) betrouwbaarheid kent de klassieke testtheorie de standaardmeetfout als (schaalafhankelijke) maat voor de nauwkeurigheid.³ De standaardmeetfout wordt gebruikt om aan te geven tussen welke grenzen de ware score van een persoon ligt (De Groot & Van Naerssen, 1969). Een belangrijke vraag is of leerlingen in de diverse opleidingstypen met eenzelfde nauwkeurigheid gemeten worden. De afsluitingstoetsen basisvorming moeten immers geschikt zijn voor de volle breedte van het voortgezet onderwijs. De standaardmeetfout is afhankelijk van de spreiding van de vaardigheid in de populatie en de betrouwbaarheid. Is de betrouwbaarheid nul, dan is de standaardmeetfout even groot als de geobserveerde spreiding van de scores. Is de betrouwbaarheid 1.00, dan is de standaardmeetfout gelijk aan nul. De Groot en Van Naerssen (1969, p. 245) vermelden dat de foutenvariantie over een breed gebied in het midden van de schaal bij benadering constant is, alleen aan de uiteinden van de schaal zal afnemen en men er in de praktijk slechts weinig naast zit als men de standaardmeetfout als constant beschouwd over (het grootste deel van) het scoregebied. Op grond hiervan veronderstellen we een zwak kromlijng verband tussen de standaardmeetfout en de moeilijkheid van de toets. Ten aanzien van de verschillen tussen de zeven opleidingstypen nemen we aan dat de standaardmeetfout kleiner zal zijn naarmate de gemiddelde moeilijkheidsgraad in het opleidingstype in kwestie verder van 50% goed afwijkt.

In dit artikel wordt de empirische relatie tussen enerzijds de betrouwbaarheid en anderzijds de moeilijkheid en de spreiding van de itemmoeilijkheden nader onderzocht. Nagegaan

wordt in hoeverre de hoge en lage moeilijkheid van de toetsen in respectievelijk de 'onderkant' en de 'bovenkant' van het voortgezet onderwijs de betrouwbaarheid heeft aangetast. Ook wordt onderzocht in hoeverre de betrouwbaarheid van de afsluitingstoetsen mogelijk te lijden heeft gehad van de welbewust aangebrachte spreiding van de itemmoeilijkheden. Het doel is informatie te verkrijgen op grond waarvan de betrouwbaarheid van de afsluitingstoetsen in de toekomst verhoogd kan worden. Om het empirische effect van moeilijkheid en itemspreiding op de betrouwbaarheid zo zuiver mogelijk vast te stellen wordt de relatie tussen betrouwbaarheid en moeilijkheid geanalyseerd onder constanthouding van de toetslengte, het percentage gesloten vragen, de scoringswijze (schriftelijke versus praktijktoets) en de totale geobserveerde scorevariantie. Ten slotte wordt verslag gedaan van een analoge analyse met de standaardmeetfout als afhankelijke variabele.

METHODE

Steekproef en proefpersonen

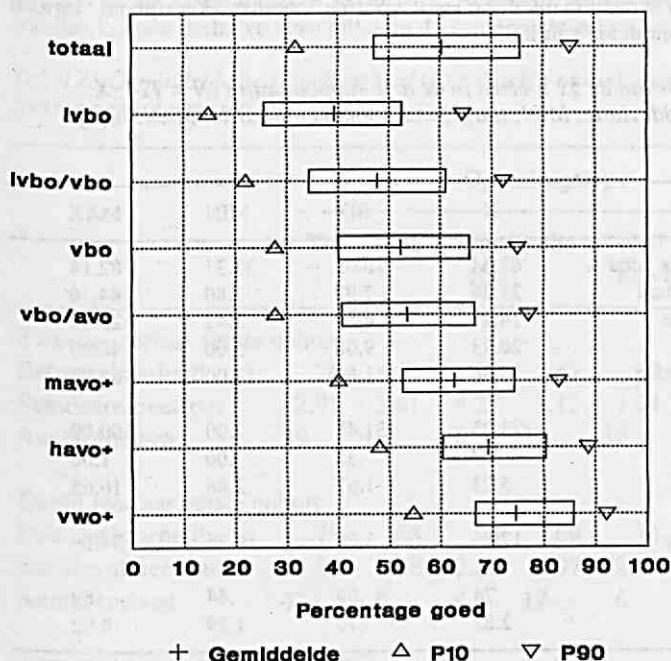
In de schooljaren 1994/95 en 1995/96 hebben bij tweede- en derdeklassers van het eerste en tweede basisvormingscohort in totaal 142915 afnames plaatsgevonden (zie Kuhlemeier, Kremers & Kleintjes, 1996; 1997). Voor dit artikel zijn gegevens van drie subpopulaties gebruikt: tweedeklassers van het eerste basisvormingscohort (schooljaar 1994/95), derdeklassers van het eerste basisvormingscohort (schooljaar 1995/96) en tweedeklassers van het tweede basisvormingscohort (eveneens schooljaar 1995/96). Van elke afname zijn de scores van één leerling voor analyse beschikbaar, te weten de alfabetisch zevende leerling binnen de klas. De afname vond plaats door de 'eigen' vakdocent. De toetsen zijn afgenomen in het kader van de afsluiting van de basisvorming. De scholen mogen het moment van afsluiting grotendeels zelf bepalen. Dientengevolge varieert ook het moment van afname tussen vakken, toetsen en opleidingstypen. In verband hiermee zijn de gegevens over de moeilijkheid en betrouwbaarheid niet representatief voor de populaties van tweede- of derdeklassers in Nederland, maar hooguit voor de populaties van tweede- en derdeklassers aan wie de afsluitingstoetsen zijn voorgelegd (vgl. Kuhlemeier, Kremers & Kleintjes, 1997). Figuur 1 laat zien hoe goed de leerlingen de opgaven gemiddeld maakten. Het gemiddelde percentage goed gemaakte opgaven per opleidingstype is weergegeven met een plusteken. De horizontale staaf met dit gemiddelde geeft het gebied aan waarbinnen zich 50% van de leerlingen bevindt. De beide driehoekjes Δ (P10) en ∇ (P90) begrenzen het gebied waarbinnen zich 80% van de leerlingen bevindt. Derhalve bevindt zich buiten de driehoekjes aan weerszijden nog 10% van de leerlingen.

De gemiddelde leerling die de toetsen maakte, heeft gemiddeld 60% van de hem of haar voorgelegde opgaven goed gemaakt. Dit percentage varieert van 39% in het ivbo tot 75% in het vwo+. In het ivbo/vbo, vbo en vbo/avo ligt de gemiddelde moeilijkheid dicht in de buurt van de toetstechnisch wenselijke 50%, maar met name in het havo+ en vwo+ ligt ze daar ver vandaan. Figuur 1 laat ook zien dat de verschillen tussen leerlingen in de 'hogere' opleidingstypen wat kleiner zijn dan in de 'lagere' opleidingstypen.

Dataset

Het onderzoek omvat in totaal 51 verschillende afsluitingstoetsen van 18 vakken van de basisvorming. Het betreft 38 schriftelijke toetsen en 13 zogeheten praktijktoetsen. De moeilijkheid en betrouwbaarheid zijn berekend per toets per subpopulatie. Daarbij zijn de gegevens van toetsen die in de desbetreffende subpopulatie minder dan vijftig keer zijn afgenomen niet meegenomen. Het voor de analyse gebruikte bestand bevat in totaal 124 'records': de gegevens van 39 toetsen in het tweede leerjaar van het eerste basisvormingscohort, van 43 toetsen in het derde leerjaar van het eerste cohort en van 42 toetsen in het tweede leerjaar van het tweede cohort.

Figuur 1: Gemiddeld percentage goed en spreiding per opleidingstype



De toetsen

Een overzicht met enkele beschrijvende toetskenmerken is weergegeven in Tabel 1.

De moeilijkheid van de toetsen - het percentage goed gemaakte opgaven - varieert van 35% tot 82% met een gemiddelde van 63% (zie Tabel 1). Dit komt overeen met een gemiddelde ruwe toetscore van 21.3 punten.

De afsluitingstoetsen verschillen sterk in de spreiding van de itemmoeilijkheden binnen de toets. Voor de 'gemiddelde' afsluitingstoets is het verschil tussen het moeilijkste en makkelijkste item 36%. Voor de 17% meest homogene toetsen is het verschil tussen het meest en minst moeilijke item gemiddeld 20% en voor de 17% meest heterogene toetsen is het overeenkomstige percentage 68%. De standaarddeviatie van de itemmoeilijkheden binnen de toets varieert van 3.4 tot 23.5 met een gemiddelde van 14.3.

Verder laat Tabel 1 onder meer zien dat de gemiddelde afsluitingstoets die is afgenomen 21 items telt, voor 23% uit gesloten vragen bestaat en in 17% van de gevallen een praktijktoets betreft.

De gemiddelde standaarddeviatie van de behaalde toetsscores (het aantal behaalde scorepunten) is 5.7. Daarnaast is een soort gestandaardiseerde standaarddeviatie gerapporteerd die vergelijkbaar is over verschillende toetsen. Die is berekend door de standaarddeviatie te nemen van het percentage goed gemaakte opgaven. Deze laatste maat varieert van 10.3 tot 34.9 met een gemiddelde van 17.9.

De betrouwbaarheid is vastgesteld aan de hand van Cronbach's alpha, een ondergrensschatting van de ware betrouwbaarheid. Deze zeer veel gebruikte maat geeft aan hoe homogeen of intern consistent de toets is. De gemiddelde betrouwbaarheid van de toetsen waarvoor in de drie subpopulaties voldoende waarnemingen beschikbaar waren, bedraagt .74. Als de vaardigheid met één toets gemeten is, wordt .85 wel genoemd als de vereiste ondergrens. Wordt de vaardigheid met meer toetsen en/of op meer momenten gemeten, dan wordt .65 wel

als de gewenste ondergrens genoemd (vgl. Veldhuijzen, Goldebelld & Sanders, 1993, p. 61). De standaardmeetfout, een schaalafhankelijke maat voor de meetnauwkeurigheid, varieert van 1.3 tot 4.9 met een gemiddelde van 2.82.

Tabel 1: Enkele kenmerken van de 51 toetsen in de drie subpopulaties (N = 124; X: gemiddelde; SD: standaarddeviatie; MIN: laagst waargenomen waarde; MAX: hoogst waargenomen waarde)

	X	SD	MIN	MAX
Moeilijkheid: percentage items goed	63.34	10.03	35.33	82.14
Moeilijkheid: behaalde toetsscore	21.26	7.97	7.80	44.56
Stand. dev. itemmoeilijkheden	14.27	4.50	3.42	23.50
Toetslengte: aantal items	20.73	9.08	5.00	40.00
Toetslengte: maximaal te behalen toetsscore	35.86	11.60	13.00	67.00
Percentage gesloten vragen	23.23	31.48	.00	100.00
Proportie praktijktoetsen	.17	.38	.00	1.00
Stand. dev. van de behaalde toetsscore	5.73	1.67	2.06	10.65
Stand. dev. van het percentage items goed	17.85	3.70	10.26	34.89
Betrouwbaarheid	.74	.09	.44	.89
Standaardmeetfout	2.82	.76	1.29	4.92

Statistische analyse

De relatie tussen enerzijds de betrouwbaarheid en de standaardmeetfout en anderzijds de moeilijkheid en de itemheterogeniteit van de afsluitingstoetsen is bestudeerd met behulp van een multiple regressie-analyse. Geanalyseerd is de regressie op de moeilijkheid en de spreiding van de itemmoeilijkheden onder constanthouding van het aantal items (de toetslengte), het percentage gesloten vragen, de wijze van scoring (schriftelijke versus praktijktoets) en de totale geobserveerde scorevariantie.⁴

Om rekentechnische redenen is Cronbach's alpha in de statistische analyse met honderd vermenigvuldigd. In de bespreking van de resultaten rapporteren we de betrouwbaarheid evenwel op de oorspronkelijke schaal van .00 tot 1.00.

RESULTATEN

Betrouwbaarheid per opleidingstype

Tabel 2 toont de betrouwbaarheden in de totale responsgroep en per opleidingstype. Op grond van de gepresenteerde prestaties per opleidingstype verwachtten we een relatief lage betrouwbaarheid in het havo+ en vwo+ en een naar verhouding hoge betrouwbaarheid in het ivbo/vbo, vbo en vbo/avo. Tabel 2 laat zien dat de gemiddelde betrouwbaarheid in het havo+ en vwo+ inderdaad het laagst is. Bovendien bevindt het mavo+ zich volgens de verwachting tussen enerzijds het havo+ en vwo+ en anderzijds het ivbo/vbo, vbo en vbo/avo. Anders dan verwacht lijkt de gemiddelde betrouwbaarheid in het ivbo niet altijd lager dan in het ivbo/vbo, vbo en vbo/avo. Toetsing van de verschillen met het opleidingstype vbo/avo als contrast wijst uit dat de betrouwbaarheid in de opleidingstypen mavo+, havo+ en vwo+ inderdaad lager is dan in het vbo/avo. Opgemerkt zij evenwel dat het aantal en de aard van de afgenomen toet-

sen niet in elk opleidingstype gelijk zijn. Achter de geconstateerde betrouwbaarheidsverschillen kunnen derhalve verschillen in de aard van de gebruikte toetsen schuil gaan.

Tabel 2: Gemiddelde betrouwbaarheid (Cronbach's alpha), standaardmeetfout en aantal toetsen per subpopulatie per opleidingstype.

	Opleidingstype							
	Tot	ivbo	ivbo/ vbo	vbo	vbo/ mavo	mavo+	havo+	vwo+
Tweede leerjaar eerste cohort								
Betrouwbaarheid	.76	.71	.69	.67	.70	.64	.61	.60
Standaardmeetfout	2.91	3.41	3.27	3.12	3.04	2.99	2.86	2.67
Aantal toetsen	39	19	20	32	30	31	23	29
Derde leerjaar eerste cohort								
Betrouwbaarheid	.73	.68	.73	.68	.74	.63	.60	.58
Standaardmeetfout	2.69	2.86	2.80	2.97	2.69	2.91	2.85	2.58
Aantal toetsen	43	6	4	17	6	34	31	31
Tweede leerjaar tweede cohort								
Betrouwbaarheid	.73	.70	.70	.67	.69	.62	.60	.61
Standaardmeetfout	2.87	3.41	3.71	3.18	3.21	2.94	2.95	2.76
Aantal toetsen	42	19	10	24	24	30	17	24

Betrouwbaarheid, moeilijkheid en itemspreiding

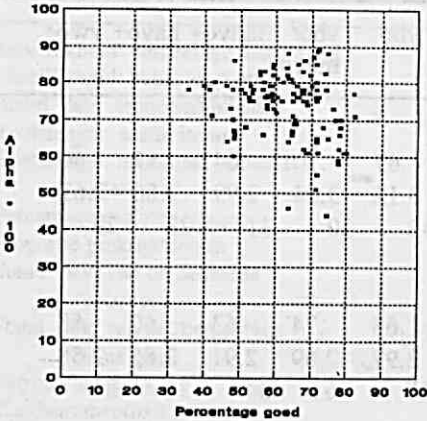
De relatie tussen de betrouwbaarheid en de moeilijkheid van de 51 toetsen in de drie subpopulaties is visueel weergegeven in Figuur 2A. Er lijkt inderdaad sprake van een omgekeerd U-vormig verband tussen betrouwbaarheid en moeilijkheid. In de regressie-analyse zal dan ook worden uitgegaan van een model met een lineaire en een kwadratische term voor de regressie van betrouwbaarheid op moeilijkheid.

Figuur 2B toont de relatie tussen de betrouwbaarheid en de standaarddeviatie van de itemmoeilijkheden. Zoals verwacht zien we een negatief lineair verband tussen beide: naarmate de items in de toets qua moeilijkheid heterogener zijn, lijkt de betrouwbaarheid te dalen.

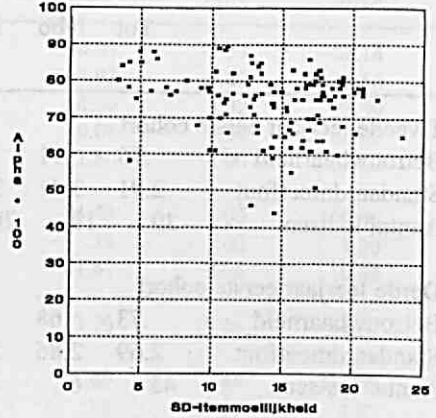
De relatie tussen de moeilijkheid en de betrouwbaarheid van de afsluitingstoetsen is geanalyseerd met behulp van een multiple regressie-analyse. Geanalyseerd is de regressie van de toetsbetrouwbaarheid op de moeilijkheid van de toets en de spreiding van de itemmoeilijkheden, onder constanthouding van het aantal items (de toetslengte), het percentage gesloten vragen, de wijze van scoring (schriftelijke versus praktijktoets) en de (gestandaardiseerde) spreiding van de toetsscores. Tabel 3 geeft de uitkomsten van het model met als predictoren de lineaire en kwadratische term voor de moeilijkheid en lineaire termen voor de standaarddeviatie van de itemmoeilijkheden, het aantal items, het percentage gesloten vragen, de scoringswijze en de standaarddeviatie van de toetsscores.

Allereerst blijkt uit Tabel 3 dat zes van de zeven predictoren significant zijn op 5%-niveau. Het regressiegewicht voor de scoringswijze is niet significant en is om deze reden uit de regressievergelijking verwijderd. Het tot zes predictoren teruggebrachte model blijkt in totaal 45% van de variantie in de betrouwbaarheid van de toetsen te verklaren.

Figuur 2A: Relatie tussen betrouwbaarheid en moeilijkheid



Figuur 2B: Relatie tussen betrouwbaarheid en spreiding itemmoeilijkheid



*Tabel 3: Regressie van de betrouwbaarheid (Cronbach's alpha * 100) op de moeilijkheid (lineair en kwadratisch effect van het percentage goed), de spreiding van de itemmoeilijkheden, het aantal items, het percentage gesloten vragen en de geobserveerde scorevariantie (de standaarddeviatie van het percentage goed) (b: regressiegewicht; se: standaardfout; t: t-toets; R²: percentage verklaarde variantie).*

	b	se	t	R ² _{totaal}	R ² _{exclusief}
Percentage verklaarde variantie				45.3%	
Constate	16.134				
Moeilijkheid lineair	1.687	.596	2.830		31.0%
Moeilijkheid kwadratisch	-.016	.005	-3.367		
Stand. dev. itemmoeilijkheden	-.956	.156	-6.139		28.3%
Aantal items	.600	.092	6.556		25.9%
Percentage gesloten vragen	-.058	.023	-2.452		43.0%
Stand. dev. (% items goed)	1.160	.176	6.584		27.5%

De laatste kolom toont de voorspelde variantie (R²_{exclusief}) na verwijdering van de desbetreffende predictor uit de regressievergelijking (waarbij het lineaire en kwadratische effect van moeilijkheid zijn samengenomen). Het verschil tussen R²_{totaal} en R²_{exclusief} geeft een indruk van het relatieve belang van de zes predictoren voor de toetsbetrouwbaarheid. Uit Tabel 3 kunnen we opmaken dat het lineaire en kwadratische effect van de moeilijkheid, de spreiding

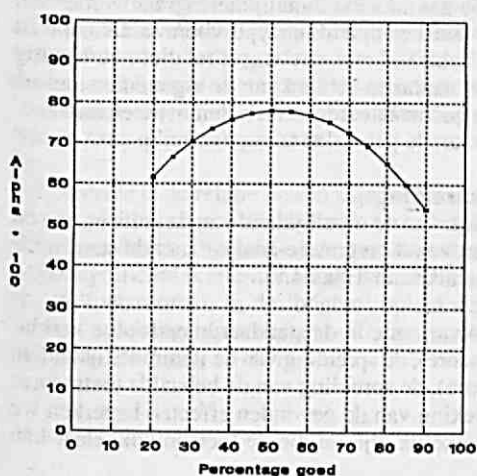
ding van de itemmoeilijkheid, het aantal items en de geobserveerde scorevariantie van relatief groot belang zijn voor de toetsbetrouwbaarheid. In de bespreking van de gevonden effecten beperken we ons tot de variabelen die de toetsontwikkelaar kan beïnvloeden.

Betrouwbaarheid en moeilijkheid. Onder constanthouding van de spreiding van de itemmoeilijkheden, het aantal items per toets, het percentage gesloten vragen en de geobserveerde scorevariantie is zowel de lineaire als de kwadratische term voor de regressie van betrouwbaarheid op moeilijkheid significant op 5%-niveau. Figuur 3A toont de op grond van het model geschatte regressie van betrouwbaarheid op moeilijkheid.

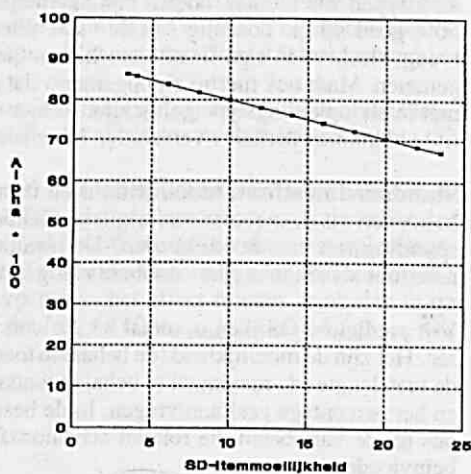
Bij de laagste waargenomen moeilijkheid (35%) bedraagt de op grond van het model geschatte betrouwbaarheid .74 en bij de hoogste waargenomen moeilijkheid (82%) is de geschatte betrouwbaarheid .63. Bij de toetstechnisch optimale moeilijkheid van 50% is de geschatte betrouwbaarheid zoals verwacht het hoogst, namelijk .78. Naarmate de afstand tot die 50% groter wordt, daalt ook de betrouwbaarheid. Zo is de geschatte betrouwbaarheid bij 30% goed .70, bij 40% goed .76, bij 60% goed .77, bij 70% goed .73, bij 80% goed .65 en bij 90% goed is de betrouwbaarheid dus duidelijk lager, maar in het zeer brede middengebied lijkt de moeilijkheid er nauwelijks toe te doen.

Ter controle is de relatie tussen betrouwbaarheid en moeilijkheid nogmaals geanalyseerd, maar nu in elk van de zeven opleidingstypen afzonderlijk. In de opleidingstypen ivbo, ivbo/vbo, vbo, vbo/avo en mavo+ bleek het algebraïsche teken van het gewicht voor de regressie van betrouwbaarheid op moeilijkheid positief, maar niet significant (op 5%-niveau). In deze opleidingstypen is er derhalve geen relatie tussen betrouwbaarheid en moeilijkheid aantoonbaar. In het havo+ en vwo+ was de regressie van betrouwbaarheid op moeilijkheid daarentegen negatief en bovendien significant (b is respectievelijk $-.33$ en $-.55$). In het havo+ en vwo+ neemt de betrouwbaarheid dus af naarmate de toets (nog) gemakkelijker is. Kennelijk is het feit dat de toetsen voor de 'bovenkant' van het voortgezet onderwijs erg gemakkelijk waren ten koste gegaan van de betrouwbaarheid.

Figuur 3A: Regressie van betrouwbaarheid op moeilijkheid



Figuur 3B: Regressie van betrouwbaarheid op spreiding itemmoeilijkheid



Betrouwbaarheid en itemspreiding. Gegeven de overige predictoren heeft de standaarddeviatie van de itemmoeilijkheden een significant effect op de betrouwbaarheid. In het algemeen daalt de betrouwbaarheid met bijna .01 als de standaarddeviatie met één punt toeneemt (zie Figuur 3B). Bij de laagste waargenomen itemspreiding (3.4) bedraagt de op grond van het model geschatte betrouwbaarheid .86 en voor de hoogste waargenomen itemspreiding (23.5) is de geschatte betrouwbaarheid .67. Wat de spreiding van de itemmoeilijkheden betekent voor de betrouwbaarheid lichten we toe aan de hand van een toets met een moeilijkheid van 60% goed. Als we aannemen dat de itemmoeilijkheden normaal verdeeld zijn, betekent een itemspreiding van 3.4 dat bijna alle itemmoeilijkheden tussen 53% en 67% liggen. De betrouwbaarheid daalt met .05 als de spreiding van de itemmoeilijkheden 5 punten hoger (dus 8.4) zou zijn. Dan liggen de itemmoeilijkheden tussen 43% en 77%. De betrouwbaarheid zou nog eens .05 dalen als de itemspreiding 13.4 zou zijn. In dat geval liggen de itemmoeilijkheden tussen 33% en 87%. Een itemspreiding van 23.5 betekent dat vrijwel alle mogelijke itemmoeilijkheden voorkomen, dus van net boven de 0% tot bijna 100%.

Betrouwbaarheid en toetslengte. Gegeven de overige predictoren heeft de toetslengte een significant effect op de betrouwbaarheid. Gemiddeld gaat een toename van de toetslengte met tien items gepaard met een stijging van de betrouwbaarheid met .06. De volgens het model geschatte betrouwbaarheid bedraagt voor de kortste toets van vijf items .66 en voor de langste toets van veertig items .87. De gemiddelde betrouwbaarheid van de toetsen was .74 bij een gemiddelde toetslengte van 20.7 items. Bij verlenging met tien items komt de betrouwbaarheid op .80. Deze waarde zou ook worden gevonden bij toepassing van de formule van Spearman-Brown voor homogene testverlenging (vgl. Veldhuijzen, Goldebeld & Sanders, 1993).

Betrouwbaarheid en vraagtype. Het percentage gesloten vragen levert nauwelijks enige extra bijdrage aan de betrouwbaarheid. Kennelijk maakt het - gegeven de moeilijkheid, de itemspreiding, het aantal items en de geobserveerde scorevariantie - weinig uit of het percentage gesloten vragen nu groot of klein is (of dat het een (semi-)objectief scorebare schriftelijke toets betreft dan wel een door beoordelaars beoordeelde praktijktoets).

Standaardmeetfout per opleidingstype

In de inleiding veronderstelden we een zwak kromlijnijs verband tussen de standaardmeetfout en de moeilijkheid van de toets in de zin dat de meetnauwkeurigheid aan de uiteinden van de schaal wat lager zou zijn. Bijgevolg zou de meetnauwkeurigheid in de onderscheiden opleidingstypen wat kleiner moeten zijn naarmate de gemiddelde moeilijkheidsgraad verder van 50% goed afligt. Toetsing van de verschillen met het opleidingstype vbo/avo als contrast brengt slechts één significant verschil aan het licht: in het vwo+ wordt relatief nauwkeurig gemeten. Maar ook hierbij zij opgemerkt dat het aantal en de aard van de afgenomen toetsen niet in elk opleidingstype gelijk zijn. Achter de geconstateerde verschillen in meetnauwkeurigheid kunnen derhalve verschillen in de aard van de gebruikte toetsen schuil gaan.

Standaardmeetfout, moeilijkheid en itemspreiding

In hoeverre is de meetnauwkeurigheid afhankelijk van de moeilijkheid van de toets en van de spreiding van itemmoeilijkheden? De resultaten van de regressie-analyse met de standaardmeetfout als afhankelijke variabele zijn gepresenteerd in Tabel 4.⁵

Vijf predictoren blijken in totaal 85.2% van de variantie in de standaardmeetfout te verklaren. Het zijn de moeilijkheid (de behaalde toetsscore), de spreiding van de itemmoeilijkheden, de toetslengte (de maximaal te behalen toetsscore), de spreiding van de behaalde toetsscores en het percentage gesloten vragen. In de bespreking van de gevonden effecten beperken we ons tot de variabelen die relatief schaalonafhankelijk zijn en die de toetsontwikkelaar kan beïnvloeden.

Standaardmeetfout en moeilijkheid. Gegeven de overige predictoren bestaat tussen de stan-

daardmeefout en de moeilijkheid van de toets (d.w.z.: de gemiddeld behaalde score op de toets) een lineair verband. Het algebraïsche teken van het regressiegewicht is negatief. In het bereik van de hogere toetscores wordt derhalve nauwkeuriger gemeten dan in de lagere regionen van de schaal. Neemt de behaalde toetscore (op een schaal van 7.8 tot 44.6) met tien punten toe, daalt de standaardmeefout met ruim tweetiende punt (op een schaal van 1.3 tot 4.9 met een gemiddelde van 2.8 en een standaarddeviatie van .76). De standaardmeefout lijkt dus relatief onafhankelijk van de behaalde toetscore ofwel relatief constant voor groepen van verschillend vaardigheidsniveau (vgl. De Groot & Van Naerssen, 1969; Mellenberg, 1996). Het in theorie zwak kromlijng verband tussen de standaardmeefout en de moeilijkheid van de toets bleek overigens niet aantoonbaar. Dit heeft mogelijk te maken met het gegeven dat uitzonderlijk moeilijke toetsen in de dataset ontbreken.

Tabel 4: Regressie van de standaardmeefout op de moeilijkheid (lineair effect van het aantal items goed), de standaarddeviatie van de itemmoeilijkheden, de maximaal haalbare toetscore, het percentage gesloten vragen en de geobserveerde scorevariantie (de standaarddeviatie van de behaalde toetscore) (b: regressiegewicht; se: standaardfout; t: t-toets; R²: percentage verklaarde variantie).

	b	se	t	R ² _{totaal}	R ² _{exclusief}
Percentage verklaarde variantie				85.2%	
Constante	.292				
Behaalde toetscore	-.022	.010	-2.241		84.7%
Stand. dev. itemmoeilijkheden	.027	.007	4.136		83.2%
Max. te beh. toetscore	.049	.010	5.154		82.0%
Stand. dev. (beh. toetscore)	.160	.034	4.701		82.6%

Standaardmeefout en itemspreiding. De spreiding van de itemmoeilijkheden heeft eveneens een effect op de standaardmeefout (gegeven de vier andere predictoren). Neemt de standaarddeviatie van de itemmoeilijkheden met één punt toe, dan stijgt de standaardmeefout met .027 punt. Het verschil in standaardmeefout tussen de minst (sd = 3.4) en meest heterogene (sd = 23.5) toets bedraagt derhalve .54 punten (op een schaal van 1.3 tot 4.9 met een gemiddelde van 2.8 en een standaarddeviatie van .76).

Standaardmeefout en vraagtype. Het effect van het vraagtype is praktisch gezien erg klein. Neemt het percentage gesloten vragen toe met 10%, dan daalt de standaardmeefout met .02 punt.

Ter controle is de relatie tussen standaardmeefout en de moeilijkheid ook binnen elk van de zeven opleidingstypen afzonderlijk geanalyseerd. Het hiervoor beschreven negatieve lineaire effect van de moeilijkheid op de meetnauwkeurigheid bleek in vijf van de zeven opleidingstypen significant (conditioneel op de maximaal te behalen toetscore en de standaarddeviatie van de toetscores). In de 'hogere' opleidingstypen was het effect minder zwak dan in de 'lagere' opleidingstypen ($b_{ivbo} = -.00$; $b_{ivbo/vbo} = -.00$; $b_{vbo} = -.02$; $b_{vbo/mavo} = -.03$; $b_{mavo+} = -.06$; $b_{havo+} = -.10$; $b_{vwo+} = -.13$).

DISCUSSIE

De gemiddelde betrouwbaarheid van de toetsen waarvoor voldoende waarnemingen beschikbaar waren, bedraagt .76 in het tweede leerjaar van het eerste basisvormingscohort en .73 in zowel het derde leerjaar van het eerste cohort als het tweede leerjaar van het tweede cohort. Bij deze bevindingen moet men bedenken dat voor de afname van een afsluitingstoets doorgaans maar één lesuur beschikbaar is. Daardoor konden de toetsen vaak maar weinig items bevatten. Zo gezien kunnen de geconstateerde gemiddelde betrouwbaarheden zeker niet als laag beoordeeld worden. Dit neemt evenwel niet weg dat de afsluitingstoetsen qua betrouwbaarheid sterk verschillen. Dit betekent dat er nog voldoende ruimte voor verbetering aanwezig is.

In opdracht van de COB heeft het Cito één toets voor de volle breedte van het voortgezet onderwijs geconstrueerd. Tijdens de constructie van de opgaven is doorgaans 'gemikt' op hetgeen een gemiddelde (mavo-)leerling in het voortgezet onderwijs na meestal drie jaar onderwijs zou moeten kennen en kunnen. Bovendien is getracht vooral die opgaven te selecteren waarbij de 'kloof' tussen het (i)vo en het vwo het kleinst is. Desalniettemin blijken de afsluitingstoetsen voor de 'onderkant' van het voortgezet onderwijs erg moeilijk en voor de 'bovenkant' erg gemakkelijk (Commissie Toetsing Basisvorming, 1996; Kuhlemeier, Kremers & Kleintjes, 1997). In dit artikel is nagegaan in hoeverre de onvolmaakte afstemming op het vaardigheidsniveau van de minst en meest vaardige leerlingen van invloed is geweest op de betrouwbaarheid. De analyse van de relatie tussen betrouwbaarheid en moeilijkheidsgraad laat zien dat de moeilijkheid in het zeer brede middengebied nauwelijks van invloed is op de toetsbetrouwbaarheid. Maar van de (extrem) gemakkelijke toetsen ($p > 70\%$ à 80% goed) is de betrouwbaarheid duidelijk lager. De aanvullende analyse per opleidingstype laat zien dat er in het (i)vo en mavo+ geen relatie tussen betrouwbaarheid en moeilijkheid aantoonbaar is. In de 'onderkant' en het 'midden' van het voortgezet onderwijs heeft de relatief hoge moeilijkheidsgraad van de afsluitingstoetsen dus geen aantoonbaar negatieve invloed op de toetsbetrouwbaarheid gehad.

In het havo+ en vwo+ is er daarentegen sprake van een negatieve regressie van betrouwbaarheid op moeilijkheid. Kennelijk is het feit dat de toetsen voor de 'bovenkant' van het voortgezet onderwijs erg gemakkelijk waren ten koste gegaan van de betrouwbaarheid. Deze resultaten zijn grotendeels in overeenstemming met de theoretische studie van Feldt (1993). Die laat zien dat men zich over het effect van de moeilijkheid van de items op de betrouwbaarheid van de toets geen al te grote zorgen hoeft te maken zolang de toetsen maar niet al te gemakkelijk of al te moeilijk zijn. Dit geldt tot op zekere hoogte ook voor de standaardmeetfout. Conditioneel op onder meer de toetslengte en de spreiding van de toetscores meten makkelijker toetsen in het algemeen iets nauwkeuriger dan moeilijker toetsen. Maar in de analyses per opleidingstype bleek dit effect alleen aantoonbaar in het vwo+.

Tijdens de item- en toetsconstructie heeft het Cito getracht een grote spreiding naar moeilijkheidsgraad van de items in de toetsen aan te brengen. De toetsen moesten immers voor 'elk wat wils' bevatten. Dus zowel gemakkelijker opgaven die ook voor het (i)vo te doen zouden zijn, als ook moeilijker opgaven die voor het vwo nog voldoende uitdaging zouden bieden. In de onderhavige studie blijkt de spreiding van de itemmoeilijkheden binnen de toets een duidelijk effect op de betrouwbaarheid te hebben. Hoe heterogener de items qua moeilijkheid, hoe lager de toetsbetrouwbaarheid en hoe groter de standaardmeetfout. Op grond hiervan kunnen we concluderen dat de welbewust aangebrachte spreiding de betrouwbaarheid en de meetnauwkeurigheid niet ten goede is gekomen. Overigens zijn er natuurlijk onderwijskundige argumenten die ervoor pleiten wat meer spreiding in de toetsen aan te brengen dan toetstechnisch gezien wellicht wenselijk zou zijn. Zo zou het opnemen van enkele relatief gemakkelijke opgaven aan het begin van de toets op wat minder vaardige leerlingen een bemoeidigend en mogelijk prestatieverhogend effect kunnen hebben.

De toetslengte blijkt zoals verwacht van belang voor de toetsbetrouwbaarheid. Binnen de

geobserveerde waarden van vijf en veertig opgaven gaat een toename van de toetslengte met tien items gemiddeld gepaard met een stijging van de geschatte betrouwbaarheid met .06. Dit geconstateerde effect van toetslengte zou men ook vinden bij toepassing van de formule van Spearman-Brown voor homogene testverlenging (vgl. Veldhuijzen, Goldeveld & Sanders, 1993).

Het percentage gesloten vragen en de scoringswijze blijken nauwelijks een extra bijdrage te leveren aan de betrouwbaarheid en de meetnauwkeurigheid. Kennelijk maakt het - gegeven de moeilijkheid, de itemspreiding, de toetslengte en de geobserveerde scorevariantie - weinig uit of de toets uit veel of weinig gesloten opgaven bestaat (of dat het een (semi-)objectief scoorbare schriftelijke toets betreft dan wel een door beoordelaars beoordeelde praktijktoets).

Op grond van deze studie kan een aantal aanbevelingen voor de ontwikkeling van volgende generaties afsluitingstoetsen worden gedaan. In de eerste plaats verdient het aanbeveling de moeilijkheid van de toetsen af te stemmen op de grote vaardigheidsverschillen tussen leerlingen die er in het voortgezet onderwijs nu eenmaal bestaan. Van elke toets zullen inhoudelijk identieke varianten met een verschillende moeilijkheidsgraad ontwikkeld moeten worden. Vanuit toetstechnisch oogpunt lijkt het daarbij raadzaam geen toetsvarianten te construeren met een moeilijkheidsgraad hoger dan 70% à 80% goed. Gelet op de beoordelingspraktijk in het voortgezet onderwijs (De Groot & Wijnen, 1981) lijkt het verstandig een gemiddelde moeilijkheid van ongeveer 60% goed na te streven. Dit leidt zonder twijfel tot een iets minder hoge betrouwbaarheid dan toetstechnisch gezien mogelijk zou zijn geweest. Maar het heeft als voordeel dat de toetsen voor meer docenten en leerlingen aanvaardbaar zullen zijn. Bovendien is de inpasbaarheid in het onderwijs groter. Als de toetsen namelijk noch te gemakkelijk noch te moeilijk zijn, kunnen ze zonder veel problemen als proefwerk gebruikt worden. De overlap van de prestatieverdelingen van het ivbo, ivbo/vbo, vbo, vbo/avo, mavo+, havo+ en vwo+ (zie Figuur 1) geeft aanleiding om de toetsvarianten niet te koppelen aan het opleidingstype. In elk opleidingstype bevinden zich immers leerlingen die qua prestaties op 'een gemiddelde afsluitingstoets' niet onderdoen voor leerlingen uit aangrenzende opleidingstypen. Ook in het (i)vbo zijn er leerlingen die zich qua prestaties op een afsluitingstoets kunnen meten met avo-leerlingen. En ook in het vwo bevinden zich leerlingen die aan een wat minder moeilijke toetsvariant nog hun handen vol zullen hebben.

Een tweede aanbeveling betreft de spreiding van de itemmoeilijkheden binnen de toets. Als het concept van 'één uniforme toets voor allen' verlaten wordt, en de moeilijkheid van de toetsvarianten wordt afgestemd op het vaardigheidsniveau van de desbetreffende doelgroep, is het aanbrengen van een sterke spreiding van de itemmoeilijkheden niet meer zo nodig. De toetsen hoeven immers niet meer 'voor elk wat wils' te bevatten. In de toekomst kan worden gestreefd naar qua itemmoeilijkheid relatief homogene toetsvarianten. Als een vuistregel zouden toetsontwikkelaars vooral die items moeten kiezen waarvan de moeilijkheid niet verder dan zeven percentagepunten afligt van de nagestreefde moeilijkheidsgraad van de toets. Bij een beoogde moeilijkheid van 60% betekent dit items in het bereik van 53% tot 67% goed. Vaak zal de voor selectie beschikbare itemverzameling echter heterogener zijn en moet men noodgedwongen concessies doen aan de betrouwbaarheid. Liggen de moeilijkheden van de geselecteerde items bijvoorbeeld tussen de 43% en 77%, dan zal men een betrouwbaarheidsverlies van .05 op de koop toe moeten nemen. Bij een nog grotere itemspreiding - van 33% tot 87% - zal de betrouwbaarheid van de toets met nog eens .05 dalen. De minder grote spreiding van de itemmoeilijkheden en de betere afstemming op het vaardigheidsniveau van de leerlingen zullen niet alleen ten goede komen aan de acceptatie van de afsluitingstoetsen in het onderwijs, maar ook aan de betrouwbaarheid.

1. Het eerste basisvormingscohort is de groep leerlingen die in september 1993 met de basisvorming startte.
2. In de rapportage aan het Cito werden de volgende twaalf schoolsoorten onderscheiden: ivbo, ivbo/vbo, vbo, vbo/mavo, vbo/mavo/havo, vbo/mavo/havo/vwo, mavo, mavo/havo, mavo/havo/vwo, havo, havo/vwo en vwo. In dit artikel is deze indeling in twaalf schoolsoorten ingedikt tot zeven opleidingstypen. Die zijn als volgt aangeduid:
- ivbo: individueel voorbereidend beroepsonderwijs;
 vbo: vorbereidend beroepsonderwijs;
 ivbo/vbo: combinatie van ivbo en vbo;
 vbo/avo: vbo/mavo en vbo/mavo/havo(+vwo);
 mavo+: mavo en mavo/havo;
 havo+: mavo/havo/vwo en havo;
 vwo+: havo/vwo en vwo.

3. Met dank aan de anonieme reviewer die ons attent maakte op het belang van de standaardmeetfout voor onze vraagstelling.

4. De correlaties tussen de variabelen zijn weergegeven in onderstaande tabel. Overigens lijkt multi-collineariteit in de onderhavige dataset geen probleem te vormen. De meeste predictoren hangen onderling nauwelijks samen.

	1	2	3	4	5	6
1 Betrouwbaarheid	1.00					
2 Moeilijkheid (% items goed)	-.21	1.00				
3 Stand. dev. itemmoeilijkheden	-.19	-.43	1.00			
4 Aantal items	.17	.02	.31	1.00		
5 Percentage gesloten vragen	.11	.04	.03	.57	1.00	
6 Schriftelijk versus praktijk	-.03	.24	-.63	-.57	-.33	1.00
7 Stand. dev. (% items goed)	.36	-.13	-.09	-.39	-.08	.08

5. In de regressie-analyse met de betrouwbaarheid als afhankelijke variabele is als indicator voor de moeilijkheid van de toets het percentage goede antwoorden genomen en als indicator voor de spreiding van de toetscores is gekozen voor de standaarddeviatie van dit percentage goed. In de analoge regressie-analyse met de standaardmeetfout als afhankelijke variabele zijn de overeenkomstige indicatoren de behaalde toetsscore en de standaarddeviatie van deze toetsscore, terwijl in plaats van aantal items de maximaal te behalen toetsscore in de regressievergelijking is ingebracht (vgl. Tabel 1). Anders dan de betrouwbaarheid is de standaardmeetfout immers een sterk schaalafhankelijke maat.

LITERATUUR

- Bergh, H. van den (1988). *Examens geëxamineerd*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.
- Commissie Opgaven Basisvorming (1994). *Werkzaamheden opgaven 1994/95 en 1995/96, conclusies en aanbevelingen, verslag per vak*. Zoetermeer: Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Commissie Toetsing Basisvorming (1996). *De toets der kritiek*. Zoetermeer: Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Ebel, R.L. & Frisbie, D.A. (1991). *Essentials of educational measurement (5th edition)*. Englewood Cliffs, NJ: Prentice Hall.
- Feldt, L.S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education*, 61, 1, 37-48.

- Frary, R.B. (1985). Multiple-choice versus free-response: A simulation study. *Journal of Educational Measurement*, 22, 21-31.
- Goldebeld, P. (1992). *Toets- en itemanalyse met TIA. Toelichting bij het lezen en interpreteren van toets- en itemanalyses voor gesloten en/of open vragen*. Arnhem: Cito.
- Goldebeld, P. (1995). *Vergelijking van de betrouwbaarheid bij gesloten en open vragen. Interne projectnotitie*. Arnhem: Instituut voor Toetsontwikkeling.
- Goldebeld, P. (1996). *De betrouwbaarheid van toetsen met open vragen. Interne projectnotitie*. Arnhem: Instituut voor Toetsontwikkeling.
- Groot, A.D. & Naerssen, R.F. van (1969). *Studietoetsen: construeren, afnemen, analyseren*. Deel II. Den Haag: Mouton.
- Groot, A.D. de & Wijnen, W.H.F.W. (1981). *Vijven en zessen. Cijfers en beslissen: het selectieproces in ons voortgezet onderwijs*. Groningen: Wolters-Noordhoff.
- Guilford, J.P. & Fruchter, B. (1978). *Fundamental statistics in psychology and education*. Tokyo: McGraw-Hill.
- Horst, P. (1953). Correcting the Kuder-Richardson reliability for dispersion of item difficulties. *Psychological Bulletin*, 50, 371-374.
- Kremers, E.J.J. (1994). Toetsen ter afsluiting van de basisvorming: actueel overzicht. *Handboek voor de basisvorming (afl. 31)*, p. 1-11.
- Kuhlemeier, J.B., Kremers, E.J.J. & Kleintjes, F. (1996). *Gebruik en moeilijkheidsgraad van de eerste generatie afsluitingstoetsen basisvorming schooljaar 1994/95*. Arnhem: Instituut voor Toetsontwikkeling.
- Kuhlemeier, J.B., Kremers, E.J.J. & Kleintjes, F. (1997). *De eerste generatie afsluitingstoetsen basisvorming: gebruik, betrouwbaarheid en maakbaarheid*. Arnhem: Instituut voor Toetsontwikkeling.
- Meuffels, B. (1994). *De verguisde beoordelaar: opstellen over opstelbeoordeling*. Amsterdam: Thesis Publisher.
- Procesmanagement Basisvorming (1995). *De afsluitende toetsing van de basisvorming. Een advies*. Den Haag: Procesmanagement Basisvorming.
- Sluijter, C. (1994). *Stappenschema voor de constructie van afsluitingstoetsen. Interne projectnotitie 94-44840-73*. Arnhem: Instituut voor Toetsontwikkeling.
- Veldhuijzen, N.H., Goldebeld, P. & Sanders, P.F. (1993). Klassieke testtheorie en generaliseerbaarheidstheorie. In T.J.H.M. Eggen & P.F. Sanders (Red.), *Psychometrie in de praktijk* (pp. 33-81). Arnhem: Instituut voor Toetsontwikkeling.

De actuele en structurele arbeidsmarktpositie van het MBO

L. Borghans, A. de Grip en A.G.M. Matheeuwsen
 (Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), Faculteit der Economische Wetenschappen en Bedrijfskunde, Universiteit Maastricht)

Middelbaar opgeleiden spelen een belangrijke rol in de Nederlandse economie. Zij voorzien in de kennis en vaardigheden die in de steeds complexer wordende samenleving nodig zijn voor de uitoefening van een groot aantal uitvoerende beroepen. Waar vroeger nog veel ongeschoolde en laaggeschoolde 'arbeiders' in deze behoefte van de arbeidsmarkt voldeden, is tegenwoordig vanwege de upgradings-tendensen een MBO- of leerlingwezendiploma een minimaal vereiste startkwalificatie voor vakarbeid. Om de arbeidsmarktpositie van de verschillende opleidingen binnen het MBO adequaat in beeld te brengen is het van belang om niet alleen naar de actuele arbeidsmarktpositie van schoolverlaters (werkloosheid, werk op het eigen opleidingsniveau, een vaste baan, deeltijdarbeid) te kijken, maar ook naar de structurele kenmerken van hun arbeidsmarktpositie, die een indicatie geven van de arbeidsmarktrisico's die het volgen van een bepaalde opleiding met zich meebrengt. Wanneer het beroependomein van bepaalde opleidingen veel vakspecifieke kennis en vaardigheden vereist, zijn de gediplomeerden van deze opleidingen immers sterk afhankelijk van de werkgelegenheidsontwikkeling in specifieke beroepen. Met name in een dergelijke situatie leiden conjuncturele schommelingen in de werkgelegenheid tot een verhoogd risico op de arbeidsmarkt. In dit artikel komt daarom de vraag aan de orde in hoeverre de actuele arbeidsmarktsituatie van schoolverlaters van MBO-opleidingen gerelateerd is aan de conjunctuur-gevoeligheid en de beroepsverspreiding van de verschillende opleidingen. Er wordt een duidelijk verband tussen conjunctuurgevoeligheid en uitwijk-mogelijkheden gevonden. Verder blijken schoolverlaters uit opleidings-richtingen die gegeven hun conjunctuurgevoeligheid veel uitwijkmogelijkheden hebben hier wel de negatieve gevolgen van te ondervinden in de vorm van een hoge mate van onderbenutting.

1. INLEIDING

Uit diverse vergelijkende onderzoeken tussen Westeuropese landen blijkt dat middelbaar opgeleiden een strategische rol spelen voor de internationale concurrentiepositie van het bedrijfsleven (zie bijvoorbeeld Ryan, 1991). Daarbij gaat het niet zozeer om het innovatie-ve potentieel van de beroepsbevolking, maar vooral om de kwaliteit van de productie en de dienstverlening. Met name in Groot Brittannië is de groep middelbaar opgeleide vakarbeiders, als gevolg van het zwak ontwikkelde middelbaar beroepsonderwijs, in veel productie- en dienstverlenende bedrijven in vergelijking met soortgelijke bedrijven in Duitsland en Frankrijk sterk ondervetegenwoordigd. Uit verschillende case-studies komt naar voren dat dit een sterk negatieve invloed heeft op de productiviteit van de bedrijven. Daarbij wordt gewezen op het vaker optreden van vertragingen in het productieproces door inadequaat han- delen en/of te weinig preventief onderhoud. Bovendien blijkt een geringere inzet van mid- delbaar opgeleiden nogal eens tot een lagere kwaliteit van de geproduceerde goederen of de

geleverde diensten te leiden (zie bijvoorbeeld Steedman en Wagner, 1987 en Mason, Van Ark en Wagner, 1994).

In Nederland heeft vooral in de jaren tachtig een *MBO-isering* van de vakarbeid plaatsgevonden (De Grip en Dekker, 1993). Desalniettemin is het percentage middelbaar opgeleiden in Nederland nog steeds aanzienlijk kleiner dan in Duitsland. Uit empirisch onderzoek blijkt dat de concurrentiekracht van het Nederlandse bedrijfsleven dan ook kan worden versterkt door het aandeel middelbaar opgeleiden in de werkzame beroepsbevolking te vergroten. Zo leidt een toename van 1% van het aantal werkenden met een middelbare beroepsopleiding in Nederland tot een stijging van de arbeidsproductiviteit met circa 1,5% (Cörvers, De Grip en Orbon, 1995).

MBO'ers voorzien in de kennis en vaardigheden die in de steeds complexer wordende samenleving nodig zijn voor de uitoefening van een groot aantal uitvoerende beroepen. Waar vroeger nog veel ongeschoolde en laaggeschoolde 'arbeiders' in deze uitvoerende functies voldeden, is tegenwoordig vanwege de upgradings van de vereiste vaardigheden een MBO- of leerlingwezendiploma een minimaal vereiste *startkwalificatie*. Vanwege het vakspecifieke karakter van deze vaardigheden zijn de betrokkenen voor hun werkgelegenheid waarschijnlijk *sterk afhankelijk van de perspectieven binnen een beperkt beroependomein*. Conjuncturele schommelingen zouden daardoor voor deze groepen een verhoogd risico op de arbeidsmarkt betekenen.

Bij de onderscheiden opleidingstypen op MBO/leerlingwezeniveau is er overigens sprake van sterk uiteenlopende werkgelegenheidsverwachtingen (zie ROA, 1995). Het werkgelegenheidsaandeel van het MBO/leerlingwezen balanceert op de rand van groei en krimp, met grote verschillen tussen de verschillende opleidingstypen. Hierbij hebben de verschuivingen in de beroepenstructuur van de verschillende bedrijfssectoren per saldo een (gering) negatief effect op de werkgelegenheidsgroei voor schoolverlaters van het middelbaar beroepsonderwijs, terwijl de toenemende opleidingseisen binnen de verschillende beroepen daarop per saldo een positief effect hebben. Het laatste duidt op een verdere voortschrijding van de genoemde *MBO-isering* van de vakarbeid ten koste van het arbeidsaanbod zonder een zogenaamde startkwalificatie voor de arbeidsmarkt. Wat dit betreft mag zonder meer verwacht worden dat het middelbaar beroepsonderwijs, met name in bepaalde opleidingsrichtingen zijn sleutelrol op de arbeidsmarkt zal blijven vervullen en daarmee een belangrijke bijdrage zal leveren aan de 'eerste klas beroepsbevolking' (Ministerie van Economische Zaken, 1995), die vereist is voor het creëren van een internationaal concurrerend vestigingsklimaat.

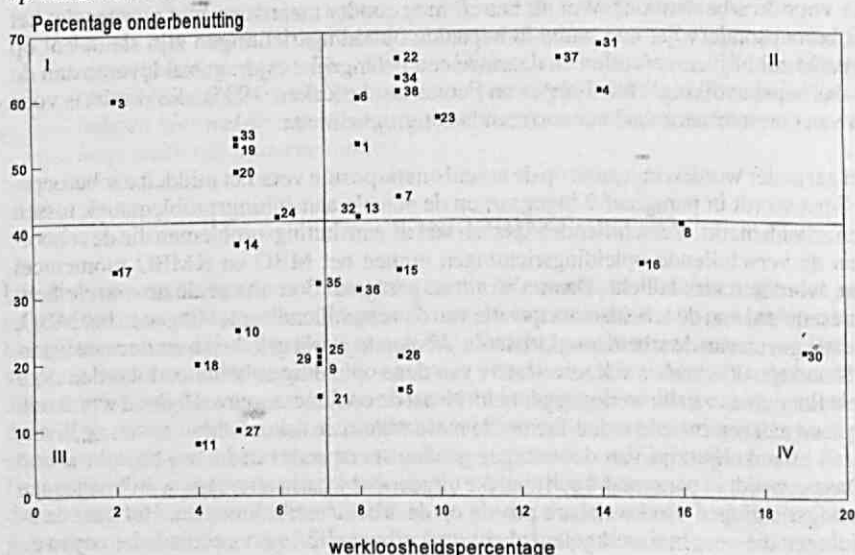
In dit artikel zal nader worden ingegaan op de arbeidsmarktpositie voor het middelbaar beroepsonderwijs. Eerst wordt in paragraaf 2 ingegaan op de actuele aansluitingsproblematiek tussen onderwijs en arbeidsmarkt. Verschillende aspecten van de aansluitingsproblemen die de schoolverlaters van de verschillende opleidingsrichtingen binnen het MBO en KMBO momenteel ondervinden, worden nader belicht. Daarna wordt in paragraaf 3 een tweetal structurele kenmerken gepresenteerd van de arbeidsmarktpositie van de verschillende opleidingen in het MBO, die een indicatie geven van de arbeidsmarkttrisco's. Als eerste wordt gekeken naar de mate waarin de arbeidsmarktpositie van de schoolverlaters van deze opleidingen beïnvloed worden door conjuncturele fluctuaties van de werkgelegenheid. Naast de conjunctuurgevoeligheid wordt ook aandacht besteed aan een tweede risico-factor: de mate waarin de schoolverlaters van de diverse opleidingen afhankelijk zijn van de werkgelegenheidsperspectieven in een beperkt aantal beroepen. Daarna wordt in paragraaf 4 een analyse uitgevoerd waarin allereerst wordt nagegaan welke opleidingsrichtingen een kwetsbare positie op de arbeidsmarkt innemen. Het gaat daarbij om opleidingen die voor hun werkgelegenheid sterk afhankelijk zijn van enkele beroepen en waarvoor de werkgelegenheid bovendien ook sterk conjunctuurgevoelig is. Vervolgens zal worden nagegaan of problemen in de actuele arbeidsmarktsituatie samenhangen met de zwakkere

structurele positie van de betreffende opleidingsrichtingen. Opleidingen die te specialistisch zijn riskeren immers een hogere werkloosheid. Daar staat tegenover dat een opleiding die gericht is op een te breed beroependomein mogelijk een negatief effect heeft op de productiviteit van de schoolverlaters. Dit zou ertoe kunnen leiden dat men vaker genoeg zou moeten nemen met een functie op een hoger niveau, waardoor er sprake is van onderbenutting. In paragraaf 5 worden de belangrijkste conclusies van het artikel kort samengevat.

2. DE ACTUELE AANSLUITINGSPROBLEMATIEK

Op basis van de schoolverlatersenquête *Registratie Uitstroom en Bestemming van Schoolverlaters* (RUBS; Van Smoorenburg en Van der Velden, 1995) kan de actuele aansluitingsproblematiek op de arbeidsmarkt voor schoolverlaters van het MBO en KMBO op een gedetailleerd niveau in beeld worden gebracht. Tussen de diverse opleidingsrichtingen binnen het MBO en KMBO blijken er aanzienlijke verschillen te bestaan in de mate waarin het aanbod van schoolverlaters aansluit op de vraag op de arbeidsmarkt. Daarbij wordt onderscheid gemaakt tussen 'open discrepanties' in de zin van het bestaan van werkloosheid onder schoolverlaters met een bepaalde opleidingsachtergrond en 'verborgen discrepanties' die betrekking hebben op het terechtkomen van schoolverlaters in functies beneden hun opleidingsniveau, de zgn. *onderbenutting* (zie ook Van Hoof en Dronkers, 1980 en De Grip, Van der Velden en Wieling, 1993). Deze onderbenutting wordt vastgesteld door een vergelijking te maken van het opleidingsniveau van de schoolverlaters met het vereiste opleidingsniveau voor de functie die men uitoefent¹. Van der Linden en Van der Velden (1998) laten zien dat de onderbenutting van schoolverlaters een duurzaam effect heeft. Uit hun analyse blijkt dat de onderbenutting van afgestudeerde van het universiteitsonderwijs 1 jaar na hun arbeidsmarktsituatie een significant positief effect heeft op de kans op onderbenutting 5 jaar later. Figuur 1 geeft een overzicht van deze actuele aansluitingsproblematiek.

Figuur 1: Actuele aansluiting tussen onderwijs en arbeidsmarkt van het MBO en KMBO op basis van werkloosheid en onderbenutting, 1994 (Bron: RUBS 1994/ROA)



* Tabel 1 geeft een overzicht van de onderscheiden opleidingsrichtingen met de bijbehorende nummers.

Tabel 1: Overzicht van de op basis van RUBS 1994 onderscheiden opleidingen met bijbehorende nummers

Nr.	Opleiding
1	KMBO agrarisch
2	MAS-A land- en tuinbouw
3	MAS-B land- en tuinbouw
4	KMLO
5	MBO chemisch laboratorium
6	KMBO metaalbewerking
7	KMBO motorvoertuigtechniek
8	KMBO elektrotechniek
9	MBO consumptieve techniek
10	MBO bouwkunde
11	MBO weg- en waterbouwkunde
12	MBO metaalkunde
13	MBO werktuigbouwkunde
14	MBO motorvoertuigtechniek
15	MBO elektrotechniek
16	MBO procestechiek
17	MBO textiel- en ledertechniek
18	MBO haven en vervoer
19	MBO verpl. en ziekenverzorging
20	MBO assistent gezondheidszorg
21	MBO medisch laboratorium
22	KMBO detailhandel
23	KMBO administratie
24	MBO handel
25	MBO administratie
26	MBO commercieel
27	MBO secretariaal
28	MBO economisch-administratief overig
29	MBO bestuurlijk
30	MBO sociale arbeid
31	KMBO verzorging
32	MBO agogisch werk
33	MBO verzorging
34	MBO mode en kleding
35	MBO activiteitenbegeleiding
36	MBO horeca
37	MBO civ. en cons.-techn. diensten
38	MBO uiterlijke verzorging

De figuur is verdeeld in vier vakken. De horizontale lijn geeft het gemiddelde percentage onderbenutting aan van alle (K)MBO-opleidingen, terwijl de verticale lijn het gemiddelde werkloosheidspercentage van schoolverlaters weergeeft. Vanuit het oogpunt van de schoolverlaters is de aansluitingsproblematiek het grootst bij de opleidingen die te kampen hebben met zowel een relatief *hoge werkloosheid* van schoolverlaters, als een relatief *hoge onderbenutting* van het kwalificatieniveau (vak II). Daarbij gaat het met name om de KMBO-opleidingsrichtingen *verzorging*, *laboratorium*, *detailhandel*, *administratie*, *metaalbewerking*, *agrarisch* en *motorvoertuigtechniek*. Even slecht is het gesteld met de arbeidsmarktsituatie voor de MBO-opleidingen *civiele en consumptief-technische diensten*, *mode en kleding* en *uiterlijke verzorging*.

Bij een aantal andere opleidingsrichtingen is er sprake van een hoge werkloosheid onder de schoolverlaters in combinatie met een relatief geringe onderbenutting (vak IV). Het gaat hier

met name om opleidingsrichtingen binnen het MBO, met als uitzondering de opleiding *KMBO elektrotechniek*. Bij een drietal van deze opleidingen is de werkloosheid het hoogst van alle MBO- en KMBO-opleidingen. Het gaat hier om *MBO sociale arbeid*, *MBO procestechiek* en *KMBO elektrotechniek*. Bij deze opleidingen is de werkloosheid waarschijnlijk mede zo hoog, omdat de schoolverlaters - gezien de geringe onderbenutting - weinig of geen mogelijkheden hebben om werk te vinden in beroepen op een lager functieniveau, of omdat het aantrekkelijker is te wachten met het accepteren van een baan totdat zich een mogelijkheid voordoet om een baan te krijgen in een bij de gevolgde opleiding aansluitende vakspecifieke functie. Het laatste zou met name het geval kunnen zijn bij de schoolverlaters van de opleidingen *elektrotechniek* en *procestechiek*, die te kampen hebben met een erg conjunctuurgevoelige werkgelegenheid, zoals verderop zal blijken.

Bij de vijf MBO-opleidingen in vak I spitst de aansluitingsproblematiek op de arbeidsmarkt zich toe op de onderbenutting van het behaalde opleidingsniveau. Daarbij gaat het om de opleidingen *MAS-B land- en tuinbouw*, *verzorging*, *assistent gezondheidszorg*, *verpleging en ziekenverzorging* en *handel*.

Daarentegen zijn de aansluitingsproblemen voor schoolverlaters gering bij tien MBO-opleidingsrichtingen en de KMBO-opleiding *consumptieve techniek* (vak III). Het laagst is de werkloosheid onder schoolverlaters *textiel- en ledertechiek*, *haven en vervoer* en *weg- en waterbouwkunde*. Laatstgenoemde opleiding kent samen met de opleiding *secretarieel*, de geringste onderbenutting van de kwalificaties van de werkende schoolverlaters. De overige MBO-opleidingen met geringe aansluitingsproblemen zijn *bouwkunde*, *medisch laboratorium*, *bestuurlijk*, *motorvoertuigtechiek*, *administratie* en *activiteitenbegeleiding*.

3. STRUCTURELE KENMERKEN: CONJUNCTURGEVOELIGHEID EN UITWIJKMOGELIJKHEDEN

Een slechte actuele arbeidsmarktsituatie, zoals die tot uiting komt in hoge werkloosheidscijfers en een hoge onderbenutting van degenen die wel werk hebben gevonden, kan het gevolg zijn van een structureel zwakkere positie van een opleiding op de arbeidsmarkt. Omgekeerd hoeft echter een opleidingsrichting met een structureel zwakke positie niet altijd problemen met de aansluiting tussen onderwijs en arbeidsmarkt te kennen, bijvoorbeeld omdat de economische situatie voor de betreffende opleiding tijdelijk gunstig is. Om een beter beeld van de mogelijke kwetsbaarheid van de verschillende opleidingstypen binnen het MBO te krijgen, wordt in deze paragraaf ingegaan op de structurele kenmerken van opleidingen, die een indicatie geven van de arbeidsmarktrisico's die het volgen van een bepaalde opleiding met zich meebrengt. Hierbij onderscheiden we de conjunctuurgevoeligheid van de werkgelegenheid en de uitwijkmogelijkheden op de arbeidsmarkt. In de eerste plaats komt het arbeidsmarktrisico tot uiting in de mate waarin de werkgelegenheid onderhevig is aan conjuncturele fluctuaties. Wanneer er sterke schommelingen optreden in de werkgelegenheid voor de schoolverlaters van een bepaalde opleiding, is er voor degenen die voor deze opleiding kiezen op het moment dat het economisch tij gunstig is een risico dat de kansen op werk op het moment dat ze hun diploma behalen slecht zijn, omdat de economie dan net in een dal zit. Bovendien is er sprake van een onzekere werkgelegenheidssituatie op langere termijn, vanwege de grotere kans op toekomstig baanverlies als er op een gegeven moment weer sprake is van een laagconjunctuur. De risico's op de arbeidsmarkt beperken zich echter niet tot de conjuncturele fluctuaties van de werkgelegenheid. Ook de mate waarin schoolverlaters afhankelijk zijn van de werkgelegenheidsperspectieven in slechts een beperkt aantal beroepen is een belangrijke risicofactor. Vooral bij opleidingen die specifiek gericht zijn op een bepaalde vakdeelmarkt is dit risico groot.

De conjunctuurgevoeligheid van de werkgelegenheid is bepaald op basis van de mate waarin de werkgelegenheid in de voor een bepaald opleidingstype relevante beroepssegmenten fluctueert ten opzichte van de trendmatige ontwikkeling². De conjunctuurgevoeligheid van de werkgelegenheid voor een bepaald opleidingstype wordt in drie stappen bepaald. Eerst wordt een indicatie berekend van de conjunctuurgevoeligheid van de werkgelegenheid in de diverse bedrijfssectoren. Daarna wordt deze verder verbijzonderd naar de verschillende beroepsklassen. Ten slotte vindt een verdere toerekening plaats naar de verschillende opleidingstypen.

De conjunctuurgevoeligheid van bedrijfssectoren wordt vastgesteld op grond van:

$$Cl_s = 100 \times \sum_t \frac{|W_s^t - \bar{W}_s^t|}{W_s^t} \quad (3.1.)$$

Hierin is

Cl_s de conjunctuurgevoeligheid van bedrijfssector s

\bar{W}_s^t de werkgelegenheid in bedrijfssector s in jaar t

\bar{W}_s^t de trend van de werkgelegenheid in bedrijfssector s in jaar t . Deze wordt berekend als:

$$\bar{W}_s^t = \frac{W_s^{t-1} + W_s^{t+1}}{2} \quad (3.2.)$$

De waarde van de indicator voor de diverse bedrijfssectoren is berekend op basis van gegevens uit de Nationale Rekeningen voor de periode 1954 tot 1994. In Dekker et al. (1993) werd vervolgens de conjunctuurgevoeligheid per beroepsklasse bepaald door per beroepsklasse een gewogen gemiddelde te nemen van de conjunctuurgevoeligheid van bedrijfssectoren op basis van de actuele werkgelegenheidsverdeling. Dit betekent dat impliciet wordt aangenomen dat iedere beroepsklasse binnen een bedrijfssector in gelijke mate de effecten ondervindt van werkgelegenheidsschommelingen. In de praktijk blijken echter de productieberoepen sterker beïnvloed te worden door deze schommelingen dan andere functies. Door het beschikbaar komen van langere tijdreeksen van de werkgelegenheid naar beroepsklasse, is het nu mogelijk de relatie tussen de werkgelegenheidsfluctuaties in een beroepsklasse en de schommelingen in een bedrijfssector vast te stellen. Daarbij is de volgende vergelijking geschat:

$$\frac{W_{bs}^t - W_{bs}^{t-1}}{W_{bs}^{t-1}} = C_{bs} + \alpha_{bs} \frac{W_s^t - W_s^{t-1}}{W_s^{t-1}} \quad (3.3.)$$

Deze vergelijking is geschat op basis van de data van de Arbeidskrachtentellingen van 1981-1985 en de Enquête Beroepsbevolking van 1988-1994. α_{bs} geeft aan in welke mate de werkgelegenheid in een bepaald beroepsklasse W_{bs} meefluctueert met de werkgelegenheid van de bedrijfssector (W_s).

De conjunctuurgevoeligheid van een beroepsklasse CI_b is vastgesteld als:

$$CI_b = \sum_s \frac{W_{bs}^{94}}{W_s^{94}} \alpha_{bs} CI_s \quad (3.4.)$$

Omdat bij een aantal beroepen waarbij het aandeel in de werkgelegenheid in een bedrijfssector vrij klein is de schattingen van α_{bs} vrij extreme waarden aannamen, is de randvoorwaarde gesteld dat α_{bs} tussen 0,3 en 3,0 moet liggen.³

Omdat het beroependomein waarin schoolverlaters met een bepaalde opleidingachtergrond werk vinden deels ook door aanbodfactoren wordt bepaald, is deze nuancering voor de berekening van de conjunctuurgevoeligheid van opleidingstypen CI_o niet zinvol. Daarom is de conjunctuurgevoeligheid van de werkgelegenheid in de diverse beroepssegmenten telkens evenredig verdeeld over de opleidingsachtergronden van de werkenden in het desbetreffende beroepssegment:

$$CI_o = \sum_b \frac{W_{ob}^{94}}{W_b^{94}} CI_b \quad (3.5.)$$

De uitwijkmogelijkheden op de arbeidsmarkt worden voorgesteld aan de hand van de Gini-Hirschman spreidingsindex (zie Borghans et al., 1995).

$$GH_o^{ber} = \left(1 - \sum_b \left(\frac{W_{ob}}{W_o} \right)^2 \right) \times \frac{N_b}{N_b - 1}$$

Hierin geeft W_{ob} het aantal werkenden met opleiding o in beroepsgroep b en W_o het totaal aantal werkenden met opleidingsachtergrond o aan. N_b weerspiegelt het aantal beroepsgroepen. Bij de spreiding van de werkenden met een bepaalde opleidingsachtergrond over de verschillende beroepen worden overigens alleen de beroepsgroepen meegerekend waarvoor minstens het niveau van de betreffende opleiding vereist is. Als alle werkenden van opleiding o in één specifiek beroep te vinden zijn, is de kwadratensom in de formule gelijk aan 1 en dus de index 0. Als de werkenden volledig gespreid zijn over alle beroepen wordt de kwadratensom bijna teruggebracht tot 0. De correctiefactor $N_b/(N_b-1)$ zorgt er voor dat in dat geval de index gelijk is aan 1.

Figuur 2 geeft een overzicht van de mate waarin schoolverlaters van de verschillende opleidingen in het MBO en KMBO geconfronteerd worden met beide risicofactoren. Evenals figuur 1 is de figuur verdeeld in vier vakken. De horizontale lijn en de verticale lijn in de figuur geven respectievelijk de gemiddelde uitwijkmogelijkheden en de gemiddelde conjunctuurgevoeligheid van alle (K)MBO-opleidingen. Vak II heeft betrekking op de opleidingsrichtingen waarbij de arbeidsmarktrisico's het grootst zijn. Bij deze opleidingen is sprake van zowel een relatief hoge conjunctuurgevoeligheid van de werkgelegenheid, als een beperkt aantal uitwijkmogelijkheden op de arbeidsmarkt. Dit is het geval bij de MBO-opleidingen,

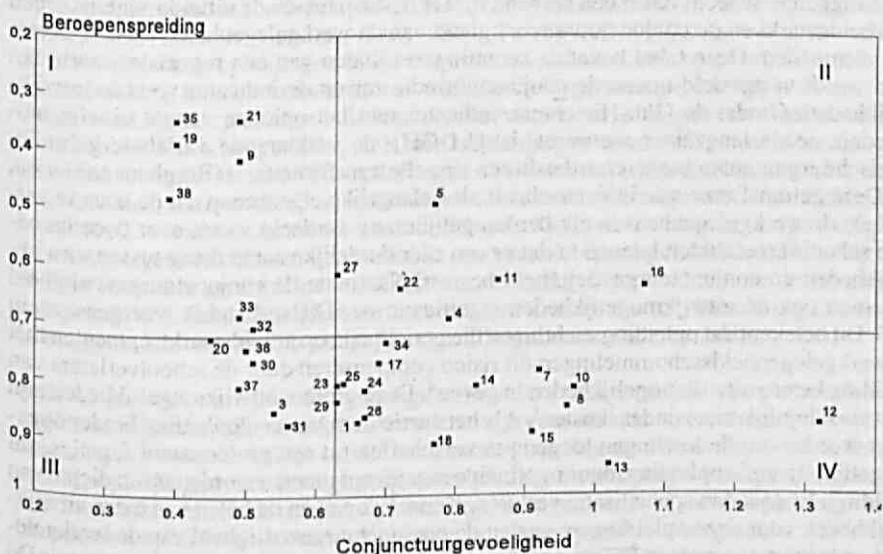
processtechniek, weg- en waterbouwkunde, chemisch laboratorium en secretariael en de KMBO-opleidingen *laboratorium en detailhandel*.

Daarentegen worden de opleidingen in vak III nauwelijks geconfronteerd met beide structurele risico's. Bij deze opleidingen is er sprake van zowel een relatief lage conjunctuurgevoeligheid van de werkgelegenheid van schoolverlaters als een relatief groot aantal uitwijkmogelijkheden naar andere beroepen. Deze gunstige risico-positie doet zich voor bij de opleidingen *civiele en consumptief-technische diensten, MAS-B land- en tuinbouw, sociale arbeid, horeca, assistent gezondheidszorg* en de KMBO-opleiding *verzorging*.

Voor de overige opleidingen kan geen eenduidige uitspraak worden gedaan over de risico's die de schoolverlaters op de arbeidsmarkt zullen ondervinden. Bij de opleiding in vak I is weliswaar sprake van een relatief stabiele werkgelegenheidsontwikkeling, maar de uitwijkmogelijkheden naar beroepen buiten het eigen beroependomein is gering. Het gaat hier om de MBO-opleidingen *agogisch werk, verzorging, uiterlijke verzorging, verpleging en ziekenverzorging, activiteitenbegeleiding* en *medisch laboratorium* en de KMBO-opleiding *consumptieve techniek*.

De opleidingsrichtingen in vak IV hebben daarentegen te maken met een instabiele werkgelegenheidsontwikkeling. Aangenomen mag echter worden dat wanneer de werkgelegenheidssituatie voor deze opleidingen op een bepaald moment ongunstig is, dit enigszins kan worden opgevangen doordat de schoolverlaters kunnen uitwijken naar andere beroepen. Deze situatie doet zich voor bij de KMBO-opleidingen *agrarisch, motorvoertuigen techniek en elektrotechniek* en de MBO-opleidingen *commercieel, administratie, handel, textiel- en leder-techniek, mode en kleding, haven en vervoer, motorvoertuigentechniek, elektrotechniek, bouwkunde, werktuigbouwkunde en metaalkunde*.

Figuur 2: Arbeidsrisico's voor schoolverlaters van het MBO en KMBO op basis van conjunctuurgevoeligheid en beroepenspreiding, 1994 (Bron: RUBS 1994/ROA)



* Tabel 1 geeft een overzicht van de onderscheiden opleidingsrichtingen met de bijbehorende nummers

4. ARBEIDSMARKTRISICO EN ARBEIDSMARKTPOSITIE

Specialisatie in de vorm van vakspecificiteit betekent dat opleidingen in het middelbaar beroepsonderwijs een belangrijke bijdrage aan de productiviteit van de Nederlandse economie kunnen leveren. In tegenstelling tot laag- of ongeschoolde arbeidskrachten die vaak in verschillende sectoren inzetbaar zijn kan deze specialisatie echter ook de kwetsbaarheid van de MBO-opleidingen vergroten. Door een sterke vorm van specialisatie worden schoolverlaters immers sterk afhankelijk van de werkgelegenheidsontwikkeling in het specifieke segment waarop zij zich richten. Zeker als deze werkgelegenheidsontwikkeling op dat arbeidsmarktsegment sterk conjunctuurgevoelig is, betekent dit een verhoogd risico op aansluitingsproblemen.

In deze paragraaf wordt daarom geanalyseerd of een sterke specialisatie, met name in conjunctuurgevoelige richtingen, de arbeidsmarktpositie van opleidingen aan lijkt te tasten. Allereerst wordt daartoe nagegaan hoe sterk of zwak de structurele positie van de verschillende opleidingen is als zowel de conjunctuurgevoeligheid als de uitwijkmogelijkheden worden verdisconteerd. Vervolgens wordt nagegaan of een zwakke structurele positie een verklaring lijkt te bieden voor actuele aansluitingsproblemen op de arbeidsmarkt.

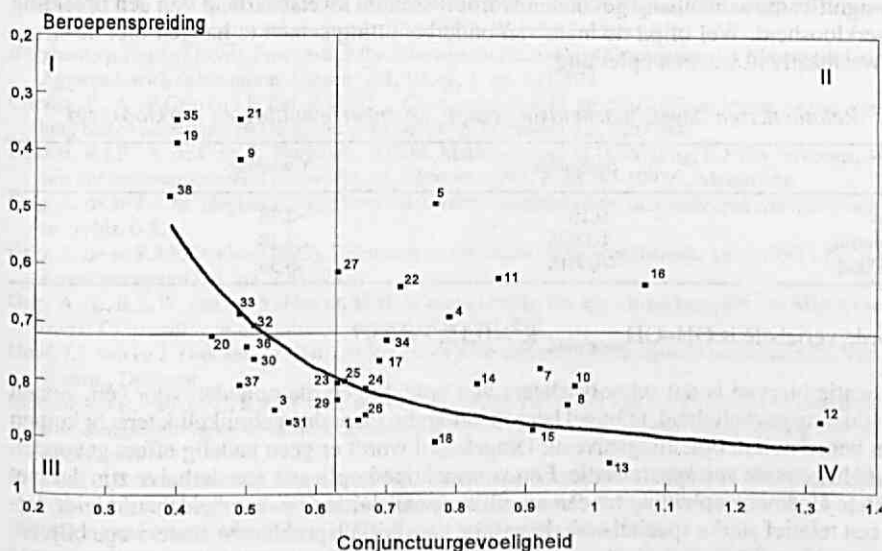
Tabel 2: Relatie tussen uitwijkmogelijkheden ($1/(1-GH)$) en conjunctuurgevoeligheid van de werkgelegenheid

	parameter	t-waarde
Constante	-2,20	-1,19
Conjunctuurgevoeligheid	11,08	4,13

$$R^2 = 0,33, F = 17,0$$

Figuur 2 suggereerde reeds dat er een verband lijkt te bestaan tussen de uitwijkmogelijkheden op de arbeidsmarkt en de conjunctuurgevoeligheid van de werkgelegenheid. Tabel 2 bevestigt dit vermoeden. Deze tabel bevat de schattingsresultaten van een regressie waarin het verband wordt vastgesteld tussen de conjunctuurindicator en de indicator voor de uitwijkmogelijkheden. Omdat de Gini-Hirschman-indicator met het oplopen van de uitwijkmogelijkheden steeds langzamer toeneemt, is $1/(1-GH)$ als verklarende variabele gebruikt. Theoretische argumenten suggereren dat dit een zinvolle transformatie is (Borghans en Heijke, 1996). Deze getransformeerde indicator heeft als belangrijke eigenschap dat de waarde verdubbelt als de werkgelegenheid in elk beroep gelijkmatig verdeeld wordt over twee beroepen. De schattingsresultaten laten zien dat er een zeer duidelijke samenhang tussen uitwijkmogelijkheden en conjunctuurgevoeligheid bestaat. Naarmate de conjunctuurgevoeligheid stijgt nemen ook de uitwijkmogelijkheden significant toe. Dit verband is weergegeven in figuur 3. Dit betekent dat opleidingsrichtingen die gericht zijn op arbeidsmarktsegmenten met sterke werkgelegenheidsschommelingen dit risico compenseren door de schoolverlaters van de opleiding betere uitwijkmogelijkheden te geven⁴. Deze grotere uitwijkmogelijkheden zijn echter waarschijnlijk niet zonder 'kosten'. Als het curriculum van een opleiding breder opgezet moet worden om de leerlingen toegang te verschaffen tot een groter aantal functies, zal bij een gelijkblijvende opleidingsduur in principe ingeleverd moeten worden op de diepte van de opleiding. Er gaat dan specialisatie verloren. Kennelijk wegen de baten van extra uitwijkmogelijkheden voor deze opleidingen, gezien de conjunctuurgevoeligheid van de werkgelegenheid, echter op tegen deze kosten van veralgemenisering van het onderwijs (zie ook De Grip en Heijke, 1991).

Figuur 3: Het verband tussen conjunctuurgevoelige en beroepenspreiding voor schoolverlaters van het MBO en KMBO, 1994



Niet alle opleidingen compenseren echter in dezelfde mate hun grotere conjunctuurgevoeligheid door een groter aantal uitwijkmogelijkheden. Opleidingen die in figuur 3 boven de 'uitruil-curve' van conjunctuurgevoeligheid en uitwijkmogelijkheden liggen hebben immers relatief weinig uitwijkmogelijkheden, gegeven de conjunctuurgevoeligheid van hun werkgelegenheid. Het gaat hier om de opleidingen: *MBO procestechniek*, *MBO weg- en waterbouwkunde*, *MBO chemisch laboratorium*, *MBO secretariael*, *KMBO laboratorium* en *KMBO detailhandel*. Deze opleidingen lijken dus extra kwetsbaar te zijn op de arbeidsmarkt. Dit zou tot uitdrukking kunnen komen in een werkloosheidspercentage dat hoger is dan normaal. Omgekeerd lijken opleidingen die beneden de 'uitruil-curve' liggen een relatief veilige positie te hebben. Dit is het geval bij de *KMBO-opleidingen agrarisch*, *motorvoertuigtechniek* en *elektrotechniek* en de *MBO-opleidingen commercieel*, *administratie*, *handel*, *textiel- en ledertechniek*, *mode en kleding*, *haven en vervoer*, *motorvoertuigtechniek*, *elektrotechniek*, *bouwkunde*, *werktuigbouwkunde* en *metaalkunde*. Gegeven de conjunctuurgevoeligheid van de werkgelegenheid kennen deze opleidingen veel uitwijkmogelijkheden. De vraag is alleen of deze meer dan gemiddelde stabiliteit van de werkgelegenheid niet ten koste gaat van de vakspecificiteit. Een dergelijk tekort aan vakspecificiteit zou tot uitdrukking kunnen komen in een grote kans op werk beneden het opleidingsniveau (onderbenutting).

Opleidingen met een relatief ongunstige verhouding tussen conjunctuurgevoeligheid en uitwijkmogelijkheden lijken dus kwetsbaarder te zijn voor arbeidsmarktschommelingen. Bij deze opleidingen kunnen daardoor eerder werkloosheidsproblemen verwacht worden. Omgekeerd valt bij opleidingen met een 'gunstige' verhouding tussen conjunctuurgevoeligheid en uitwijkmogelijkheden te verwachten dat vanwege de algemeenheid van de opleiding vaker werk beneden het opleidingsniveau wordt verkregen. Deze hypothese kan getoetst worden door het verschil van de indicator voor de uitwijkmogelijkheden en de voorspelling hiervan op grond van de schatting in tabel 2, te beschouwen als een indicator van deze kwetsbaarheid.

Tabel 3 laat de resultaten zien van een regressie van deze 'kwetsbaarheidsindicator' op de verklarende variabelen onderbenutting en werkloosheid. Uit de schattingsresultaten blijkt dat er geen significante samenhang gevonden wordt tussen de kwetsbaarheid van een opleiding en de werkloosheid. Wel blijkt de mate van onderbenutting samen te hangen met de structurele kwetsbaarheid van een opleiding.

Tabel 3: Relatie tussen 'kwetsbaarheidsindicator' en onderbenutting en werkloosheid

	parameter	t-waarde
Constante	-0,16	-2,78
Onderbenutting	0,0026	2,28
Werkloosheid	-0,0018	-0,33

Verklaarde variabele is $GH - GH_{\text{voorspeld}}$, $R^2 = 0,15$, $F = 2,97$

De implicatie hiervan is dat schoolverlaters van opleidingen die opleiden voor een, gezien hun conjunctuurgevoeligheid, te breed beroependomein meer dan gebruikelijk terecht komen in banen beneden hun opleidingsniveau. Omgekeerd wordt er geen nadelig effect gevonden van een te hoge mate van specialisatie. Een voorzichtige implicatie zou derhalve zijn dat met name een te algemene opleiding tot een moeilijker aansluiting op de arbeidsmarkt leidt, terwijl bij een relatief sterke specialisatie dergelijke aansluitingsproblemen achterwege blijven.

5. BESLUIT

In dit artikel is ingegaan op de actuele arbeidsmarktpositie en de structurele situatie op de arbeidsmarkt van opleidingen in het MBO. Vanwege het sterke vakgerichte karakter van het MBO, zouden met name bij opleidingen die zich richten op beroepen die een grote mate van conjunctuurgevoeligheid kennen, problemen verwacht kunnen worden in de aansluiting tussen onderwijs en arbeidsmarkt. Uit de analyses blijken de schoolverlaters van dergelijke opleidingsrichtingen echter doorgaans op een veel breder segment van de arbeidsmarkt werk te vinden waardoor de risico's van schommelingen in de werkgelegenheid grotendeels worden ondervangen. Er blijkt dus een trade-off te bestaan tussen specialisatie en flexibiliteit. Naarmate de werkgelegenheidsrisico's groter zijn leidt een optimale afweging van specialisatie en flexibiliteit tot een opleiding die zich richt op een breder beroependomein.

Voor opleidingen die specialistischer zijn dan op grond van deze trade-off verwacht zou mogen worden, valt een vergroot werkloosheidsrisico te verwachten. Dit verband werd echter niet aangetroffen. Omgekeerd kan verwacht worden dat bij opleidingen die breder geïntendeerd zijn dan op grond van deze trade-off verwacht zou mogen worden, de productiviteit van de schoolverlaters aangetast zou kunnen worden. Dit verband wordt wel aangetroffen. Dergelijke opleidingsrichtingen blijken een hogere mate van onderbenutting te kennen dan gemiddeld, hoewel ook dit verband strikt genomen niet significant is.

Concluderend kan gesteld worden dat binnen het MBO een zeer sterke samenhang bestaat tussen uitwijkmogelijkheden en conjunctuurgevoeligheid, waardoor een groot deel van de risico's van schommelingen in de werkgelegenheid wordt vermeden. Met name bij een opleiding die zich, gezien de conjunctuurgevoeligheid van de werkgelegenheid van de afgestudeerden, richt op een te breed beroependomein, lijkt de verminderde specialisatie tot een hogere mate van onderbenutting te leiden.

LITERATUUR

- Borghans, L., A. de Grip, R. Dekker, A. Matheeuwsen, W. Smits, E. Willems (1995), *Methodiek van het Informatiesysteem Onderwijs-Arbeidsmarkt 1995*, ROA-W-1995/3, Maastricht.
- Borghans en Heijke (1996), Forecasting the Educational Structure of Occupations: A Manpower Requirement Approach with Substitution, *Labour*, vol. 10, no. 1, pp. 151-192.
- Cörvers, F., A. de Grip en J.P. Orbon (1995), 'Concurrentiekracht, produktiviteit en human capital: een vergelijking tussen Nederland en Duitsland', *Maandschrift Economie*, blz. 221-241.
- Dekker, R.J.P., A. de Grip, L. Borghans, A.G.M. Matheeuwsen, M.H. Wieling, E.J.T.A. Willems, *Methodiek van het informatiesysteem onderwijs-arbeidsmarkt 1993*, ROA-W-1993/3, Maastricht.
- Grip, A. de en J.A.M. Heijke (1991), Flexibiliteit versus produktiviteit, in: *Onderzoek van Onderwijs*, vol. 20, no. 1, blz. 6-8.
- Grip, A. de en R.J.P. Dekker (1993), 'Winnaars en verliezers op de arbeidsmarkt 1985-1990', *Tijdschrift voor Arbeidsvraagstukken*, blz. 220-229.
- Grip, A. de, R.K.W. van der Velden en M.H. Wieling (1993), De arbeidsmarktpositie van MDGO-schoolverlaters, *Tijdschrift voor Arbeidsvraagstukken*, blz. 241-252.
- Hoof, J.J. van en J. Dronkers (1980), *Onderwijs en arbeidsmarkt*, Sociologische monografieën, Van Loghum Slaterus, Deventer.
- Linden, A.S.R. van der en R.K.W. van der Velden (1998), Het effect van werkloosheid en onderbenutting bij arbeidsmarktintrede op de latere loopbaan en inkomen, *De Nederlandse Arbeidsmarkt 1997* (verschijnt binnenkort).
- Mason, G., B. van Ark en K. Wagner (1994), 'Productivity, Product, Quality and Workforce Skills: Food Processing in Four European Countries', *National Institute Economic Review*, blz. 62-83.
- Ministerie van Economische Zaken (1995), *Kennis in beweging*, Den Haag.
- Researchcentrum voor Onderwijs en Arbeidsmarkt (1995), *De arbeidsmarkt naar opleiding en beroep tot 2000*, ROA-R-1995/3, Maastricht.
- Ryan, R. (ed.) (1991), *International Comparisons of Vocational Education and Training for Intermediate Skills*, Falmer Press, London.
- Sicherman, N. (1991), Overeducation in the Labour Market, *Journal of Labor Economics*, vol. 9, blz. 101-122.
- Smooenburg, M.S.M. van en R.K.W. van der Velden (1995), *Schoolverlaters op de arbeidsmarkt; uitstroom en bestemming van het schooljaar 1992-1993*, Stichting Landelijk Dienstverlenend Centrum voor Studietoelating en Beroepskeuzevoorlichting, Leeuwarden.
- Steedman, H. en K. Wagner (1987), 'A second look at productivity, machinery and skills in Britain and Germany', *National Institute Economic Review*, blz. 84-97.

Notities en Commentaren

LANGETERMIJNONTWIKKELINGEN IN DE OPBRENGSTEN VAN DIPLOMA'S OP DE NEDERLANDSE ARBEIDSMARKT; DOET SOCIALE HERKOMST ER WAT TOE?

Maarten H.J. Wolbers

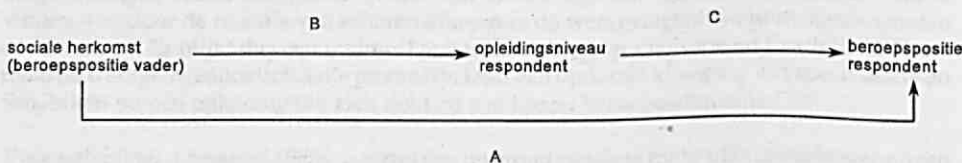
Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), Postbus 616, 6200 MD Maastricht

Uit eerder onderzoek is bekend dat hetzelfde diploma tegenwoordig tot een lagere beroepspositie leidt dan vroeger (Huijgen, 1989; Van der Ploeg, 1992; Wolbers & De Graaf, 1996). Vooral de lagere opleidingsniveaus hebben veel last van diploma-inflatie. In dit onderzoek is voorbij gegaan aan de gewoonte in de sociologie van de sociale stratificatie en mobiliteit om bij het bestuderen van (veranderingen in) de relatie tussen opleiding en beroep rekening te houden met de (veranderde) invloed van sociale herkomst. Dit is van belang omdat zowel het hoogstbereikte onderwijsniveau als het niveau van het uitgeoefende beroep rechtstreeks afhankelijk zijn van sociale herkomst (Blau & Duncan, 1967). Daarmee zou de bivariate samenhang tussen iemands opleiding en beroepspositie weleens gedeeltelijk schijn kunnen zijn (Bowles & Gintis, 1976). In deze notitie ga ik na in hoeverre de eerder gevonden ontwikkelingen in de opbrengsten van diploma's op de Nederlandse arbeidsmarkt zijn overschat, omdat er geen rekening is gehouden met de (veranderde) invloed van sociale herkomst.

Theoretische achtergronden

Een klassiek model binnen het sociale stratificatie- en mobiliteitsonderzoek is het zogenoemde statusverwervingsmodel (Blau & Duncan, 1967). Dit causale model maakt duidelijk hoe de maatschappelijke posities van ouders en hun kinderen met elkaar in verband staan. Het basismodel voor het proces van statusverwerving ziet er als volgt uit:

Figuur 1: De vereenvoudigde versie van het statusverwervingsmodel van Blau en Duncan



Links in figuur 1 is de statuspositie van de ouders van de respondent weergegeven (geïndiceerd door het beroepsniveau van de vader) en rechts de positie die de respondent zelf heeft bereikt (het beroepsniveau van de respondent). De samenhang tussen de statusposities van ouders en hun kinderen komt via twee wegen tot stand. Enerzijds is er sprake van directe statusovererving van ouders op kinderen (pad A), anderzijds verloopt de invloed van sociale herkomst indirect, waarbij het onderwijsniveau van de respondent een sleutelpositie inneemt (pad B-C).

Aan de hand van dit statusverwervingsmodel hebben Blau en Duncan hypothesen geformuleerd over historische ontwikkelingen in het proces van statusoverdracht: de directe invloed van sociale herkomst op het bereikte onderwijs- en beroepsniveau (de effecten A en B) zou zijn afgenomen, terwijl de directe invloed van opleiding op beroep (effect C) zou zijn toegenomen. De verklaring voor deze veranderingen zou zijn gelegen in het verdelingsregime van moderne, (post-) industriële samenlevingen (Treiman, 1970). Volgens de moderniseringstheorie laten hedendaagse samenlevingen zich steeds sterker karakteriseren door doelmatige

selectieprocessen die niet langer plaatsvinden op basis van toegeschreven ('ascribed') kenmerken zoals sociale herkomst. In plaats daarvan berust de toewijzing van schaarse goederen (zoals het bereikte beroep) tegenwoordig op individueel verworven ('achieved') kenmerken. Tot deze kenmerken behoren bekwaamheden, intelligentie en motivatie die alle in zekere mate zijn af te lezen aan iemands onderwijsniveau. Om deze reden zouden diploma's het belangrijkste selectiecriteria zijn geworden in moderne, meritocratische samenlevingen.

In Nederland blijkt ondersteuning voor de moderniseringstheorie (De Graaf & Luijckx, 1992). Ten eerste is de rol van het ouderlijk sociaal milieu op de onderwijskansen van kinderen sterk gedaald. Ten tweede is de directe overerving van beroepsposities tussen generaties afgenomen. Ten derde is de directe invloed van de behaalde onderwijskwalificaties op het bereikte beroepsniveau enigszins toegenomen. Deze drie bevindingen tezamen bevestigen de veronderstelde verschuiving van ascription naar achievement.

Het inzicht dat zowel het onderwijsniveau als de beroepspositie direct afhankelijk zijn van sociale herkomst en dat de samenhang tussen sociale herkomst en bereikte posities is verzwakt heeft twee gevolgen voor het bestuderen van langetermijntoewikkelingen in de opbrengsten van diploma's op de Nederlandse arbeidsmarkt. Allereerst veronderstel ik dat de eerder gevonden invloed van opleiding op het bereikte beroepsniveau is overschat; wanneer rekening wordt gehouden met sociale herkomst zijn de opbrengsten van diploma's geringer dan wanneer de rol van sociale herkomst buiten beschouwing wordt gelaten. Daarnaast formuleer ik de hypothese dat de overschatting in de opbrengsten van diploma's tegenwoordig geringer is dan vroeger. De reden hiervoor is dat de invloed van sociale herkomst op zowel de behaalde diploma's als het bereikte beroepsniveau is afgenomen. Dit betekent dat de gedocumenteerde diploma-inflatie is overschat.

Onderzoeksopzet

Om beide hypothesen te toetsen gebruik ik één groot gegevensbestand dat is samengesteld uit 16 Nederlandse (cross-sectionele) datasets verzameld in de periode 1970-1995. In dit samengestelde gegevensbestand, het Nederlandse Stratificatie- en Mobiliteitsbestand genoemd, is informatie aanwezig over het onderwijsniveau en de beroepspositie van de respondenten en (retrospectief) over de beroepspositie van hun vaders. Vanwege het feit dat de afhankelijke variabele het bereikte beroepsniveau is en de te bestuderen periode zeer ver teruggaat in de tijd, beperkt de analyse zich tot mannen (15-64 jaar). Vrouwen met een beroepsloopbaan waren in het verleden zeldzaam en bovendien vormden zij een selecte groep van hoogopgeleiden. Langetermijntrends in de samenhang tussen opleiding en beroepspositie schat ik door een cohortbenadering. Hiertoe worden opeenvolgende generaties respondenten via arbeidsmarktcohorten onderscheiden. Een arbeidsmarktcohort bestaat uit personen die in hetzelfde jaar de arbeidsmarkt hebben betreden. Het jaar van intrede - bij gebrek aan rechtstreekse informatie - heb ik gelijkgesteld aan het jaar van schoolverlating. Zodoende kan ik voor een zeer lang tijdvak (1921-1995) ontwikkelingen in de opbrengsten van diploma's bestuderen.

Het bereikte beroepsniveau bij intrede op de arbeidsmarkt wordt modelmatig gereconstrueerd door in de analyse rekening te houden met de hoeveelheid arbeidsmarktervaring. Oudere werknemers bezetten vanwege hun ervaring vaak hogere posities op de arbeidsmarkt dan recente schoolverlaters en daarom dient men te corrigeren voor de arbeidsmarktervaring van respondenten. Zowel de beroepspositie van de respondenten als die van hun vaders (als indicator voor sociale herkomst) wordt gemeten door het bereikte functieniveau. Ik gebruik daartoe de functieniveau-indeling van Huijgen (Huijgen, 1989) die via een hercodering van de 4-cijferige beroepenclassificatie van het Centraal Bureau voor de Statistiek (CBS) tot stand komt. Deze functieniveau-indeling is samengesteld aan de hand van een aantal kenmerken (aanleertijd, noodzakelijk geachte scholing en zelfstandigheid voor adequate beroepsuitoefening) en bestaat uit zeven categorieën gerangschikt van laag naar hoog niveau. Een nadeel van de indeling is dat ze zich beperkt tot de personen die werkzaam zijn in loondienst. De analyse berust dan ook alleen op deze categorie van werkenden. Hierdoor is het mogelijk dat

de effecten van sociale herkomst enigszins worden onderschat, want men kan verwachten dat personen in loondienst een kleiner herkomsteffect hebben dan zelfstandigen.

Het onderwijsniveau ten slotte is gebaseerd op de Standaard OnderwijsIndeling (SOI) van het CBS.

Resultaten

Met behulp van cumulatieve logit analyse (Agresti, 1990) breng ik langetermijntontwikkelingen in de opbrengsten van diploma's op de Nederlandse arbeidsmarkt in kaart. In deze analyse worden de opbrengsten van een diploma uitgedrukt in termen van relatieve kansen (logits) op het bereiken van een willekeurig hoger functieniveau. De waarde van ieder diploma vergelijk ik met de waarde van alleen lager onderwijs. Er zijn twee modellen onderscheiden. Het eerste model schat lineaire trends in de opbrengsten van diploma's. Ik houd hierbij geen rekening met sociale herkomst. In het tweede model is dat wel het geval. Het model schat de parameters van het eerste model opnieuw, maar nu na controle voor de invloed van het functieniveau van de vader. De resultaten van beide modellen vat ik samen aan de hand van figuur 2.

In figuur 2 wordt zichtbaar dat de opbrengsten van alle diploma's zijn afgenomen in de periode 1921-1995. Dit bevestigt eerdere onderzoeken waarin diploma-inflatie is aangetoond. De geobserveerde diploma-inflatie is niet gelijk voor ieder onderwijsniveau. De opbrengsten van diploma's in het voortgezet onderwijs (LBO/MAVO en MBO/HAVO/VWO) zijn het sterkst verminderd. Zo is voor LBO/MAVO het effect op de logit in vijftenzeventig jaar tijd gedaald van iets meer dan 2,0 punten naar 0,0 punten. Tegenwoordig zijn personen met een diploma op LBO/MAVO-niveau in even hoge functies werkzaam als individuen die alleen de lagere school hebben voltooid. De daling voor MBO/HAVO/VWO is ook aanzienlijk; de waarde van een diploma op dit niveau is van ongeveer 3,5 punten in 1921 gedaald naar iets meer dan 1,0 punten in 1995. De waarde van een diploma in het hoger onderwijs is minder sterk achteruit gegaan. In de onderzochte periode is het effect op de logit voor HBO gedaald van ongeveer 4,3 punten naar ruim 2,5 punten; voor een academische opleiding van bijna 6,0 punten naar zo'n 5,0 punten.

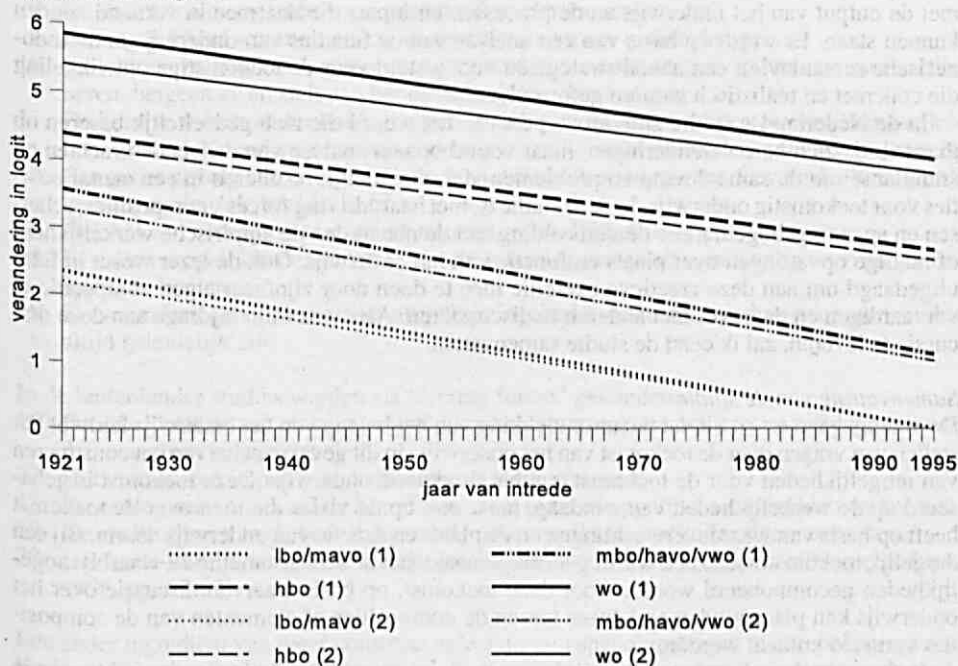
Figuur 2 maakt verder zichtbaar dat de ontwikkelingen in de opbrengsten van diploma's worden vertekend, als men geen rekening houdt met sociale herkomst. Ten eerste wordt de waarde van diploma's enigszins overschat zonder correctie voor sociale herkomst. Gemiddeld over de gehele onderzochte periode bedraagt het effect op de logit voor LBO/MAVO ongecontroleerd voor sociale herkomst 1,0661 en gecontroleerd 0,9875; dit betekent dat het oorspronkelijke effect voor 7,4 procent ($1 - 0,9875 / 1,0661 = 0,0737$) kan worden toegeschreven aan de rechtstreekse invloed van sociale herkomst. Voor MBO/HAVO/VWO, HBO en WO zijn deze percentages respectievelijk 7,9 procent, 6,3 procent en 6,6 procent. Ten tweede is de overschatting in de opbrengsten van diploma's tegenwoordig geringer dan vroeger. Wanneer de werking van sociale herkomst uit het opleidingseffect wordt verwijderd, blijkt de waardevermindering van diploma's geringer te zijn dan wordt verondersteld. In figuur 2 is goed te zien dat de trendlijnen die zijn gecorrigeerd voor de effecten van sociale herkomst vlakker lopen dan de overeenkomstige trendlijnen zonder correctie voor sociale herkomst.

Besluit

Op basis van deze resultaten stel ik dat de twee opgestelde hypothesen over de invloed van sociale herkomst op de mate van diploma-inflatie op de arbeidsmarkt worden bekrachtigd. In de eerste plaats is gevonden dat de opbrengsten van diploma's in beperkte mate worden overschat als er geen rekening wordt gehouden met de invloed van sociale herkomst. Ongeveer zeven procent van de oorspronkelijke samenhang tussen opleiding en beroep kan worden toegeschreven aan de directe invloed van sociale herkomst. In de tweede plaats heb ik laten zien dat de toch al geringe overschatting in de opbrengsten van diploma's is afgenomen in de onderzochte periode en in de jaren negentig nauwelijks nog bestaat. In het verleden was de waar-

de van een diploma gedeeltelijk toe te schrijven aan sociale herkomst, maar tegenwoordig is dit (bijna) niet meer het geval. Daarmee is de in ander onderzoek geconstateerde diploma-inflatie wat overschat. Dat individuen tegenwoordig niet meer zo eenvoudig met hun diploma een hoge positie op de arbeidsmarkt weten in te nemen als vroeger, komt niet alleen doordat de onderwijsexpansie zo ver is voortgeschreden, maar ook doordat ouders minder gemakkelijk hun gunstige posities rechtstreeks hebben kunnen overdragen op hun kinderen. Deze overschatting is echter zeer gering. Ook als we rekening houden met de inzichten uit het sociale stratificatie- en mobiliteitsonderzoek over de (veranderde) invloed van sociale herkomst op de onderwijs- en beroepsloopbaan, dan luidt de conclusie dat de Nederlandse arbeidsmarkt te kampen heeft met diploma-inflatie.

Figuur 2: Langetermijntwikkelingen in de opbrengsten van diploma's op de Nederlandse arbeidsmarkt: (1) zonder controle voor sociale herkomst en (2) met controle voor sociale herkomst. (Bron: Nederlandse Mobiliteits- en Stratificatiebestand 1970-1995, eigen berekeningen)



Literatuur

- Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley & Sons.
- Blau, P. & Duncan, O. (1967). *The American occupational structure*. New York: John Wiley & Sons.
- Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America*. New York: Basic.
- Graaf, P. de, & Luijckx, R. (1992). Van 'ascription' naar 'achievement'? Trends in statusverwerving in Nederland tussen 1930 en 1980. *Mens en Maatschappij*, 67, 412-433.
- Huijgen, F. (1989). *De kwalitatieve structuur van de werkgelegenheid in Nederland, Deel III*. OSA-voorstudie nr. V33. 's-Gravenhage: OSA.
- Ploeg, S. van der (1992). Verspreiding en opbrengsten van diploma's. *Sociale Wetenschappen*, 35, 44-60.
- Treiman, D. (1970). Industrialization and social stratification. In E. Laumann (Ed.), *Social stratification, research and theory for the 1970s* (pp. 207-234). Indianapolis: Bobbs-Merill.
- Wolbers, M., & Graaf, P. de (1996). Langetermijntwikkelingen in de opbrengsten van diploma's op de Nederlandse arbeidsmarkt. *Tijdschrift voor Arbeidsvraagstukken*, 12, 296-309.

WAARZEGGEN EN LUCHTFIETSEN, COMMENTAAR OP 'TOEKOMSTEN VOOR HET FUNDEREND ONDERWIJSBELEID'

Bert P. M. Creemers

Inleiding

Op het moment dat de redactie van het *Tijdschrift voor Onderwijsresearch* mij vroeg om commentaar op de studie 'Toekomst voor het funderend onderwijsbeleid' (In 't Veld, De Bruijn, & Lips, 1996) was ik bezig met de afwerking van een rapport over de evaluatie van het 'Primary Education Quality Improvement Project' in Indonesië. In dat kader nam ik kennis van een rapport van het Ministerie van Onderwijs en Cultuur over de ontwikkelingen in de komende 25 jaar in het Indonesische onderwijs (Ministry of Education and Culture, 1997).

Beide studies zijn gericht op het uitzetten van lijnen voor de toekomst van het onderwijsbeleid, maar dat is ook bijna de enige overeenkomst. Van het Indonesische rapport is driekwart gewijd aan een analyse van problemen in het onderwijs: aantallen leerlingen, problemen met de output van het onderwijs en de processen en inputs die daarmee in verband zouden kunnen staan. Er wordt op basis van een analyse van de functies van onderwijs in de Indonesische samenleving een aantal strategieën voorgesteld voor de toekomstige ontwikkeling die concreet en realistisch worden geformuleerd.

In de Nederlandse studie zijn ontwerpers aan het woord die zich gedeeltelijk baseren op thans reeds zichtbare veranderingen, maar vooral op een analyse van drijvende krachten en stimulansen in de samenleving en problemen, die uiteindelijk resulteren in een aantal posities voor toekomstig onderwijs. In deze studie is, met haar 'driving forces', composities, schetsen en spanningsbogen, meer de verbeelding aan de macht dan de empirische werkelijkheid of huidige opvattingen over plaats en functie van het onderwijs. Ook de lezer wordt in feite uitgedaagd om aan deze creatieve exercitie mee te doen door zijn/haar eigen compositie te vervaardigen en daarover met anderen te discussiëren. Alvorens mijn bijdrage aan deze discussie te leveren, zal ik eerst de studie samenvatten.

Samenvatting van de studie

De auteurs gaan ervan uit dat de herontdekking van het belang van het onderwijs leidt tot het stellen van vragen over de toekomst van het onderwijs. In dit geval gaat het om het construeren van mogelijkheden voor de toekomst van het funderend onderwijs. Deze toekomst is gebaseerd op de werkelijkheden van vandaag, maar ook op de visies die men over de toekomst heeft op basis van waardeverwachtingen en de plaats en functie van onderwijs daarin. Bij een dergelijk toekomstonderzoek wordt gebruikgemaakt van de scenariomethode, waarbij mogelijkheden gecomponeerd worden voor deze toekomst, op basis waarvan discussie over het onderwijs kan plaatsvinden en keuzes tussen de composities of elementen van de composities gemaakt kunnen worden.

In de maatschappij en het onderwijs kan men al een aantal veranderingen zien aankomen in het funderend onderwijs. Deze zijn door de auteurs opgespoord en beschreven. Het betreft de veranderingen ten aanzien van de inhoud van het onderwijs, de betrokkenheid en de keuze van de ouders, de werkwijzen binnen het onderwijs, de positie en de rol van docenten en de bestuurlijke en financiële voorwaarden. Deze zogenaamde 'gaandeweg veranderingen' zijn door de auteurs geformuleerd, waarschijnlijk op basis van een visie op en interpretatie van de ontwikkelingen in het onderwijs op dit moment. Deze gaandeweg veranderingen zijn in een werkconferentie met deskundigen uit het onderwijsveld besproken, waarbij nuancering en inschattingen van de ontwikkelingen door de deelnemers werden ingebracht. In de gepubliceerde studie zijn de bijlagen, die wel in de tekst zijn vermeld, vervallen, zodat onbekend is wie deze deskundigen zijn, zolang men de foto's die in het rapport zijn opgenomen kan ontcijferen.

Bij de veranderingen gaandeweg gaat het om een beschrijving van de huidige situatie en inschattingen over de ontwikkelingen op redelijk korte termijn. Dat is minder het geval bij de

andere ingrediënten voor de compositie, die meer het resultaat zijn van verbeeldingskracht. Deze ingrediënten zijn de zogenaamde drijvende krachten in de samenleving en in het onderwijs en de verwachte maatschappelijke ontwikkelingen, die van invloed zouden kunnen zijn op het funderend onderwijs, in deze studie 'driving forces' genoemd. In vier werkplaatsen over 1) samenlevingsintegratie en veiligheid, 2) samenlevingsvormen, 3) zingeving, kennis en kunst en 4) communicatietechnologie en media, en op basis van de analyse van een aantal buitenlandse studies is een twintigtal 'driving forces' onderscheiden:

1. De trend van individualisering zal doorzetten.
2. Grotere aandacht voor het lokale.
3. Toenemende aandacht voor zingevingsvraagstukken, waarden en normen.
4. Het kind zal steeds meer een 'black box' worden.
5. Het streven naar maatschappelijke integratie zal toenemen, terwijl tegelijkertijd de onhaalbaarheid daarvan ook duidelijker wordt.
6. De schaarste van een aantal praktische voorzieningen zal anders zijn dan voorheen.
7. De maatschappelijke uitvalsverschijnselen zullen toenemen.
8. De variëteit aan samenlevingsvormen zal verder toenemen.
9. De ongelijkheid van mensen zal als constatering meer nadruk krijgen.
10. Individuen zullen in een groter aantal verschillende contexten en rollen moeten functioneren, hetgeen eisen stelt aan het communicatieve vermogen.
11. De huidige informatietechnologische ontwikkelingen zullen tot grote veranderingen leiden, waarbij de aard en de richting van de veranderingen grotendeels onvoorspelbaar is.
12. Een ander gevolg van de technologische ontwikkeling zal zijn dat de woordcultuur wordt vervangen door de beeldcultuur.
13. Ook zal een vervaging plaatsvinden tussen werkelijkheid en fictie.
14. Er zullen mogelijkheden ontstaan om tot nieuwe gemeenschapsvorming en vrijetijdsbeleving te komen.
15. De veranderingen zullen echter altijd zijn ingebed in bestaande gedragspatronen en dus altijd geleidelijk zijn.

In de buitenlandse studies werden als 'driving forces' gevonden:

16. De arbeidsmarkt zal grote segmenten kennen waar mensen werken op een lager niveau dan waarvoor ze zijn opgeleid.
17. Het vertrouwen in publieke instituties zal sterk afnemen.
18. De ecologische crisis zal de agenda van veel maatschappelijke vraagstukken beheersen.
19. De snelle groei van de bevolking zal de druk op allerlei instituties doen toenemen.
20. Het korte-termijndenken zal steeds dominanter worden (In 't Veld, De Bruijn, & Lips, 1996, pp. 57-63).

Een ander ingrediënt van meer construerende aard zijn spanningsbogen tussen concurrerende beelden. Bij spanningsbogen gaat het om de constructie van een aantal concurrerende beelden en waarden rondom een bepaald thema die in concurrentie staan met elkaar zodat de discussie en constructie over toekomstige ontwikkelingen scherper wordt. Rondom tien thema's, zoals waarden pluralisme, de invloed van de technologie, de verantwoordelijkheid en financiering, een brede of een smalle school, het curriculum, werkwijze in het onderwijs, rol van de omgeving, rol van de school, rol van de docent, de veranderingen in het funderend onderwijs. In de filosofie achter het funderend onderwijsbestel zijn de werkelijke spanningsbogen van concurrerende waarden geformuleerd.

Een voorbeeld van een spanningsboog is die over de rol van de docenten (zie volgende pag.).

Een volgende stap of element in de constructie is de vertaling van de 'driving forces' en een aantal spanningsbogen in schetsen voor de school. Daarbij wordt geen rekening gehouden met de uitvoerbaarheid of de haalbaarheid. Deze schetsen, beelden, zijn bedoeld om te inspi-

renen, terwijl ze tegelijkertijd concreter zijn doordat ze een type school construeren. De schetsen waartoe de auteurs zijn gekomen zijn:

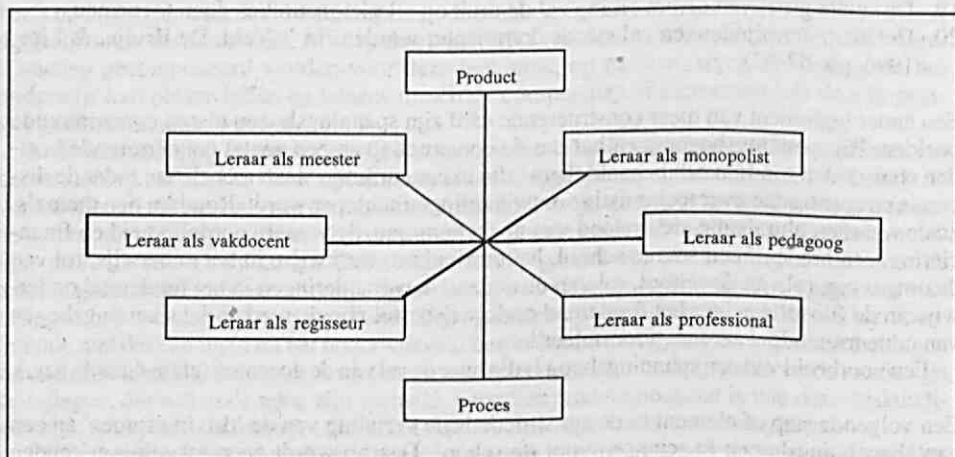
1. de gedifferentieerde school, waarin onderwijs op maat wordt aangeboden;
2. de prestatieschool, ook gericht op individuele leertrajecten, maar sterk cognitief gericht;
3. de praktijkschool, op vaardigheden en met name sociale vorming gericht;
4. de school als leerhuis, waarin zingeving- en nuttigheidsvakken in ruime mate aanwezig zijn;
5. de school als thuis, waarbij de school als verlengstuk van het gezin wordt gezien;
6. de pluriforme school, waar de nadruk wordt gelegd op de culturele pluriformiteit.

Deze gedachtenlijnen uit het rapport, de 'driving forces', de veranderingen gaandeweg, de spanningsbogen en de schetsen zijn uiteindelijk samengevoegd in drie composities voor toekomstig funderend onderwijs. Deze drie composities zijn:

1. uniek kind, krachtig bestel;
2. ouders en school, evenwichtige diversiteit;
3. samenhang bestel, gelijke voorwaarden voor scholen.

Deze composities worden uitvoerig beschreven aan de hand van uitgangspunten, kenmerken van het bestel, kenmerken van het beleid en de centrale waarden. In compositie 1 is het kind het centrale aangrijpingspunt: de leraar is begeleider, met veel nadruk op collectieve maatschappelijke verantwoordelijkheden. In compositie 2 krijgen de ouders een belangrijke rol toebedacht: zij zullen waarschijnlijk meer belangstelling krijgen voor het onderwijs in het algemeen en de leerresultaten van het eigen kind in het bijzonder, hetgeen waarschijnlijk zal leiden tot een verdere benadrukking van diversiteit. In de derde compositie staat de school centraal, waarbij er grote aandacht is voor het tegengaan van ongelijkheid door standaardisatie van het overheidsbeleid. Scholen hebben een breed curriculum met concrete eindtermen, onderwijsleerstofgericht, leraren zijn monopolist en vakdocent en hebben door de concreetheid van de eindtermen weinig invloed op de inhoud van het curriculum. Het gaat vooral om uniformiteit, gelijke financiële toerusting, waarbij extra middelen worden toegekend voor het wegwerken van individuele of sociale achterstanden en conformiteit aan een op nationaal niveau overeengekomen curriculum.

Spanningsbogen rond het thema 'de rol van docenten'
(Bron: In 't Veld, de Bruijn, & Lips) 1996, p. 113.



Commentaar

De neiging is groot, zeker wanneer de auteurs daar uitdrukkelijk toe uitnodigen, om ook een duit in het zakje te doen en, ofwel op basis van nieuwe bouwstenen ofwel met de aangereikte ingrediënten, een compositie te maken of wijzigingen binnen de composities aan te brengen. Degenen die enigszins bekend zijn met mijn positie binnen het onderwijsonderzoek zal het duidelijk zijn welke opmerkingen dan gemaakt worden over de functie van de school, de doelstellingen van het onderwijs, de beperkte 'range' die daarvoor nodig is gezien de tijd en andere factoren die in het onderwijs een rol spelen, eindigend in een pleidooi voor een effectieve, efficiënte school (modificatie van compositie 3?). Dat is waarschijnlijk ook de reden waarom de redactie mij voor een reactie heeft benaderd.

De vraag is echter of er niet wat kanttekeningen geplaatst moeten worden bij de operatie op zich en de suggestie die de uitkomsten ervan wekken. Natuurlijk, beleid is op de toekomst gericht. Zelfs wanneer op dit moment uitspraken worden gedaan over de wenselijkheid van veranderingen in het onderwijs, zal het een hele tijd en veel activiteiten kosten alvorens deze wensen omgezet zijn in onderwijspraktijk. Daarmee is het al een operatie in de toekomst geworden. Wanneer het dan een operatie in de toekomst is, is het zeker zinvol om te kijken of niet een aantal te verwachten ontwikkelingen kan worden meegenomen. Maar dit is naar mijn mening geen vrijbrief voor exercities waarbij het aantal vrijheidsgraden toch wel erg groot wordt en het gevaar aanwezig is dat scenario's en schetsen zich loszingen van de werkelijkheid.

Daarbij wordt uitgegaan van de maakbaarheid van de samenleving en van het onderwijs, maar tevens van de veronderstelling dat de werkelijkheid en het onderwijs als onderdeel ervan grotendeels zichzelf maken en min of meer 'vanzelf' veranderen. En dat is maar goed ook. Het betekent dat, naast wensen, ideeën over de toekomst van de samenleving en de plaats van het onderwijs daarin, een heel ander gegeven bepalend is, namelijk de feitelijkheid van het onderwijs van vandaag, de problemen die zich daarin voordoen en de pogingen om bij te dragen aan het oplossen van deze problemen. De grootste en duidelijkste relatie die de studie met de werkelijkheid van het onderwijs heeft is, behalve een aantal referenties naar de huidige situatie, de conferentie over 'gaandeweg veranderingen', waar in elk geval een relatie met veranderingen in het huidige onderwijs uitgangspunt is geweest. Daarbij is ook, gezien een aantal opmerkingen in dit verband, duidelijk geworden dat een aantal veranderingen in het onderwijs niet zo snel verloopt en dat sommige veranderingen een lange doorwerkijd hebben, veel tijd nemen, maar ook, hetgeen overigens door de auteurs wordt onderkend, onderdelen van het onderwijs, factoren, karakteristieken, componenten, samenhang hebben en dat wijzigingen ten aanzien van bepaalde onderdelen vaak ook invloed hebben op overige onderdelen van het systeem.

Ondanks het feit dat over 'veranderingen gaandeweg' wordt gesproken tegenover radicale veranderingen, die ook in het kader van deze studie mogelijk worden geacht, kan ik me niet aan de indruk onttrekken dat de feitelijkheid van het onderwijs en de maatschappij in het geheel van de studie toch een tamelijk onbelangrijke plaats inneemt; in die zin is het ontwerpen van deze scenario's vergelijkbaar met de doelstellingsonderzoeken van enkele decennia terug, waarin aan 'de maatschappij' werd gevraagd welke doelen door het onderwijs zouden moeten worden bereikt. Deze discussie werd niet gehinderd door de beperkingen die in de werkelijkheid aanwezig zijn, met betrekking tot de tijd en andere randvoorwaarden voor onderwijs. Immers, men kan wel van alles willen, maar wanneer daar de tijd, de schooltijd, de capaciteiten van onderwijs en van leerkrachten tegenover staan, dan is het formuleren van doelstellingen en van scenario's een tamelijk irreële bezigheid.

In het doelstellingendebat werd deze situatie doorbroken doordat duidelijk gesteld werd dat datgene wat bereikt werd in het onderwijs als uitgangspunt genomen zou moeten worden voor de verdere formulering van doelstellingen. Dat wil zeggen dat duidelijk moest worden wat op een bepaald moment bereikt werd, gegeven de omstandigheden, en wat gezien datgene wat bereikt werd en wat wenselijk werd geacht mogelijk in de toekomst onder welke condities bereikt zou kunnen worden.

Ook de scenariomethode zou naar mijn mening heel wat realistischer worden, wanneer de werkelijkheid van alledag en met name de schoolse werkelijkheid als uitgangspunt zou worden genomen en vervolgens voor elk scenario zou worden aangegeven onder welke condities keuzes zouden kunnen worden gerealiseerd, wat ze vereisen aan inputs en processen. Nu wekken ze de indruk dat de verbeelding aan de macht is.

Een nog belangrijker vraag in dit verband is of het wel realistisch is om, telkens wanneer er eventueel veranderingen in de maatschappij - want ook dat is niet duidelijk - plaatsvinden, te veronderstellen dat het onderwijs daardoor wijzigt en daardoor min of meer op zijn kop komt te staan. Ook al veranderen de opvattingen over bijvoorbeeld geneeskunde, dat betekent nog niet dat alle instituten rondom de gezondheidszorg ineens zouden moeten veranderen. Over het geheel genomen blijft de functie van instituten binnen de gezondheidszorg hetzelfde. Dit geldt ook voor het onderwijs. Uiteindelijk blijft de functie van onderwijs hetzelfde: ontwikkeling van individuele gaven en krachten - dus individuele capaciteiten - en introductie in de maatschappij. Dit impliceert dat er aan de ene kant moet worden nagegaan op welke wijze deze capaciteiten ontwikkeld kunnen worden en aan de andere kant dat er een socialisatie en introductie in de maatschappij plaatsvindt. Dat kan in meer of mindere mate aan verandering onderhevig zijn in de loop van de tijd, omdat de maatschappij verandert, en zeker in de arbeidsmarkt misschien andere functies eisen, maar over het geheel genomen blijft de functie van de school in deze min of meer gelijk.

In dit verband dient er ook op gewezen te worden dat langzamerhand ook duidelijk is dat het onderwijs in de tijd die haar ter beschikking staat niet alles kan leveren wat mensen vervolgens als adolescenten, volwassenen en ouderen gedurende hun hele leven nodig zullen hebben. In feite wordt er in toenemende mate van uitgegaan dat er basisvaardigheden en basis-kennis op de diverse terreinen wordt verworven en dat leerlingen met deze kennis en vaardigheden vervolgens in de maatschappij kunnen functioneren. Dit betekent - en dat gebeurt ook al - dat meer dan vroeger de nadruk wordt gelegd op de transfer van kennis, leren leren en dergelijke vaardigheden. Maar dat is iets anders dan wat uit dit rapport opdoemt, dat onderwijs telkens andere functies heeft en andere kennis en andere vaardigheden aangeleerd zouden moeten worden.

Ter vermijding van telkens opnieuw veranderingsprocessen in de school, waardoor rust in het onderwijs een illusie wordt, zou het wellicht verstandig zijn om te wijzen op deze constante taak binnen het onderwijs en de duurzaamheid van de school die de tand des tijds wel kan doorstaan, omdat ze in staat is om, daar waar nodig, modificaties op het grondpatroon aan te brengen. Scholen blijken hoe langer hoe meer in staat te zijn om als lerende organisatie te functioneren.

Tegen een dergelijke achtergrond is het van belang te pleiten om kwaliteit in het onderwijs te vertalen als de kwaliteit van de onderwijsuitkomsten op het niveau van de leerlingen; dat betekent effectiviteit van het onderwijs en die processen en inputs van het onderwijs waardoor leerlinguitkomsten over een breed terrein kunnen worden bereikt. Uiteindelijk blijft het ook in dat nieuwe onderwijs gaan om de prestaties van leerlingen op een breed terrein. De school zal zich ook in de toekomst met name moeten richten op het aanleren van die vaardigheden en het overbrengen van die kennis die niet elders - in het gezin, op straat - kan worden aangeleerd. Dat betekent cognitieve vaardigheden en het cognitieve substraat voor andere kennis en vaardigheden.

In het kader van hun poging om tot onderscheiden scenario's te komen zijn de auteurs naar mijn mening veel verder gegaan dan in de werkelijkheid van het onderwijs aan de orde is. In de werkelijkheid van het onderwijs zal het altijd gaan om diverse zaken tegelijkertijd: én om invloed van de ouders én om aansluiting bij het gezin; én om cognitieve vaardigheden én om affectieve en sociale vaardigheden. Bovendien zijn scenario's, schetsen en 'driving forces' wat minder scherp uit elkaar te houden dan in deze exercitie wordt gesuggereerd. Dat betekent, tegen de achtergrond van de brede functie en taken van het onderwijs, dat er ook meer gemeenschappelijks zit in de schetsen en composities dan in de uitwerking wordt gesuggereerd.

Het voorgaande impliceert naar mijn mening dat het ontwikkelen van scenario's voor de toekomst meer dan in deze studie dient uit te gaan van de werkelijkheid van het huidige onderwijs. Wat wordt er in het onderwijs bereikt, welke uitkomsten worden er behaald, welke doelstellingen worden nagestreefd? Welke processen en inputs dragen bij aan het verklaren van de variantie in leerlingresultaten en wat zou op basis daarvan in de toekomst gewenst zijn? Hoe kan die toekomst dan vormgegeven worden in wijzigingen van inputs, van context en van processen? Daarbij zal het altijd gaan om variaties op hetzelfde thema, namelijk individuele ontplooiing en maatschappelijke introductie. Dat is nu net wat in het rapport van het Ministerie van Onderwijs en Cultuur in Indonesië wordt aangehaald, wanneer het gaat over de ontwikkelingen in de komende 25 jaar in Indonesië. Welke veranderingen doen zich voor in de maatschappij, met name in de arbeidsmarktsituatie, en hoe kan de intellectuele uitrusting van leerlingen worden bevorderd?

Het rapport en het commentaar wekken de indruk dat deskundigen veel meer dan in het verleden een rol willen en kunnen spelen in scenario's en de doelstellingendiscussie rondom het onderwijs. Naar mijn mening is dat terecht, al was het alleen maar vanwege het feit dat, wanneer de empirische werkelijkheid als uitgangspunt wordt genomen, van onderwijsonderzoekers en onderwijskundigen mag worden verwacht dat ze daarvoor het materiaal kunnen aandragen. Uiteindelijk beslissen zij niet over het doel van het onderwijs - dat blijft inderdaad een beleidskeuze - maar ze kunnen wel informatie verschaffen over de huidige stand van zaken in het onderwijs: wat realistische opties zijn en wat de condities zijn waaronder eventuele veranderingen bereikt kunnen worden en via welke processen dat met meer of minder succes kan worden gerealiseerd (zie Creemers, 1996). Dat ontbreekt in dit rapport ten enenmale en de schetsen blijven daardoor 'juweeltjes van vergezichten'.

Literatuurreferenties

- Creemers, B.P.M. (1996). The goals of school effectiveness and school improvement. In D. Reynolds, R. Bolton, B. Creemers, D. Hopkins, L. Stoll, & N. Lagerweij (Eds.), *Making good schools* (pp. 21-35). London: Routledge.
- In 't Veld, R., de Bruijn, H., & Lips, M. (1996). *Toekomst voor het funderend onderwijsbeleid*. Den Haag: Sdu.
- Ministry of Education and Culture (1997). *Education development in Indonesia*. Jakarta: Office of Educational and Cultural Research and Development.

COMMENTAAR BIJ ARTIKEL 'GEZIN EN SCHOOLLOOPBAAN'

A.C. Glebbeek

ICS/Vakgroep Sociologie, Grote Rozenstraat 31, 9712 TG Groningen

Van Veen c.s. stellen dat zij met hun onderzoek de volgende hypothese willen toetsen: 'De invloed van het cultureel kapitaal van de ouders op de schoolloopbaan van het kind is meer positief naarmate er meer sociaal kapitaal aanwezig is in de relatie tussen de ouders en het kind. Deze invloed is minder positief naarmate er minder sociaal kapitaal in deze relatie aanwezig is. Met andere woorden, er is sprake van een interactie-effect tussen sociaal en cultureel kapitaal op de schoolloopbaan van het kind.' (artikel, p. 000.) Hun conclusie is dat deze hypothese niet wordt bevestigd. Ik meen dat dit een onterechte conclusie is.

Ik beperk mij in dit commentaar tot het meningsverschil dat is gerezen omtrent de interpretatie van de onderzoeksresultaten. Dat is in dit geval een belangrijk meningsverschil, gezien de sterke conclusies die de auteurs zelf aan hun onderzoek verbinden. Ik citeer: 'De mate van aanwezigheid blijkt in die zin van invloed te zijn op de schoolloopbaan dat juist ouders die weinig aanwezig zijn een positievere invloed hebben op de schoolloopbaan van hun kinderen

dan ouders die veel aanwezig zijn. (...) Colemans theorie over sociaal kapitaal in het gezin, meer specifiek dat sociaal kapitaal een belangrijke voorwaarde is voor de invloed van cultureel kapitaal, gaat dus niet op voor deze Nederlandse steekproef (artikel, p. 000, mijn cursiveringen). Het is duidelijk dat de auteurs van mening zijn dat de hypothese is weerlegd, sterker nog: dat eerder het tegendeel het geval is. Dat is ook duidelijk de boodschap die spreekt uit het *abstract*, waarmee de onderzoeksresultaten in de onderwijskundige literatuur worden vastgelegd.

Het zou jammer zijn wanneer de onderzoeksresultaten op deze wijze herinnerd zullen worden, want dezelfde uitkomsten laten mijns inziens ook een geheel andere conclusie toe. De conclusies van Van Veen c.s. zijn naar mijn oordeel het gevolg van een overinterpretatie van de uitkomsten voor de vaders en een onderinterpretatie van de uitkomsten voor de moeders. Het gewicht dat wordt gehecht aan het negatieve resultaat bij de vaders komt voort uit een onbereflecteerde toepassing van de theorie van het sociale kapitaal. Aan het gevonden positieve resultaat bij de moeders wordt in het artikel vrijwel geheel voorbijgegaan.

Wat betreft de resultaten voor de vaders vragen Van Veen c.s. zich onvoldoende af of de betreffende waarnemingen wel geschikt zijn om de hypothese op te toetsen. Uit de in Tabel 2 vermelde frequenties wordt duidelijk dat het aantal waarnemingen van hoger opgeleide en veel aanwezige vaders zeer beperkt is, zodat de betreffende combinaties wellicht te weinig voorkomen om over deze kwestie uitsluitel te bieden. Het normale patroon lijkt te zijn dat vaders veel tijd aan hun werk besteden en weinig thuis zijn als hun kind van school komt (dat geldt voor 219 van de 235 vaders). De 16 vaders van cluster 3 vormen duidelijk een atypische groep, wier aanwezigheid thuis waarschijnlijk niet als een blijk van veel sociaal kapitaal mag worden opgevat.¹ Bij de overige 'drukke' vaders blijkt meer aandacht voor het kind er inderdaad niet toe te doen. Dat is vanuit de theorie gezien ook niet verwonderlijk, omdat deze vaders op voorhand niet de meest relevante helft van het ouderpaar vormen om het hier geselecteerde aspect van sociaal kapitaal te onderzoeken. De sociaal kapitaaltheorie veronderstelt dat, gegeven de rolpatronen in onze samenleving, vaders vooral het *gezinsexterne* sociaal kapitaal inbrengen, wat niet loopt via het mechanisme van hun aandacht of aanwezigheid binnen het gezin. Alleen gezinnen waarin in het geheel geen vader aanwezig is (eenoudergezinnen dus) blijven verstoken van deze sociaal-kapitaal-inbreng, maar deze gezinnen zijn hier niet aan de orde. Binnen twee-oudergezinnen is het vooral de moeder die de dagelijkse gezinsopvoeding ter hand neemt en dus het sterkst het *gezinsinterne* aspect van sociaal kapitaal tot gelding brengt. In de theorie zelf ligt daarom besloten dat, voor de in dit onderzoek geselecteerde aspecten (aandacht en aanwezigheid thuis), de resultaten van de moeders bij voorbaat veelzeggender zijn.

Welnu, kijk naar de resultaten in de tabellen 5 en 6. In Tabel 5 blijkt een overheersend effect van de interactieterm van het opleidingsniveau en de aanwezigheid van de moeder. Hoe meer de moeder aanwezig is, hoe positiever de invloed van haar opleidingsniveau. De vier desbetreffende interactietermen wijzen alle in dezelfde richting. Het interactie-effect is zelfs zo sterk, dat dit het primaire effect van het opleidingsniveau van de moeder overheerst: haar beide opleidingsdummies zijn - zéér ongebruikelijk - negatief.² Hetzelfde sterke interactie-effect kan men aflezen in Tabel 6: de positieve invloed van het opleidingsniveau van de moeder is het duidelijkst in cluster 3, iets minder geprononceerd in cluster 2 en afwezig in cluster 1. Dit alles is vrijwel letterlijk in overeenstemming met hetgeen in de hypothese wordt gesteld. *Door de auteurs wordt dit resultaat echter geheel genegeerd.*

Wat kan de reden voor deze merkwaardige omissie zijn? Als ik het goed zie, lijken er twee oorzaken in het spel te zijn. Bij de interpretatie van Tabel 5 maken de auteurs waarschijnlijk de fout 'niet-significant' gelijk te stellen aan 'niet'. Drie van de hier gepresenteerde interac-

tietermen hebben weliswaar het juiste teken, maar zijn statistisch niet significant. Een niet-significante coëfficiënt staat hier echter niet op zichzelf, maar maakt deel uit van een consistent patroon. De toevalskans dat het in omvang bescheiden effect in de populatie niet bestaat neemt daardoor af en de kans op een onterechte aanvaarding van de nulhypothese toe. Ik bestrijd niet dat het effect bij de hoog-opgeleide moeders weinig spectaculair is, maar het patroon in zijn totaliteit is niet in strijd met de theorie.

De tweede oorzaak is dat de auteurs de heldere uitkomsten proberen te compliceren door, *via horizontaal lezen*, verstoringen op te sporen.³ Zo zou de bevinding dat bij de laag-opgeleide moeders juist de kinderen van de minst aanwezige moeders de beste resultaten halen een tegenindicatie zijn. Dat is het echter niet in het licht van de hier te toetsen theorie: de laag opgeleide moeders die het minst aanwezig zijn doen kennelijk het minste kwaad, omdat sociaal kapitaal hier het minst het deficiet in cultureel kapitaal kan overdragen. Het omgekeerde voorbeeld heeft betrekking op de lage scores van kinderen van veel aanwezige laag-opgeleide moeders en kan met hetzelfde argument bestreden worden. Principiëler is echter nog het volgende. De beide door Van Veen c.s. gemaakte tegenwerpingen duiden op het onvermogen directe effecten en interactie-effecten te onderscheiden en daarmee hun eigen hypothese te doorgronden. In de beide tegenvoorbeelden stappen zij ongemerkt over op een andere hypothese: 'meer aanwezigheid is altijd beter dan minder aanwezigheid'. Wat daar ook voor gezegd mag worden, *het is niet de door de auteurs zelf geformuleerde hypothese*. Die hypothese benadrukt een interactie-effect, niet het directe effect van sociaal kapitaal.

Afrondend: de gevonden resultaten kunnen de sterke conclusies van de auteurs niet dragen. Op grond van hun resultaten kan de hypothese, en daarmee de achterliggende theorie, niet worden verworpen.

1. Zijn hier werklozen bij of arbeidsongeschikten? Uit onderzoek is bekend dat langdurige inactiviteit van de ouders geen gunstige uitgangssituatie biedt voor de schoolloopbanen van de kinderen. Vgl. H. te Grotenhuis, *Bijstandskinderen*, Amsterdam University Press, 1993.

2. Dit komt door de overheersende invloed van het interactie-effect. Tabel 3 zonder de interactietermen laat voor de moeders wel de normale positieve opleidingscoëfficiënten zien.

3. Voor het toetsen van het in de hypothese gestelde interactie-effect moet men de resultaten in Tabel 6 uiteraard *verticaal* lezen: binnen welke clusters van 'aanwezigheid' blijkt het verband tussen de opleiding van de ouders en de schoolloopbaan van de kinderen het duidelijkst te bestaan? Horizontale lezing heeft betrekking op het *directe effect* van de aanwezigheid van de ouders: maakt gegeven het opleidingsniveau de mate van aanwezigheid uit?

HOE DOMINANT IS DE INVLOED VAN MOEDERS OP DE SCHOOLLOOPBAAN? EEN ANTWOORD OP DE REACTIE VAN GLEBBEEK

Klaas van Veen, Eddie Denessen, Peter van der Kley en Jan Gerris

Een centrale vooronderstelling in het commentaar van Glebbeek is dat voor de invloed van het 'gezinsinterne' sociaal kapitaal, de moeder belangrijker zou zijn dan de vader, omdat, aldus Glebbeek, *'vaders op voorhand niet de meest relevante helft van het ouderpaar vormen om het hier geselecteerde aspect van sociaal kapitaal te onderzoeken. De sociaal kapitaaltheorie veronderstelt dat, gegeven de rolpatronen in onze samenleving, vaders vooral het gezinsexterne sociaal kapitaal inbrengen, wat niet loopt via het mechanisme van hun aan-*

dacht of aanwezigheid binnen het gezin. Alleen gezinnen waarin in het geheel geen vader aanwezig is (eenoudergezinnen dus) blijven verstoken van deze sociaalkapitaalbreng, maar deze gezinnen zijn hier niet aan de orde. Binnen tweeoudergezinnen is het vooral de moeder die de dagelijkse gezinsopvoeding ter hand neemt en dus het sterkst het gezinsinterne aspect van sociaal kapitaal tot gelding brengt. In de theorie zelf ligt daarom besloten dat, voor de in dit onderzoek geselecteerde aspecten (aandacht en aanwezigheid thuis), de resultaten van de moeders bij voorbaat veelzeggender zijn'.

Het kan betwijfeld worden of deze vooronderstelling over de dominantie van moeders terecht is en niet wordt tegengesproken door sociologisch onderzoek. Uit ander onderzoek (De Graaf & Nieuwbeerta, 1995)* blijkt dat er geen sprake is van mannelijke danwel vrouwelijke dominantie.

Ook op basis van onze resultaten is het de vraag of er een verschil gemaakt mag worden tussen de invloed van vaders en moeders op de schoolloopbaan van kinderen. Tabel 5 laat zien dat de totaal verklaarde variantie voor vaders en moeders nagenoeg gelijk is, dus hun invloed is even groot. Bovendien kan op grond van Tabel 3 Glebbeeks assumptie, dat vaders vooral verantwoordelijk zijn voor extern kapitaal, worden betwijfeld gezien dat alleen het effect van de culturele oriëntatie van de vader op de schoolloopbaan significant is. De culturele oriëntatie heeft zowel betrekking op activiteiten die zich binnen het gezin afspelen (zoals het praten over politiek) als externe activiteiten (samen naar musea gaan).

Als nu de dominantie-assumptie van Glebbeek niet opgaat, dus de invloed van vader en moeder in het gezin op de schoolloopbaan verschilt niet qua zwaarte van elkaar, dan moeten de resultaten voor vaders en moeders gelijk geïnterpreteerd worden, en wel als volgt: kijkend naar Tabel 5 zien we dat de significante interactie-effecten van de aanwezigheid en opleiding van vaders negatief zijn. Daarover verschillen we ook niet van mening met Glebbeek, alleen over de zwaarte die aan dit resultaat moet worden gegeven. Zoals hierboven weergegeven, zegt hij op grond van de veronderstelde 'gezinsinterne' dominantie van moeders dat dit resultaat van de vaders niet overschat moet worden, want vaders zijn wat betreft de gezinsinterne invloed minder relevant. Wij onderschrijven deze assumptie niet en stellen dat vaders even relevant zijn als moeders wat betreft hun invloed op de schoolloopbaan van hun kinderen.

Als we vervolgens naar de moeders kijken, is er slechts één significant effect waar te nemen, namelijk de positieve interactie tussen de combinatie 'midden' opgeleid en een sterke mate van aanwezigheid (cluster 3) in vergelijking met de combinatie laag opgeleid en weinig aanwezig (cluster 1). Deze relatie komt inderdaad overeen met de hypothese. Echter, waar op grond van de hypothese positieve effecten verwacht zouden mogen worden - namelijk bij de combinaties tussen aanwezigheid en een hoog opleidingsniveau - worden geen significante effecten aangetroffen. De gevonden resultaten mogen in het licht van de theorie of andere onderzoeksresultaten een zekere relevantie toegedicht worden, empirisch berusten zij eerder op toeval dan dat zij een 'consistent patroon' laten zien. Oftewel, de gevonden (significante en de niet significante effecten) resultaten geven volgens ons een onvoldoende onderbouwing van de hypothese. Dit gecombineerd met de resultaten van de vader, die in onze ogen even zwaar wegen, doet ons concluderen dat de hypothese niet opgaat voor deze steekproef.

Een aanvullend argument dat er voor pleit de waarde van het gevonden interactie-effect bij moeders niet te overschatten en het niet als bewijs te zien voor de hypothese, is het ontbreken van enige significante effecten als het gaat om de ondersteuningsvariabelen (zie Tabel 4). Het gevonden signifiante interactie-effect betreft alleen de fysieke aanwezigheid. (Tabel 4 laat ook negatieve effecten op de schoolloopbaan zien wat betreft het hoofdeffect van de aanwezigheid van vaders en moeders).

Een verdere theoretische vraag die uit deze resultaten naar voren komt, heeft betrekking

op de psychologische betekenis van de ouders voor de schoolloopbaan van hun kind. In deze steekproef heeft de fysieke aanwezigheid en de psychologische ondersteuning niet het verwachte effect. Daarvoor zijn verschillende verklaringen te geven, zoals ook geformuleerd in het artikel. Veronderstellend dat de ouders een belangrijk effect hebben op de schoolloopbaan van hun kinderen, zou nagedacht moeten worden over een alternatieve operationalisering van de door Coleman in zijn sociaal kapitaaltheorie gegeven variabelen, aanwezigheid en ondersteuning. Wellicht dat alternatieven in de vorm van opvoedingsstijlen (Steinberg, 1989)** adequater zijn om de relatie tussen ouders en kinderen en het effect daarvan op de schoolloopbaan vast te stellen.

* Graaf, N.D. de & Nieuwebeerta, P. (1995). Sociale klasse, mobiliteit, trouwpatronen en politieke voorkeur. In J. Dronkers & W.C. Ultee (Eds.). *Verschuivende ongelijkheid in Nederland: sociale gelaagdheid en mobiliteit*. Assen: Van Gorcum.

** Steinberg, L., Elmen, J.D. & Mounts, N.S. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Development*, 60, 1424-1436.

NOOT VAN DE REDACTIE

In het laatste nummer van TOR, 22-4, is een deel van de notitie van Oosterbeek en Webbink wege gevallen. Daarom wordt deze nogmaals afgedrukt samen met de reactie van Bosker.

IS VERKLEINING VAN KLASSEN EEN GOEDE INVESTERING?

Hessel Oosterbeek & Dinand Webbink

In oktober 1996 bood de voorzitter van de Commissie Kwalitatieve Aspecten van Groepsgrootte in het Basisonderwijs (Van Eijndhoven) het eindrapport 'Klassenverkleining' aan, aan staatssecretaris Netelenbos.¹ 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 lag de zaak 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.²

De unanimitéit 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 overtui-

gend 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 leerprestatie 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 rekenvaardigheid, taalvaardigheid, welbevinden van leerlingen, en het functioneren van leerkrachten (werkbeleving en attitude) anderzijds. Wie de hoofdttekst 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 Twee- als leerlingen in groepen van 25-29, 30-34, 35-39 leerlingen scoren slechter dan leerlingen in de 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 leerlingen 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 analyse 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.

In tabel 2.2 van het rapport worden voor 27 landen de gemiddelde leerlingsscores voor begrijpend lezen van 9-jarigen gegeven. Nederland neemt in deze tabel een 21ste positie in: vlak achter Hongarije en Slovenië en vlak voor Cyprus en Portugal. Finland, de Verenigde Staten en Zweden voeren deze ranglijst aan.

In de conclusie van het betreffende hoofdstuk verbindt de commissie de resultaten uit deze twee tabellen op de volgende manier: 'Vast staat dat de Nederlandse rijksoverheid financieel verhoudingsgewijs matig in het basisonderwijs investeert, ook al is deze sector dankzij gericht beleid bij bezuinigingen ontzien. Dit leidt in internationaal opzicht tot een vrij ongunstige leerling/leerkracht-ratio en gemiddeld grote klassen. Internationale peilingen laten bovendien zien dat de prestaties van Nederlandse kinderen op cruciale basisvaardigheden nog wel wat te wensen laten. De Commissie concludeert derhalve dat de gemiddelde leerling/leerkracht-ratio in Nederland te hoog is' (p.37). Dit is gebaseerd op de positie van een enkele waarneming (Nederland) in de twee tabellen. Alle landen uit de eerste tabel komen echter ook in de tweede tabel voor, en het is verleidelijk om na te gaan of het gesuggereerde verband ook voor de 14 andere landen opgaat. Daartoe hebben we een regressievergelijking geschat met de gemiddelde leerlingsscore begrijpend lezen als afhankelijke variabele en klassengrootte en uitgaven per leerling als verklarende variabelen. De uitkomsten staan in tabel 1.

Anders dan de suggestie die gewekt wordt door de plaats van Nederland in de tabellen, vinden we nu geen negatieve samenhang tussen klassengrootte en prestatie-score. Zelfs is het zo dat bij gegeven uitgaven per leerling, leerlingen in landen waarin de gemiddelde klassengrootte groot is, beter scoren dan leerlingen in landen waarin de gemiddelde klassengrootte klein is. Wel komt uit de tabel de significante samenhang tussen uitgaven per leerling en prestaties tot uitdrukking: uitgaven per leerling die 35% boven het OECD gemiddelde liggen (zoals in de Verenigde Staten het geval is) gaan gepaard met een gemiddelde leerlingsscore op begrijpend lezen, die een standaardafwijking boven het gemiddelde ligt. Overigens willen we geens-

zins suggereren dat op basis van een analyse van geaggregeerde gegevens van vijftien zeer verschillende landen veel zinvol kan worden beweerd over het verband tussen klassengrootte, uitgaven en leesprestaties. Maar dat is wel precies wat de commissie doet door een verband te suggereren op grond van de positie van Nederland in beide tabellen. De resultaten in tabel 1 laten zien dat de conclusies anders zijn wanneer je alle informatie en niet die van slechts één land gebruikt. Verder is van belang om te vermelden dat, na het weglaten van het uitga-venniveau in de regressievergelijking, het effect van klassengrootte op leesprestatie niet sig-nificant van nul verschilt.

Ook ten aanzien van dit bewijsstuk is de conclusie duidelijk: vergelijking van internatio-nale gegevens over klassengrootte en leerlingsscores wijzen niet op een negatieve samenhang tussen beide grootheden. Eerder lijken er aanwijzingen voor een positieve samenhang.

Tabel 1: Verklaring van gemiddelde leerlingsscore begrijpend lezen; gewone kleinste kwadraten

variabele	coëfficiënt	t-waarde
constante	399,28	7,92**
klassengrootte	2,80	1,83*
uitgaven	0,67	2,66**
adj. R square	0,268	

Toelichting: Het gemiddelde van de afhankelijke variabele is 517, met standaardafwijking 24,3. Gemid-delde klassengrootte loopt van 16 in Italië tot 31 in Ierland. De uitgaven per leerling zijn genormeerd op 100 per leerling in alle OECD landen te zamen. ** duidt op significantie op 5%-niveau; * duidt op signi-ficantie op 10%-niveau.

Resumé

We hebben nu de vier stukken van bewijs die de commissie aandraagt voor een negatieve samenhang tussen leerlingprestaties en klassengrootte nader bekeken. De conclusie is onver-mijdelijk: er is geen bewijs. De commissie heeft naar onze mening veel te eenzijdig gegevens verzameld en geïnterpreteerd. Wat in het betoog te pas komt wordt breed uitgemeten; wat er niet past wordt op z'n best in een bijlage weggestopt. Een lid van de commissie heeft de aan-gedragen evidentie wel vergeleken met de stukjes van een puzzel; de gevonden stukjes lijken bij elkaar te passen⁹. Om in die analogie te blijven, volgens ons is de commissie bij het ver-zamelen van puzzelstukjes uitgegaan van een voorbeeldplaatje van een kleine klas. Stukjes die niet bij dat voorbeeldplaatje passen zijn niet gebruikt. Het vraagstuk van de klassengrootte is echter een puzzel zonder voorbeeldplaatje.

Maar al is er geen bewijs dat klassenverkleining gewenst is, dat laat onverlet dat er wel dege-lijk problemen in het basisonderwijs zijn, die om aanpak vragen. We zijn het met de com-missie eens dat verdere professionalisering van leerkrachten betreffende moderne didactiek voor jonge kinderen (p. 9) een goede maatregel is. Maar we verschillen met de commissie van mening dat klassenverkleining daar een noodzakelijke voorwaarde voor is. Veel meer zien wij in verlenging en verzwarening van de opleiding tot onderwijzer. Nu duurt zo'n opleiding 4 jaar en is er betrekkelijk weinig tijd ingeruimd voor stages en het leren gebruiken van moder-ne media. Weliswaar constateert de commissie dat er bij de opleidingen voor onderwijzers problemen zijn bij de specialisatie voor de onderbouw (p.50), maar voor de oplossing daar-van is klaarblijkelijk geen extra geld nodig. Verder zou er naar onze mening veel meer aan-nascholing en opfrissing van leerkrachten gedaan kunnen worden. Sporadisch volgen onder-wijzers die al jaren in het vak zitten een cursus over nieuwe ontwikkelingen. In veel scholen staan inmiddels computers, maar het ontbreekt vaak aan docenten met genoeg deskundigheid

om de mogelijkheden die dit medium biedt ten volle te benutten. Uit Amerikaans onderzoek komt naar voren dat onderwijs met behulp van computers veel kosten-effectiever is dan klassenverkleining.¹⁰ Dit onderzoek is inmiddels wat gedateerd, hetgeen naar alle waarschijnlijkheid de resultaten alleen maar versterkt omdat de kosten van computergebruik zijn gedaald.

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. Teveel mensen met gevestigde reputaties hebben hun naam reeds aan dit idee verbonden. Maar als er dan toch fors geïnvesteerd gaat worden, dan hopen we dat Nederlandse onderwijsbeleidsmakers iets hebben geleerd van de evaluatie van het Onderwijsvoor-rangsbeleid. Met de invoering van het OVB is veel geld uitgetrokken voor de evaluatie van dat beleid. Die evaluatie is echter niet goed opgezet. Het elementaire idee dat het voor evaluatie van beleid nodig is een controle-groep te creëren, is over het hoofd gezien of bewust genegeerd. Bij gevolg kan bij voorbeeld niet worden uitgemaakt of allochtone leerlingen in de loop van de jaren beter zijn gaan presteren als gevolg van het OVB-beleid, of als gevolg van het feit dat hun ouders gemiddeld langer in Nederland verblijven. We pleiten er daarom voor dat als er geld wordt besteed aan verkleining van klassen, dat daarbij voor een werkelijk experimentele opzet wordt gekozen. Een opzet waarbij de maximale klassengrootte per regio verschilt, waarbij het lot bepaald heeft welke regio welke maximale klassengrootte heeft. Ook pleiten wij ervoor om niet alleen te kijken naar de onderwijsprestaties van leerlingen maar om ook de kosten gedetailleerd in kaart te brengen. Pas dan kan vastgesteld worden of de middelen goed zijn besteed. Als het geld dan over de balk is gesmeten, kan in elk geval wetenschappelijk worden onderbouwd dat dat is gebeurd.

(Hessel Oosterbeek is universitair docent bij de vakgroep algemene economie van de Universiteit van Amsterdam. Dinand Webbink werkte ten tijde van het schrijven van dit artikel als senior-onderzoeker bij de Stichting voor Economisch Onderzoek (SEO). Inmiddels is hij werkzaam bij de Onderwijsraad.)

1. 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.
2. Inmiddels is vanaf augustus 1997 jaarlijks 270 miljoen gulden extra beschikbaar voor klassenverkleining en kwaliteitsverbetering in de onderbouw.
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 derde graads 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 Onderwijsvoor-rangsbeleid.
7. Sociaal Cultureel Planbureau, Sociaal en Cultureel Rapport 1994, Rijswijk, 1994, blz. 310.
8. Voor leerlingen in jaar groep 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

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 stil staan. 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 ter 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 hoofdtekst 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. 71-76). 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 mee gewogen 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 doorslaggevend 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.

Overigens zijn de scholen in het STAR-experiment inderdaad niet aselekt toegewezen aan de experimentele en controle conditie, 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 voor 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 so wie so 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 dichtten. 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.

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 gebruik gemaakt 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 2^{de} of 3^{de} graads polynomen gemodelleerd 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 ze 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.
- Oosterbeek, H. & Webbink, D. (1997). Is verkleining van klassen een goede investering? *Tijdschrift van Onderwijsresearch*, p. 71-77
- 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).

Boekbesprekingen

J.J.C.M. Visser

Lezen op AVI-niveau. De schaling, normering en validering van de nieuwe AVI-toetskaarten
Dissertatie KUN, 's-Hertogenbosch: KPC, 1997, ISBN 90-6755-118-X

Lezen op AVI-niveau is de verantwoording van een serie door het Katholiek Pedagogisch Centrum uitgegeven AVI-toetskaarten voor het onderwijs in technisch lezen op de basisschool. Het doel van het instrument is het afstemmen van het niveau van de leesteknik van leerlingen op de moeilijkheid van de leesstof. Op de kaarten staan teksten die zo snel en nauwkeurig mogelijk door leerlingen gelezen moeten worden. Elke kaart staat voor een zogenaamd AVI-niveau. Als een leerling bijvoorbeeld de kaart met AVI-niveau 5 voldoende snel en nauwkeurig kan lezen, en de kaart met AVI-niveau 6 nog niet, dan kan deze leerling, zo is de claim van de ontwikkelaars, alle teksten met AVI-niveau 5 lezen. De door de leesstof veronderstelde leesteknik wordt met Leesindex A (Brouwer, 1963), een leesbaarheidsformule, bepaald en uitgedrukt in AVI-niveaus.

Visser doet in haar dissertatie verslag van de constructie van de toetskaarten en van het calibratie-, normerings- en valideringsonderzoek. In het *calibratie-onderzoek* gaat Visser na of de leestijden op de in moeilijkheid verschillende kaarten op één schaal kunnen worden uitgedrukt. Het *normeringsonderzoek* levert referentiegegevens op voor de relatieve of normgerichte interpretatie van de in AVI-niveaus uitgedrukte snelheid en nauwkeurigheid. De relatieve niveau-indicaties lopen van 'goed tot zeer goed' (leerlingen in het hoogste kwartiel van de verdeling van de leesteknik in een gegeven leerjaar) tot 'zwak tot matig' (laagste kwartiel). In het kader van het normeringsonderzoek bepaalt Visser ook de criteria volgens welke leerlingen in één van de AVI-niveaus worden ingedeeld. In het *valideringsonderzoek* gaat zij na of het AVI-instrumentarium inderdaad technisch lezen meet. Maar ze beschouwt ook andere vraagstukken die direct te maken hebben met het doel van het AVI-instrumentarium en met het begrip 'leesteknik'. De belangrijkste zijn de bepaling van de grenswaarden en de leesbaarheidsformule. Bij de bepaling van de grenswaarden gaat het om de vraag wanneer de snelheid en de nauwkeurigheid van de leesteknische verwerking van een toetskaart als voldoende kunnen worden aangemerkt. Leesindex A speelt in de AVI-opzet een centrale rol. Deze voorspelt de door teksten vereiste leesteknik op het continuüm waarop de toetskaarten de leesteknik van leerlingen meten, dat wil zeggen op de 'AVI-niveau-schaal'. De formule en de toetskaarten behoren dus hetzelfde begrip 'leesteknik' te meten.

In *hoofdstuk 1* bespreekt Visser eerst de geschiedenis en de opzet van de oude AVI-toets. Verder gaat ze in op de overwegingen die geleid hebben tot de vernieuwing van 'de AVI'. De belangrijkste reden was dat het oude - zo'n 20 jaar in het basisonderwijs gebruikte - KPC-instrument nooit verantwoord was: over de fundering van de normen, de validiteit en betrouwbaarheid waren geen adequate empirische gegevens verzameld of bekend. Vervolgens gaat Visser in op het begrip 'technisch lezen', de verwerving van de leesteknik en het onderwijs in technisch lezen. Ze definieert het begrip 'technisch lezen' als het nauwkeurig, automatisch en met een passende snelheid herkennen van woorden. Tot slot bestudeert ze met het oog op haar validiteitsonderzoek de relatie tussen technisch lezen, de woordenschat en begrijpend lezen.

In *hoofdstuk 2* beschrijft Visser de constructie van de nieuwe AVI-kaarten. Ze maakt daar zoals ze dat zelf noemt een uitstapje naar de Leesindex A van Brouwer, de leesbaarheidsformule waarmee de AVI-constructeurs de door teksten vereiste leesteknik bepalen. De niveaus 5 tot en met 9 zijn geheel op deze formule gebaseerd, de niveaus 1 tot en met 4 worden behalve

door de formule ook door de inhoud van het aanvankelijk leesonderwijs bepaald. Hoewel Leesindex A een formule is voor de begrijpelijkheid van teksten - Visser merkt dat ook uitdrukkelijk op (blz.61) - gebruikt ze deze formule om de door teksten veronderstelde lees-techniek te voorspellen. Wat de constructie van de toetskaarten betreft, meldt ze dat de auteurs en hun screeners de opdracht hadden rekening te houden met het feit dat er behoefte was aan korte afgeronde teksten die aan alle criteria van een bepaald AVI-niveau moesten voldoen (blz 62). Zij besluit hoofdstuk 2 met een overzicht van onderzoeksvragen.

In *hoofdstuk 3* komen de eerste 3 onderzoeksvragen aan de orde. Voor de calibratie van de 18 toetskaarten (2 series van 9: 1A t/m 9A en 1B t/m 9B) heeft Visser gebruik gemaakt van de aanpassing van het 'Raschmodel voor leestijden' door Jansen (1997). Uit de resultaten blijkt dat de moeilijkheid van de AVI-toetskaarten zoals bepaald met Leesindex A niet spoort met de in het calibratie-onderzoek op basis van de leestijden bepaalde moeilijkheid (*onderzoeksvraag 1*). Dit negatieve resultaat, van wat kan worden beschouwd als een poging tot controle-validatie, accepteert Visser uit praktische overwegingen. Nieuwe kaarten maken is op praktische gronden niet mogelijk en een niveau-indeling overeenkomstig de resultaten van het calibratie-onderzoek kan evenmin, omdat de leesstof met behulp van Leesindex A naar AVI-niveaus is ingedeeld.

In hoofdstuk 3 komt vervolgens het bepalen van de indelingscriteria voor de AVI-niveaus aan de orde. Visser betreft daarbij behalve de leestijd ook het aantal leesfouten. De vraag waar het om draait is: bij welk aantal fouten en bij welke leestijd hoort welk AVI-niveau? (blz 76). Bij de vaststelling van de grenswaarden is rekening gehouden met de discrepantie tussen de voorspelde moeilijkheid (via Leesindex A) en de empirische moeilijkheid (uitkomsten calibratie-onderzoek).

Een volgend thema in hoofdstuk 3 is de betrouwbaarheid van de toetskaarten (*onderzoeksvraag 2*). Het uiteindelijke AVI-niveau, bepaald met de A-versie, kwam in ruim 56% van de gevallen overeen met het niveau zoals bepaald met de B-versie. Visser noemt dit percentage teleurstellend. Maar ook de consistentie van de uitkomsten tussen opeenvolgende kaarten laat te wensen over, aldus Visser (blz. 100).

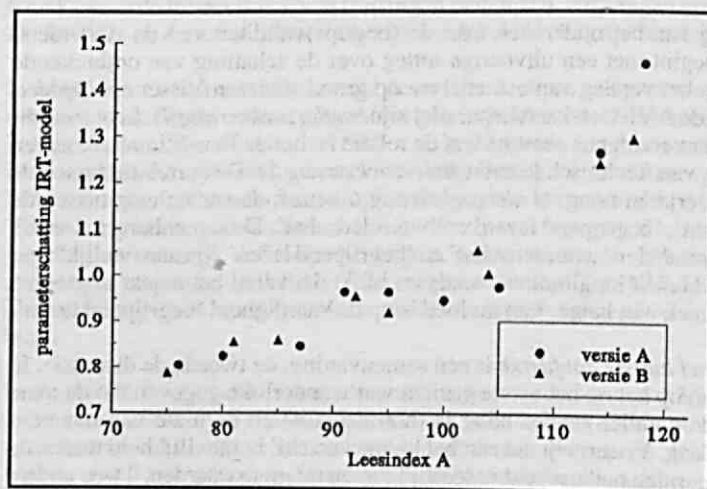
Tot slot van het hoofdstuk presenteert Visser de normeringstabel, de tabel ten behoeve van de relatieve interpretatie van de bepaalde AVI-niveaus. Uit de bestudering van de verdeling van de leerlingen in de groepen 3 tot en met 6 over de AVI-niveaus (*onderzoeksvraag 3*) blijkt dat volgens verwachting de leestechniek in de loop der leerjaren toeneemt, in het begin sterker dan in de latere leerjaren.

Hoofdstuk 4 doet verslag van het onderzoek naar de (begrips)validiteit van de AVI-toetskaarten. De rapportage begint met een uitvoerige uitleg over de schatting van ontbrekende gegevens in de dataset en het verslag van een analyse op grond waarvan Visser concludeert dat de A- en B-versie van de AVI-toetskaarten parallel zijn (*onderzoeksvraag 4*). In het validiteitsonderzoek gaat Visser vervolgens eerst na wat de relatie is met de Een-Minuu-Test, een andere operationalisering van 'technisch lezen' (*onderzoeksvraag 5*). De correlatie tussen de EMT en AVI is in alle leerjaren hoog. *Onderzoeksvraag 6* betreft de samenhang tussen de variabelen 'technisch lezen', 'begrijpend lezen' en 'woordenschat'. De samenhang tussen de AVI-toetskaarten en de variabelen 'woordenschat' en 'begrijpend lezen' zijn aanzienlijk lager dan tussen AVI en de EMT. Uit longitudinale analyses blijkt dat vanaf het najaar in groep 5 een *toenemende* leestechniek niet langer van invloed is op de vaardigheid 'begrijpend lezen'.

De eerste paragraaf van *het laatste hoofdstuk* is een samenvatting, de tweede de discussie. In de discussie gaat Visser in op het op het eerste gezicht wat wonderlijke gegeven dat de mate van samenhang tussen de parallel-versies hoog is (hoofdstuk 4) en de mate van overeenstemming (hoofdstuk 3) laag. Visser wijt dat aan het kleine verschil in moeilijkheid tussen de AVI-toetskaarten en het geringe onderscheid in leestijd tussen de grenswaarden. Twee andere

discussiepunten betreffen de validiteit van de AVI-toetskaarten. De punten zijn (1) de discrepantie tussen de moeilijkheid uitgedrukt in Leesindex A en de moeilijkheid op basis van leestijden en (2) de validiteit van de AVI-toetskaarten. Wat het laatste punt betreft concludeert Visser dat de sterkte van de samenhang met soortgenoten evenals de relatie met de variabelen 'woordenschat' en 'begrijpend lezen' overeenkwamen met de voorspellingen op basis van ander onderzoek en op grond van theoretische overwegingen. Wat de discrepantie betreft tussen de twee 'soorten' moeilijkheid zegt Visser: 'Deze resultaten ten aanzien van de schaling van de toetskaarten roepen vragen op over de koppeling tussen (de moeilijkheidsgraad van de) leesstof en de AVI-niveaus'. (blz. 171).

Ik wil mijn *commentaar* met het resultaat van de schaling beginnen. Bij Visser klinkt enige teleurstelling door over de discrepantie tussen de twee soorten moeilijkheid van haar toetskaarten. Het zou echter eerder verbazingwekkend geweest zijn als er geen verschil zou hebben bestaan. Leesindex A kan beschouwd worden als een aanpassing van de 'Reading Ease'-formule van Flesch (1948). Het criteriummateriaal van Flesch bestaat uit de zogenaamde 'Standard Test Lessons in Reading', een serie leestaken uit 1925, bestemd voor het Amerikaanse basis- en voortgezet onderwijs in die tijd. Het criterium voor 'leesbaarheid' dat Flesch hanteert, is een zogenaamde 'grade equivalent score', in dit geval het leerjaar waarin de leerlingen gemiddeld 75% (opgevat als een 'beheersings'-percentage) van de begripsvragen bij een 'Test Lesson' juist beantwoorden. Maar ook als er niets op het criterium aan te merken zou zijn geweest dan zou een volledige overeenkomst tussen de twee soorten moeilijkheid mij hebben verbaasd. Ook op adequaat leesbaarheidsonderzoek gebaseerde formules zijn bruikbare, maar zeker geen perfecte voorspellers. En dat de AVI-toetskaarten wat dat betreft niet onder doen voor Leesindex A, heeft Visser in haar dissertatie duidelijk laten zien. Natuurlijk, perfecte metingen zijn ons aller ideaal, maar de praktijk wijst dikwijls uit dat we het met iets minder moeten doen. Visser had volledig kunnen profiteren van de eigenschappen van het model van Jansen (1997) als ze de discrepantie tussen de twee soorten moeilijkheid als een 'normaal' verschijnsel zou hebben geaccepteerd en gebruik had gemaakt van het verband dat er tussen de volgens het latente trek model geschatte moeilijkheid van de 18 toetskaarten en Leesindex A bestaat. Tussen haakjes: ik ga er hierbij gemakshalve van uit dat de assumpties onder het model inderdaad juist zijn. Ter illustratie van de (sterke) samenhang tussen de twee 'soorten' leesmoeilijkheid geef ik Figuur 1. Het illustreert ook de paralleliteit van de A- en de B-versie. Het figuur is gebaseerd op de in het proefschrift in Tabel 3.2 (blz. 80) gepresenteerde leesindexen en IRT-parameters.



Leesindex A is, zegt Visser, een formule die de begrijpelijkheid van teksten voorspelt. Ze gebruikt de formule omdat zij in de oude AVI al gebruikt werd en omdat er geen formule voor de 'technische' moeilijkheid van teksten bestond. Dat klopt: de Cito Index voor de Leestechiek (CILT; Staphorsius en Verhelst, 1997) is gepubliceerd toen Visser haar onderzoek al had afgerond. Wat echter wel eens niet zou kunnen kloppen, is haar vermoeden dat formules voor de leestechiek hetzelfde voorspellen als formules voor de begrijpelijkheid van teksten. Die veronderstelling is ook niet in overeenstemming met de resultaten van haar validiteitsonderzoek. Dat wijst net als veel ander onderzoek op het gebied van het lezen uit: technisch lezen is een andere vaardigheid dan begrijpend lezen. De logische consequentie daarvan is dat een criterium voor de leestechiek iets anders moet zijn dan een criterium voor leesbegrip. Visser merkt op dat de Leesindex A, net als vele andere leesbaarheidsformules juist de technische aspecten van een tekst gebruikt om de begrijpelijkheid te bepalen (blz. 61) en ze doelt dan op kenmerken als woord- en zinslengte. Maar de gewichten van deze predictoren kunnen zich in een multipele regressie analyse waarin een criterium voor leestechiek gehanteerd wordt, onderling weleens heel anders verhouden dan in een analyse met leesbegrip als criteriumvariabele. Het is bovendien niet uitgesloten dat in de ene analyse blijkt dat er tussen deze kenmerken een veel grotere overlap bestaat dan in de andere analyse.

Een ander punt dat direct met de validiteit van de AVI-toetskaarten te maken heeft, is de bepaling van de grens tussen 'voldoende' en 'onvoldoende'. Waarom is (bijvoorbeeld) op kaart 7A een leessnelheid van maximaal 2 minuten en een maximaal aantal fouten van 6 voldoende? Het antwoord van Visser luidt dan: omdat die grenswaarde optimaal bijdraagt aan de consistentie van de uitkomsten op de verschillende toetskaarten. Maar voor de beoogde toepassing is een psychologisch relevant criterium noodzakelijk. Het mooiste zou het zijn als dat werd gespecificeerd door een theorie over 'technisch lezen', maar zo'n theorie is niet voorhanden. Bij gebrek aan beter kan er gebruik worden gemaakt van het oordeel van deskundigen. Aan leerkrachten kan bijvoorbeeld gevraagd worden voor welke populatie leerlingen ze gegeven teksten 'technisch' leesbaar achten (Zie Staphorsius en Verhelst, 1997). Dat levert geen subjectief criterium (*Lezen op AVI-niveau*, blz. 180) op, maar - bij een adequate onderzoeksopzet - een intersubjectief bepaald criterium.

Deze kanttekeningen bij het constructie- en onderzoekswerk van Visser nemen niet weg dat het haar niet geringe verdienste is dat er nu 20 jaar na de eerste uitgave een gedegen en goed leesbare verantwoording beschikbaar is voor een in het basisonderwijs veel gebruikt meetinstrument. Het nieuwe instrument mag dan op een aantal punten nog tekortkomingen hebben, het is een goede zaak - en een grote vooruitgang in vergelijking met het oude AVI-instrument - dat die nu bekend zijn.

Gerrit Staphorsius
Cito,
Postbus 1034
6801 MG Arnhem

Literatuur

- Brouwer, R.H.M. (1963). Onderzoek naar de leesmoeilijkheid van Nederlands proza. *Pedagogische Studiën*, 40, 454 - 464.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221 - 233.
- Jansen, M.G.H. (1997). The Rasch Model for speed tests and some extensions with applications to incomplete designs. *Journal of Educational and Behavioral Statistics*, 22, summer issue.
- Staphorsius, G. & Verhelst, N. (1997). Indexering van de leestechiek. *Pedagogische Studiën*, 74, 154 - 164.

Anne Bert Dijkstra, Jaap Dronkers, Roelande Hofman (red.)
Verzuiling in het onderwijs. Actuele verklaringen en analyse
Groningen: Wolters-Noordhoff 1997, ISBN 90 01 25840 9, prijs: f 59,-

Misschien wel op de valreep, nu de onderwijs-verzuiling op voorstel van de Onderwijsraad zelf van kleurenpalet lijkt te gaan verschieten (scholen hoeven zich niet langer te identificeren in termen van de bestaande denominaties, de identiteit wordt dus 'vrij'), schetst een aantal onderwijssociologen in een omvangrijke bundel hun beeld van het hoe en waarom: hoe kan toch de verzuiling voortbestaan in een zo grootschalig ontkerkelijkt context? Het antwoord op die vraag vormt het middenstuk van het boek, voorafgegaan door een aanzienlijk omvangrijker deel dat gericht is op de praktijk en achtergronden en gevolgd door een wel heel kort deel drie: 'blijft de onderwijsverzuiling bestaan?'

Geen van de auteurs is - zo zegt de inleiding - een geharnast voor- of tegenstander van de verzuiling. Dat lijkt dan ook maar beter, want belangrijker dan de vraag voor of tegen, is de vraag, hoe de verzuiling in de toekomst haar weg zal gaan in een nader te beschrijven krachtenveld. Misschien dat na zo'n analyse de vraag naar voor of tegen er ook heel anders uitziet.

In deel 1 worden behandeld:

- het verschijnsel verzuiling in sociaal en politiek perspectief
- de juridische vormgeving
- de relatie tot schoolkeuze
- een economische analyse
- de relatie tot onderwijskansen en
- het verband met religieuze socialisatie.

In aanzet formuleert Karsten, die zich overwegend tot een beschrijving beperkt, de vraag, of de verzuiling het produkt is van een niet-homogene natie-staat. De vraag die hij daarmee eigenlijk opwerpt is die naar de verhouding tussen de grondwettelijke garantie voor degelijk en in kwaliteit gelijk onderwijs voor allen en de, ook zonder migranten, multiculturele randcondities die onze overheid maar beperkt greep op het onderwijs gunnen. Zo'n vraag had gepast in actuele theorieën over de veranderde rol van de natie-staat en over de met schoolautonomie en deregulering dreigende tweedeling. En ze had ruimte geboden voor een lokale verbijzondering van het verzuilingsverhaal (Vgl. het schitterende proefschrift van Dolly Verhoeven (1994) *Ter vorming van verstand en hart. Lager onderwijs in oostelijk Noord-Brabant ca. 1770 - 1920*. Hilversum: Verloren).

Interessant is ook de kennelijke rekbaarheid van de wetgeving, gedemonstreerd door Akkermans in de actuele context waarin de 'moderniserende' overheid meent, dat ze de financiële middelen niet kan vergaren voor een ongewijzigde onderwijsbekostiging. Een om marxistische verklaringen smekend fenomeen dat zich niet beperkt tot de onderwijsbegroting, en dat daarom breder moet worden geanalyseerd, als onderdeel van de economische en ideologische crisis van de welvaarts- en welzijnsstaat. In de bijdrage van Blank en Koelman over economische aspecten van de verzuiling komt zo'n macro-analyse niet voor; wel maken ze duidelijk dat er niet per se meerkosten aan verzuiling verbonden zijn.

Tetwat gedwongen-rechtlijnig doen analyses aan, waarin de onevenredigheid tussen de verzuilde voorzieningen-structuur aan de ene kant en de feitelijke kerkelijkheid aan de andere kant wordt beklemtoond. Wat Van Wieringen en Van Kessel in dit verband presenteren, is alleszins het bestuderen waard, maar het analysekader moet nodig verder worden aangescherpt: waarom kiezen katholieke ouders bij mij om de hoek in meerderheid voor een officieel openbare school (honderd meter verder ligt een katholieke), ook al worden ze er (ze zijn ook op die openbare school, die door een katholiek college wordt bestuurd, in de meerderheid) door de lokale bisschop gedwongen om de voorbereiding op de eerste communie door een leek en niet door de lokale pastor te laten doen? Zoiets vraagt om meer analyse-factoren

dan alleen de denominatie van de school en de kerkelijkheid van de ouders. Je zou ook hier wat meer aandacht willen zien voor het multi-level-probleem dat de nationale wetgeving lokaal veelal volstrekt onbedoeld uitpakt, doordat de godsdienstige segregatie die er landelijk is, zich lokaal vrijwel nergens zo herhaalt.

Boeiend, ook al levert ze daarmee het antwoord niet op bovenstaande vraag, is de door Hofman beschreven verhouding tussen overheidsbesturing en private besturing. Een onderdeel van de algemenere vraag naar een representatieve dan wel corporatistische bestuursvorm: in welke verhouding staat de kwaliteit, de maatschappelijke positie en de feitelijke daadkracht van de mensen die via gemeenteraden bestuurlijk scholen moeten 'regeren' tot degenen die dat doen als belangenbehartiger voor een heel bepaalde maatschappelijke groep en met alleen dat ene belang voor ogen en niet tegelijk dat van het lokale zwembad, de straatverlichting en de schouwburg? Hofman erkent nadrukkelijk de functie van het maatschappelijk middenkader en wijst op de problematische gevolgen voor hun rol, van de door de overheid ingezette schaalvergroting.

In dat verband stellen Van Wieringen en Van Kessel vast, dat schoolmanagers toenemend bestuurlijke verantwoordelijkheid nemen. In de context van schoolautonomie en budget-rationaliteit had die constatering tot wat sceptische predicties mogen leiden, eventueel via een uitstapje naar de gezondheidszorg, waar opeenvolgende management-moderniseringën beslist niet resulteren in de verhoopte kostenreductie en rationalisering, en waar (als dat geen voorbode is van hoe het de verzuiling zal vergaan?) pijnlijk duidelijk wordt, dat de som van vele private belangen niets garandeert voor het gezamenlijke belang van allen. Wat betekent zo'n constellatie van bestuurlijke 'moderniseringën' voor de grondwettelijke garanties voor gelijke zorg voor allen en wat betekent ze voor de mate waarin het parlement straks überhaupt nog weet wat er in het land gebeurt? De vraag naar de sociologische determinanten van zo'n groot-schalige omslag in de overheidsideologie (van super-regulerende betweter naar alle verantwoordelijkheid en budgetbeheer afschuivende deregulator, Vgl. de ontmaskerende analyses van de oude en wijze voormalige Hessische onderwijsminister en -socioloog Von Friedeburg: *Schulentwicklung zur Ungleichheit*. In: *Mitteilungen des Instituts für Sozialforschung* (1995, 6, 53-68) mag eigenlijk niet terzijde blijven in een analyse die ook de toekomst van de verzuiling probeert te tackelen.

Uitertaard ontbreekt in dit boek de vraag niet (Vreeburg), hoe het staat met de religieuze socialisatie in het licht van de verzuiling; ze mag gesteld worden, hoewel ze wat oubollig aandoet; ze mag onderhand wel worden aangevuld met de vraag, of er door de zuilen verschildend wordt omgegaan met prestatiedruk, met disciplineren en met talent-ontplooiing of met conflict-regulering en aangeboden levensfilosofiën. Dat het voor zo'n analyse aan bruikbare data ontbreekt is evident, maar blijf de vraag er naar alsjeblijft benadrukken, zeker in een boek dat menig student zal doorwerken!

Voor een helder overzicht van de relatie tussen zuil en ongelijke onderwijskansen is deze bundel eigenlijk te vroeg verschenen. Dijkstra beschikte nog niet over bevindingen uit het PRIMA-cohort, die de geringere effectiviteit van het openbaar onderwijs grotendeels herleidt tot zijn oververtegenwoordiging in het ineffectievere deel van het land, het Noorden.

In het tweede deel van de bundel dragen Hofman, Roeleveld en Dijkstra verklaringen aan voor het voortbestaan van de onderwijsverzuiling. Roeleveld beperkt zich tot een heldere analyse, waarin lokaal wisselende zuil-verhoudingen verband blijken te houden met wisselende schooleffectiviteit. Het is de vraag of de concurrentieslag tussen scholen zich ooit om werkelijke effectiviteit zal bekommeren; scholen zijn in staat om zich - terecht of niet - te profileren via het type publiek waarop ze zich beroemen en via de kwaliteit die ze suggereren. En diegenen die hun loopbaankansen minder aan thuis dan aan school moeten ontlenen, zullen weinig schoolkeus hebben.

Hofman toont zich een perfecte specialiste bij de beschrijving van de institutionalisering van het onderwijsmanagement: een veelvormig gebeuren met een eigen logica, waarzonder actuele onderwijsontwikkeling niet te begrijpen valt. Met wat meer durf en cynisme, een must

voor een socioloog, had ze nare prognoses kunnen formuleren over de bestuurlijke ontwikkeling naar lokaal onderwijsbeleid.

Dijkstra's aandacht voor sociale netwerken rond scholen daagt uit tot reflectie: wat moeten we ons daar straks bij voorstellen, als scholen erin slagen om liefst in verschillende gebouwen voor verschillend publiek een in kwaliteit uiteenlopend aanbod te serveren? Hoe verhoudt dat zich tot de wens van diverse min of meer etnische categorieën om het belang van de eigen ingroup te dienen? Hoe definiëren onderwijsmanagers straks hun markten en sociale netwerken?

Het derde deel is, als gezegd, wel heel kort; het is geschreven door de redacteurs en het schetst een ontwikkeling naar 'een verzuiling die gebaseerd is op private non-profitorganisaties'. Hier wrekt zich nogmaals, dat geen van de auteurs aandacht heeft gehad voor de principiële onverenigbaarheid van de verzuiling met het ideaal van Comenius' eenheidsschool (uitgewerkt door Tews in de Humboldt-Akademie in Berlijn, enkele jaren voor de Nederlandse parlementaire debatten over de verzuiling van het lager onderwijs): een voor allen gelijk en kwalitatief hoogstaand onderwijssysteem, onverenigbaar met de sociale klasse-segregatie die later door de verzuiling aan het oog en de aandacht werd onttrokken. Nederlandse politici waren zich - blijkens de parlementaire discussies van 1919 en 1920 terdege van die interdependentie van sociale-klasse-segregatie en verzuiling bewust (Vgl. de parlementair-politieke beschouwingen van Oud). Achteraf mogen zij gedeeltelijk ongelijk hebben gehad en heeft de godsdienstige verzuiling - via de simpele logica dat je leerlingen maar naar weinig criteria kunt indelen zonder dat je getalsmatig het recht op een eigen school verliest - in de praktijk de sociale segregatie soms zelfs geremd, maar nergens zijn de principiële gevaren van verzuiling voor sociale ongelijkheid zo helder verwoord. Nu de verzuiling van kleurenpalet lijkt te gaan wisselen, had in het slot van deze bundel die link naar OPNIEUW de dreiging van ongelijke voorzieningen voor ongelijk publiek, als toekomstige verzuilingsvariant niet misstaan.

Januari 1998
Paul Jungbluth
Netherlands Institute for Advanced Studies
Meijboomlaan 1
2242 PR Wassenaar, The Netherlands

M.M.H. Frederiks

Beslissen over kwaliteit: verklaring van besluitvorming over het gebruik van onderwijs-visitatierapporten in Nederlandse universiteiten.

(Dissertatie). CSHOB, Universiteit Twente. December 1996.

1. Samenvatting van de inhoud

De auteur heeft met deze dissertatie twee doelen nagestreefd: een wetenschappelijk en een maatschappelijk doel. 'Het wetenschappelijk doel is het vergroten van onze kennis op het gebied van de externe kwaliteitszorg in het hoger onderwijs, in het bijzonder ten aanzien van de verklaring van academische besluitvorming over het gebruik van oordelen van visitatiecommissies over de kwaliteit van het wetenschappelijk onderwijs' (p. 19).

Het maatschappelijk doel bestaat hierin, dat de resultaten van het onderzoek 'bij de universiteiten zouden kunnen leiden tot een kritische reflectie op de besluitvorming over het gebruik van visitatierapporten'. Ook andere instanties zouden deze resultaten kunnen gebruiken bij hun beleidsbepaling ten aanzien van kwaliteitszorg in het hoger onderwijs (VSNU, HBORaad, OCW, Inspectie).

Als centrale vraag formuleert de auteur: "Hoe kan de besluitvorming over het gebruik van onderwijsvisitatierapporten in Nederlandse universiteiten worden verklaard" (p. 20). Deze

centrale vraag wordt uiteengelegd in vier onderzoeksvragen: welke theorie is het meest van toepassing, welke hypothesen kunnen uit deze theorie worden afgeleid, hoe verloopt de besluitvorming over het gebruik van de visitatierapporten en, tenslotte, in welke mate verklaren de hypothesen het feitelijk gebruik.

Het boek volgt deze vier onderzoeksvragen heel systematisch, waarbij de 'zoektocht naar een theorie' (hoofdstuk 2) het omvangrijkste deel is. In dit hoofdstuk presenteert Frederiks een nieuwe cybernetische-incongruentietheorie, waarin het optreden en reduceren van mentale incongruenties wordt beschouwd als een fundamentele eigenschap van sociale, cybernetische systemen. Hij komt tot deze nieuwe theorie na een keuze voor algemene, atomistische gedrags-theorieën (c.q. na verwerping van specifieke en holistische). Vervolgens verwerpt hij de rationale keuze-theorie ten gunste van de theorie der mentale incongruenties (TMI). Kritiek op onderdelen van deze laatste theorie (o.a. begripsprecisie) en aanpassing voor toepassing op sociale systemen brengt Frederiks tot formulering van een nieuwe theorie, die hij dus de cybernetische-incongruentietheorie noemt. Cybernetisch slaat op een open systeembenadering en op het zelfregulerende, responsieve en adaptieve karakter.

In de opvatting van Frederiks bepalen een aantal onafhankelijke variabelen de omvang van de incongruentie. Deze onafhankelijke variabelen zijn: legitimiteit, uniciteit, centraliteit, dependentie, incongruentieduur. Het systeem heeft vijf mogelijke reacties op incongruentie: interne aanpassing, externe aanpassing, realiteitsontkenning, geen aanpassing en dissatisfactie. Het voorkomen van dergelijke reacties is afhankelijk van de combinatie van de onafhankelijke variabelen.

In hoofdstuk 3 wordt deze theorie geconcretiseerd in hypothesen ten aanzien van het gebruik door opleidingen van aanbevelingen van visitatie-commissies. De aanbevelingen zijn voor de opleiding in kwestie op te vatten als incongruenties, die men zal willen reduceren ofwel door actief gebruik (overnemen), passief gebruik (bespreken, maar nog niet invoeren), schijngebruik, niet-gebruik en dissatisfactie. De eerder genoemde onafhankelijke variabelen (mate van legitimiteit, uniciteit, centraliteit, afhankelijkheid, duur) bepalen de wijze van gebruik van de aanbevelingen. Hierover heeft Frederiks hypothesen geformuleerd.

Ter toetsing van deze hypothesen is een survey-onderzoek verricht onder een aantal opleidingen die tussen 1992-1994 zijn gevisiteerd (hoofdstuk 4). De vormen van gebruik worden geoperationaliseerd door aan opleidingen te vragen naar de reactie op een aantal uitgebrachte aanbevelingen. De onafhankelijke variabelen zijn als volgt geoperationaliseerd: wie nam de maatregelen en was die het eens met de aanbevelingen (legitimiteit), het onderwerp van de aanbeveling (uniciteit), de mate van urgentie van de aanbeveling (centraliteit), de ontwikkeling van de studentinstroom (dependentie), het aantal maanden dat is verstreken na het bezoek van de visitatiecommissie (incongruentieduur).

In hoofdstuk 5 en 6 worden de resultaten beschreven. Hoofdstuk 5 doet melding van het gebruik van aanbevelingen van de visitatiecommissies: 53% van de aanbevelingen worden actief gebruikt. Vooral de opleidingscommissies blijken daarin een belangrijke rol te spelen. Het centrale niveau heeft een zeer beperkte rol in het hele proces.

De toetsing van de hypothesen (hoofdstuk 6) krijgt uiteraard de meeste aandacht. Van de vijf hypothesen is er één in zijn geheel gefalsificeerd (betreffende schijngebruik), terwijl de overigen slechts gedeeltelijk worden gesteund door de feiten. Van de dertig deelhypothesen zijn er elf gecorroboereerd. 'Op het eerste gezicht is de empirische ondersteuning dus mager', constateert Frederiks (p. 202).

De legitimiteit blijkt de belangrijkste verklarende variabele, waarbij vrijwel uitsluitend het aantal actoren dat met de aanbevelingen instemt relevant is en niet hun invloed. Dit leidt Frederiks tot de conclusie, dat het collegiale besluitvormingsmodel waarschijnlijk een betere beschrijving geeft van de academische besluitvormingspraktijk dan het rationale, politieke of anarchistische model.

Fredriks zoekt de verklaring voor de tegenvallende resultaten vooral in de (matige) operationalisering van de centrale begrippen, zoals uniciteit, centraliteit, dependentie. Bovendien vermoedt hij, dat zijn hypothese over schijngebruik niet juist is afgeleid uit de theorie.

Kortom, Frederiks houdt vertrouwen in zijn theorie, zij het dat die verder uitgewerkt en verfijnd moet worden. "Hoewel de correlaties tussen de onafhankelijke en afhankelijke variabelen laag zijn, bieden de uit de cybernetische-incongruentietheorie afgeleide hypothesen toch een redelijke verklaring van gebruik en dissatisfactie, met uitzondering van de hypothesen aangaande schijngebruik en centraliteit. Betere verklaringen hebben zich bovendien niet aangediend" (p. 215).

2. Maatschappelijke doelstelling gerealiseerd?

Zoals boven aangegeven, stelde Frederiks zich twee doelen met zijn dissertatie: een wetenschappelijke en een maatschappelijke. Hoewel geen ongebruikelijke taakstelling, is deze lastiger te realiseren naarmate de wetenschappelijke doelstelling fundamenteler van aard is.

De benutting van de aanbevelingen van visitatiecommissies in het hoger onderwijs is een casus of zelfs een bijzaak geworden door het te plaatsen in een debat over een fundamenteel theoretisch model. Wie in deze casus is geïnteresseerd, kan weinig met de belangrijkste conclusie dat een aanbeveling van een visitatiecommissie meer kans van invoering heeft naarmate meer actoren van de opleiding het met die aanbeveling eens zijn. Het is niet alleen geen verrassende conclusie, het geeft bovendien geen enkel houvast voor beleidmakers van de instellingen of overkoepelende organen, noch voor de Inspectie of de overheid.

De invloedrijke rol van de opleidingscommissie is daarentegen wel van belang. Hier speelt echter de recente beleidsontwikkeling een versturende rol. Direct na het verschijnen van het proefschrift is de wet 'Modernisering Universitaire Bestuursorganisatie' ingevoerd, waarmee de interne besluitvormingsprocessen binnen de universiteit als geheel en de afzonderlijke faculteiten aanzienlijk zijn gewijzigd. De opleidingscommissies blijven weliswaar bestaan, maar de context waarbinnen ze zullen functioneren zal de komende jaren aanzienlijk veranderen. Integraal management van de decanen en de komst van opleidingsdirecteuren zal hun rol beïnvloeden. Frederiks refereert op geen enkele plaats naar de invoering van de MUB, die in 1996 al wel (voor)ontwerp van wet was.

Het bovenstaande leidt mij tot de conclusie, dat deze dissertatie zeker geen 'must' is voor beleidmakers. Frederiks trekt zelf ook geen beleidsconclusies, noch geeft hij aanbevelingen voor actoren die betrokken zijn bij de opzet en het functioneren van de externe kwaliteitszorg.

3. De cybernetische-incongruentietheorie

Laat ik op voorhand duidelijk zijn. Iedere wetenschappelijke bijdrage in een wetenschappelijk debat, zeker als het gaat over het hoger onderwijs, vind ik nuttig en waardevol. Het functioneren van wetenschappelijke instellingen of onderdelen en aspecten daarvan wordt nog onvoldoende onderzocht vanuit een theoretisch kader.

Het belang van een wetenschappelijke bijdrage is o.m. gelegen in de vraag aan welk debat wordt deelgenomen. Frederiks participeert met zijn dissertatie hoofdzakelijk in het debat over het begrijpen van gedrag van complexe sociale systemen. Hij pleit voor een benadering van de universiteit, en naar ik mag aannemen van andere professionele organisaties, als een cybernetisch systeem dat streeft naar reductie van incongruentie. De toetsing van deze benadering vindt in dit proefschrift plaats aan de hand van één casus. De vraag is of dat niet een te eenvoudig toetsingsmodel is.

Ik ben met Frederiks van mening, dat de door hem gekozen operationalisatie van de begrippen uniciteit, dependentie en centraliteit voor verbetering vatbaar is. Maar levert dat een voldoende verklaring voor de tegenvallende resultaten? Het zou juist ook een grond voor ernstige twijfel aan de kwaliteit van de theorie kunnen zijn. De theorie was immers de basis voor de geformuleerde hypothesen en de operationalisatie.

Frederiks concludeert op p. 215 dat zich geen betere verklaringen dan de door hem gecon-

strueerde theorie hebben aangediend. Zijn onderzoeksopzet en literatuurstudie waren echter niet gericht op het vinden daarvan. Theoretisch gezien, zijn er voldoende concurrerende modellen beschikbaar. Frederiks verwijst zelf terloops naar politiek-economische theorieën: zijn referentie naar een consensus-denken binnen Nederlandse universiteiten past in dat kader. Verklaringen voor het gebruik van de aanbevelingen van visitatiecommissies zoeken in theorieën die organisatie-cultuur en symbolisch management centraal stellen hadden zeer voor de hand gelegen, temeer daar deze in onderzoek naar universitaire organisaties niet ongebruikelijk zijn (o.a. Cameron & Whetten, 1996). Frederiks heeft gekozen voor een toetsend onderzoek. Dat is zijn goed recht, maar de bovenstaande conclusie kan daarmee niet onderbouwd worden. Uitspraken over welke theorie de beste verklaring geeft, vraagt om een vergelijkend of exploratief onderzoek. Daaraan zou ik overigens gezien de status van de nieuwe theorie verre de voorkeur gegeven hebben.

Frederiks pleit tenslotte voor verder onderzoek vanuit het cybernetisch-incongruentie-model. Hij zelf heeft reeds enige verfijningen voorgesteld. De voorliggende studie heeft mij echter onvoldoende overtuigd van het nut daarvan. De theorie is zo abstract, geeft zo weinig richting aan het formuleren van bruikbare hypothesen en relevante operationalisaties en leidt tot zo weinig opzienbarende conclusies, dat een grondige herbezinning op de mogelijke betekenis van deze theorie voor het begrijpen van besluitvormingsprocessen in het hoger onderwijs mij meer had aangesproken.

Literatuur

Cameron, K.S., & Whetten, D.A. (1996). Organizational effectiveness and quality: The second generation. In: J. C. Smart (Ed.). *Higher Education: Handbook of theory and research*. (Vol. XI). (265-306). New York: Agathon.

J.F.M.J. van Hout
Faculteit Pedagogische en Onderwijskundige Wetenschappen
Universiteit van Amsterdam

Evaluatie van de mogelijkheden voor doorstroming binnen het MBO

Ard W. Lazonder en Frank P.C.M. de Jong
Landbouwwuniversiteit Wageningen, Leerstoelgroep Agrarische Onderwijskunde
Postbus 8130, 6700 EW Wageningen, tel. 0317 - 482710 / 484343
E-mail: Ard.Lazonder@Alg.AO.WAU.NL

Trefwoorden: secundair beroepsonderwijs, onderwijsvernieuwing, onderwijs-systeem

ABSTRACT

Vocational education should be flexible in order to respond to the changes of society. Moreover, the amount of flexibility that is realized should be in proportion to the amount of flexibility that is actually being used. This 'balance' between realized and used flexibility was studied in Dutch agricultural senior secondary vocational education (MBO). Because flexibility is a comprehensive concept, the study focused on horizontal and vertical transition, one of the key features of flexible education systems. Fifty-seven schools for agricultural MBO participated in the study by completing two questionnaires. The results of this study indicated that the schools put considerable effort in facilitating transition, whereas relatively few (i.e., less than 5%) of the students made use of the flexibility they were provided with. Based on these outcomes, it is concluded that the balance between realized and used flexibility may have gone awry. In the discussion, possible explanations are advanced and suggestions for further research are identified.

1. INLEIDING

Door de snelle technologische ontwikkelingen is de 'houdbaarheidsdatum' van kennis steeds sneller verstreken. Beroepsspecifieke kennis en vaardigheden die op school geleerd worden, kunnen kort na het betreden van de arbeidsmarkt al niet meer actueel zijn. Hierdoor worden andere eisen aan het onderwijs gesteld dan zo'n tien jaar geleden. Onderwijs, en met name beroepsonderwijs, moet tegenwoordig flexibel zijn om te kunnen voldoen aan de snel veranderende eisen van de samenleving.

In het MBO is de aansluiting onderwijs-arbeidsmarkt al geruime tijd een punt van aandacht. Sinds het begin van de jaren '80 wordt gewerkt aan de verbetering van de kwalitatieve aansluiting vanuit het beroepsonderwijs. Vanaf 1985 worden deze en andere inspanningen gebundeld onder de noemer landelijke kwalificatiestructuur. Met de invoering van de kwalificatiestructuur worden naast een verbeterde aansluiting van het onderwijs op de arbeidsmarkt ook andere doelen nagestreefd. Kort gezegd gaat het om verbetering van het interne rendement, de erkenning van behaalde diploma's, de mogelijkheden voor interne doorstroming en een grotere transparantie van opleidingen, certificaten en diploma's (Ministerie van OC&W, 1994). De kwalificatiestructuur is op haar beurt onderdeel van de veranderingen die in het kader van de Wet Educatie en Beroepsonderwijs (WEB) worden ingevoerd.

Vanaf dit schooljaar (1997/1998) is de kwalificatiestructuur in alle sectoren van het MBO

ingevoerd. Hierdoor rijst de vraag in hoeverre het onderwijsveld in staat is gebleken de doelen die met de kwalificatiestructuur beoogd werden te realiseren. Anders gezegd: heeft de invoering van de kwalificatiestructuur het gewenste effect opgeleverd? Deze vraag is moeilijk te beantwoorden. Niet alleen ontbreken objectieve criteria waaraan de invoering van de kwalificatiestructuur getoetst kan worden, ook is de invoering zo recent dat de scholen hun onderwijs wellicht nog niet volledig hebben kunnen aanpassen aan de eisen van de kwalificatiestructuur.

Dit laatste argument geldt niet - of in mindere mate - voor het agrarisch MBO. In deze sector is de kwalificatiestructuur in al 1992 ingevoerd, waardoor de agrarische MBO-scholen relatief veel tijd hebben gehad om de implicaties van de kwalificatiestructuur in hun onderwijs door te voeren. Een evaluatie van de agrarische kwalificatiestructuur kan dus belangrijke implicaties hebben voor de overige MBO-sectoren.

De evaluatie richt zich op een van de doelen van de kwalificatiestructuur: het verbeteren van de mogelijkheden voor interne doorstroming. Doorstroming moet hier opgevat worden als verandering van leerweg of kwalificatieniveau binnen het MBO, waarbij aansluitproblemen en onnodig tijdverlies zo veel mogelijk voorkomen moeten worden (Ministerie van LNV, 1989; Ministerie van OC&W, 1994). Verandering van leerweg wordt horizontale doorstroming genoemd. Hierbij wisselt een leerling van beroepsopleidende leerweg (dagschool) naar beroepsbegeleidende leerweg (leerlingwezen) of omgekeerd. Bij verticale doorstroming verandert het kwalificatieniveau van de opleiding. Leerlingen kunnen binnen het MBO doorstromen naar een opleiding met een hoger niveau (opstroom) of een lager niveau (afstroom). Bij beide vormen van doorstroming blijft het soort opleiding (bijvoorbeeld akkerbouw, bloemschikken, dierenzorg) gelijk.

Horizontale en verticale doorstroming vergroten de flexibiliteit van het onderwijs (Aitken, Lilley & Wardman-Browne, 1991; Collis, Vingerhoets & Moonen, 1997; Nasta, 1994; Raffae, 1994; Warwick, 1987). Voorwaarde hiervoor is wel dat de scholen goede mogelijkheden voor doorstroming hebben weten te creëren. Het is niet ondenkbaar dat de scholen hierbij op problemen zijn gestuit: de invoering van flexibele onderwijssystemen als de kwalificatiestructuur blijkt een kostbare en complexe operatie te zijn (o.a. De Bruijn, 1992; Raffae, 1994). Uit deze onderzoeken blijkt bovendien dat de wensen en behoeften van leerlingen relatief homogeen zijn. Zij maken hierdoor weinig gebruik van de flexibiliteit die het onderwijssysteem hen biedt (zie ook Babeliowsky, 1995). Er kan kortom sprake zijn van een zekere scheefgroei tussen de hoeveelheid gerealiseerde en benutte flexibiliteit. De vraag of deze scheefgroei zich ook voordoet bij horizontale en verticale doorstroming in de kwalificatiestructuur staat centraal in dit onderzoek.

2. HORIZONTALE EN VERTICALE DOORSTROMING

Horizontale en verticale doorstroming zijn bedoeld om leerlingen van opleiding te laten veranderen zonder aansluitproblemen en zonder teveel tijdsverlies als gevolg van herhalingen. De kwalificatiestructuur kent vier principes om dit te realiseren. Allereerst zorgt de erkenning van deelkwalificaties ervoor dat leerlingen reeds behaalde kwalificaties kunnen meenemen naar hun nieuwe opleiding. Dit voorkomt onnodige herhalingen en daardoor tijdsverlies. Door het modulaire karakter van de kwalificatiestructuur kunnen leerlingen die tijdens het schooljaar doorstromen reeds behaalde leseenheden (modulen) meenemen naar de nieuwe opleiding. Hierbij maakt de leerwegaafhankelijkheid van certificaten het voor leerlingen mogelijk om hun opleiding in een andere leerweg te vervolgen zonder al teveel tijdsverlies. Op dezelfde manier zorgt het inclusiviteitsprincipe (kwalificaties op hogere niveaus omvatten kwalificaties op lagere niveaus) ervoor dat doorstroming naar een opleiding op een hoger of lager niveau zonder onnodig tijdsverlies kan plaatsvinden. (vgl. Aitken *et al*, 1991; Nasta, 1994).

Op het macro niveau zijn deze principes verankerd in het ontwerp van de kwalificatiestructuur. Binnen het hierin geschetste kader van overheidsregels kunnen de scholen zelf bepa-

len hoe zij doorstromende leerlingen willen vrijwaren van aansluitproblemen en tijdsverlies. Als gevolg van dit dereguleringsbeleid (Ministerie van OC&W, 1992) mogen scholen hun eigen beleid voeren om doorstroommogelijkheden te realiseren. Dit doet een beroep op het beleidsvoerend vermogen van de scholen (Peters, Slegers & Bergen, 1995). Slegers (1991) onderscheidt 22 maatregelen die scholen kunnen nemen om beleid te voeren. Deze maatregelen zijn geselecteerd aan de hand van de bevoegdheden van schoolbesturen en zijn ingedeeld in drie domeinen: beheersmatig, onderwijskundig en contested (vgl. Hanson, 1979). Jongmans en Beijaard (1995) hebben deze maatregelen gevalideerd voor het agrarisch onderwijs.

Deze maatregelen vormen de basis voor dit onderzoek. Omdat het onderzoek zich richt op een concrete, onderwijskundige innovatie, en niet zoals bij Slegers (1991) op beleidsvoering in algemene zin, zijn alleen de maatregelen uit het onderwijskundig domein relevant. De maatregel 'opzet en werkwijze brugperiode' vormt hierop een uitzondering. Het MBO kent geen brugperiode, vandaar dat deze maatregel niet in het onderzoek wordt meegenomen. Eveneens ingegeven door de context van het onderzoek is, in navolging van Jongmans en Beijaard (1995), de maatregel 'afstemming over de praktijk- of stagecomponent' toegevoegd. De maatregelen met betrekking tot roostering zijn toegevoegd vanwege hun veronderstelde relevantie voor horizontale doorstroming. Voor verticale doorstroming zijn deze maatregelen niet van toepassing: de praktijk wijst uit dat, om organisatorische redenen, verticale doorstroming op vrijwel elke school alleen aan het einde van het schooljaar plaatsvindt. Een overzicht van de maatregelen die in dit onderzoek zijn gebruikt staat in Figuur 1.

Figuur 1: Maatregelen om mogelijkheden voor doorstroming te realiseren

COLLECTIEVE MAATREGELEN	INDIVIDUELE MAATREGELEN
<i>Curriculum-maatregelen</i>	<i>Roostering</i>
Afstemming van de organisatie van toetsing	Gespreide roostering voor parallelklassen*
Afstemming van de roostering*	Mogelijkheden voor remedial teaching*
Afstemming van de praktijk- of stagecomponent	Mogelijkheden voor zelfstudie*
Afstemming van de profiel- en keuzecertificaten	
<i>Instructie-maatregelen</i>	<i>Begeleiding</i>
Afstemming van de inhoud van het onderwijs	Leerlingen informeren over doorstroming
Afstemming van de te gebruiken leermiddelen	Leerlingen stimuleren/afraden om door te stromen
Afstemming van de didactische werkvormen	Leerlingen begeleiden bij keuze nieuwe opleiding
Afstemming van de onderwijsdoelen v.d. school	Leerlingen begeleiden na de doorstroming
	Leerlingen begeleiden bij remedial teaching*

Noot: De met een * gemarkeerde maatregelen gelden alleen voor horizontale doorstroming.

De maatregelen uit Figuur 1 zijn a-priori geassocieerd als collectief of individueel. Collectieve maatregelen hebben betrekking op de afstemming die binnen en tussen scholen kan plaatsvinden. Collectieve maatregelen betreffen altijd meer dan één opleiding en zorgen zodoende voor een zo optimaal mogelijke doorstroming tussen opleidingen. Individuele maatregelen worden binnen de eigen school genomen. Deze maatregelen zijn niet zozeer gericht op het afstemmen van opleidingen onderling, maar op het afstemmen van opleidingen op de wensen en behoeften van de leerlingen.

Bij de collectieve maatregelen is verder onderscheid gemaakt tussen maatregelen die betrekking hebben op het curriculum en de instructie. Curriculum-maatregelen liggen 'ver van het onderwijsleerproces' en betreffen de organisatie van het onderwijs, de toetsing, de stage en de keuzecertificaten. Deze maatregelen zijn vooral bedoeld om het tijdsverlies terug te dringen. Wanneer een doorgestroomde leerling bijvoorbeeld zijn/haar stage kan vervolgen

(in plaats van een nieuwe stage te beginnen) en dezelfde keuzecertificaten in zijn/haar studieprogramma kan opnemen (in plaats van nieuwe keuzecertificaten te gaan volgen) voorkomt dit onnodige herhaling en dus tijdsverlies. Instructie-maatregelen liggen 'dichtbij het onderwijsleerproces'; ze houden direct verband met de leerinhoud, de leermiddelen, de didactische werkvormen en de onderwijsdoelen van de school. Deze maatregelen zijn vooral gericht op het voorkomen van aansluitproblemen. Bij goede afstemming houden leerlingen na hun overstap dezelfde leerinhouden en leermiddelen en zijn de werkvormen en onderwijsdoelen toegesneden op de aard en het niveau van de opleiding.

De individuele maatregelen hebben betrekking op de roostering van het onderwijs en de begeleiding van leerlingen. Maatregelen met betrekking tot roostering helpen om onnodig tijdsverlies te voorkomen, bijvoorbeeld door doorstromers in een klas te plaatsen die wat betreft onderwijsaanbod het best past bij de deelkwalificaties die nog gehaald moeten worden. Wanneer doorgestroomde leerlingen buiten het rooster om modules kunnen halen of bijgespijkerd kunnen worden, voorkomt dit (een vergroting van) de studieachterstand. Door het begeleiden van leerlingen kunnen problemen in de huidige opleiding tijdig gesignaleerd worden; na de doorstroming zorgt begeleiding voor een snelle aanpak van eventuele aansluitproblemen.

Op basis van deze operationalisatie kunnen de centrale begrippen van dit onderzoek als volgt omschreven worden. De hoeveelheid gerealiseerde flexibiliteit wordt gedefinieerd als het aantal maatregelen dat scholen hebben genomen om mogelijkheden voor horizontale en verticale doorstroming te realiseren. Het aantal leerlingen dat van deze mogelijkheden gebruik maakt vormt de maat voor de benutte flexibiliteit. In deze evaluatie worden deze vormen van flexibiliteit vergeleken. Deze vergelijking kan drie uitkomsten hebben. De eerste mogelijkheid betreft een 'surplus' aan gerealiseerde flexibiliteit, veroorzaakt doordat slechts weinig leerlingen gebruik maken van de mogelijkheden om door te stromen. De tweede mogelijkheid sluit aan bij de bevinding van De Bruijn (1992) en Raffe (1994) dat de invoering van flexibele onderwijssystemen als de kwalificatiestructuur zo omvangrijk en complex is dat de scholen nog niet voldoende maatregelen hebben kunnen nemen. De derde mogelijkheid is dat de hoeveelheid gerealiseerde flexibiliteit is afgestemd op de hoeveelheid benutte flexibiliteit.

3. METHODE

3.1 Scholen

Alle 77 scholen voor agrarisch MBO zijn gevraagd aan het onderzoek deel te nemen. Uiteindelijk hebben 57 scholen hun medewerking verleend, een respons van 74%. Deze scholen verzorgen middelbaar agrarisch beroepsonderwijs op verschillende kwalificatieniveaus en in verschillende leerwegen (zie Tabel 1). Non-respons onderzoek toonde geen verschil tussen scholen die wel en niet aan het onderzoek deelnemen voor wat betreft hoogste kwalificatieniveau, leerweg en aantal leerlingen.

Organisatorisch gezien maakt elke school deel uit van een AOC (Agrarisch Opleidings Centrum: een uni-sectorale integratie van scholen voor agrarisch VBO en MBO). In totaal zijn er 18 AOC's; van elk AOC heeft minimaal 1 school aan het onderzoek meegewerkt.

3.2 Materialen

Voor elke vorm van doorstroming is een aparte vragenlijst gemaakt. De vragenlijst voor horizontale doorstroming bevatte 17 vragen; de vragenlijst voor verticale doorstroming bestond uit 12 vragen. De opzet van beide vragenlijsten is identiek.

De meeste vragen hebben betrekking op de maatregelen die scholen kunnen nemen om mogelijkheden voor doorstroming te creëren. Daarnaast is gevraagd naar het aantal leerlingen dat van de mogelijkheden om door te stromen gebruik heeft gemaakt. Per vragenlijst zijn de vragen toegespitst op horizontale dan wel verticale doorstroming, waarbij zoveel moge-

lijk dezelfde formulering is gebruikt. Beide vragenlijsten bevatten uitsluitend gesloten vragen.

De relevantie van de maatregelen uit de vragenlijst (zie Figuur 1) voor het creëren van doorstroommogelijkheden is op twee manieren gevalideerd. Voorafgaand aan het onderzoek hebben een aantal deskundigen (4 onderzoekers en 6 docenten uit het middelbaar agrarisch onderwijs) de relevantie van de maatregelen beoordeeld. Zij gaven aan dat alle maatregelen relevant geacht worden en gaven bovendien aan dat geen maatregelen gemist werden.

De tweede validering vond plaats tijdens het onderzoek zelf. De respondenten is gevraagd het belang van de maatregelen voor doorstroming aan te geven op een 4-punt schaal. Uit de gemiddelde scores blijkt dat alle maatregelen redelijk belangrijk tot belangrijk gevonden worden (range = 2.5 - 3.5). De maatregel 'gespreide roostering voor parallelklassen' vormt hierop een uitzondering. De reden hiervoor is dat slechts weinig scholen parallelklassen hebben.

Tabel 1: Aantal deelnemende scholen

LEERWEG ¹	HOOGSTE KWALIFICATIENIVEAU ²		
	BB (2-JARIG)	VF (3-JARIG)	MK/S (4-JARIG)
BOO	1	0	3
BBO	6	6	3
Beide leerwegen	1	2	35

Noot. Door het inclusiviteitsprincipe omvatten opleidingen op een hoger niveau de opleidingen op een lager niveau.

¹) BOO = beroepsopleidende leerweg (dagschool); BBO = beroepsbegeleidende leerweg (leerlingwezen).

²) BB = basisberoepsbeoefenaar; VF = vakfunctionaris; MK/S = middenkaderfunctionaris/specialist.

3.3 Procedure

Dataverzameling

De gegevens zijn verzameld in maart en april 1997. Hiervoor zijn alle scholen van alle AOC's aangeschreven ($N = 77$). Bij het selecteren van de scholen is gebruik gemaakt van de informatiegids *Opleidingen Middelbaar Agrarisch Onderwijs 1997*.

Tijdens de validering is aan de docenten gevraagd wie zij het meest geschikt vonden om de vragenlijsten in te vullen. Afgaande op hun reacties is besloten om de vragenlijst voor horizontale doorstroming door de onderwijscoördinator te laten invullen. Voor de vragenlijst voor verticale doorstroming bleek de decaan de meest aangewezen persoon te zijn.

Beide vragenlijsten zijn naar de directeur van de school gezonden met het verzoek deze door de onderwijscoördinator en de decaan te laten invullen. Inspelend op de mogelijkheid dat een school geen onderwijscoördinator en/of decaan heeft, werd de directeur gevraagd om in dat geval de betreffende vragenlijst zelf in te vullen. Na twee weken is een herinneringsbrief gestuurd naar alle scholen waarvan nog geen reactie was ontvangen. De scholen die niet op deze herinnering reageerden, zijn na nog eens drie weken telefonisch benaderd. Uiteindelijk hebben 57 scholen aan het onderzoek meegewerkt. Deze scholen hebben in totaal 99 vragenlijsten ingevuld; 50 voor horizontale en 49 voor verticale doorstroming.

Operationalisatie

In het onderzoek worden drie variabelen gebruikt: realisatiegraad, gerealiseerde flexibiliteit en benutte flexibiliteit. De variabele realisatiegraad geeft aan op hoeveel scholen beleid wordt ontwikkeld voor doorstroming. De realisatiegraad is gedefinieerd als het aantal scholen dat maatregelen neemt voor horizontale en verticale doorstroming. De hoeveelheid gerealiseerde flexibiliteit wordt geoperationaliseerd door het aantal gerealiseerde collectieve en individuele maatregelen (zie Figuur 1). Voor de hoeveelheid benutte flexibiliteit is per vorm van

doorstroming geregistreerd hoeveel procent van de leerlingen van een school doorstroomt. Bij verticale doorstroming is hierbij onderscheid gemaakt tussen opstroom en afstroom. Analooq hieraan is bij horizontale doorstroming ook gekeken naar de overstap van BOO naar BBO en omgekeerd. Alle leerlingaantallen hebben betrekking op het schooljaar 1995/1996.

Onderzoeksopzet en statistische analyses

Allereerst is de realisatiegraad beschreven. De realisatiegraad geeft aan hoeveel instellingen mogelijkheden voor horizontale en verticale doorstroming hebben gerealiseerd en geeft zo een indruk van de generaliseerbaarheid van de resultaten.

De hoeveelheid gerealiseerde flexibiliteit is onderzocht aan de hand van het aantal collectieve en individuele maatregelen dat scholen hebben genomen. Bij de collectieve maatregelen is gekeken naar de verschillen tussen het aantal curriculum-maatregelen en instructie-maatregelen. Ook is onderzocht of de scholen gemiddeld meer collectieve maatregelen nemen (1) binnen de eigen school; (2) met een school van het eigen AOC; of (3) met een school van een ander AOC. De analyses zijn uitgevoerd met gepaarde t-testen of Wilcoxon matched pair signed rank testen (toetsingsgrootheid T).

Bij de individuele maatregelen is nagegaan of verschillende soorten scholen evenveel individuele maatregelen genomen hebben. Voor horizontale doorstroming is hiertoe een MANOVA uitgevoerd met schooltype (BBO, BOO, beide leerwegen) als between-subject factor en soort maatregel (roostering, begeleiding) als afhankelijke variabele. Voor verticale doorstroming is een ANOVA uitgevoerd. De between-subject factor 'schooltype' wordt hier bepaald door het hoogste kwalificatieniveau.

De hoeveelheid benutte flexibiliteit is onderzocht door het percentage doorstromende leerlingen van een school te vergelijken. Deze vergelijking betreft zowel de manier van doorstroming (binnen de eigen school, binnen het eigen AOC, van of naar een ander AOC) als het soort doorstroming. Voor horizontale doorstroming heeft soort doorstroming twee waarden: 'van BOO naar BBO' en 'van BBO naar BOO'. Voor verticale doorstroming wordt het soort doorstroming bepaald door opstroom en afstroom. De analyses zijn uitgevoerd met MANOVAs met soort en manier van doorstroming als within-subject factors. Tot slot zijn correlaties tussen de hoeveelheid benutte en gerealiseerde flexibiliteit berekend.

4. RESULTATEN

4.1 Realisatiegraad

De realisatiegraad is hoog. Tabel 2 laat zien dat vrijwel alle scholen (98%) individuele maatregelen nemen voor horizontale doorstroming. Voor verticale doorstroming is de realisatiegraad 100%. Collectieve maatregelen worden eveneens door een ruime meerderheid van de scholen genomen. Slechts 5 scholen nemen geen collectieve maatregelen voor horizontale doorstroming; 4 scholen nemen geen collectieve maatregelen voor verticale doorstroming.

Bij horizontale doorstroming verschilt de mate waarin collectieve maatregelen worden genomen (Cochran $Q(2) = 32.76, p < .01$). Collectieve maatregelen worden vaker dan gemiddeld binnen de eigen school genomen, terwijl collectieve maatregelen met scholen van een ander AOC minder dan gemiddeld voorkomen. De data voor verticale doorstroming laten een soortgelijk beeld zien. Ook hier zijn de onderlinge verschillen significant (Cochran $Q(2) = 23.21, p < .01$).

4.2 Gerealiseerde flexibiliteit

Collectieve maatregelen

Tabel 3 geeft een overzicht van het aantal gerealiseerde collectieve maatregelen. Bij horizontale doorstroming worden de meeste collectieve maatregelen binnen de eigen school genomen. Het totaal aantal collectieve maatregelen dat scholen intern nemen is groter dan het aantal

maatregelen dat samen met scholen uit het eigen AOC wordt gerealiseerd ($T(18) = 9, p < .01$). Het aantal intern gerealiseerde collectieve maatregelen verschilt ook van het aantal maatregelen dat met scholen van een ander AOC wordt genomen ($T(6) = 0, p < .05$). Het verschil tussen het aantal maatregelen 'in AOC' en 'extern' is niet significant.

Ook bij verticale doorstroming is de hoeveelheid collectieve maatregelen binnen de eigen school het grootst. Intern neemt men meer collectieve maatregelen dan met scholen uit het eigen AOC ($T(19) = 9.5, p < .05$) of met scholen van een ander AOC ($T(9) = 1, p < .01$). Bovendien is het verschil tussen het aantal maatregelen met scholen van binnen en buiten het eigen AOC significant ($T(8) = 1, p < .05$).

In Tabel 3 wordt ook onderscheid gemaakt naar het soort collectieve maatregel (curriculum, instructie). Bij horizontale doorstroming worden alleen bij afstemming met scholen uit het eigen AOC meer curriculum-maatregelen dan instructie-maatregelen genomen ($t(31) = 2.44, p < .05$). Bij de andere vormen van afstemming verschilt het aantal curriculum- en instructie-maatregelen niet van elkaar. Ook bij verticale doorstroming bestaat geen verschil tussen het aantal gerealiseerde curriculum- en instructie-maatregelen.

Tabel 2: Realisatiegraad

	INTERN	IN AOC	EXTERN
<i>Horizontale doorstroming</i>			
Collectieve maatregelen	31 (88.6%) ¹	32 (64.0%)	9 (18.0%)
Individuele maatregelen	49 (98.0%)	-	-
<i>Verticale doorstroming</i>			
Collectieve maatregelen	38 (77.6%)	29 (59.2%)	14 (28.6%)
Individuele maatregelen	49 (100%)	-	-

¹) alleen berekend voor scholen met beide leerwegen.

Tabel 3: Gemiddeld aantal gerealiseerde collectieve maatregelen

	INTERN	IN AOC	EXTERN
<i>Horizontale doorstroming</i>			
Curriculum-maatregelen	2.9 (1.3) ¹	1.8 (1.1)	1.1 (0.3)
Instructie-maatregelen	2.7 (1.3) ¹	1.4 (1.0)	0.9 (0.9)
<i>Verticale doorstroming</i>			
Curriculum-maatregelen	2.4 (0.9)	1.6 (1.1)	0.9 (1.0)
Instructie-maatregelen	2.3 (1.2)	1.4 (1.0)	0.8 (0.6)

Noot. Standaarddeviaties tussen haakjes.

¹) alleen berekend voor scholen met beide leerwegen.

Individuele maatregelen

Het gemiddeld aantal gerealiseerde individuele maatregelen is vergeleken tussen de verschillende soorten scholen (zie Tabel 4). Bij horizontale doorstroming heeft het soort school een multivariaat effect op het totaal aantal gerealiseerde individuele maatregelen ($F(4,90) = 4.91, p < .01$). Kennelijk nemen scholen met verschillende leerwegen niet evenveel individuele

maatregelen. Soort school heeft bovendien een univariaat effect op het aantal maatregelen dat betrekking heeft op de roostering van het onderwijs ($F(2,47) = 10.11, p < .01$). Uit post-hoc Scheffé analyse blijkt dat BBO-scholen minder maatregelen voor roostering nemen dan scholen met beide leerwegen. Het soort school heeft geen effect op het aantal maatregelen voor begeleiding ($F(2,47) = 1.3$).

Tabel 4: Gemiddeld aantal gerealiseerde individuele maatregelen

	SOORT INSTELLING		
	BOO	BBO	BEIDE
<i>Horizontale doorstroming</i>			
Roostering	1.8 (.5)	0.5 (0.8)	1.8 (0.9)
Begeleiding	3.3 (1.7)	2.9 (1.1)	3.6 (1.2)
	BB	VF	MK/S
<i>Verticale doorstroming*</i>			
Begeleiding	3.0 (0.8)	2.8 (0.8)	3.1 (0.8)

Noot. Standaarddeviaties tussen haakjes.

* Maatregelen met betrekking tot roostering zijn niet van toepassing omdat verticale doorstroming in de praktijk vrijwel uitsluitend aan het einde van een schooljaar plaatsvindt.

Bij verticale doorstroming heeft het soort school geen effect op het aantal maatregelen voor de begeleiding van leerlingen ($F(2,46) = .21$). Scholen die een BB-, VF- of MK/S-opleiding verzorgen nemen even veel maatregelen om leerlingen te begeleiden bij verticale doorstroming.

Bovendien is nagegaan of het aantal individuele maatregelen verband houdt met het aantal collectieve maatregelen. Bij horizontale doorstroming zijn de aantallen gerealiseerde individuele en collectieve maatregelen niet aan elkaar gerelateerd. Het aantal collectieve maatregelen dat een school neemt zegt dus niets over het aantal individuele maatregelen dat wordt genomen en vice versa.

Bij verticale doorstroming bestaat coherentie voor wat betreft de flexibiliteit die binnen de eigen school wordt gerealiseerd. Het aantal gerealiseerde individuele maatregelen hangt samen met het aantal collectieve maatregelen dat intern wordt genomen ($r = .28, p < .05$). Dit verband bestaat niet voor de flexibiliteit die met scholen binnen of buiten het eigen AOC wordt gerealiseerd.

4.3 Benutte flexibiliteit

Uit Tabel 5 blijkt dat weinig leerlingen gebruik maken van de mogelijkheden om horizontaal en verticaal door te stromen. Alle percentages leerlingen die doorstromen zijn lager dan 5%.

De drie manieren om horizontaal door te stromen worden relatief gezien even vaak gebruikt ($F(2,50) = 1.08$). Het soort horizontale doorstroming verschilt wel: meer leerlingen verhuizen de dagschool voor het leerlingwezen dan omgekeerd ($F(1,25) = 5.45, p < .05$). Uit het interactie-effect tussen manier en soort doorstroming blijkt dat dit verschil niet voor elke manier van doorstroming even groot is ($F(2,50) = 4.84, p < .05$). Een post-hoc Tukey test wijst uit dat het verschil 'intern' significant groter is dan 'in AOC' of 'extern'.

De manieren om verticaal door te stromen worden niet door evenveel leerlingen gebruikt

($F(2,30) = 11.00, p < .01$). Uit Tukey testen blijkt dat verticale doorstroming binnen de eigen school vaker voorkomt dan verticale doorstroming binnen het eigen AOC of naar een ander AOC. Tussen het gebruik van deze laatste twee manieren bestaat geen verschil. Wel verschilt het soort verticale doorstroming ($F(1,31) = 15.35, p < .01$). Opstroom komt vaker voor dan afstroom. Omdat ook hier een interactie-effect is gevonden ($F(2,30) = 3.60, p < .05$), is dit verschil niet altijd even groot. Net als bij horizontale doorstroming geven Tukey testen aan dat het verschil intern groter is dan 'inAOC' of 'extern'.

Tabel 5: Gemiddeld percentage leerlingen dat doorstroomt

	INTERN	IN AOC	EXTERN
<i>Horizontale doorstroming</i> ¹			
Van BOO naar BBO	1.0 (1.4)	0.3 (0.5)	0.5 (0.9)
Van BBO naar BOO	0.1 (0.2)	0.3 (1.4)	0.1 (0.5)
<i>Verticale doorstroming</i>			
Opstroom	4.5 (5.9)	1.1 (2.2)	0.7 (1.3)
Afstroom	0.9 (1.1)	0.2 (0.5)	0.1 (0.2)

Noot. Scores zijn gecorrigeerd voor 'out-of-range' values. Standaarddeviaties tussen haakjes.

¹) alleen berekend voor scholen met beide leerwegen.

Tot slot is onderzocht of de benutte flexibiliteit verband houdt met de gerealiseerde flexibiliteit. Hiervoor zijn correlaties berekend tussen het percentage doorstromers en de indicatoren van gerealiseerde flexibiliteit (zie Tabel 6). Uit deze correlaties blijkt dat de maatregelen die scholen nemen voor horizontale doorstroming geheel onafhankelijk zijn van het aantal leerlingen dat horizontaal doorstroomt. Ook bij verticale doorstroming bestaat geen verband tussen het aantal gerealiseerde maatregelen en het aantal doorstromende leerlingen.

Tabel 6: Relaties tussen het percentage leerlingen dat doorstroomt en de gerealiseerde maatregelen

GEREALISEERDE MAATREGELN	PERCENTAGE LEERLINGEN		
	INTERN	IN AOC	EXTERN
<i>Horizontale doorstroming</i>			
Collectieve maatregelen	-.03 (26) ¹	-.04 (29)	.10 (8)
Individuele maatregelen	.04 (28)	-.21 (45)	-.17 (45)
<i>Verticale doorstroming</i>			
Collectieve maatregelen	.04 (35)	.09 (39)	-.03 (39)
Individuele maatregelen	-.05 (35)	.04 (39)	-.04 (39)

Noot. Pearson correlatie, *n*-waarden tussen haakjes.

¹) alleen berekend voor scholen met beide leerwegen.

5. DISCUSSIE

In de kwalificatiestructuur wordt gestreefd naar goede mogelijkheden voor horizontale en verticale doorstroming. De mate waarin scholen deze mogelijkheden gecreëerd hebben geeft een indicatie van de hoeveelheid gerealiseerde flexibiliteit. De mate waarin leerlingen van deze mogelijkheden gebruik maken bepaalt de hoeveelheid benutte flexibiliteit. In dit onderzoek is nagegaan hoe deze vormen van flexibiliteit zich tot elkaar verhouden in het agrarisch MBO.

Uit het onderzoek blijkt dat het realiseren van mogelijkheden voor horizontale en verticale doorstroming een breed draagvlak heeft in het agrarisch MBO. Alle onderzochte scholen nemen individuele maatregelen om mogelijkheden voor doorstroming te creëren. Ook neemt vrijwel elke school collectieve maatregelen om horizontale en verticale doorstroming zo goed mogelijk te laten verlopen. Ondanks de hoge realisatiegraad zijn de resultaten niet zonder meer te generaliseren. De veelal hoge standaarddeviaties wijzen op een grote spreiding van de antwoorden, waardoor het beeld dat de resultaten schetsen niet op elke school van toepassing hoeft te zijn.

De meeste flexibiliteit wordt binnen de eigen school gerealiseerd. Afstemming binnen de eigen school komt gemiddeld vaker voor dan de andere vormen van afstemming. Ook worden intern de meeste collectieve maatregelen genomen. Daarnaast dragen de individuele maatregelen - die per definitie betrekking hebben op de eigen school - bij aan de hoeveelheid intern gerealiseerde flexibiliteit.

Deze resultaten kunnen verklaard worden door de voorwaarden voor het nemen van collectieve maatregelen. Volgens Ros *et al* (1996) is regelmatig intercollegiaal overleg essentieel voor het realiseren van dit soort maatregelen. Binnen de eigen school zijn de condities hiervoor gunstig: de afstand tot collega's is klein en de communicatielijnen zijn kort. Bij overleg met collega's van andere scholen is de situatie minder ideaal, waardoor het moeilijker wordt om (veel) collectieve maatregelen te realiseren.

De flexibiliteit die in AOC-verband wordt gerealiseerd verschilt niet op alle punten van de flexibiliteit die met scholen van een ander AOC wordt gecreëerd. Afstemming binnen het eigen AOC komt gemiddeld vaker voor dan afstemming buiten het eigen AOC. Bij verticale doorstroming worden 'inAOC' meer collectieve maatregelen genomen dan met een school van een ander AOC. Bij horizontale doorstroming doet dit verschil zich echter niet voor.

Dat in AOC-verband niet altijd meer flexibiliteit gerealiseerd wordt dan met scholen van een ander AOC kan verklaard worden door de lage *n*-waarden in de analyses. Een meer inhoudelijke verklaring ligt in de aard van de samenwerking in AOC's. Volgens Wassink (1996) is deze samenwerking eerder federatief dan geïntegreerd van aard. De meeste AOC's bestaan uit relatief zelfstandige scholen, waardoor het mogelijk is dat scholen van oudsher samenwerken met scholen die momenteel niet tot het eigen AOC behoren. Deze samenwerking zal vooral zijn blijven bestaan wanneer een school opleiding(en) aanbiedt die niet door andere scholen in het AOC worden verzorgd (vgl. Knip, 1982). Voor het nemen van collectieve maatregelen voor horizontale en verticale doorstroming (waarbij het soort opleiding gelijk blijft) zijn deze scholen dan aangewezen op scholen buiten het eigen AOC.

Naast het aantal maatregelen dat scholen nemen om doorstroming mogelijk te maken is ook gekeken naar het soort maatregelen. De collectieve maatregelen zijn onderverdeeld in maatregelen die betrekking hebben op het curriculum en op de instructie. Scholen realiseren deze maatregelen vrijwel in gelijke mate, alleen voor horizontale doorstroming binnen het eigen AOC worden meer curriculum- dan instructie-maatregelen genomen. Dit betekent dat scholen in vrijwel alle gevallen op zowel organisatorisch als didactisch gebied voorwaarden voor doorstroming scheppen.

De individuele maatregelen hebben betrekking op de roostering van het onderwijs en de begeleiding van leerlingen. Bij horizontale doorstroming nemen BBO-scholen minder maatregelen voor de roostering van het onderwijs dan scholen met beide leerwegen. Dit verschil kan verklaard worden door de aard van de leerweg. Van oudsher gaan BBO-leerlingen gemid-

deld een dag per week naar school. Hierdoor is de noodzaak voor interne afspraken over roostering geringer dan wanneer leerlingen vier of vijf dagen per week op school doorbrengen. Met de invoering van de WEB krijgen de scholen echter grotere vrijheid in het bepalen van het schoolse deel van een BBO-opleiding. Mogelijk verandert hierdoor het aantal maatregelen dat BBO-scholen nemen over de roostering. In ieder geval betekent het een punt van aandacht voor die BBO-scholen die hun schoolse aandeel willen gaan veranderen.

Op grond van het aantal leerlingen dat doorstroomt kan geconcludeerd worden dat de hoeveelheid benutte flexibiliteit klein is. Ronduit weinig leerlingen maken gebruik van de mogelijkheid om horizontaal en verticaal door te stromen. Hierdoor lijkt er sprake te zijn van een onevenwichtige verhouding tussen de hoeveelheid gerealiseerde en benutte flexibiliteit. Een vergelijking in absolute zin is echter niet mogelijk. Er is immers geen 'norm' voor het aantal leerlingen dat per jaar moet doorstromen en aan het aantal te realiseren maatregelen wordt evenmin eisen gesteld. Desondanks geven de resultaten de indruk dat scholen op relatief veel terreinen actief zijn om horizontale en verticale doorstroming mogelijk te maken, terwijl relatief weinig leerlingen van deze mogelijkheden gebruik maken. Deze conclusie wordt versterkt doordat deze maten geen enkele samenhang vertonen. Wat de scholen 'doen' om mogelijkheden voor doorstroming te creëren is geheel onafhankelijk van het aantal leerlingen dat doorstroomt.

Op grond van deze mogelijke scheefgroei kunnen twee voorlopige conclusies getrokken worden. Ten eerste blijkt het creëren van goede doorstroommogelijkheden in de kwalificatiestructuur niet zo complex en tijdrovend te zijn dat de scholen nog niet voldoende flexibiliteit hebben gerealiseerd. De scholen lijken eerder teveel flexibiliteit gecreëerd te hebben, zeker gezien het geringe aantal leerlingen dat doorstroomt. Dit laatste vormt de tweede voorlopige conclusie. Net als in andere onderzoeken naar flexibiliteit (Babeliowsky, 1995; De Bruijn, 1992; Raffe, 1994) maken de leerlingen weinig gebruik van de flexibiliteit die het onderwijssysteem hen biedt (i.e., de mogelijkheden voor doorstroming).

Deze conclusies zijn voorlopig: de geconstateerde scheefgroei kan ook op andere manieren verklaard worden. Door de invoering van de WEB zijn scholen min of meer verplicht om mogelijkheden voor doorstroming te creëren. Het feit dat de scholen zich hierbij ogenschijnlijk niet laten leiden door de feitelijke vraag naar doorstroming kan betekenen dat scholen het belang dat in de WEB hecht wordt aan doorstroming onderkennen. De geconstateerde onafhankelijkheid kan ook opgevat worden als een teken van een surplus aan flexibiliteit. De scholen scheppen voorwaarden voor doorstroming zonder rekening te houden met het aantal leerlingen dat feitelijk doorstroomt. Omdat dit aantal klein is, lijkt het op voorhand creëren van flexibiliteit niet alleen overbodig, maar ook inefficiënt. Vooral het realiseren van collectieve maatregelen lijkt dermate arbeidsintensief te zijn dat het de vraag is of de geleverde inspanningen in verhouding staan tot het aantal leerlingen dat doorstroomt.

Aan de andere kant worden collectieve en individuele maatregelen niet alleen genomen voor horizontale of verticale doorstroming. Zo kan afstemming met andere scholen de efficiency verhogen, bijvoorbeeld wanneer overlegd wordt over stagebedrijven, leermiddelen uitgewisseld worden of gezamenlijk profiel- en keuzecertificaten aangeboden worden. Ook de individuele maatregelen hebben een bredere toepassing. Remedial teaching en zelfstudie zijn manieren om het tempo en de moeilijkheid van het onderwijs aan individuele leerlingen aan te passen. Hiervan profiteren ook de leerlingen die niet van plan zijn door te stromen.

Een derde mogelijke verklaring betreft de geringe hoeveelheid benutte flexibiliteit. Mogelijk wordt dit veroorzaakt door het ontbreken van bekendheid met en behoefte aan doorstroming. In het eerste geval zijn leerlingen onvoldoende op de hoogte van de mogelijkheden voor doorstroming. Hoewel uit een post-hoc analyse blijkt dat 96% van de scholen bekendheid geeft aan de mogelijkheden voor doorstroming, is het denkbaar dat deze informatie bij de leerlingen niet overkomt. Een tweede mogelijkheid is dat leerlingen geen behoefte hebben om door te stromen. Dit kan komen doordat leerlingen (dankzij goede begeleiding) een passende opleiding hebben gekozen, of juist een totaal verkeerde opleiding hebben gekozen en uitstromen in plaats van doorstromen. Ook zou de geringe behoefte aan doorstroming veroorzaakt

kunnen zijn door de overige vernieuwingen die als gevolg van de kwalificatiestructuur zijn ingevoerd. Zaken als tempodifferentiatie en individuele leertrajecten maken het onderwijs flexibeler, waardoor de leerlingen meer vrijheid hebben om hun eigen leerproces vorm en inhoud te geven en hierdoor wellicht minder behoefte hebben om door te stromen.

Deze discussie maakt duidelijk dat de uitkomsten van dit onderzoek niet kunnen leiden tot een definitief oordeel over de mate waarin de invoering van de kwalificatiestructuur de mogelijkheden voor doorstroming heeft verbeterd. Hoewel de scholen veel in het werk hebben gesteld om doorstroming mogelijk te maken, is het effect van deze inspanningen moeilijk op waarde te schatten. Slechts weinig leerlingen stromen door, zodat geconcludeerd moet worden dat de verhouding tussen gerealiseerde en benutte flexibiliteit mogelijk scheefgegroeid is.

Deze conclusie leidt tot twee suggesties voor verder onderzoek. Ten eerste moet onderzoek worden of de scholen tevreden zijn met de hoeveelheid gerealiseerde flexibiliteit. Het gaat hierbij om eventuele verschillen tussen het huidige en gewenste aantal gerealiseerde maatregelen voor doorstroming. Daarnaast moeten de mogelijke oorzaken voor de geringe hoeveelheid benutte flexibiliteit onderzocht worden. Dit kan door leerlingen uit het agrarisch MBO te bevragen over hun bekendheid met en behoefte aan doorstroming. Op basis van deze twee onderzoeken kan bepaald worden of de mogelijkheden voor horizontale en verticale doorstroming dusdanig verbeterd zijn dat de invoering van de kwalificatiestructuur op dit punt succesvol genoemd kan worden.

6. LITERATUUR

- Aitken, B., Lilley, D. & Wardman-Browne, S. (1991). NVQ's: Delivering quality with a flexible modular approach. In D. Müller & P. Funnel (eds.), *New developments in vocational education* (pp. 144 - 152). Londen: Kogan Page.
- Babeliowsky, M. (1995). *Schoolloopbanen en studierendenment in het voortgezet onderwijs: Resultaten voor het OVG Noordwest Overijssel*. Amsterdam: SCO Kohnstamm Instituut.
- Bruijn, E. de (1992). *Modularisation in Dutch vocational education and training*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Collis, B., Vingerhoets, J. & Moonen, J. (1997). Flexibility as a key construct in European training: Experiences from the telescopia project. *British Journal of Educational Technology*, 28(3), 199 - 217.
- Hanson, M.E. (1979). *Educational administration and organisational behavior*. Boston: Allyn & Bacon.
- Jongmans, C.T. & Beijaard, D. (1995). *Schoolbeleid en professionalisering van leraren: Betrokkenheid van leraren bij schoolbeleid, professionele oriëntatie van leraren en ontwikkeling van scholingsbeleid in scholen voor voortgezet agrarisch onderwijs*. Wageningen: Landbouwniversiteit Wageningen, Vakgroep Agrarische Onderwijskunde.
- Knip, H. (1982). *Organisatiestudies in het onderwijs*. Den Haag: SVO.
- Ministerie van LNV (1989). *Discussienota diploma's en certificaten*. Den Haag: DLO.
- Ministerie van OC&W (1992). *Opleiding schoolleiders primair onderwijs: meewerkende of leidende directoren*. Den Haag: Ministerie van OC&W.
- Ministerie van OC&W (1994). *Wet educatie en beroepsonderwijs*. Den Haag: SDU.
- Nasta, T. (1994). *How to design a vocational curriculum: A practical guide for schools and colleges*. Londen: Kogan Page.
- Opleidingen middelbaar agrarisch onderwijs '97: Informatiegids voor studie- en beroepskeuzebegeleiding in het voortgezet onderwijs en de arbeidsvoorziening* (1997). Helmond: SILO.
- Peeters, K., Slegers, P. & Bergen, Th. (1995). De taakperceptie van schoolleiders. *Pedagogische Studiën*, 72, 424 - 436.
- Raffe, D. (1994). The new flexibility in vocational education. In W.J. Nijhof & J.N. Streumer (eds.), *Flexibility in training and vocational education* (pp. 13 - 32). Utrecht: Lemma.
- Ros, A.A., Swinkels-Kuijlaars, H.P.C., Theunissen, J.J.C.M., Visser, J.J.C.M., Jongmans, C.T. & Geijsel, F. (1996). *Over vernieuwingen gesproken*. Wageningen: Landbouwniversiteit Wageningen, Vakgroep Agrarische Onderwijskunde/Kompactgroep.
- Slegers, P.J.C. (1991). *School en beleidsvoering: Een onderzoek naar de relatie tussen het beleidsvoerend*

vermogen van scholen en het benutten van de beleidsruimte door scholen. Proefschrift Katholieke Universiteit Nijmegen, Nijmegen.

Warwick, D. (1987). *The modular curriculum.* Oxford: Basil Blackwell.

Wassink, H. (1996). *De veranderende rol van het management van AOC's.* Wageningen: Landbouwniversiteit Wageningen, Vakgroep Agrarische Onderwijskunde.

Leerkrachtwisselingen en verschillen in toetsscores tussen leerlingcohorten¹

Hans Luyten, *Universiteit Twente, Vakgroep Onderwijsorganisatie en -management*
 Simone Doolaard, *Universiteit Twente, Vakgroep Onderwijsorganisatie en -management*

Results from school effectiveness studies that have taken into account differences between grades within schools indicate that differences between teachers may outweigh the differences between schools. Comparisons between parallel classes, however, have yielded much smaller differences. For teachers in different grades it is very difficult to compare the achievements of their pupils, while this is relatively easy for teachers in parallel classes. The present study relates to another situation that allows for a comparison between teachers within schools, namely when in different years different teachers have taught the same grade, the final grade of elementary education. In that case teachers can compare the achievements of their pupils to those in previous years. Our analyses yield small differences between teachers within schools.

1. INLEIDING

Wat maakt meer verschil voor de leerprestaties die een leerling behaalt, de school waar hij naar toe gaat of de leerkracht van wie hij les krijgt? Anders geformuleerd: hoe verhouden zich de verschillen in effectiviteit tussen scholen tot de verschillen tussen leerkrachten binnen een school? Zijn de verschillen tussen leerkrachten erg groot of bestaat er juist een sterke mate van overeenkomst binnen scholen? Naarmate de verschillen tussen leerkrachten de verschillen tussen scholen overtreffen wordt het minder belangrijk welke school een leerling bezoekt. Immers, als in dat geval een leerling op een effectieve school een van de minst effectieve leerkrachten treft is hij al gauw slechter af dan wanneer hij op een minder effectieve school les zou krijgen van een van de meest effectieve leerkrachten. De betekenis van de verschillen tussen scholen neemt af naarmate de verschillen tussen leerkrachten binnen scholen groter zijn.

Studies naar de verschillen in effectiviteit tussen leerkrachten binnen basisscholen geven geen eenduidig antwoord op de vraag naar de omvang van leerkrachteffecten. Als parallelgroepen binnen scholen worden vergeleken, zijn de verschillen meestal klein, maar vergelijkingen tussen jaargroepen laten juist grote verschillen zien tussen leerkrachten. Als een school goede resultaten behaalt met de leerlingen in groep 6, dan zegt dit bijvoorbeeld betrekkelijk weinig over de resultaten in groep 8.

In tabel 1 worden de resultaten samengevat van de effectiviteitsstudies in het Nederlandse basisonderwijs waarin zowel school- als leerkrachteffecten zijn vastgesteld. Hierbij wordt een onderscheid gemaakt tussen uitkomsten die zijn gebaseerd op vergelijkingen tussen jaargroepen en uitkomsten gebaseerd op vergelijkingen tussen parallelgroepen. De tabel laat zowel de resultaten zien met betrekking tot leerprestaties die niet zijn gecorrigeerd voor leerlingkenmerken (zoals intelligentie, pre-test, OVB-leerlinggewicht, opleidingsniveau ouders en sexe), als de uitkomsten voor gecorrigeerde scores. Behalve de variantie-componenten op school- en leerkrachtniveau is ook steeds de tussen-classes binnen-scholen correlatie (ρ) vermeld. Deze parameter geeft aan of de school- dan wel de leerkracht-variantie meer gewicht in de schaal legt en is berekend volgens onderstaande formule (vgl. Bosker, 1991):

$$\rho = \frac{\text{Schoolvariantie}}{\text{Schoolvariantie} + \text{Leerkrachtvariantie}}$$

Als de schoolvariantie groot is en de leerkrachtvariantie klein, komt ρ in de buurt van 1. Is de schoolvariantie klein is en de leerkrachtvariantie groot, dan komt ρ dichtbij 0 te liggen. Zolang de schoolvariantie groter is dan de leerkrachtvariantie, is ρ groter dan .50.

Tabel 1: School- en leerkrachteffecten in Nederlands effectiviteitenonderzoek

STUDIE	Variantie op schoolniveau		Leerkracht-Variantie		ρ	
	ruwe scores	gecorrigeerde scores	ruwe scores	gecorrigeerde scores	ruwe scores	gecorrigeerde scores
Jaargroepen						
Bosker, 1991						
Rekenen	15%	5%	11%	6%	.58	.47
Taal	13%	1%	6%	4%	.69	.26
Luyten & Snijders, 1996						
Rekenen	--	5%	--	14%	--	.26
Taal	--	0%	--	6%	--	.00
Van der Werf & Guldemond, 1996						
Rekenen 1988	10%	4%	6%	5%	.61	.41
Rekenen 1990	11%	5%	6%	5%	.65	.48
Rekenen 1992	9%	3%	7%	5%	.58	.39
Taal 1988	8%	2%	11%	10%	.44	.14
Taal 1990	11%	2%	7%	6%	.62	.22
Taal 1992	13%	3%	6%	6%	.67	.36
Bosker, Lam, Dekkers & Vierke, 1997						
Rekenen	--	5%	--	8%	--	.38
Taal	--	3%	--	5%	--	.38
Parallelgroepen						
Bosker, 1991						
Rekenen	23%	8%	5%	4%	.81	.69
Taal	14%	4%	0%	0%	1.00	.96
Bosker, Lam, Dekkers & Vierke, 1997						
Rekenen	--	5%	--	4%	--	.76*
Taal	--	3%	--	4%	--	.67*

*In deze studie is een model met vier niveaus gefit (school, jaargroep, parallelklas, leerling). De formule voor de tussen-classes binnen-scholen correlatie wordt dan:

$$\rho = \frac{\text{Schoolvariantie} + \text{Jaargroepvariantie}}{\text{Schoolvariantie} + \text{Jaargroepvariantie} + \text{Parallelklasvariantie}}$$

De tabel laat hogere waarden voor ρ zien als de leerkrachtvariantie is gebaseerd op vergelijkingen tussen parallelgroepen. Met andere woorden: als parallelgroepen worden vergeleken, dan zijn de verschillen tussen leerkrachten relatief klein. Als ρ wordt bepaald op basis van verschillen tussen jaargroepen, dan vindt men voor de ruwe scores waarden tussen .44 en .69. Op basis van verschillen tussen parallelgroepen worden hogere waarden voor ρ gevonden, namelijk .81 en 1.00. Ook voor de gecorrigeerde scores geldt dat de hoogste waarden worden gevonden als parallelklassen met elkaar worden vergeleken. Op basis van verschillen tussen jaargroepen varieert ρ van .00 tot .48, terwijl vergelijkingen tussen parallelgroepen waarden opleveren tussen .67 en .96.

Mandeville en Anderson (1987) bieden als verklaring voor de verschillen tussen jaargroepen de schoolspecifieke curriculumopbouw. Dat zou betekenen dat de verschillen het gevolg zijn van een bewuste keuze. Zo zou een leerkrachtenteam kunnen besluiten om het laatste leerjaar voornamelijk te gebruiken voor herhaling en verdieping van de basisstof. In zo'n geval is de stof aan het eind van groep 7 al grotendeels behandeld. De leerlingen in groep 7 van zo'n school zouden op een test ver boven het gemiddelde kunnen scoren. In groep 8 zouden ze dan waarschijnlijk weer dichterbij het gemiddelde uitkomen omdat de leerlingen in de overige scholen hun achterstand dan grotendeels hebben ingehaald. Bosker (1991) spreekt in dit verband van een sterke en een zwakke bewijsvoering. Verschillen tussen parallelgroepen binnen een school vormen in deze gedachtegang een sterk bewijs voor het bestaan van leerkrachteffecten omdat alternatieve verklaringen in dat geval nauwelijks voorhanden zijn. Naar onze mening geeft een vergelijking tussen parallelgroepen inderdaad een beter beeld van de omvang van leerkrachteffecten dan een vergelijking tussen jaargroepen. Dat vergelijkingen tussen jaargroepen grote verschillen opleveren is waarschijnlijk slechts ten dele te wijten aan verschillen tussen leerkrachteffectiviteit.

Wat wij wel betwijfelen is of de opbouw van het curriculum de belangrijkste versturende variabele is. Gezien de uitkomsten van een studie door Luyten & Sniijders (1996) is dit niet erg waarschijnlijk. Hun analyses hebben betrekking op een groep leerlingen die gedurende de laatste twee jaar in het basisonderwijs zijn gevolgd. Iets minder dan de helft van deze leerlingen kreeg gedurende de laatste twee jaar les van dezelfde leerkracht, terwijl de andere leerlingen bij de overgang van groep 7 naar groep 8 een nieuwe leerkracht kregen. Zowel voor rekenen als voor taal bleken de verschillen tussen leerkrachten veel groter dan die tussen scholen (zie ook tabel 1). Als verschillen tussen jaargroepen het gevolg zijn van een gemeenschappelijk besluit om het curriculum op een bepaalde manier te doorlopen, dan zouden in deze onderzoeksopzet geen sterke leerkrachteffecten zijn gevonden.

Ons onderzoek is ingegeven door het vermoeden dat, wanneer het gaat om de verschillen in effectiviteit tussen leerkrachten binnen een school, de vergelijkbaarheid van de leerprestaties een grote rol speelt. Verschillen in leerkrachteffectiviteit worden vermoedelijk versterkt als het voor de leerkrachten moeilijk is om de prestaties van elkaars leerlingen te vergelijken. Bepaling van leerkrachteffecten aan de hand van verschillen tussen jaargroepen levert zodoende een vertekend beeld op. Voor een onderzoeker is het betrekkelijk gemakkelijk om vast te stellen of op een school de leerprestaties in de ene jaargroep even hoog liggen als in de andere. Er wordt dan simpelweg per school nagegaan hoe ver de prestaties in de ene jaargroep boven of beneden het gemiddelde in de steekproef liggen. Vervolgens wordt berekend in hoeverre de afwijking in de andere jaargroep daarmee overeenkomt. Voor leerkrachten is het wel erg moeilijk, zo niet onmogelijk, om op een dergelijke manier de leerprestaties in verschillende jaargroepen met elkaar te vergelijken. Eigenlijk is het niet verwonderlijk dat effectiviteit van leerkrachten binnen scholen nogal uiteenloopt, als ze van elkaar niet eens (kunnen) weten of ze vergelijkbare resultaten boeken. Voor parallelgroepen geldt echter dat de leerkrachten zelf de voortgang van elkaars leerlingen goed kunnen vergelijken. Zodoende kunnen de leerkrachten ervoor zorgen dat de leerprestaties niet te zeer gaan verschillen van klas tot klas. Ze kunnen in ieder geval voorkomen dat men in de ene klas langzamer door de stof heengaat dan in de andere klas(sen). Ook kan men ervoor zorgen dat in elke klas dezelfde normen worden gehanteerd bij de beoordeling.

In dit artikel wordt ons vermoeden over het belang van de vergelijkbaarheid van de leerprestaties op een indirecte manier getoetst. De onderlinge vergelijkbaarheid is het grootst in het geval van parallelgroepen en ze is zeer beperkt als het gaat om leerkrachten die aan verschillende jaargroepen lesgeven. In dit artikel gaat het over een vergelijking tussen leerkrachten die in verschillende jaren aan groep 8 les hebben gegeven. Deze leerkrachten beschikken evenals leerkrachten van parallelgroepen over een goede vergelijkingsbasis, namelijk de prestaties die andere leerkrachten in de school in voorgaande jaren met groep 8 hebben behaald. Voor leerkrachten die aan verschillende jaargroepen les geven ontbreekt een dergelijke vergelijkingsbasis. Onze verwachting is dan ook dat wanneer we op deze wijze de omvang van leerkrachteffecten bepalen de uitkomsten meer overeenkomst vertonen met de resultaten die zijn gevonden bij vergelijkingen tussen parallelgroepen en minder met de uitkomsten gebaseerd op vergelijkingen tussen jaargroepen. Bij vergelijkingen tussen parallelgroepen zijn relatief kleine verschillen tussen leerkrachten gevonden en relatief grote verschillen tussen scholen. We verwachten dat voor de ruwe scores de waarde van p hoger zal zijn dan .80. Een uitkomst lager dan .70 zou duidelijk in strijd zijn met wat wij verwachten. Dergelijke uitkomsten zijn immers gevonden bij vergelijkingen tussen jaargroepen. Voor gecorrigeerde scores ligt een uitkomst boven de .65 in de lijn der verwachting. Een p -waarde beneden de .50 zou duidelijk in strijd zijn met wat wij verwachten.

In een viertal eerdere studies is nagegaan in hoeverre de leerprestaties in de Nederlandse basisscholen van jaar tot jaar verschillen (Van Batenburg, 1990; Blok & Hoeksma, 1993; Van der Werf & Guldemond 1996; Doolaard, 1996 en Doolaard & Bosker, 1997). De onderzoeksuitkomsten laten een vrij hoge mate van stabiliteit zien. Alleen Blok & Hoeksma (1993) hebben het verband tussen leerkrachtwisselingen en de stabiliteit van schooleffecten onderzocht. Zij concluderen dat "de leerkracht een onbelangrijke factor lijkt" (blz. 342). Zij hebben voor drie deelttoetsen van de Eindtoets Basisonderwijs (rekenen, taal en lezen) onderzocht in hoeverre een leerkrachtwisseling van invloed is op de toetsscores. "Per toets komt dit effect neer op ongeveer één punt hogere score voor scholen waar de leerkracht van het ene op het andere jaar gewisseld is" (blz. 340). Met één punt wordt bedoeld één item dat correct is beantwoord. De reken- en taaltoetsen bestaan uit 60 items, de leestoets uit 40. Toch geeft hun analyse slechts in zeer beperkte zin het effect van verschillen tussen leerkrachten weer. Wat zij hebben onderzocht is of leerlingen een hogere score behalen als er een leerkrachtwisseling heeft plaatsgevonden. Echter, als we ervan uitgaan dat er effectieve en minder effectieve leerkrachten bestaan, dan moet een leerkrachtwisseling soms een verbetering, maar soms ook een verslechtering inhouden. Het zwakke effect dat Blok & Hoeksma hebben gevonden zou heel goed de optelsom kunnen zijn van sterk positieve en sterk negatieve effecten. In de multilevel-analyses die door ons zijn uitgevoerd is het effect van de leerkracht dan ook gemodelleerd als een random effect dat van school tot school kan variëren. In de door Blok & Hoeksma uitgevoerde analyses is leerkrachtwisseling als een fixed effect gemodelleerd.

2. ONDERZOEKSVRAAG, DATASETS EN ANALYSE

De vraag die we met behulp van de analyses willen beantwoorden luidt als volgt: Hoe groot zijn de verschillen tussen leerkrachten binnen scholen die in verschillende jaren aan groep 8 van het basisonderwijs les hebben gegeven wat betreft de prestaties van hun leerlingen en hoe verhouden zich deze verschillen tot die tussen scholen?

Onze analyses hebben betrekking op twee datasets. Ten eerste is een secundaire analyse uitgevoerd op de dataset die door Blok & Hoeksma (1993) is aangelegd voor hun studie naar de stabiliteit van schooleffecten (Dataset A). De dataset, zoals die door ons is geanalyseerd bevat 12.734 leerlingen, verdeeld over 132 scholen die in vijf opeenvolgende jaren (1987-1991) aan de Eindtoets Basisonderwijs hebben deelgenomen. Twee scholen die in het oorspronkelijke bestand waren vertegenwoordigd zijn in onze analyses niet opgenomen. Het

betreft hier scholen met parallelgroepen, waarbij niet duidelijk is welke leerlingen in welke klas hebben gezeten. Blok & Hoeksma (1993) hebben laten zien dat de scores op de Eindtoets Basisonderwijs in de steekproef voor alle vijf de jaren heel dicht in de buurt liggen bij die van de totale populatie van scholen die aan de toets hebben meegedaan. In de analyses die wij hebben uitgevoerd zijn alleen de scores op de deelttoetsen taal en rekenen meegenomen. De dataset van Blok & Hoeksma bevat verder nog gegevens die betrekking hebben op leesvaardigheid, wat weer een subset is uit de deelttoets Informatieverwerking, maar deze gegevens zijn buiten beschouwing gelaten.

De tweede dataset (Dataset B) is opgebouwd in twee delen. Het eerste deel, verzameld in 1987/88 door Brandsma (1993) is in 1995/96 door Doolaard gerepliceerd om te onderzoeken of veranderingen in school- en klaskenmerken samengaan met veranderingen in leerlingprestaties. Van de oorspronkelijke 200 scholen uit 1987/88 hebben 120 scholen deelgenomen aan het vervolg. Beide steekproeven zijn voldoende representatief, er is een lichte oververtegenwoordiging van RK scholen uit het zuiden van het land. De dataset die nu door ons geanalyseerd is bestaat uit 97 scholen en 3.500 leerlingen. Van alle leerlingen zijn gegevens bekend van een reken- en taaltoets (inclusief begrijpend lezen). Voor deze toetsen is gebruik gemaakt van gedeelten van de Eindtoets van de betreffende jaren. De taaltoets bestond in 1988 uit 57 items en in 1996 uit 35 items, de rekentoets in 1988 uit 30 items en in 1996 uit 18 items. De betrouwbaarheidscoëfficiënt varieert van .79 tot .88 (Cronbach's α).

In de 132 scholen van de eerste dataset hebben in de periode van 1987 tot 1991 232 verschillende leerkrachten lesgegeven aan groep 8. Samen hebben zij lesgegeven aan 660 "klassen" of cohorten. Voor iedere school hebben we immers vijf keer een groep 8. In de tweede dataset, waar we voor elk van de 97 scholen maar twee keer een groep 8 hebben, komen 140 verschillende leerkrachten voor (zie ook tabel 2). De cijfers in tabel 3 laten zien dat over vrij lange periodes toch vaak dezelfde leerkracht les blijft geven aan groep 8. In ruim 45% van de scholen van de eerste dataset heeft steeds dezelfde leerkracht deze groep voor zijn rekening genomen. In 36% van de scholen zijn het twee verschillende leerkrachten geweest. Een aantal van drie of vier verschillende leerkrachten komt vrij zelden voor. Voor de tweede dataset geldt dat in meer dan de helft van de scholen in 1996 nog steeds of alweer dezelfde leerkracht aan groep 8 heeft lesgegeven als in 1988. Van de 660 "klassen" in de eerste dataset hebben er 83 (12.6%) les gehad van een leerkrachtenteam (bijna altijd een duo). Als de samenstelling van zo'n team van het ene op het andere jaar is veranderd, is dit ook opgevat als een leerkrachtwisseling. Ook in dataset B komen duo-banen voor; in 1988 gold dat voor 9% van de klassen, in 1996 voor 28% van de klassen. Degene die de meeste uren maakt wordt beschouwd als de leerkracht van groep 8, als dat in 1996 een ander was dan in 1988 dan is dat opgevat als een leerkrachtwisseling.

Tabel 2: Beschrijving datasets

	Dataset A, 5 cohorten ('87 - '91)	Dataset B, 2 cohorten ('88 en '96)
Aantal scholen	132	97
Aantal leerkrachten	232	140
Aantal "klassen"	660	194
Aantal leerlingen	12,734	3,500

De eerste dataset bevat geen achtergrondgegevens over individuele leerlingen. Wel zijn op schoolniveau dergelijke gegevens voorhanden. Het betreft hier de zgn. achterstandsindex, die is gebaseerd op het aantal 11- tot en met 13-jarigen in de verschillende achterstandscategorieën in de jaren 1987 tot en met 1991. De achterstandsindex geeft voor elke school per jaar het gemiddelde OVB-leerlinggewicht weer. Voor scholen zonder leerlingen uit een achterstandscategorie is het gemiddelde gewicht gelijk aan 1.00. Voor scholen met uitsluitend

allochtone leerlingen bedraagt het gemiddelde gewicht 1.90. De gegevens zijn afkomstig van het Ministerie van Onderwijs en Wetenschappen en hebben betrekking op een grotere groep leerlingen dan alleen degenen die aan de Eindtoets Basisonderwijs hebben meegedaan. De criteria op basis waarvan de leerlinggewichten worden toegekend zijn na 1987 gewijzigd met als gevolg dat de spreiding en het gemiddelde van de index aanmerkelijk zijn afgenomen (Blok & Hoeksma, 1993; blz. 334-335). In tabel 4 zijn voor de achterstandsindex de intercorrelaties tussen de verschillende jaren weergegeven. Voordat deze correlaties zijn berekend, zijn de scores op de achterstandsindex evenals de toetsscores per jaar gestandaardiseerd door ze om te zetten in N-scores². Omdat de intercorrelaties tussen de verschillende jaren zo hoog zijn (.80 - .92) en omdat het gemiddelde leerlinggewicht niet is gebaseerd op precies dezelfde groep leerlingen als degenen die aan de Eindtoets hebben deelgenomen, is niet de score per jaar als controle-variabele in de analyses opgenomen, maar is in plaats daarvan voor elke school het gemiddelde over de hele periode van vijf jaar berekend en als controle-variabele gebruikt. Overigens komen de intercorrelaties in tabel 5 niet helemaal overeen met de waarden die Blok & Hoeksma (1993, blz. 335) rapporteren. Dit is vooral te wijten aan het feit dat het hier gaat om N-scores, terwijl de cijfers van Blok & Hoeksma betrekking hebben op de ruwe scores. Bovendien hebben wij twee scholen uit de dataset verwijderd.

Tabel 3: Aantal leerkrachten per school

Aantal	Dataset A		Dataset B	
	Frequentie	Percentage	Frequentie	Percentage
1	60	45.5	56	57.7
2	48	36.4	43	42.3
3	20	15.2	--	--
4	4	3.0	--	--

Tabel 4: Achterstandsindex - intercorrelaties 1987-1991 (Dataset A)

	1987	1988	1989	1990
1988	.80			
1989	.82	.92		
1990	.82	.90	.90	
1991	.83	.90	.90	.90

De tweede dataset bevat wel achtergrondgegevens op leerlingniveau. Het gaat hier om sociaal-economische status (SES), een performale intelligentiemeting, pre-tests voor taal en rekenen en het geslacht van de leerling. In de analyses is gecontroleerd voor de invloed van deze variabelen. De SES-score is gebaseerd op opleiding en beroep van vader en moeder, voor de intelligentiemeting is gebruik gemaakt van de performale gedeelten van de ISI-test (van Boxtel, Sijnders & Welten, 1982). In groep 7 is bij alle leerlingen een deel van de Entreetoets afgenomen als voortoets. In 1987 bestond de voortoets taal uit 47 items, in 1995 uit 39 items. De voortoets rekenen bestond in 1987 uit 20 items en in 1995 uit 35 items. De betrouwbaarheidscoëfficiënten (Cronbach's α) variëren van .68 tot .84. Daarnaast is ook het klasgemiddelde over beide pre-tests als controle-variabele meegenomen. Behalve voor de variabele geslacht geldt ook voor deze variabelen dat ze zijn omgezet in N-scores. In de analyses van beide datasets is een samengestelde score, het gemiddelde van de reken- en taalscores, gebruikt als maat voor de leerprestaties³.

De analyses moeten antwoord geven op de vraag naar de omvang van verschillen tussen leerkrachten en die tussen scholen. Aan de hand van een aantal multilevel-analyses wordt de totale variantie in de toetsscores opgesplitst in een school-, leerkracht-, cohort- en leerling-

component. In beide datasets zijn deze vier niveaus te onderscheiden. Leerlingen maken deel uit van een cohort, terwijl een leerkracht aan één of meer cohorten heeft lesgegeven. De leerkrachten zijn tenslotte weer gegroepeerd binnen scholen.

Overigens geven Blok & Hoeksma (1993) aan dat naar hun mening de hiërarchische structuur tussen scholen en cohorten (zij spreken van "afnamejaar") niet vastligt. Hun kernoverweging is dat de Eindtoets Basisonderwijs ieder jaar andere items bevat. Wij stellen ons op het standpunt dat het wel verantwoord is om het cohortniveau te beschouwen als een niveau dat in een hiërarchische structuur moet worden gezien als genest binnen scholen en leerkrachten. De Eindtoets basisonderwijs bestaat van jaar tot jaar weliswaar uit andere items, maar de stof die door de toets wordt gedekt is ieder jaar dezelfde. Bovendien is door de scores te standaardiseren bewerkstelligd dat het gemiddelde en de standaarddeviatie per jaar gelijk zijn.

De volgende multilevel-analyses worden uitgevoerd. Op de eerste dataset worden vier modellen gefit: twee modellen met drie niveaus (leerling, cohort en school) en twee modellen waarin het leerkrachtniveau als extra niveau is opgenomen. Zodoende kan worden nagegaan in hoeverre een model met leerkrachtniveau een beter fit oplevert dan een model waarin geen apart leerkrachtniveau wordt onderscheiden. Zowel voor de modellen met drie als vier niveaus wordt een model met en een model zonder controle-variabelen gefit. In het geval van de eerste dataset is de achterstandsindex de enige beschikbare controle-variabele.

Op de tweede dataset worden zes modellen gefit, drie modellen met drie niveaus en drie modellen met vier niveaus. In het eerste model met drie niveaus (model 0) is geen controle-variabele opgenomen en in het tweede model (model 1) wordt alleen gecorrigeerd voor sociaal-economisch status. Dit model komt tot op zekere hoogte overeen met het model waarin gecontroleerd wordt voor de achterstandsindex, maar een belangrijk verschil is natuurlijk wel dat in de eerste dataset uitsluitend geaggregeerde achtergrondgegevens beschikbaar zijn. In het derde model (model 2) worden alle overige controle-variabelen toegevoegd (geslacht, IQ en pre-tests). Voor de modellen met vier niveaus geldt hetzelfde.

Bij de modellen met vier niveaus is ook steeds de tussen-klassen binnen-scholen correlatie (ρ) berekend. Deze parameter geeft aan of de school- dan wel de leerkrachtvariantie meer gewicht in de schaal legt.

3. RESULTATEN

In tabel 5 worden de resultaten weergegeven van de analyses die betrekking hebben op de eerste dataset. Het is duidelijk dat het opnemen van het leerkrachtniveau in de multilevel-modellen een significant betere fit oplevert. Belangrijker nog dan de vraag of verschillen tussen leerkrachten statistisch significant zijn, is de vraag wat deze verschillen nu betekenen in termen van leerprestaties. Op het eerste gezicht lijkt een percentage van iets meer dan 2% variantie op leerkrachtniveau wellicht bescheiden. De waarden voor ρ (.88 en .80) geven in ieder geval aan dat de variantie op schoolniveau de leerkrachtvariantie in ruime mate overtreft. Bevestigen deze uitkomsten daarmee de conclusie dat "de leerkracht een onbelangrijke factor lijkt" (Blok & Hoeksma, 1993, blz. 342) of werpt onze analyse toch een nieuw licht op de zaak?

Om te kunnen beoordelen of de ruim 2% variantie die uit onze analyses naar voren komt duidt op een sterker leerkrachteffect dan het door Blok & Hoeksma gerapporteerde effect dient dit percentage te worden terugvertaald naar de originele toetsscores. Blok & Hoeksma geven immers aan dat het effect van een leerkrachtwisseling per deelttoets neerkomt op een hogere score van ongeveer één punt. Onze analyses hebben betrekking op twee deelttoetsen, namelijk taal en rekenen. De standaarddeviatie voor de som van reken- en taalscores bedraagt 19.5. Deze waarde is gebaseerd op alle toetsscores in de periode 1987-1991. Het gemiddelde over vijf jaar is 84.9 op een toets die 120 items beslaat (60 taal- en 60 rekenitems). De totale variantie bedraagt dus 380.25 en 2.2% hiervan komt neer op 8.37, wat weer correspondeert met een

standaarddeviatie van 2.89. Aan de hand van de standaarddeviatie kan bepaald worden hoe groot het verschil is in toetscores tussen de 20% meest en 20% minst effectieve leerkrachten. De verwachte score voor een leerkracht die precies in het midden van de bovenste 20% zit (dus op het negende deciel) is gelijk aan het gemiddelde plus 1.28 maal de standaarddeviatie. Voor een leerkracht uit de minst effectieve 20% geldt dat hij 1.28 maal de standaarddeviatie beneden het gemiddelde zit. Een effectieve leerkracht zit dus ruim 3.6 punten boven het gemiddelde en een leerkracht uit de minst effectieve 20% zit ruim 3.6 punten beneden het gemiddelde. Met andere woorden: het verschil tussen de meest en minst effectieve leerkrachten bedraagt ruim 7 punten. De verschillen tussen leerkrachten zoals die uit onze analyses naar voren komen zijn derhalve duidelijk groter dan de uitkomsten van Blok & Hoeksma (1993) suggereren. Daarbij moet wel worden bedacht dat het effect zoals door hen gerapporteerd (één punt) betrekking heeft op deelttoetsen, terwijl het in onze analyses gaat over de som van twee deelttoetsen (rekenen en taal). In vergelijking met verschillen op school- en leerlingniveau blijven de verschillen tussen leerkrachten bescheiden. Als we op een soortgelijke manier de verschillen tussen de meest en minst effectieve scholen berekenen, komen we uit op een verschil van ruim 15 punten, terwijl het verschil tussen de meest en minst succesvolle leerlingen binnen scholen neerkomt op bijna 45 punten.

Tabel 5: Variantie op school-, leerkracht-, cohort- and leerlingniveau - Dataset A

	DRIE NIVEAUS		VIER NIVEAUS	
	Model 0	Model 1	Model 0	Model 1
FIXED EFFECTEN				
(standard errors)				
Intercept	-.052(.038)	-.051(.031)	-.058(.038)	-.058(.031)
Achterstandsindex	---	-.255(.031)	---	-.254(.031)
RANDOM EFFECTEN:				
Variantiecomponenten				
Schoolniveau	17.1%	10.7%	15.7%	9.3%
Leerkrachtniveau	---	---	2.2%	2.3%
Cohortniveau	2.6%	2.6%	1.8%	1.8%
Leerlingniveau	80.3%	80.3%	80.3%	80.3%
Verklaarde variantie	---	6.3%	---	6.4%
ρ	---	---	.88	.80
SIGNIFICANTIE VAN LEERKRACHTNIVEAU				
χ^2 (df = 1)	---	---	10.40	10.10
p-waarde	---	---	.001	.001

132 scholen; 232 leerkrachten; 5 "klassen" of cohorten per school ('87 - '91); 12,734 leerlingen

Tabel 6 laat de uitkomsten zien met betrekking tot de tweede dataset. De variantie op schoolniveau in model 0, met drie niveaus, verschilt nauwelijks tussen beide datasets (17.1% en 17.9%). De variantie op cohortniveau is echter duidelijk groter in de tweede dataset (7.3% versus 2.6%). Waarschijnlijk is dit te wijten aan het feit dat de tweede dataset betrekking heeft op twee cohorten met een lange periode tussen de beide afnamejaren. Wat verder opvalt is

dat de variantie op cohortniveau niet afneemt als gecontroleerd wordt voor sociaal-economische status (model 1). Ze wordt dan zelfs wat groter. Kennelijk kunnen de verschillen tussen cohorten in sociaal-economische status de variantie in leerprestaties op cohortniveau niet verklaren. Vergelijkbare resultaten zijn gerapporteerd door Goldstein (1987) en Willms & Raudenbush (1989). Als mogelijke oorzaak wordt genoemd dat leerkrachten bij veranderingen in de achtergrond van de leerlingpopulatie toch proberen om dezelfde prestaties als voorheen te behalen. Volgens deze redenering wordt het streefniveau dus niet (meteen) aangepast aan veranderingen in de achtergrond van de leerlingen. Wanneer in model 2 de overige controlevariabelen worden toegevoegd, wordt uiteindelijk ruim 65% van de totale variantie verklaard. De effecten van de individuele scores op de beide pre-tests zijn het sterkst. De afname van variantie is relatief het grootst op het schoolniveau en het kleinst op het cohortniveau. In model 2 is de variantie op cohortniveau zelfs groter dan die op schoolniveau (5.7% versus 3.7%). Wat betreft de leerprestaties in groep 8 zijn na controle voor de scores op de pre-tests dus relatief grote verschillen waar te nemen tussen cohorten binnen een school.

Tabel 6: Variantie op school-, leerkracht-, cohort- and leerlingniveau - Dataset B

	DRIE NIVEAUS			VIER NIVEAUS		
	Model 0	Model 1	Model 2	Model 0	Model 1	Model 2
FIXED EFFECTEN (standard errors)						
Intercept	-.053(.050)	-.044(.047)	-.053(.029)	-.053(.050)	-.042(.047)	-.051(.029)
SES	---	.379(.017)	.107(.011)	---	.379(.017)	.107(.011)
Meisjes	---	---	.071(.018)	---	---	.071(.018)
IQ-score	---	---	.151(.011)	---	---	.150(.011)
Pre-test taal	---	---	.384(.012)	---	---	.384(.012)
Pre-test rekenen	---	---	.346(.013)	---	---	.346(.013)
Klasgemiddelde beide pre-tests	---	---	.052(.024)	---	---	.053(.024)
RANDOM EFFECTEN: Variantiecomponenten						
Schoolniveau	17.9%	14.5%	3.7%	17.9%	13.5%	2.5%
Leerkrachtniveau	---	---	---	0.0%	2.3%	2.4%
Cohortniveau	7.3%	8.4%	5.7%	7.3%	7.3%	4.4%
Leerlingniveau	74.8%	65.0%	25.6%	74.8%	65.0%	25.6%
Verklaarde variantie	---	12.1%	65.1%	---	12.0%	65.0%
ρ	---	---	---	1.00	.86	.51
SIGNIFICANTIE VAN LEERKRACHTNIVEAU						
χ^2 (df = 1)	---	---	---	0.00	0.42	1.64
p-waarde	---	---	---	1.000	.517	.200

97 scholen; 140 leerkrachten; 2 "klassen" of cohorten per school ('88 en '96); 3,500 leerlingen

Als we de modellen met vier niveaus vergelijken met de drie-niveau-modellen, dan blijkt dat in de tweede dataset het opnemen van een leerkrachtniveau statistisch gezien veel minder overtuigende uitkomsten geeft dan in de eerste dataset. Voor model 0 kan helemaal geen variantie op leerkrachtniveau worden vastgesteld. Het meest significante resultaat wordt gevonden bij model 2. De overschrijdingskans (p-waarde) bedraagt in dat geval echter .20. Als we de leerkrachtgegevens in de dataset zouden vervangen door toevalsgetallen, dan zouden we dus in 20% van de gevallen een leerkrachtvariantie van minstens 2.4% vinden. De waarden voor ρ bedragen in de modellen 0 t/m 2 respectievelijk 1.00, .86 en .51. Deze laatste is groter dan de waarden die zijn gevonden bij vergelijkingen tussen jaargroepen, maar kleiner dan de uitkomsten met betrekking tot parallelklassen (zie tabel 1).

Dat in de tweede dataset geen significante leerkrachteffecten zijn gevonden is in de eerste plaats te wijten aan hun geringe omvang. Daarnaast is het geringe aantal leerkrachten per school en het geringe aantal cohorten per leerkracht van belang. Er is altijd een wisselwerking tussen steekproefomvang en het onderscheidingsvermogen van een toets. Als we de beschikking hadden gehad over een zeer grote dataset met bijvoorbeeld 20 cohorten in plaats van 2, dan waren zelfs de meest minuscule leerkrachteffecten statistisch significant geweest. In de tweede dataset zijn de standard errors voor de variantiecomponenten echter vrij hoog. De geschatte componenten moeten dan ook niet te klein worden om nog significant te zijn. In tabel 7 worden de standard errors vermeld bij de variantiecomponenten voor het meest uitgebreide model van zowel de eerste als de tweede dataset. Daarnaast zijn ook betrouwbaarheidsintervallen (95%) berekend. Het is duidelijk dat de standard error voor de variantiecomponent op leerkrachtniveau in de tweede dataset aanmerkelijk groter is dan in de eerste dataset. Gezien de standard error van 1.9 zou de variantie op leerkrachtniveau in de tweede dataset minstens 3.2% moeten bedragen om significant te zijn voor $\alpha < .05$ bij een eenzijdige toetsing. Een leerkrachtvariantie van meer dan 3.2% zou bij een schoolvariantie van 2.5%, zoals gevonden in de multilevel-analyses (zie tabel 6), een opleveren van minder dan .44. Een dergelijk resultaat zou in strijd zijn met het vermoeden dat de aanleiding vormt voor deze studie. Dat vermoeden is immers dat een vergelijking tussen leerkrachten die in verschillende jaren aan groep 8 les hebben gegeven een uitkomst oplevert die lijkt op de resultaten die zijn gevonden bij vergelijkingen tussen parallelgroepen. Een ρ lager dan .44 past heel goed bij de resultaten gebaseerd op vergelijkingen tussen jaargroepen, maar niet bij de resultaten die zijn gevonden bij vergelijkingen tussen parallelgroepen (zie tabel 1). Als ons vermoeden onjuist was, zou dit waarschijnlijk zijn gebleken uit analyses van de tweede dataset. Dat de verschillen tussen leerkrachten zo klein zijn dat ze statistisch niet significant zijn is eerder gevonden in onderzoek bij parallelgroepen (Bosker, 1991).

Tabel 7: Variantie op school-, leerkracht-, cohort- and leerlingniveau - Dataset A

	Schoolniveau	Leerkrachtniveau	Cohortniveau	Leerlingniveau
Dataset A, model 1				
vier niveaus				
Puntschatting	9.3%	2.3%	1.8%	80.3%
Standard error	1.6%	0.8%	0.4%	1.0%
Betrouwbaarheids-interval (95%)	6.2% - 12.4%	0.7% - 3.8%	0.9% - 2.6%	78.3% - 82.3%
Dataset B, model 2				
vier niveaus				
Puntschatting	2.5%	2.4%	4.4%	25.6%
Standard error	1.6%	1.9%	1.1%	0.6%
Betrouwbaarheids-interval (95%)	0.0% - 5.7%	0.0% - 6.2%	2.2% - 6.6%	24.4% - 26.9%

Wat verder opvalt in tabel 7 is dat zelfs de variantie op schoolniveau niet significant afwijkt van 0. Hieruit moet niet geconcludeerd worden dat er geen significante verschillen tussen scholen zijn. In het model met drie niveaus, zonder leerkrachtniveau, is de schoolvariantie namelijk wel significant. Het betekent echter wel dat in de tweede dataset voor model 2, waar gecorrigeerd is voor meerdere achtergrondkenmerken, een precieze schatting van school- en leerkrachtvariantie problematisch wordt gezien de beperkte omvang van beide effecten. Het is duidelijk dat een dataset met meer dan twee leerlingcohorten vereist is om op betrouw-

baarheidsinterval voor de leerkrachtvariantie uit te komen dat zo smal is dat het de waarde 0 niet omvat. Overigens blijkt uit tabel 7 dat ook in de eerste dataset, die vijf cohorten omvat, het betrouwbaarheidsinterval voor het leerkrachteffect dicht in de buurt komt van 0%. Zelfs een zeer omvangrijke dataset biedt dus bepaald geen garantie dat het 95%-betrouwbaarheidsinterval volledig rechts van 0% komt te liggen.

4. DISCUSSIE

Dit artikel is ingegeven door het vermoeden dat de verschillen in effectiviteit tussen leerkrachten binnen scholen niet per se groter zijn dan de verschillen tussen scholen. Het gaat er maar om welke leerkrachten met elkaar worden vergeleken. Als men leerkrachten met elkaar vergelijkt voor wie het zelf ook goed mogelijk is om de leerprestaties van elkaars leerlingen te vergelijken, dan zouden de verschillen in effectiviteit veel kleiner kunnen zijn. Dit zou verklaren waarom bij vergelijkingen tussen parallelgroepen betrekkelijke kleine verschillen zijn gevonden. Om ons vermoeden te toetsen hebben wij een situatie onder de loep genomen die lijkt op de situatie van leerkrachten die aan parallelklassen lesgeven. Het betreft een vergelijking tussen leerkrachten die in verschillende jaren aan groep 8 les hebben gegeven.

De verwachting dat deze aanpak een beeld zou opleveren dat vooral overeenkomt met de resultaten die zijn gevonden bij vergelijkingen tussen parallelklassen wordt door de uitkomsten ondersteund. De verhouding tussen school- en leerkrachtvariantie zoals die wordt uitgedrukt door de tussen-klassen binnen-scholen correlatie (ρ) past beter bij de resultaten die gevonden zijn bij vergelijkingen tussen parallelklassen dan bij de resultaten gebaseerd op vergelijkingen tussen jaargroepen. Voor de ruwe scores hebben wij waarden voor ρ gevonden van .88 en 1.00. Vergelijkingen tussen parallelklassen in eerder onderzoek hebben waarden voor opgeleverd van .81 en 1.00, terwijl bij vergelijkingen tussen jaargroepen de waarden uiteenliepen van .44 tot .69 (zie ook tabel 1). Ook de resultaten voor gecorrigeerde scores vertonen meer overeenkomst met eerdere resultaten die zijn gebaseerd op vergelijkingen tussen parallelgroepen dan met uitkomsten die betrekking hebben op verschillen tussen jaargroepen. In onze analyses zijn de volgende waarden voor ρ gevonden: .80, .86 en .51. In eerdere vergelijkingen tussen parallelgroepen varieerde ρ van .67 tot .96, terwijl bij vergelijkingen tussen jaargroepen de waarden uiteenliepen van .00 tot .48. Eerdere vergelijkingen tussen parallelklassen hebben ook laten zien dat de verschillen tussen leerkrachten zo klein kunnen zijn dat ze statistisch niet significant zijn (Bosker, 1991). Onze onderzoeksuitkomsten ondersteunen dus het vermoeden dat de verschillen tussen leerkrachten binnen scholen klein zijn als de leerkrachten zelf de prestaties van elkaars leerlingen goed kunnen vergelijken.

Toekomstig onderzoek zal echter moeten uitwijzen in hoeverre ons vermoeden nader bevestigd wordt. In de eerste plaats zouden de analyses waarover wij hier hebben gerapporteerd gerepliceerd kunnen worden op een dataset die minstens vijf cohorten omvat en waarin tevens de nodige achtergrondgegevens over de leerlingen voorhanden zijn. De bestanden die zijn aangelegd ten behoeve van de evaluatie van het onderwijsvoorrrangsbeleid lijken daarvoor geschikt. Deze bestanden bevatten bovendien niet alleen informatie over de leerprestaties in groep 8, maar ook in de groepen 4 en 6. Het is nog maar de vraag of de leerkrachteffecten in deze twee groepen even klein zijn als in groep 8. Groep 8 vormt immers de afsluiting van de basisschool en juist dan worden er op school bepaalde standaarden nagestreefd wat betreft het te bereiken niveau. Daarnaast zou onderzocht moeten worden wanneer leerkrachten die aan parallelgroepen lesgeven juist heel erg uiteenlopende resultaten bereiken en wanneer leerkrachten die aan verschillende jaargroepen lesgeven juist zeer gelijksoortige resultaten behalen. Naar onze mening kan daarbij niet worden volstaan met de in effectiviteitsonderzoek nogal populaire aanpak om alle variabelen zo kwantitatief mogelijk te meten bij voorkeur aan de hand van schriftelijke vragenformulieren met voorgestructureerde antwoordcategorieën. Juist als het gaat om samenwerking tussen twee of meer personen zijn het al gauw details,

finesses en timing die de doorslag geven, kortom; het gaat om aspecten die zich slecht lenen voor kwantitatief onderzoek. De afhankelijke variabelen in effectiviteitsonderzoek, de leerprestaties, zijn doorgaans wel goed te operationaliseren als numerieke variabelen, voor de verklarende variabelen ligt dit vaak veel moeilijker.

Naar onze mening is de feedback die een leerkracht krijgt over de resultaten van het onderwijs dat hij geeft van groot belang voor het niveau dat hij met zijn leerlingen weet te bereiken. In sommige opzichten is er sprake van een "natuurlijke" feedback, zoals in het geval van parallelgroepen. Ook in de situatie die in dit onderzoek is geanalyseerd is hiervan sprake. De resultaten in voorgaande jaren zullen voor een (nieuwe) leerkracht al gauw als een richtsnoer fungeren. Dit lijkt met name waarschijnlijk als scholen meedoen aan de Eindtoets Basisonderwijs. De scholen waar deze toets wordt afgenomen krijgen elk jaar een duidelijk beeld van het prestatieniveau van hun leerlingen in vergelijking met het niveau in andere scholen. Maar ook scholen die niet aan de Eindtoets deelnemen zullen bepaalde standaarden hanteren met betrekking tot de kennis en vaardigheden waarmee hun leerlingen na acht jaar basisonderwijs moeten zijn toegerust. In andere situaties is er geen sprake van natuurlijke feedback. Het is echter wel mogelijk om leerkrachten informatie te verschaffen die aangeeft hoe zij het doen in vergelijking met hun collega's die lesgeven aan andere jaargroepen en/of andere vakken onderwijs. Een leerkracht die lesgeeft aan groep 6 kan zodoende te weten komen hoe ver zijn leerlingen boven of onder het nationaal gemiddelde presteren en vooral of de andere jaargroepen in zijn school even ver boven of beneden dat gemiddelde zitten.

NOTEN

1. De auteurs willen Henk Blok (Universiteit van Amsterdam, SCO-Kohnstamm Instituut voor Onderzoek van Opvoeding en Onderwijs) bedanken voor het beschikbaar stellen van de een van twee bestanden die in het kader van deze studie zijn geanalyseerd.

2. Dit betekent dat op basis van de rangorde van de originele scores aan elke score de verwachte waarde is toegekend binnen de standaardnormale verdeling. Als we bijvoorbeeld 20 scores zouden hebben, dan krijgt de hoogste score de waarde 1.96 en de laagste de waarde -1.96. Het gemiddelde van een verdeling van N-scores is altijd gelijk aan 0, terwijl de standaarddeviatie 1 bedraagt. Het voordeel van deze transformatie boven een transformatie naar Z-scores is dat op deze manier scheve verdelingen meer symmetrisch worden gemaakt, zodat een verdeling van N-scores doorgaans een goede benadering oplevert van de standaardnormale verdeling. Als de scores op de afhankelijke variabele standaardnormaal verdeeld zijn en de steekproef zeer groot is, is het verantwoord om op basis van de standaard errors die het multilevel-programma MLN berekent een betrouwbaarheidsinterval op te stellen voor de variantiecomponenten (Goldstein, 1995).

3. Afzonderlijke analyses voor taal-, rekenen- en leesprestaties zijn eveneens uitgevoerd, maar hierover zal niet worden gerapporteerd. Wij volstaan met de mededeling dat de afzonderlijke analyses soortgelijke, maar wat minder betrouwbare resultaten opleveren dan de analyses met een samengestelde prestatiemaat als afhankelijke variabele.

4. De berekening zoals die hier wordt weergegeven is gebaseerd op uitkomsten van analyses op de ruwe scores, waarover hier verder niet wordt gerapporteerd. De variantiecomponenten die op die manier zijn verkregen wijken echter niet noemenswaardig af van de componenten die het resultaat zijn van de analyses op de N-scores.

LITERATUUR

- Batenburg, Th.A. van (1990). Variatie in schoolgemiddelden op de Cito-eindtoets basis-onderwijs. *Tijdschrift voor Onderwijsresearch*, 15, 362-369.
- Blok, H. & Hoeksma, J.B. (1993). De stabiliteit van het schooleffect in de tijd: een analyse op basis van vijf jaar Eindtoets van het Cito. *Tijdschrift voor Onderwijsresearch*, 18, 331-342.

- Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16, 206-218.
- Bosker, R.J., Lam, J.F., Dekkers, H. & Vierke, H. (1997). *De betekenis van kwaliteits-verschillen tussen scholen*. Enschede/Nijmegen: Vakgroep Onderwijsorganisatie & -management Faculteit der Toegepaste Onderwijskunde Universiteit Twente/ Instituut voor Toegepaste Sociale Wetenschappen Katholieke Universiteit Nijmegen.
- Boxtel, H.W. van, Snijders, J. Th. & Welten, V.J. (1982). *ISI: Interesse, Schoolvorderingen, Intelligentie*. Publicatie 7. Vorm III. Groningen: Wolters-Noordhoff.
- Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs*. Groningen: RION.
- Doolaard, S. (1996). *Changes in characteristics and effects of school leadership over time*. Paperpresentatie ECER, Sevilla.
- Doolaard, S. & Bosker, R.J. (1997). *Changes in schools and in their effects; a ten-year longitudinal study*. Paperpresentatie ICSEI, Memphis.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Charles Griffin & Co.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Luyten, H. & Snijders, T.A.B. (1996). School effects and teacher effects in Dutch elementary education. *Educational Research and Evaluation*, 2, 1-24.
- Mandeville, G.K. & Anderson, L.W. (1987). The stability of school effectiveness indices across grade levels and subject areas. *Journal of Educational Measurement*, 24, 203-216.
- Werf, M.P.C. van der & Guldemond, H. (1996). *Omvang, stabiliteit en consistentie van schooleffecten in het basisonderwijs*. Groningen: GION.
- Willms, J.D. & Raudenbush, S.W. (1989) A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 209-232.

Pedagogisch klimaat in het basisonderwijs

Wim J.C.M. van de Grift¹
Hennie P. Brandsma²

SUMMARY

This article focuses on the construction of an instrument to measure the pedagogical quality of school climate. After conceptual deliberations, and constructing consequently the domain matrix of a total of 211 indicators for the measurement of pedagogical climate are discerned. Applying Item Response Theory (Rasch scaling) a selection of this intempool is used to construct a short instrument that can be put to use in the work of the inspectorate. In the psychometric analyses specific attention is paid to construct validity, calibration, homogeneity, unidimensionality, local stochastic independency, parallel item curves, and interrater reliability. The instrument constructed appears to be reliable and valid. Furthermore the cut-off score to determine a threshold below which climate has to be evaluated as of insufficient quality is discussed, and thereafter determined. Applying this criterion leads to the conclusion that in 31% of the primary schools pedagogical climate lacks sufficient quality.

1. INLEIDING

In het schooljaar 1993/1994 is door de Inspectie van het Onderwijs in samenwerking met het OCTO van de Universiteit Twente het pedagogisch klimaat in het basisonderwijs geëvalueerd (Inspectie van het Onderwijs, 1994). Ten behoeve van deze evaluatie is een uitgebreid instrumentarium geconstrueerd, dat geschikt is voor gebruik in de schoolbezoeken, die door de inspectie uitgevoerd worden (Brandsma & Bos, 1994). Op basis van de resultaten van dit evaluatieproject is door de inspectie een kort instrumentarium geconstrueerd (Van de Grift, 1994). Het verslag van het psychometrisch onderzoek naar de kwaliteit van dit beknopte instrumentarium staat centraal in deze publikatie.

2. HET EVALUEREN VAN: 'DE PEDAGOGISCHE KWALITEIT VAN HET KLIMAAT'

In deze paragraaf worden de conceptuele achtergronden van het instrument voor het evalueren van de pedagogische kwaliteit van het klimaat in de leerlinggroepen van het basisonderwijs behandeld.

2.1 Theoretische achtergronden

Aan het klimaat in de leerlinggroepen van het basisonderwijs kan als eis gesteld worden dat het van voldoende pedagogische kwaliteit is. Met de pedagogische kwaliteit van het klimaat wordt de mate bedoeld waarin de onderwijsgevende een situatie creëert die geschikt is voor het leren van de leerlingen. Hierbij kan gedacht worden aan het creëren van een veilige en positieve werksfeer, waarbinnen de inzet en inbreng van leerlingen gewaardeerd en gestimuleerd wordt. De evaluatiestandaard die voor het evalueren van de pedagogische kwaliteit van

het klimaat in de leerlinggroepen van het basisonderwijs is gebruikt, luidt als volgt:

Het klimaat in de leerlinggroepen van het basisonderwijs is van voldoende pedagogische kwaliteit.

Deze standaard is uitgewerkt in een dimensie, die net als een meetlat een aantal gradaties kent en waarop de cesuur tussen voldoende en onvoldoende is aangegeven. In deze paragraaf wordt de dimensie van de standaard behandeld. De bepaling van de cesuur tussen voldoende en onvoldoende (de norm) komt in paragraaf 5.1 aan de orde.

Het pedagogisch klimaat moet onderscheiden worden van de pedagogische doelen die de school beoogt. Bij dit laatste kan bijvoorbeeld gedacht worden aan de vorming van de harmonisch geïntegreerde en gereguleerde persoonlijkheid (Waterink, 1951), de bekwaming tot zedelijke zelfbepaling in een morele orde die mensen zelf de verantwoordelijkheid laat (Langeveld, 1971), of de uitgroei van het kind tot verantwoordelijke volwassen persoon (Perquin, 1964). Het is uitdrukkelijk niet de bedoeling om met het onderhavige instrumentarium te treden in deze thematiek van de pedagogische doelen. Hierin kunnen en mogen scholen verschillen. In deze publikatie gaat het om het klimaat, de sfeer waarin het onderwijzen en opvoeden vorm krijgt.

In de onderzoeksliteratuur worden diverse en vooral verschillende termen gehanteerd als het gaat om het klimaat. Er worden onderscheidingen gemaakt in school- en klasseklimaat (Fraser, 1986ab). Anderen prefereren termen als sociaal klimaat (Moos, 1979) of onderwijsleerklimateit (Houtveen e.a., 1993). Soms gaat dit gepaard met verschillen in visie, vaker is er sprake van een meer beperkte dan wel een bredere scope van het betreffende onderzoeksproject. In sommige onderzoeken is het klimaat in school en klas een doel op zich (zie bijvoorbeeld Fraser, 1986AB), in andere onderzoeken (bijvoorbeeld Brookover e.a., 1979, Purkey & Smith, 1983) staat het klimaat in dienst van een zo effectief mogelijke school.

Het pedagogisch klimaat verwijst naar de sfeer van de pedagogische situatie op school met betrekking tot de opvang, verzorging, omgang en ontspanning van de leerlingen. Het gaat daarbij om kenmerken van een sfeer, waarin het kind zich thuis voelt, een sfeer van acceptatie, belangstelling en tolerantie die van invloed is op het sociaal emotioneel welbevinden op school.

Diverse auteurs hebben gepoogd structuur aan te brengen in de diverse domeinen en aspecten van het klimaat. Moos (1979) gebruikt de term sociaal klimaat, omdat zijn model ook van toepassing is op andere organisaties dan scholen. Hij werkt met de categorieën: onderlinge relaties, persoonlijke ontwikkeling en systeemkenmerken. Arter (1987), die uitgaat van het begrip schoolklimaat breidt dit systeem van Moos uit met een vierde categorie: fysieke omgeving. Brandsma en Bos (1993) hebben de diverse domeinen en aspecten van het pedagogisch school- en klasseklimaat in kaart gebracht. (Zie tabel 1.)

Deze publikatie is uitsluitend gericht op het pedagogisch handelen in de klas. Het klimaat in de ene leerlinggroep van het basisonderwijs heeft meer pedagogische kwaliteit dan het klimaat in de andere leerlinggroep. Deze verschillen komen tot uitdrukking in een dimensie die de pedagogische kwaliteit van het klimaat genoemd wordt. De zinsnede 'de pedagogische kwaliteit van het klimaat' is echter een hoog inferentieel abstractum, dat zelf niet direct observeerbaar is. Het klimaat in een leerlinggroep is een resultante van zowel het gedrag van de leerlingen als van de onderwijsgevende. Juist vanwege de nadruk op het bijvoeglijk naamwoord 'pedagogisch' komt het accent te liggen op het handelen van de onderwijsgevende: Hij of zij is immers verantwoordelijk voor de pedagogische kwaliteit van het klimaat. Daarom wordt ervoor gekozen om de pedagogische kwaliteit van het klimaat af te meten aan het handelen van de onderwijsgevende. Dit gedrag van de onderwijsgevende kan geformuleerd worden in indicatoren die direct observeerbaar zijn. Aan de hand van deze indicatoren kan het klimaat niet alleen gekend, maar ook geëvalueerd worden. Hoe hoger de pedagogische kwaliteit van het klimaat is, hoe meer indicatoren er zichtbaar worden. Sommige indicatoren, zoals bijvoorbeeld het stimuleren van de inzet van de leerlingen, hebben een hogere indicatiewaarde dan andere, zoals bijvoorbeeld het ontspannen zijn van de onderwijsgevende ten opzichte van

de leerlingen. Dit betekent dat niet alleen het aantal indicatoren indicatief is voor de pedagogische kwaliteit van het klimaat in de leerlinggroep, maar ook de aan- en afwezigheid van symptomen met een hoge indicatiewaarde.

Tabel 1: Domeinmatrijs pedagogisch school- en klasseklimaat

domein	aspecten	indicatoren
pedagogische visie en doelstellingen	houdingen van individuele leerlingen die de school bezoekt op het gebied van de cognitieve basis	7
	houdingen van individuele leerlingen die de school bezoekt op het gebied van de sociale basis (verantwoordelijkheid tonen)	5
	houdingen van individuele leerlingen die de school bezoekt op het gebied van de sociale basis (sociaal functioneren)	5
	houdingen van individuele leerlingen die de school bezoekt op het gebied van de sociale basis (normen en waarden)	4
	houdingen van individuele leerlingen die de school bezoekt op het gebied van de emotionele basis	3
kenmerken van de school als systeem	afspraken over pedagogisch handelen	13
	aandacht voor gedragsregels	5
	omgaan met probleemleerlingen	14
	aandacht voor allochtone leerlingen	8
	kennis en vaardigheden van leerkrachten	15
	lesmateriaal en werkvormen	12
	evaluatie pedagogisch handelen	12
	rol van de schoolleiding	13
	relatie school met ouders	17
relatie school met externe instanties	7	
fysieke omgeving	schoolplein en gangen	18
	overblijfruimte, documentatieruimte en personeelskamer	6
	de inrichting van de onderwijsruimte	9
pedagogisch handelen in de klas	creëren van een ontspannen sfeer in de klas	8
	belangstelling hebben voor alle leerlingen	3
	positieve feedback geven aan de leerlingen	5
	regels en ordehandhaving	4
	structuur bieden	5
	aan leerlingen ruimte bieden om zelfontdekkend bezig te zijn	5
	voorbeeldgedrag van leerkracht naar leerlingen	4
reacties van leerkracht op interacties tussen leerlingen	4	
totaal aantal indicatoren		211

De pedagogische kwaliteit van het klimaat in de leerlinggroep is een dimensie die loopt van eenvoudige kenmerken waaraan een klimaat al snel voldoet, tot kenmerken die een groot appèl doen op de pedagogische vaardigheden van de onderwijsgevende. In de dimensie 'pedagogisch klimaat' kan een hiërarchie van vier stadia onderscheiden worden: Een minimaal ver-eiste voor een goed pedagogisch klimaat bestaat in een ontspannen houding van de onderwijsgevende die gekenmerkt wordt door respect en tolerantie voor alle leerlingen. Dit eerste stadium, een respectvolle, tolerante en ontspannen houding is echter nog maar het begin. Een goed pedagogisch klimaat wordt ook gekenmerkt door een duidelijke, veilige en positieve werksfeer waar alle leerlingen bij betrokken zijn. Dit vormt het tweede stadium. Het derde stadium van het pedagogisch klimaat betreft het waarderen van de inzet en eigen inbreng

van de leerlingen. De kroon op het pedagogisch klimaat, het vierde stadium, bestaat in het aanbieden van een rijke en stimulerende leeromgeving. Kortom bij het pedagogisch klimaat begint de dimensie bij relatief eenvoudig te realiseren leerkrachtgedrag: respect tonen voor alle leerlingen en de dimensie eindigt bij het realiseren van zeer complex leerkrachtgedrag: het stimuleren van leerlingen tot het zelf zoeken van oplossingen. Hoe beter het pedagogisch klimaat is, hoe hoger de onderwijsgevende op deze dimensie gewaardeerd wordt.

In globale zin is nu helder geworden wat er bedoeld wordt met de pedagogische kwaliteit van het klimaat in de leerlinggroepen van basisonderwijs. De volgende stap is nu het selecteren van zichtbare aspecten van het pedagogisch klimaat, die gebruikt kunnen worden als indicator voor dat klimaat. Daarover gaat de volgende paragraaf.

2.2 De selectie van indicatoren

In dit project is gewerkt met een groep constructeurs, die bestaat uit vier onderzoekers van de Universiteit Twente, negen inspecteurs basisonderwijs, een stafinspecteur en een speciaal voor dit werk aangetrokken gespreksleider. Door de groep constructeurs is in een aantal fasen diverse taken vervuld: In de eerste fase zijn in een open 'brainstorm' de belangrijkste vraag- en observatiepunten voor het pedagogisch klimaat geïnventariseerd. In de tweede fase hebben de onderzoekers van de Universiteit Twente een literatuurverzameling aangelegd. Zij hebben ten behoeve van het aanleggen van deze literatuurverzameling enkele deskundigen op het gebied van school- en klasseklimaat, prof. dr B. Fraser en dr P. van der Sijde, geraadpleegd. De literatuurverzameling is gecompleteerd met behulp van de resultaten van searches op de automatisch toegankelijke literatuurbestanden ADION, British Education Thesaurus, ERIC en Psychological Index Terms. Daarnaast is gebruik gemaakt van de Social Science Citation en literatuuroverzichten van Gottfredson, e.a. (1986) en Arter (1987). In de derde fase hebben de inspecteurs uit de groep constructeurs gefungeerd als een besluitvormend filter voor de observeerbaarheid en representativiteit van de indicatoren, die door de onderzoekers van de Universiteit Twente zijn geformuleerd. Op deze wijze is een pool van 211 indicatoren aangelegd, die in een domeinmatrijs (zie tabel 1) is ingedeeld. Van deze 211 indicatoren hebben er 24 betrekking op de pedagogische visie en doelstellingen zoals die in schooldocumenten zoals het schoolplan zijn neergelegd. Voor de fysieke omgeving zijn er 33 indicatoren. Voor het pedagogisch handelen zoals dat in de leerlinggroepen van het basisonderwijs vorm krijgt, zijn er 38 indicatoren. Er zijn 116 indicatoren die kenmerken vormen van de school als systeem. Het onderhavige project is uitsluitend gericht op het pedagogisch klimaat in de leerlinggroepen van het basisonderwijs.

Er zijn in de domeinmatrijs 38 indicatoren die betrekking hebben op het pedagogisch handelen in de klas. Het gaat hierbij om de volgende aspecten van het pedagogisch handelen van de onderwijsgevers: het creëren van een ontspannen sfeer in de klas, belangstelling hebben voor alle leerlingen, positieve feedback geven aan de leerlingen, regels en ordehandhaving, structuur bieden, aan leerlingen ruimte bieden om zelfontdekkend bezig te zijn, voorbeeldgedrag van leerkracht naar leerlingen en reacties van de leerkracht op interacties tussen leerlingen. Deze aspecten staan centraal in deze rapportage.

Het doel van het onderhavige project is het construeren van een instrumentarium dat geschikt is voor observatie in de groepen van het basisonderwijs. Echter lang niet alle indicatoren uit de domeinmatrijs zijn ook daadwerkelijk geschikt voor observatie in de groepen. Voor sommige indicatoren geldt dat daarvoor schooldocumenten, zoals het schoolplan geraadpleegd moeten worden. Andere indicatoren zijn meer gevoelig voor interview of enquête dan voor directe observatie. Aan de hand van de onderzoeksuitkomsten waarover Brandsma en Bos (1994) rapporteerden kan vastgesteld worden in welke mate deze 38 indicatoren gevoelig zijn voor observatie. Er blijken vijftien indicatoren te zijn waarvoor geldt dat ze bij gebruik in observatie méér dan 20% missende waarden 'oogsten'. Dit is een aanwijzing dat deze indicatoren niet of minder goed vatbaar zijn voor observatie. Deze vijftien indicatoren zijn daarom uit de verzameling verwijderd. De verwijderde indicatoren zijn niet gelijkelijk verdeeld

over de negen genoemde aspecten van het pedagogisch handelen in de klas. Voor twee aspecten ('regels en ordehandhaving' en 'reacties van de leerkracht op interacties van de leerlingen') blijkt te gelden dat alle indicatoren (veel) meer dan 20% missende waarden hebben. Deze aspecten maken daarom geen deel uit van het nieuwe instrument.

3. DE STEEKPROEF

De evaluatie heeft plaatsgevonden bij een aselekt getrokken steekproef van 120 basisscholen. Hiervan zijn 118 basisscholen door de inspectie bezocht. Twee scholen zijn door overmacht niet bezocht. Dit is een respons van ruim 98%. Met een steekproefgrootte van 118 scholen wordt op schoolniveau een nauwkeurigheid bereikt van 9% voor de schattingen van proporties in de populatie basisscholen waarnaar gegeneraliseerd moet worden. Dit is voor het doel van deze evaluatie door de Inspectie van het Onderwijs, het bieden van een beeld van de toestand van het pedagogisch klimaat in het basisonderwijs, ruim voldoende.³

De inspecteurs hebben met het oog op een evenwichtige spreiding in elke school zowel kleuter-, onder-, midden- als bovenbouwgroepen bezocht. De scholen konden hun normale planning volgen. De reguliere gang van zaken op school werd zo weinig mogelijk door dit schoolbezoek verstoord. Het aantal bezochte lessen is gelijkelijk verdeeld over de kennisvakken, de instrumentele vakken en de expressievakken. In de kleutergroepen heeft het schoolbezoek plaats gevonden tijdens de werklessen. In totaal is het pedagogisch klimaat van 472 leerlinggroepen geëvalueerd.

De gegevens die in deze steekproef verzameld zijn, kunnen representatief geacht kunnen worden voor uiteenlopende leer- en vormingsgebieden en voor de verschillende leeftijdsgroepen van leerlingen die in de populatie van basisscholen aangetroffen worden.

4. RESULTATEN PSYCHOMETRISCH ONDERZOEK

De resultaten van het psychometrisch onderzoek worden hieronder besproken. Achtereenvolgens komen aan de orde:

- de inhoudsvaliditeit van het instrument,
- de calibratie van de indicatoren,
- de homogeniteit van het instrument,
- de unidimensionaliteit van het instrument,
- de lokale stochastische onafhankelijkheid van de indicatoren,
- de paralleliteit van de indicator karakteristieke curves,
- de interbeoordelaarsbetrouwbaarheid van het instrument.

4.1 De inhoudsvaliditeit

De vraag naar de inhoudsvaliditeit is hier opgevat als de vraag naar de mate waarin de indicatoren van het instrument een goede dekking vormen van het geheel van operationele karakteristieken van het pedagogisch klimaat. Linn (1980) en ook Rovinelli & Hambleton (1977) bieden enkele alternatieven voor het beantwoorden van deze vraag. In een van deze alternatieven wordt gebruik gemaakt van oordelen van inhoudsspecialisten, die gevraagd worden om een oordeel uit te spreken over de relatie tussen de indicatoren van een proefinstrument en het 'werkelijkheidsgebied' van het instrument. Deze procedure is ook in dit project gevolgd. Voor de inrichting van een panel inhoudsspecialisten kan men gebruik maken van de deskundigheid van diverse specialisten uit diverse geledingen: practici zoals: onderwijsgeveinden en schoolleiders en deskundigen zoals onderwijsbegeleiders, onderzoekers van het onderwijs en inspecteurs van het onderwijs. Het raadplegen van deskundigen op het gebied van het pedagogisch klimaat had reeds plaats gevonden bij de samenstelling van de itempool.

Het was dus niet noodzakelijk om hen in het panel op te nemen. Er is voor gekozen om een panel samen te stellen van inspecteurs van het onderwijs: Zij zijn door hun dagelijkse observaties in vele klassen en scholen bij uitstek deskundig. Er is van afgezien om gebruik te maken van de oordelen van onderwijsgevend en schoolleiders, omdat de ervaring met het pedagogisch klimaat in klassen bij veel van hen niet verder zal reiken dan het klimaat in de klassen van hun eigen school. Achteraf gezien wordt dit echter als een gemis ervaren. Bij projecten die de inspectie recenter in uitvoering had en bij op dit moment lopende projecten wordt daarom wel van hun oordelen gebruik gemaakt bij de samenstelling van evaluatie instrumenten.

Voor dit project is een panel ingericht bestaande uit 14 inspecteurs basisonderwijs. De 23 in het voorafgaande geselecteerde indicatoren zijn voorgelegd aan dit panel van 14 inhoudspecialisten. Deze hebben onafhankelijk van elkaar de indicatoren voor de pedagogische kwaliteit van het klimaat in de klas beoordeeld op het criterium: de mate waarin de indicator een goede aanwijzing vormt voor de kwaliteit van het pedagogisch klimaat in de klas. Als eis is gesteld dat de betreffende indicator door tenminste 75% van het panel beoordeeld moest worden als goede indicator voor de kwaliteit van het pedagogisch handelen van de onderwijsgevende. Het resultaat was dat 11 van de 23 indicatoren aan deze eis voldeed. De indeling van deze indicatoren over de oorspronkelijke aspecten van het domein is weergegeven in tabel 2.

Deze elf indicatoren zijn zowel op observeerbaarheid als op indiceerbaarheid geselecteerd. Dit kan uiteraard gevolgen hebben voor de inhoudvaliditeit van het instrumentarium. De scores op het nieuwe instrument moeten een goede voorspellende waarde hebben voor de scores op het oorspronkelijke instrumentarium.

Aangezien dit korte instrumentarium gebaseerd is op een secundaire analyse van gegevens uit een veel breder opgezet onderzoek naar het pedagogisch klimaat (zie Brandsma & Bos, 1994) kan deze voorspellende waarde vastgesteld worden door het berekenen van de correlaties tussen de elf indicatoren van het nieuwe instrument en de indicatoren van de oorspronkelijke instrumenten. Hierbij treedt, gedeeltelijk door overlap van indicatoren, een zekere autocorrelatie op. Het is mede daarom niet ongebruikelijk om als eis te stellen, dat in deze gevallen niet voor attenuatie⁴ gecorrigeerde correlaties tenminste .50 bedragen. Men kan ook corrigeren voor autocorrelatie door het verwijderen van de dubbel voorkomende items uit de oude schalen. In dat geval is het gebruikelijk om als minimumeis te verlangen dat de niet voor attenuatie gecorrigeerde correlaties tenminste .35 zijn. Als dat niet het geval is heeft het nieuwe instrument onvoldoende inhoudvaliditeit. Dit stelt ons in de gelegenheid om na te gaan wat de voorspellende waarde van het nieuwe instrumentarium is voor het 'oude' instrumentarium. In tabel 2 wordt een overzicht gegeven van de 'ruwe' correlaties tussen het nieuwe instrument en de oorspronkelijke instrumenten. In de linker kolom zijn de correlaties gecorrigeerd voor autocorrelatie en in de rechter kolom niet.

Tabel 2: *Correlaties tussen de schaal 'pedagogische kwaliteit van het klimaat in de leerlinggroepen van het basisonderwijs' en de oorspronkelijke aspecten van het pedagogisch handelen in de klas uit de domeinmatrijs*

aspecten	aantal indicatoren	R met nieuwe schaal	
		gecorrigeerd	ongecorrigeerd
creëren van een ontspannen sfeer in de klas	3	.78	.86
belangstelling hebben voor alle leerlingen	2	.43	.80
positieve feedback geven aan de leerlingen	1	.77	.82
regels en ordehandhaving ¹³	0	.74	.74
structuur bieden	1	.47	.56
aan leerlingen ruimte bieden om zelfonitdekkend bezig te zijn	2	.69	.80
voorbeeldgedrag van leerkracht naar leerlingen	2	.67	.79
reacties van leerkracht op interacties tussen leerlingen ¹⁴	0	.98	.98

Vastgesteld kan worden dat alle correlaties van het verkorte instrument met de diverse aspecten van het oorspronkelijke instrumentarium royaal boven het vereiste van respectievelijk .35 en .50 liggen. Dit geldt ook voor de twee domeinen waaruit geen indicatoren gerecrueteerd voor de nieuwe schaal. Daarmee is de inhoudsvaliditeit van het nieuwe instrument voldoende gewaarborgd.

4.2 De calibratie van de indicatoren

De indicatoren hadden in het oorspronkelijke instrumentarium (zie Brandsma en Bos, 1994) een tetrachotoom karakter. Ten behoeve van het verkorte instrument zijn de indicatoren gedichotomiseerd. Ze worden nu gescoord op 0 en 1. De score 0 wordt gebruikt als de indicator niet of slechts enigszins van toepassing is op de betreffende leerkracht. De score 1 wordt gebruikt als de indicator (vrijwel) volledig van toepassing is op de geobserveerde leerkracht.

Elke indicator uit de schaal 'pedagogisch klimaat' moet gecalibreerd worden. Dat wil zeggen: Voor elke indicator uit deze schaal moet de plaats op de dimensie berekend worden. De σ -waarde markeert de plaats die de indicator inneemt op de dimensie van de schaal. De σ -waarden van een dimensie worden uitgedrukt in een logaritmische schaal. De schaal kan in theorie uiteenlopen van $-\infty$ tot $+\infty$. De gemiddelde σ -waarde van alle indicatoren wordt vastgezet op 0. Voor het calibreren van indicatoren kan men diverse rekentechnieken gebruiken (cf. Fischer, 1974; Wright & Stone, 1979; Gustafsson, 1979). Meestal wordt voor de calibratie gebruik gemaakt van de CML (conditional maximum likelihood) methode. Dat wordt ook hier gedaan.

In tabel 3 worden de σ -waarden voor de indicatoren en hun bijbehorende standaard error (s.e.⁵) gepresenteerd.

Tabel 3: Calibratie van de indicatoren

Indicatoren	Dif (log) σ -waarde	s.e.
G1 de leerkracht toont respect voor alle leerlingen	-1.77	.18
G2 de leerkracht is tolerant ten opzichte van alle leerlingen	-1.37	.16
A1 de leerkracht is ontspannen naar leerlingen toe	-1.20	.16
E1 de leerkracht formuleert opdrachten zodanig dat leerlingen weten wat van hen verwacht wordt	-1.17	.16
A7 in de klas hangt een positieve werksfeer	-.57	.14
B2 de leerkracht betreft alle leerlingen bij het onderwijs	-.31	.14
A3 de leerkracht creëert een sfeer van veiligheid en geborgenheid	-.15	.13
C3 de leerkracht waardeert getoonde inzet	.47	.13
B3 de leerkracht gaat in op de eigen inbreng van de leerlingen	1.52	.12
F2 de leerkracht zorgt voor een rijke leeromgeving	1.90	.13
F3 de leerkracht stimuleert leerlingen zelf oplossingen te zoeken	2.65	.14

In tabel 3 kan afgelezen worden dat de indicatorparameters (de σ -waarden) goed gespreid zijn over de dimensie. Ze lopen ze van -1.77 tot en met 2.65. Elke indicator heeft nu zijn plaats op de dimensie 'Pedagogisch klimaat' gekregen. Daarmee is niet alleen de volgorde van de indicatoren vastgelegd, maar ook hun onderlinge afstand. Zo kan bijvoorbeeld geconstateerd worden dat de indicatoren A1 en E1 vrij dicht bijeen liggen. De indicatoren C3 en B3 liggen daarentegen veel verder uit elkaar.

Deze interpretatie van deze plaatsbepaling van de indicatoren ten opzichte van elkaar heeft echter alleen zin als deze indicatoren ook daadwerkelijk één dimensie vormen. Voorafgaand aan het beantwoorden van de vraag of dit het geval is, gaan we echter eerst na of de schaal voldoet aan de eisen, die in de klassieke testtheorie gesteld worden.

4.3 De homogeniteit van de schaal

In de klassieke testtheorie (Nunnally & Bernstein, 1994) wordt als eis gesteld dat een homogene schaal ten minste een KR-20 heeft van .70. De KR-20 van de schaal blijkt met .83 ruim

boven de grens van .70 te liggen. Een andere belangrijke eis die in de klassieke testtheorie aan een homogeen instrument gesteld wordt is dat de correlaties van de indicatoren met de somscore op de schaal tenminste .50 bedragen. In de tabel 4 wordt een overzicht geboden van de biseriële indicator-totaal-correlaties van elke indicator uit de schaal. Hier blijkt dat de indicator-totaal-correlaties ruimschoots aan de gestelde eis voldoen.

Tabel 4: Biseriële indicator-totaal-correlaties

Indicatoren	R_{jit}
G1 de leerkracht toont respect voor alle leerlingen	.99
G2 de leerkracht is tolerant ten opzichte van alle leerlingen	.85
A1 de leerkracht is ontspannen naar leerlingen toe	.93
E1 de leerkracht formuleert opdrachten zodanig dat leerlingen weten wat van hen verwacht wordt	.79
A7 in de klas hangt een positieve werksfeer	.95
B2 de leerkracht betreft alle leerlingen bij het onderwijs	.89
A3 de leerkracht creëert een sfeer van veiligheid en geborgenheid	.92
C3 de leerkracht waardeert getoonde inzet	.76
B3 de leerkracht gaat in op de eigen inbreng van de leerlingen	.78
F2 de leerkracht zorgt voor een rijke leeromgeving	.80
F3 de leerkracht stimuleert leerlingen zelf oplossingen te zoeken	.68
$x = .85$	
$sd = .09$	
range = .68 - .99	

Aan deze gegevens die normaal gesproken bij de klassieke testtheorie gebruikt worden, kunnen ook aanwijzingen ontleend worden voor de paralleliteit van de indicator karakteristieke curves (zie ook paragraaf 4.6). Als de biseriële indicator-totaal correlaties gelijk zijn, dan is dat een aanwijzing voor de gelijkheid van de discriminatieparameters. In tabel 4 kan geconstateerd worden dat de gemiddelde indicator-totaal-correlatie zeer hoog te noemen is. Het gemiddelde bedraagt .85⁶. Dat de biseriële indicator-totaal-correlaties redelijk gelijk zijn kan afgeleid worden uit twee aanwijzingen. Ten eerste is de standaard deviatie zeer gering te noemen. Een tweede, hier overigens mee samenhangende, aanwijzing kan ontleend worden aan een suggestie van Hambleton & Swaminathan (1985). Zij stellen voor om als criterium een range van .15 rondom het gemiddelde te hanteren. In ons geval zouden de biseriële indicator-totaal correlaties zich dan tussen de .70 en de 1.00 moeten bewegen. Dit is duidelijk het geval voor alle indicatoren, behalve indicator F3. De laagste correlatie bedraagt .68 en de hoogste .99. Ook dit biedt een positief beeld van de paralleliteit van de indicator karakteristieke curves.

Dit is een eerste aanwijzing dat deze verzameling van elf indicatoren voldoet aan de eis van het hebben van parallelle indicator karakteristieke curves. We komen hier op terug in paragraaf 4.6.

4.4 De unidimensionaliteit van het instrument

Een belangrijke eis, die vanuit de itemresponstheorie aan een schaal gesteld wordt, is dat een schaal unidimensioneel is. Dit brengt met zich mee dat de maatverdeling zoals die door de σ -waarde van de indicatoren aangegeven wordt, niet beïnvloed mag worden door andere zaken dan de eigenschap die met de schaal gemeten moet worden. In dit project is nagegaan of de maatverdeling van de dimensie 'Pedagogisch klimaat' beïnvloed wordt door de leeftijdsgroep waarin de onderwijsgevende werkt en het leer- en vormingsgebied dat aan de orde is als de pedagogische kwaliteit van het klimaat wordt geëvalueerd. Om de hiervoor benodigde analyses uit te kunnen voeren is een onderscheid gemaakt tussen de diverse leeftijdsgroepen van leerlingen in het basisonderwijs en de verschillende leer- en vormingsgebieden, die op het programma van het basisonderwijs staan. De leeftijdsgroepen zijn onderscheiden naar bouw:

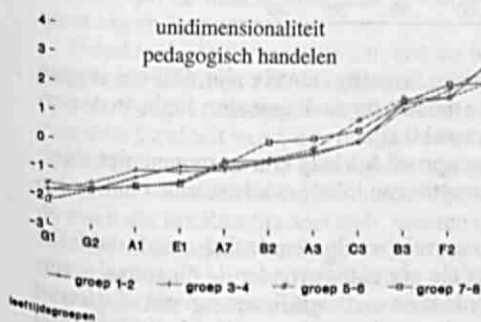
de kleuterbouw (groep 1/2), de onderbouw (groep 3/4), de middenbouw (groep 5/6) en de bovenbouw (groep 7/8). De leer- en vormingsgebieden zijn ingedeeld naar de kennisvakken (wereld-oriëntatie, biologie e.d.), instrumentele vakken (taal, rekenen) en expressievakken (handvaardigheid, lichamelijke oefening) en werklessen die aan kinderen in de kleuterbouw gegeven worden.

Voor elk van deze beide groepen kenmerken zijn groepen lessen onderscheiden. Voor elke groep zijn apart de σ -waarden of indicatorparameters berekend. Als de σ -waarden van elke indicator bij elke onderscheiden groep (binnen het 95% betrouwbaarheidsinterval) gelijk blijven, mag de schaal unidimensioneel genoemd worden. De indicatorparameters zijn met behulp van de Andersentoets getoetst op gelijkheid. Met deze toets (Andersen, 1973) kan nagegaan worden of de kwaliteit van het pedagogisch klimaat zoals dat in de σ -waarde van de indicatoren tot uitdrukking komt, afgezien van fluctuatie door kans, invariant is voor verschillende geobserveerde groepen onderwijsgevenden. De toets bestaat in het berekenen van een χ^2 met $(g-1)(k-1)$ vrijheidsgraden.

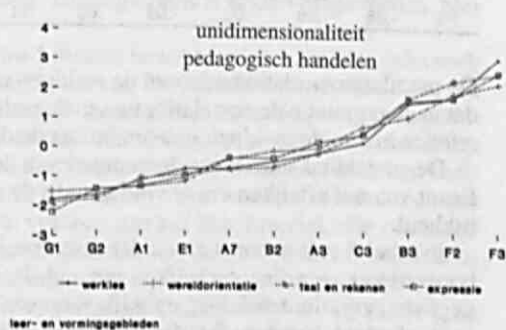
De χ^2 van de Andersentoets die berekend is voor de verschillende leeftijdsgroepen bedraagt bij 30 vrijheidsgraden 43.86. De bijbehorende overschrijdingskans is .05. Deze resultaten wijzen erop dat de unidimensionaliteit van de schaal 'pedagogisch klimaat' niet beïnvloed wordt door verschillen in leeftijdsgroepen.

De χ^2 voor de verschillende leer- en vormingsgebieden bedraagt bij 30 vrijheidsgraden 36.5; de overschrijdingskans is .19. Deze resultaten wijzen erop dat de unidimensionaliteit van de schaal 'pedagogisch klimaat' niet beïnvloed wordt door verschillen in leer- en vormingsgebied.

Figuur 1



Figuur 2



De bij dit unidimensionaliteitsonderzoek gebruikte χ^2 -toets (de Andersentoets) is gevoelig voor de grootte van de steekproef. Bij grote steekproeven kunnen de indicatorparameters met zeer grote nauwkeurigheid geschat worden. (De standaard error is dan heel klein.) Daardoor wordt bij grote steekproeven de grenswaarde van de χ^2 -toets al bereikt bij heel geringe afwijkingen van de indicatorparameters. (Hoe groter de steekproef, hoe sneller de toets significant wordt.) Daarom worden deze toetsgegevens aangevuld met grafische gegevens (zie figuur 1 en 2), waarin de indicatorparameters van de verschillende groepen geplotted worden. In de bovenstaande grafieken is voor elke leeftijdsgroep en voor de vier onderscheiden groepen leer- en vormingsgebieden de σ -waarde voor elk van de indicatoren geplotted. Ook hieruit blijkt dat de maatverdeling van dimensie nauwelijks verandert bij de verschillende leeftijdsgroepen en bij de verschillende leer- en vormingsgebieden. Dit is een bevredigend resultaat.

4.5 De lokale stochastische onafhankelijkheid van de indicatoren

De score op elke indicator van het instrument dient onafhankelijk te zijn van de score op de overige indicatoren. De correlaties tussen de indicatoren van een schaal mogen dus niet veroorzaakt worden door afhankelijkheden van de indicatoren onderling, maar moeten hun verklaring vinden in het effect van de werking van het klimaat dat in de indicatoren tot uitdrukking komt. In tabel 5 zijn boven de diagonaal de eerste orde correlaties tussen de elf indicatoren gepresenteerd. We kunnen vaststellen dat er een duidelijke en positieve samenhang tussen de indicatoren bestaat.

Het effect van het klimaat op de onderlinge correlaties van de indicatoren kan met behulp van regressie-analyse uitgeparialiseerd worden. Tussen de aldus ontstane residuen kunnen de correlaties berekend worden. Deze partiële correlaties zijn in tabel 5 onder de diagonaal weergegeven.

Tabel 5: De eerste orde correlaties en partiële correlaties tussen de indicatoren

	G1	G2	A1	E1	A7	B2	A3	C3	B3	F2	F3
G1		.60	.48	.35	.46	.37	.47	.40	.31	.29	.23
G2	.45		.46	.24	.39	.27	.41	.28	.26	.26	.19
A1	.17	.23		.33	.41	.36	.48	.34	.33	.27	.22
E1	.07	.04	.07		.38	.44	.33	.30	.23	.25	.19
A7	.16	.13	.18	.12		.44	.52	.36	.42	.33	.26
B2	.02	-.02	.09	.17	.15		.37	.40	.41	.34	.32
A3	.12	.20	.23	.07	.35	.07		.40	.37	.37	.26
C3	-.08	.00	.08	.03	.10	.23	.18		.47	.33	.28
B3	.04	-.05	.10	-.07	.14	.24	.14	.27		.43	.49
F2	-.02	.00	.00	-.01	.08	.10	.18	.18	.25		.45
F3	-.08	-.08	-.02	-.05	.00	.11	.00	.09	.30	.30	

De populatiecorrelatiematrix van de residuen moet een 'identity' matrix zijn. Dat wil zeggen dat in de populatie de correlaties tussen de residuen moeten op de diagonalen 1 zijn en de correlaties tussen de residuen moeten buiten de diagonaal 0 zijn.

De correlaties van de residuen moeten in de steekproef dus laag zijn en mogen niet significant van nul afwijken om te voldoen aan de assumptie van lokale stochastische onafhankelijkheid.

In tabel 5 valt te constateren dat de eerste orde correlaties (boven de diagonaal) duidelijk hoog zijn en significant afwijken van nul. De partiële correlaties (onder de diagonaal weergegeven) zijn duidelijk laag en wijken op een incident na niet significant van nul af. Hieruit mag afgeleid worden dat de assumptie van lokale stochastische onafhankelijkheid niet geschonden wordt.

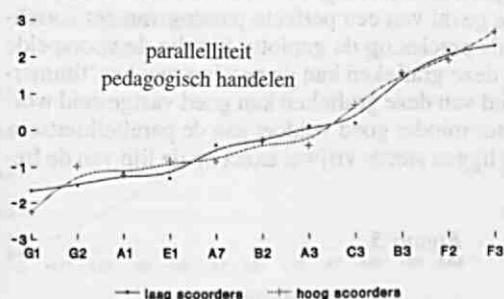
4.6 De paralleliteit van de indicator karakteristieke curves

In het paralleliteitsonderzoek staat de vraag centraal of de indicatoren uit de schaal een gelijk discriminerend vermogen hebben. Wanneer dit niet het geval is fluctueert de plaats van de indicatoren op de schaalverdeling (de σ -waarde) afhankelijk van het niveau van de personen in de steekproef. Of de schaal hiervoor gevoelig is kan nagegaan worden met behulp van twee toetsen: de Andersentoets en de Martin-Löftoets.

Met behulp van de Andersentoets kan nagegaan worden of de σ -waarde van de indicatoren, zoals deze op de maatverdeling van de dimensie staat, gelijk blijft onafhankelijk van het feit of dit nu is vastgesteld aan de hand van de onderwijsgevenden met een zeer goed pedagogisch klimaat of aan de hand van onderwijsgevenden met een matig pedagogisch klimaat.

De χ^2 van de Andersentoets is met 16.77 bij 10 vrijheidsgraden en een overschrijdingskans van .08 laag genoeg om te mogen stellen, dat deze verzameling indicatoren in zijn totaliteit voldoet aan de eis van het hebben van parallelle indicator karakteristieke curves.

Figuur 3



Ook voor deze toepassing van de Andersentoets kan een grafische voorstelling gemaakt worden (zie figuur 3). Dit is nuttig omdat, zoals reeds opgemerkt is, de resultaten van de Andersentoets gevoelig zijn voor de grootte van de steekproef. Bij deze toepassing van de Andersentoets gaat het om een vergelijking van de indicatorparameters bij respectievelijk hoog- en laagscorende onderwijsgevenden. In de bijgaande grafiek zijn de indicatorparameters van de hoogscorende onderwijsgevenden geplott ten opzichte van de indicatorparameters die berekend zijn bij de onderwijsgevenden met een matig pedagogisch klimaat. Wanneer de indicatorparameters die vastgesteld worden bij de hoogscorende onderwijsgevenden niet (binnen de grenzen van kansfluctuatie) gelijk zijn, dan is dat een aanwijzing dat de hellingen van de indicator karakteristieke curves niet gelijk zijn. In de bijgaande grafiek kan geconstateerd worden, dat de indicatorparameters van de hoog- en laagscorende onderwijsgevenden zeer dicht bijeen liggen.

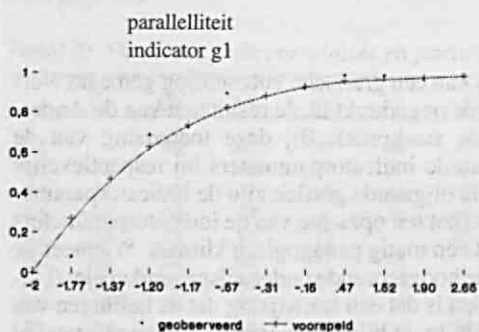
Behalve de Andersentoets kan ook de Martin-Löftoets benut worden voor het onderzoek naar de paralleliteit van de indicator karakteristieke curves. De Martin-Löftoets is gebaseerd op het volgende principe. Voor elke onderwijsgevende die een klimaat realiseert met een pedagogische kwaliteit van ξ kan de kans berekend worden dat hij of zij een positieve score krijgt op een indicator met een moeilijkheidsgraad van σ . Deze kans wordt uitgedrukt in een indicator karakteristieke curve. De indicator karakteristieke curves van een schaal, die voldoen aan de eisen die het Raschmodel stelt, moeten om te voldoen aan het Raschmodel, alle even steil zijn. De Martin-Löftoets is gevoelig voor de heterogeniteit van de hellingen van de indicator karakteristieke curves. Met behulp van de formule voor het Raschmodel (zie noot 9) kunnen voor elke groep onderwijsgevenden met een pedagogisch klimaat van kwaliteit ξ de frequenties van het 'een positieve score krijgen' op een bepaalde indicator voorspeld worden. Deze voorspelde frequenties kunnen vergeleken worden met de geobserveerde frequenties van elke groep onderwijsgevenden met een pedagogisch klimaat van gelijke kwaliteit. Aan de hand van de afwijkingen tussen deze voorspelde en geobserveerde frequenties wordt een χ^2 -som opgebouwd met $(k-1)(k-2)$ vrijheidsgraden. Wanneer de met behulp van de Martin-Löftoets gevonden χ^2 kleiner of gelijk is aan het aantal vrijheidsgraden, dan kunnen we aannemen dat de indicator karakteristieke curves van de schaal parallel zijn.

De χ^2 van de Martin-Löftoets is met 118.76 bij 90 vrijheidsgraden en een overschrijdingskans van .02 net laag genoeg om te mogen stellen, dat deze verzameling indicatoren in zijn totaliteit voldoet aan de eis van het hebben van parallelle indicator karakteristieke curves.

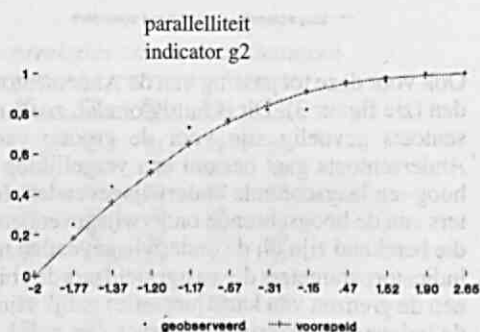
Aangezien niet alleen de Andersentoets, maar ook de Martin-Löftoets gevoelig is voor de grootte van de steekproef, worden de resultaten van ook deze toets met 'grafische toetsen' ondersteund. Begonnen wordt met het plotten van enkele grafieken voor de Martin-Löftoets. Met behulp van de formule voor het Raschmodel kunnen voor elke vaardigheidsgroep de voorspelde frequenties van de positieve antwoorden op elke indicator berekend worden. Deze voorspelde frequenties kunnen vergeleken worden met de geobserveerde frequenties. In de

onderstaande grafieken (4 tot en met 14) is voor elke indicator steeds per groep onderwijsgevenden met dezelfde score op het pedagogisch klimaat de geobserveerde en de voorspelde frequentie tegen elkaar afgezet. In het ideale geval van een perfecte passing van het Rasch-model liggen de geobserveerde punten steeds precies op de geplote lijn die de voorspelde punten met elkaar verbindt. Met behulp van deze grafieken kan de passing met het 'timmermansoog' geïnspecteerd worden. Aan de hand van deze grafieken kan goed vastgesteld worden welke indicatoren beter en welke indicator minder goed voldoet aan de parallelleiteitsis. De punten van de geobserveerde frequenties liggen steeds vrijwel exact op de lijn van de frequenties.

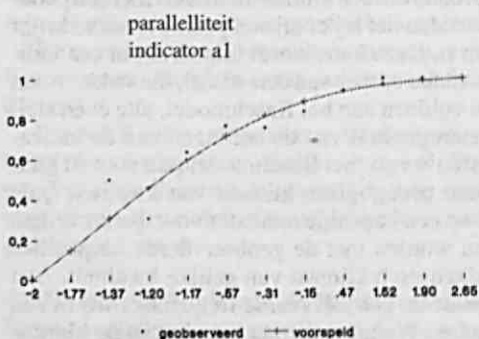
Figuur 4



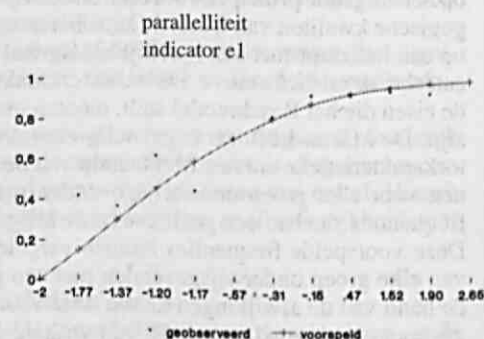
Figuur 5



Figuur 6

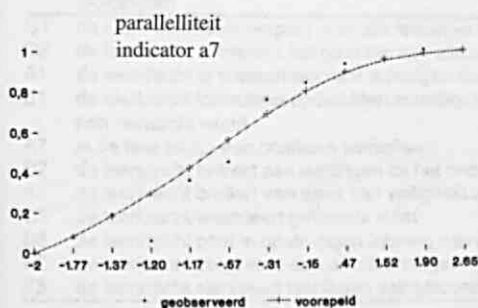


Figuur 7

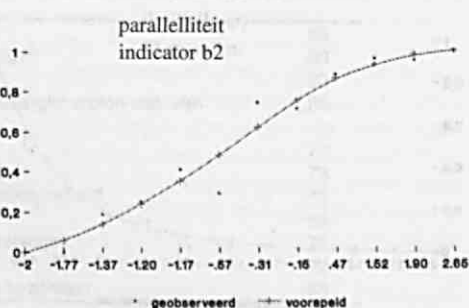


Een derde grafische toets die uitgevoerd kan worden om de parallelleiteit van de indicator karakteristieke curves te kunnen onderzoeken bestaat in de volgende aanpak. Allereerst wordt een datamatrix opgebouwd waarin de rijen de scoregroepen vormen en de kolommen de cumulatieve frequenties van de indicatoren (in procenten). Op basis van deze datamatrix kunnen de cumulatieve frequentiepolygonen van de indicatoren geplote worden. Dit is in de volgende twee grafieken gebeurd (zie figuur 15 en 16). In deze grafieken zijn de cumulatieve frequentiepolygonen van de indicatoren geplote. Deze cumulatieve frequentieverdelingen vormen een goede indicatie van hoe de geobserveerde indicator karakteristieke curves eruit zien.

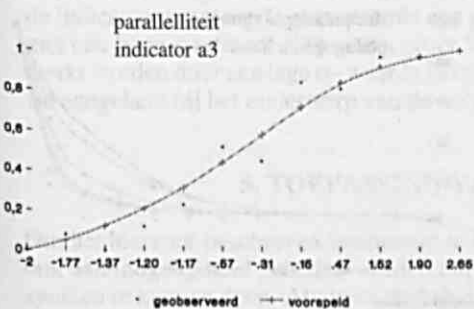
Figuur 8



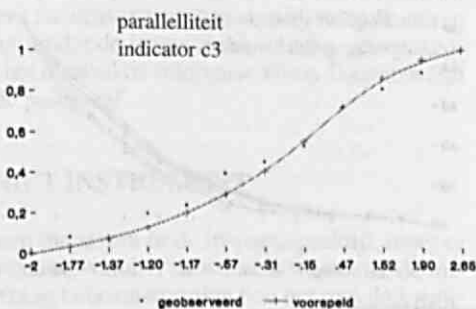
Figuur 9



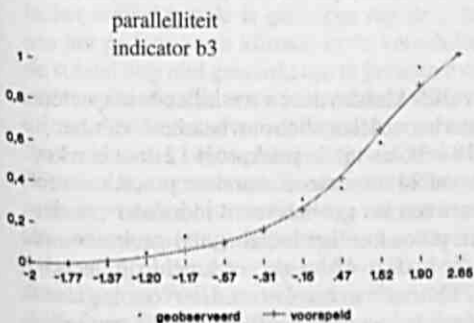
Figuur 10



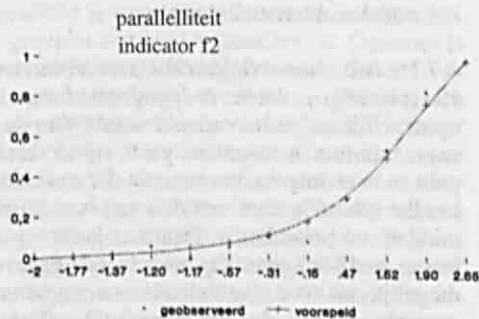
Figuur 11



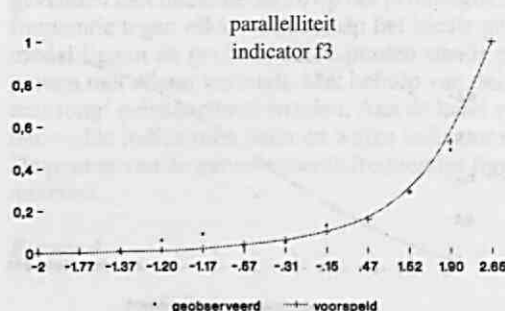
Figuur 12



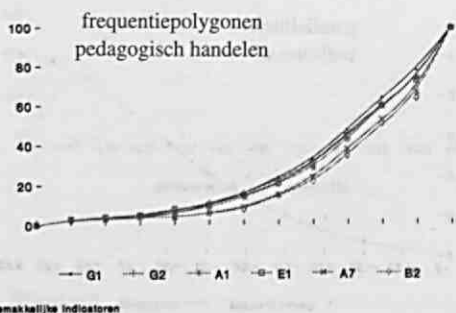
Figuur 13



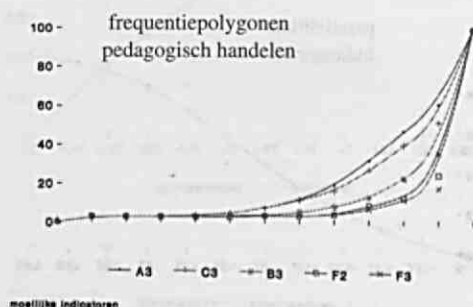
Figuur 14



Figuur 15



Figuur 16



Geconstateerd kan worden, dat de frequentieverdelingen elkaar niet snijden, al liggen ze soms heel dicht bij elkaar. Dit betekent dat er geen schendingen zijn van de eis van de parallele indicator karakteristieke curves.

4.7 De interbeoordelaarsbetrouwbaarheid

Het is belangrijk dat het pedagogisch klimaat in dezelfde klassen door verschillende inspecteurs op dezelfde wijze beoordeeld wordt. Om deze interbeoordelaarsbetrouwbaarheid van het instrumentarium na te kunnen gaan, zijn er van de 118 scholen uit de steekproef 12 door een koppel van twee inspecteurs bezocht. Er werkten in totaal 24 inspecteurs aan deze procedure mee. In elke school is door een duo van twee inspecteurs een les geobserveerd in kleuter-, onder-, midden- en bovenbouw. Daarmee komt voor deze procedure het totaal aantal geobserveerde lessen op 48 te liggen. Op basis van de gegevens, die in deze 48 lessen verzameld zijn, was het mogelijk om voor elke indicator een consensuscoëfficiënt¹⁰ te berekenen. Hiervoor is gebruik gemaakt van de volgende formule $\{\sum o_p / \sum o\}$. De hoogte van deze coëfficiënt geeft per indicator het percentage consensus aan dat door de 24 beoordelaars bereikt is in de 48 lessituaties. Deze coëfficiënt is niet gecorrigeerd voor kans. Voor het gehele instrument is de gemiddelde interbeoordelaarsbetrouwbaarheidscoëfficiënt per indicator berekend $\{(\sum o_p / \sum o) / k\}$. Ook deze coëfficiënt is ongecorrigeerd voor kans. Bij dichotome antwoordmodaliteiten wordt er doorgaans naar gestreefd om geen lagere coëfficiënten te accepteren dan .80. Bij deze procedure voor het berekenen van de consensus zijn echter de oorspronkelijke tetrachotome antwoordmodaliteiten gebruikt. Hierbij is de kans op consensus bij willekeurige invulling tweemaal zo klein. Daarom wordt de cesuur voor een acceptabele overeenstemming gelegd bij .70¹¹.

Tabel 6: *Interbeoordelaarsbetrouwbaarheid*

Indicatoren		
G1	de leerkracht toont respect voor alle leerlingen	.85
G2	de leerkracht is tolerant ten opzichte van alle leerlingen	.90
A1	de leerkracht is ontspannen naar leerlingen toe	.90
E1	de leerkracht formuleert opdrachten zodanig dat leerlingen weten wat van hen verwacht wordt	.86
A7	in de klas hangt een positieve werksfeer	.79
B2	de leerkracht betreft alle leerlingen bij het onderwijs	.73
A3	de leerkracht creëert een sfeer van veiligheid en geborgenheid	.71
C3	de leerkracht waardeert getoonde inzet	.83
B3	de leerkracht gaat in op de eigen inbreng van de leerlingen	.71
F2	de leerkracht zorgt voor een rijke leeromgeving	.77
F3	de leerkracht stimuleert leerlingen zelf oplossingen te zoeken	.65
consensus		.79

Geconstateerd kan worden dat de consensuscoëfficiënt voor het totale instrumentarium ruimschoots de ondergrens van .70 overschrijdt (zie tabel 6). Opvallend is dat de beoordelaars bij de indicatoren met een lage σ -waarde een grotere consensus bereiken dan bij de indicatoren met een hoge σ -waarde. Dit is van groot belang omdat de inspectie in situaties, die gekenmerkt worden door een lage σ -waarde vaker tot het oordeel onvoldoende komt. Daarmee zijn we aangeland bij het onderwerp van de volgende paragraaf.

5. TOEPASSING VAN HET INSTRUMENT

Om het hiervoor beschreven instrument te kunnen toepassen in de inspectiepraktijk moet er ook een mogelijkheid geboden worden om met behulp van het instrument waarderende uitspraken te kunnen doen. Alvorens derhalve de vraag te beantwoorden hoe het met de kwaliteit van het pedagogisch klimaat in de Nederlandse scholen voor basisonderwijs gestelde (5.2) zal eerst de normeringskwestie aan de orde moeten worden gesteld (5.1)

5.1 De normering van het instrument

In het voorafgaande is gebleken dat de schaal geschikt is om verschillen vast te stellen tussen het pedagogisch klimaat in de verschillende groepen van het basisonderwijs. Daarmee is de schaal nog niet geschikt om te kunnen evalueren. Evalueren wil zeggen: nagaan of een specifieke situatie aan een norm voldoet. Hiervoor is het nodig dat op de dimensie 'Pedagogisch klimaat' een cesuur wordt bepaald, waarmee vastgesteld kan worden of een bepaald klimaat in een klas van voldoende pedagogische kwaliteit is of niet. Voor het bepalen van zo'n cesuur of norm kunnen verschillende methoden gebruikt worden. Deze methoden worden onderscheiden in absolute en relatieve normeringstechnieken. Hier wordt gebruik gemaakt van een absolute normeringstechniek: Er wordt een vast punt bepaald op de dimensie, dat de grens aangeeft tussen voldoende en onvoldoende. De vraag is nu: Hoe bepalen we waar deze cesuur moet liggen? Wanneer kan men met recht en rede zeggen, dat een pedagogisch klimaat onvoldoende is. Het is duidelijk dat aan het beantwoorden van deze vraag inhoudelijke overwegingen ten grondslag moeten liggen, die het welzijn van de leerlingen en de kwaliteit van het onderwijs betreffen. Tegelijkertijd moet het zo zijn, dat er geen normen gesteld worden, waaraan geen onderwijsgevende in redelijkheid kan voldoen. De norm moet dus niet alleen redelijk (inhoudelijke overwegingen), maar ook rechtvaardig (onderwijsgevend) moeten er aan kunnen voldoen zijn. Het terrein van 'rede en recht' in het onderwijs is bij uitstek het werkterrein van de inspectie. De inspectie kent als geen ander de toestand van het onderwijs. Daarom is bij het vaststellen van de cesuur tussen voldoende en onvoldoende gebruik gemaakt van de oordelen van inspecteurs. Waar ligt de cesuur tussen voldoende en onvoldoende? Om deze

vraag te kunnen beantwoorden zijn aan een groep van 20 inspecteurs die werkzaam zijn in het basisonderwijs de elf indicatoren van de schaal voorgelegd met de vraag aan hoeveel indicatoren er voldaan diende te zijn om te kunnen spreken van een klimaat dat van voldoende pedagogische kwaliteit is.

Het bleek dat de inspecteurs gemiddeld gezien verlangden dat het klimaat in de leerlinggroepen van het basisonderwijs aan tenminste 7 van de 11 indicatoren diende te voldoen voor een klimaat van voldoende pedagogische kwaliteit.

Een klimaat van voldoende pedagogische kwaliteit wordt uitgaande van deze cesuur gekarakteriseerd door respectvol en tolerant gedrag ten opzichte van de leerlingen, door een ontspannen houding van de onderwijsgevende en door een duidelijke, veilige en positieve werksfeer waar alle leerlingen bij betrokken zijn.

Een klimaat dat van voldoende pedagogische kwaliteit is, is overigens wel verbeterbaar. De onderwijsgevende die de score 'voldoende' krijgt, kan het klimaat verbeteren door meer waardering te tonen voor de inzet en eigen inbreng van de leerlingen en door de leeromgeving voor de leerlingen rijker en stimulerender te maken.

In tabel 7 is de normering van de schaal beknopt uitgewerkt.

Tabel 7: Normering van de schaal

score	kwalificatie	beschrijving
0-6	het klimaat is van onvoldoende pedagogische kwaliteit	Er zijn op korte termijn aanzienlijke verbeteringen noodzakelijk
7	het klimaat is van voldoende pedagogische kwaliteit	De onderwijsgevende vertoont respectvol en tolerant gedrag ten opzichte van de leerlingen, heeft een ontspannen houding ten opzichte van de leerlingen en biedt een duidelijke, veilige en positieve werksfeer waarin alle leerlingen bij het onderwijs betrokken zijn. Er zijn verbeteringen mogelijk op het punt van de inzet en inbreng van leerlingen en de leeromgeving kan voor de leerlingen rijker en stimulerender gemaakt worden.
8-11	het klimaat is van goede pedagogische kwaliteit	Kleine verbeteringen zijn mogelijk.

De normering in tabel 7 is gebaseerd op de situatie in de afzonderlijke leerlinggroepen. Daarmee beschikken we nog niet over een norm voor het evalueren van de pedagogische kwaliteit van het klimaat in de school. Dat is van belang, omdat de inspectie niet zozeer individuele onderwijsgegenden, als wel scholen evalueert. Uiteraard wordt het oordeel over de scholen gebaseerd op waarnemingen bij de individuele onderwijsgegenden. Daarom is er een procedure nodig om de waarnemingen bij individuele onderwijsgegenden te kunnen aggregeren naar schoolniveau. Bij het vaststellen van de cesuur tussen voldoende en onvoldoende voor het oordeel per school is eveneens gebruik gemaakt van de oordelen van inspecteurs. Het bleek dat de inspecteurs gemiddeld gezien als norm voor een *voldoende* pedagogisch klimaat op een basisschool stelden, dat de gemiddelde score van de onderwijsgegenden waar het pedagogisch klimaat is geëvalueerd aan ten minste zeven van de elf indicatoren diende te voldoen. In dit geval kunnen onderwijsgegenden voor elkaar compenseren. Het is bijvoorbeeld mogelijk dat op een school een relatief goed pedagogisch klimaat bij de kleuters compenseert voor een relatief zwak pedagogisch klimaat in de bovenbouw. De norm voor een *goed* pedagogisch klimaat is beduidend strenger. Hiervoor is het nodig dat alle onderwijsgegenden op een school voldoen aan de norm van zeven indicatoren.

De volgende vraag is nu bij hoeveel procent van de scholen er bij de onderwijsgegenden in sprake van een voldoende of goed pedagogisch klimaat. Dit komt aan de orde in de volgende paragraaf.

5.2 De eigenschappen van de frequentieverdeling

De kwaliteit van het pedagogisch klimaat is geëvalueerd op 118 basisscholen. Hiervoor is op elk van deze scholen gebruik gemaakt van observaties bij onderwijsgegenden die in respectievelijk kleuter-, onder-, midden- en bovenbouw werken van deze basisscholen. Deze steekproef is representatief voor de populatie en voldoende groot om nauwkeurige schattingen te maken van het pedagogisch klimaat in de leerlinggroepen van het basisonderwijs. Dit maakt het van belang om specifiek in te gaan op de eigenschappen van de frequentieverdeling. Hierbij kunnen twee wegen bewandeld worden. We kunnen zowel de school de individuele onderwijsgegende als eenheid van analyse nemen. We doen beide en beginnen met het pedagogisch klimaat in de school. In tabel 8 zijn de belangrijkste resultaten samengevat.

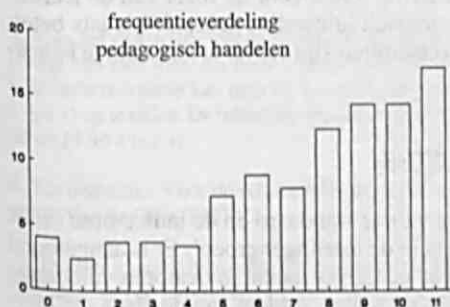
Tabel 8: Het percentage scholen dat aan de norm voldoet

score	kwalificatie	% scholen
de gemiddelde score van de onderwijsgegenden is < 7 (ξ -waarde < .71)	het klimaat is van onvoldoende pedagogische kwaliteit	31
de gemiddelde score van de onderwijsgegenden is tenminste 7 (ξ -waarde \geq .71)	het klimaat is van voldoende pedagogische kwaliteit	29
de laagste score van een onderwijsgegende op school is ten minste 7 (ξ -waarde \geq .71)	het klimaat is van goede pedagogische kwaliteit	40

Waar staat de gemiddelde school op de dimensie van het pedagogisch klimaat? De norm voor een voldoende pedagogisch klimaat op schoolniveau vereist dat het gemiddelde van de onderwijsgegenden van een school aan ten minste zeven indicatoren voldoet. Dit blijkt bij 69% van de basisscholen het geval te zijn. Hierbij dient bedacht te worden dat onderwijsgegenden hier voor elkaar kunnen compenseren. De veel strengere norm voor een goed pedagogisch klimaat op schoolniveau is dat alle onderwijsgegenden op een school voldoen aan zeven indicatoren. Aan deze veel strengere norm voor een goed pedagogisch klimaat blijkt 40% van de scholen te voldoen.

Men kan ook een goed beeld krijgen van de landelijke situatie aan de hand van het pedagogisch klimaat dat bij de gemiddelde onderwijsgegende wordt aangetroffen.

Figuur 17



In de nevenstaande grafiek wordt een overzicht gegeven van het percentage onderwijsgegenden, dat aan de indicatoren van de gestelde standaard voldoet. Het blijkt dan dat bijna tweederde van de onderwijsgegende voldoet aan zeven van de elf indicatoren van het pedagogisch klimaat. De gemiddelde score uitgedrukt in de ξ -waarde ligt bij 1.17 en de spreiding (standaard deviatie) bedraagt 1.99. Vier procent van de onderwijsgegenden blijkt aan geen enkele indicator van een kwalitatief goed pedagogisch klimaat te voldoen.

In hoeverre verschilt het pedagogisch klimaat in de klassen?

Het klimaat in de taal en rekenlessen verschilt nauwelijks van dat in de lessen in expressie, maar scoort iets lager dan het klimaat in de wereldoriëntatielessen. De verschillen zijn echter te klein om van betekenis (niet significant op het 95% niveau) te zijn. Wel zijn er significante (95% niveau) verschillen tussen het klimaat in de leeftijdsgroep 1/2 enerzijds en het klimaat in de overige leeftijdsgroepen. Het pedagogisch klimaat is in groep 1/2 iets hoger. De verschillen in het klimaat bij groep 3/4, groep 5/6 en groep 7/8 zijn minimaal en van geen betekenis.

We hebben nu een inzicht gekregen in de eigenschappen van de frequentieverdeling. Daarmee weten we echter nog niet hoeveel procent van de onderwijsgeevenden erin slagen om in hun leerlinggroep een klimaat te realiseren, dat voldoende pedagogische kwaliteit heeft. In tabel 9 wordt een overzicht gegeven van het percentage onderwijsgeevenden dat aan de gestelde norm voldoet.

Tabel 9: Het percentage onderwijsgeevenden dat aan de norm voldoet

	% onderwijsgeevenden dat aan de norm voldoet	gemiddelde kwaliteit van het pedagogisch klimaat (de ξ -waarde)	s.d.
groep 1/2 (werkles)	77	1.86	1.60
groep 3/4	64	1.13	1.86
groep 5/6	63	0.77	2.20
groep 7/8	59	0.86	2.15
wereldoriëntatie	64	0.94	2.03
taal en rekenen	61	0.84	2.06
expressie	59	0.85	2.12
totaal	66	1.17	1.99

Ongeveer tweederde van de onderwijsgeevenden voldoet aan de norm, die door de inspectie voor het pedagogisch klimaat gesteld is. In groep 1/2 voldoet ruim driekwart van de onderwijsgeevenden aan de norm. Bij de overige leeftijdsgroepen voldoet bijna tweederde van de onderwijsgeevenden aan de norm.

Groep 1/2 heeft een significant¹² beter pedagogisch klimaat dan de overige groepen in het basisonderwijs. Het verschil komt met name tot uitdrukking in het feit dat de onderwijsgeevenden in groep 1/2 gemiddeld gezien meer waardering tonen voor de inzet van de leerlingen, dan in de overige groepen gebeurt. Het pedagogisch klimaat in groep 3/4 is iets beter, dan in de groepen met de oudere leerlingen. Deze verschillen zijn echter te klein om te benoemen en niet significant.

6. CONCLUSIES

In het voorafgaande is er een beschrijving gegeven van de standaard en de indicatoren op de dimensie 'de pedagogische kwaliteit van het klimaat in de leerlingengroep'. Er is aangetoond dat de inhoudsvaliditeit ondanks het verwijderen van een groot aantal indicatoren voldoende is. Verder is er aangetoond dat de homogeniteit van de schaal voldoet aan de eisen die hieraan in de klassieke testtheorie gesteld worden. Ook bleek de schaal te voldoen aan de drie assumpties, die vanuit het Raschmodel aan een goede schaal gesteld worden (de indicatoren bleken tezamen één dimensie vormen, de indicatoren bleken in hoge mate lokaal stochastische onafhankelijk en de indicator karakteristieke curves bleken parallel). Voorts heeft het instrument een voldoende interbeoordelaarsbetrouwbaarheid.

Geconcludeerd kan worden dat het gelukt is om een instrument te construeren, dat voldoet aan de stringente eisen van het Raschmodel en dat voldoende inhoudsvaliditeit en interbeoor-

delaarsbetrouwbaarheid heeft. Hierbij dient bedacht te worden, dat het instrument ontworpen is voor inspecteurs van het onderwijs. Het is allerminst gegarandeerd dat met het instrument bij gebruik door niet-professionele observatoren en beoordelaars van het onderwijs betrouwbare en valide resultaten bereikt kunnen worden. Het instrument bestaat uit elf indicatoren en is genormeerd voor het gebruik in de diverse groepen van het basisonderwijs, zodat het voor evaluatiedoeleinden gebruikt kan worden.

De 'gemiddelde onderwijsgevende op de gemiddelde basisschool' blijkt aan ten minste zeven van de elf indicatoren te voldoen. Daarmee blijkt de gemiddelde basisschool royaal aan de norm voor een voldoende pedagogisch klimaat te voldoen. 40% van de basisscholen blijkt te voldoen aan de stringentere norm voor een goed pedagogisch klimaat.

Het pedagogisch klimaat in de taal- en rekenlessen verschilt nauwelijks van dat in de lessen expressie, maar scoort iets lager dan het klimaat in de wereldoriëntatielessen. De verschillen zijn echter te klein om van betekenis te zijn.

Wel zijn er significante verschillen tussen het klimaat in de leeftijdsgroep 1/2 en het klimaat in de overige leeftijdsgroepen. In groep 1/2 voldoet ruim driekwart van de onderwijsgevers aan de norm voor een voldoende pedagogisch klimaat. Bij de overige leeftijdsgroepen voldoet bijna tweederde van de onderwijsgevers aan deze norm. De verschillen in het klimaat bij groep 3/4, groep 5/6 en groep 7/8 zijn minimaal en van geen betekenis.

NOTEN

1. Werkadres: Inspectie van het Onderwijs, Park Voorn 4, postbus 2725, 3500 GS Utrecht. Telefoon: 030 6665704. E-mail: W.vandeGrift@owinsp.nl.

2. Werkadres: Universiteit Twente, OCTO, postbus 217, 7500 AE Enschede.

3. De nauwkeurigheid (ϵ) waarmee op basis van een steekproef een proportie in de populatie geschat kan worden, kan berekend worden met gebruikmaking van formule $\{n=N/(1+N(\epsilon)^2)\}$, waarbij (N) de populatie-omvang en (n) de steekproefomvang uitdrukt.

4. Bij de correctie voor attenuatie wordt de correlatie die in de steekproef tussen twee variabelen is berekend, gecorrigeerd voor de meetonbetrouwbaarheid. Hiervoor wordt de volgende formule gebruikt:

$\{R_{12}^c = R_{12} / \sqrt{(R_{11} * R_{22})}\}$, waarbij R_{12} de 'ruwe' correlatie is en R_{11} en R_{22} de betrouwbaarheid van respectievelijk de eerste en de tweede variabele.

5. Voor het vaststellen van de σ -waarde van de indicatoren is gebruik gemaakt van een stochastisch model. Dit brengt met zich mee, dat de σ -waarde vastgesteld wordt, gegeven een bepaald betrouwbaarheidsinterval. De maximale σ -waarde kan men bij een 95% betrouwbaarheidsinterval berekenen door bij elke σ -waarde $\{1.96 \times (\text{s.e.})\}$ op te tellen. De minimale waarde bedraagt bij dit betrouwbaarheidsinterval het verschil van de σ -waarde en $\{1.96 \times (\text{s.e.})\}$.

6. Ter oriëntatie: Voor de schoolleiderinterventieschaal bedroeg de gemiddelde biseriële correlatie .76. (Vergelijk Van de Grift, 1987.)

7. De letter g wordt hier gebruikt voor het aantal groepen waartussen vergelijkingen worden gemaakt van de σ -waarde van de indicatoren en de letter k wordt gebruikt voor het aantal indicatoren in de schaal.

8. De formule voor het Raschmodel ziet er als volgt uit: $\frac{2.71^{(2p - \alpha)}}{1 + 2.71^{(2p - \alpha)}}$

9. De letter k wordt gebruikt voor het aantal indicatoren in de schaal.

10. Het is in deze situatie helaas niet mogelijk om de coëfficiënt Cohens κ te berekenen. Deze in Nederland

veelgebruikte techniek om de betrouwbaarheid van onder meer opstelbeoordeling na te gaan is alleen te gebruiken, wanneer alle beoordelaars alle objecten beoordelen. In het onderhavige geval is het pedagogisch handelen in de lessen beoordeeld door wisselende duo's van beoordelaars.

11. Er wordt hier een stringenter eis geformuleerd, dan gebruikelijk. Hiervoor zijn twee redenen. De eis van consensus kan stringenter zijn, omdat de inspecteurs beroepsmatige evaluatoren van het onderwijs zijn. De consensus eis moet stringenter zijn omdat vermeden moet worden, dat de inspectie in gelijke situaties tot ongelijke oordelen zou komen.

12. Met gebruikmaking van de formule $\{(\mu_1 - \mu_2) / \sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}\}$ is nagegaan of verschillen significant zijn op het 95% niveau.

13. Door missende waarden (zie de paragraaf over de inhoudsvaliditeit) kon deze correlatiecoëfficiënt slechts over een beperkt aantal respondenten berekend worden.

14. Door missende waarden (zie de paragraaf over de inhoudsvaliditeit) kon deze correlatiecoëfficiënt slechts over een beperkt aantal respondenten berekend worden.

LITERATUUR

- Allerup, P. & G. Sorber (1977). *The Rasch model for questionnaires, with a computer program*. Kopenhagen: Danmarks Paedagogiske Institut.
- Andersen, E.B. (1973). A Goodness of fit test for the Rasch model. *Psychometrika*, 38, 123- 140.
- Arter, J.A. (1986). *Assessing school and classroom climate. A consumer's guide*. Portland: Northwest Regional Educational Lab.
- Brandsma H.P. & K.T. Bos (1994). *Vaststelling en waardering van het pedagogisch klimaat in basisscholen*. Enschede: OCTO.
- Brookover, W.B. & J.M. Schneider (1975). Academic environments and elementary schoolachievement. In: *Journal of research and development in education*, vol 9, nr 1.
- Cuban, L. (1983). Effective schools: a friendly but cautionary note. In: *Phi Delta Kappan*.
- Eggen, T.J.H.M. & P.F. Sanders (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Wenen: Verlag Hans Huber.
- Fischer, G.H. & L.W. Molenaar (1975). *Rasch models, foundations, recent developments and applications*. New York: Springer Verlag.
- Fisher, D.L. (1991). *The study of learning environments*. Tasmania: Department of Education, University of Tasmania at Launceston.
- Fraser, B.J. (1986A). Two decades of research on classroom psychosocial environment. In Fraser, B.J. (1986). *The study of learning environments*. Vol. 1. Salem: Assessment Research.
- Fraser, B. (1986B). *Classroom environment*. London: Croom Helm.
- Gottfredson, G.D. & M.W. Stewart (1989). *The internal structure and correlates of the South Carolina needs assessment instruments*. Washington D.C.: Office of educational research and improvement.
- Gottfredson, G.D. & D.C. Gottfredson (1989). *School climate, academic performance, attendance and dropout*. Washington D.C.: Office of educational research and improvement.
- Grift, W. van de (1987). *De rol van de schoolleider bij onderwijsvernieuwingen*. 's-Gravenhage: VUGA uitgeverij.
- Grift, W. van de (1994). *Het evalueren van het pedagogisch klimaat in 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). *PML: a computer program for conditional estimation and testing in the Rasch model for dichotomous items*. Göteborg: Institute of education, University of Goteborg.
- Gustafsson, J.E. (1979b). *Testing and obtaining fit of data to the Rasch model*. Göteborg: Institute of Education, University of Göteborg.
- Hambleton, R.K. & L. Murray (1983). Some goodness of fit investigations for item response models. In: Hambleton, R. K. (1983). *Applications of item response theory*. Vancouver: Educational Research Institute of British Columbia.

- Hambleton R.K. & H. Swaminathan (1985). *Item response theory: principles and applications*. Boston, Dordrecht, Lancaster: Kluwer/Nijhoff Publishing.
- Houtveen, Th. & Vermeulen, C. & Grift, W. van de (1993). *Bouwstenen voor onderzoek naar de kwaliteit van scholen*. Utrecht: ISOR.
- Kuder, G.F. & M.W. Richardson (1937). The theory of the estimation of test reliability. *Psychometrika*, 1937, 2, 151-160.
- Lamberigts, R.J.A.G. (1987). Sociaal klimaat in de klas. *Losbladig onderwijskundig lexicon*. Alphen aan den Rijn: Samsom.
- Langeveld, M.J. (1971). *Beknopte theoretische pedagogiek*. Groningen: Wolters Noordhoff.
- Linn, R.L. (1980). Issues of validity for criterion-referenced measures. *Applied Psychological Measurement*, 1980, 4, 4.
- Martin-Löf, P. (1973). *Statistiska Modeller Anteckningar från seminarier lasaret 1969-1970 utarbetade af Rolf Sundberg*. Stockholm: Institutet för forsakringsmatematik och matematisk statistik.
- Mokken, R.J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton.
- Molenaar, I.W. (1989). *PML-PC, gebruikers handleiding*. Groningen: Progamma.
- Moos, R. H. (1973). Conceptualizations of human environments. *American psychologist*, 28, 652-665.
- Moos, R. H. & E.J. Trickett (1987). *Manual for the classroom environment inventory*. Palo Alto (CA): Consulting psychologists press.
- Noordam, N.F. (1970). *Inleiding in de historische pedagogiek*. Groningen: Wolters Noordhoff.
- Nunnally, J.C. & I.H. Bernstein (1994). *Psychometric theory*. New York: McGraw-Hill.
- Perquin, N. (1964). *Pedagogiek*. Roermond/Maaseik: J.J. Romen & Zonen.
- Purkey, S. C. & M.S. Smith (1983). Effective schools: a review. In: *The elementary school journal*, 83, 4.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagiske Institut.
- Rovinelli, R.J. & R.K. Hambleton (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Tijdschrift voor Onderwijsresearch*, 1977, 2, (49-60).
- Rutter, M., B. Maughan, P. Mortimore, J. Ouston & A. Smith (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Sommersets: Open books.
- Rutter, M. (1980). School influences on children behaviour and development. In: *Pediatrics*, vol. 65, nr 2, (208-220).
- Saldern, M. van (1992). *Social climate in the classroom*. New York: Waxmann Studies.
- Verhelst, N.D. (1992). *Het eenparameter logistisch model (OPLM)*. Arnhem: Cito.
- Waterink, J. (1951). *Theorie der opvoeding*. Kampen: Kok.
- Wright, B.D. & M.H. Stone (1979). *Best test design, Rasch measurement*. Chicago: MESA press.

BIJLAGE: EVALUATIE INSTRUMENT PEDAGOGISCH KLIMAAT

Standaard:

Het klimaat in de leerlinggroepen van de basisschool is van voldoende pedagogische kwaliteit.

Observeer elke leerkracht gedurende een les. Scoor op elke indicator een: **0** als de indicator niet of slechts enigszins van toepassing is op de geobserveerde leerkracht en scoor een: **1** als de indicator volledig van toepassing is op de geobserveerde leerkracht

indicatoren	groep			
	1/2	3/4	5/6	7/8
1 de leerkracht toont respect voor alle leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 de leerkracht is tolerant ten opzichte van alle leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 de leerkracht is ontspannen naar leerlingen toe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 de leerkracht formuleert opdrachten zo dat leerlingen weten wat van hen verwacht wordt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 in de klas hangt een positieve werksfeer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 de leerkracht betreft alle leerlingen bij het onderwijs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 de leerkracht creëert een sfeer van veiligheid en geborgenheid	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8 de leerkracht waardeert getoonde inzet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9 de leerkracht gaat in op de eigen inbreng van de leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10 de leerkracht zorgt voor een rijke leeromgeving	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11 de leerkracht stimuleert leerlingen zelf oplossingen te zoeken	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
totaal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

normering: Het pedagogisch klimaat op de school is ...	
onvoldoende	als het gemiddelde van de leerkrachten lager is dan 7
voldoende	als het gemiddelde van de leerkrachten tenminste 7 is
goed	als alle leerkrachten tenminste 7 scoren

Vijfentwintig jaar onderzoek naar "Leren en instructie": een persoonlijk perspectief

Erik De Corte¹

Centrum voor Instructiepsychologie en -Technologie (CIP&T), Universiteit te Leuven, België

Plenaire lezing gepresenteerd tijdens de Onderwijsresearchdagen 1998 georganiseerd aan de Universiteit Twente, Enschede, 13-15 mei 1998

ABSTRACT

This article reviews mainstreams and developments in inquiry on "learning and instruction" over the past twenty five years, using as a guideline the research on mathematics learning and problem solving of the Leuven Center for Instructional Psychology and Technology. Those headlines in the field of instructional psychology in the Dutch language region (The Netherlands and Flanders) are situated in the broader context of the major international trends in this domain of research.

The genesis of instructional psychology as a major subdomain of instructional science in the Dutch language region is situated in the 1970s. Initially the field was dominated by the Russian activity theory, but by the end of the decade the information-processing approach became more and more influential. Consolidation and internationalization were the two major trends during the 1980s; this is illustrated by a description of the extensive research carried out in Leuven on children's solution processes of simple addition and subtraction word problems. Finally, the developments during the 1990s are outlined focusing on new, main challenges for research on learning and instruction at the threshold of the 21st century, namely a deeper understanding of the processes of constructive, self-regulated and situated learning, and of the nature and the role of powerful instructional interventions to elicit those acquisition processes, and furthermore the elaboration of a strategy for bridging the research/theory-practice gap.

INLEIDING

Een overzicht geven van vijfentwintig jaar onderzoek naar "Leren en instructie" is uiteraard binnen het bestek van een enkele lezing niet mogelijk. Bovendien is mijn zicht op het gebied hoe dan ook toch partieel en selectief. Ik wil derhalve van bij de aanvang stellen dat het inderdaad, zoals in de titel staat, om een persoonlijk perspectief zal gaan, waarbij ik dan nog het eigen Leuvens onderzoek als leidraad zal gebruiken.

In dit eigen onderzoek lag de klemtoon steeds op het domein van het reken/wiskunde-onderwijs op de basisschool. Enige verklaring hiervoor is te vinden in mijn persoonlijke geschiedenis. Mijn basisopleiding is immers deze van onderwijzer en na mijn studies te Leuven in de Pedagogische Wetenschappen (met een zwaar accent op de psychologie), heb ik mijn doctoraatsproefschrift gemaakt op het gebied van precies het reken/wiskunde-onderwijs. Overigens lag dit doctoraatsonderzoek niet zo direct op het gebied van leren en instructie,

maar de centrale probleemstelling betrof het inventariseren en formuleren van onderwijsdoelstellingen in het perspectief van grootschalig peilingsonderzoek, een thema dat nu weer actueel is.

Na het doctoraatsonderzoek is mijn aandacht evenwel vrij snel verschoven naar de problematiek van leren en instructie, mede omdat ik aan de Leuvense Alma Mater belast werd met wat thans nog steeds noemt "Pedagogische psychologie". Meer bepaald ging de interesse vooral naar de cognitieve processen en structuren die aan de grondslag liggen van vaardig denken en probleemoplossen en naar de aard van de onderwijsleerprocessen die bij leerlingen het verwerven van deze hogere cognitieve vaardigheden faciliteren. Derhalve wordt sinds de jaren zeventig in het Centrum voor Instructiepsychologie en -Technologie (CIP&T) gewerkt aan de ontwikkeling van een empirisch onderbouwde onderwijsleertheorie van het probleemoplossend denken bij leerlingen van de basisschool. Omwille van de complexiteit van de problematiek en de overtuiging dat deze het best vakgebonden kan bestudeerd worden, werd - in de lijn van mijn vroegere interesse - het onderzoek toegespitst op het wiskundig probleemoplossen. Bij de uitwerking van het onderzoeksprogramma, waarbij sinds 1977 nauw samenwerkt wordt met Lieven Verschaffel, zijn er steeds goede en constructieve contacten geweest met Nederlandse onderzoekers.

DE JAREN ZEVENTIG: WORDING VAN HET ONDERZOEK NAAR LEREN EN INSTRUCTIE IN HET NEDERLANDSE TAALGEBIED

Het onderzoek naar leren en instructie - ook wel aangeduid met de term onderwijspsychologie - heeft zich tijdens de jaren zeventig in ons taalgebied geleidelijk ontwikkeld tot een min of meer zelfstandig onderzoeks domein op het kruispunt van psychologie en onderwijskunde. Dit loopt min of meer parallel met de ontwikkeling in de Verenigde Staten, alwaar Gagné en Rohwer in 1969 voor het eerst een overzichtsartikel publiceerden in de *Annual Review of Psychology* onder de titel "Instructional psychology".

Overigens werd het domein in die aanvangsperiode in belangrijke mate beheerst door de Utrechtse handelingspsychologische school onder leiding van Van Parreren (zie o.m. Van Parreren & Carpay, 1972). Tijdens die periode zijn er trouwens vrij intensieve contacten geweest tussen de Utrechtse groep en de toenmalige Leuvense Afdeling Didactiek en Psychopedagogiek. Zo werden er in 1975, 1976 en 1978 drie Neerpeltconferenties georganiseerd over (de Utrechtse) onderwijsproceeskunde (Van Parreren, 1981), (de Leuvense) didaxologie (De Corte, Geerligts, Lagerweij, Peters, & Vandenberghe, 1981) en de uitbouw van een theoretisch en methodologisch kader voor het onderzoek van onderwijsleerprocessen. Toen Van Parreren in het academiejaar 1978-1979 titularis was van de buitenlandse Francqui-leerstoel aan de Vrije Universiteit te Brussel werd aldaar een interuniversitair colloquium georganiseerd met deelname van een groot aantal Neerpelters. In 1980 en 1982 hadden nog twee conferenties plaats respectievelijk in Utrecht en Leuven, waar de recent afgesloten en lopende onderzoeksprojecten van beide groepen werden voorgesteld en besproken (De Corte & Span, 1983).

In de loop van die jaren zeventig had er zich evenwel inmiddels een belangrijke ontwikkeling voorgedaan. De Amerikaanse cognitief-psychologische benadering - overigens in haar ontstaan zelf beïnvloed door het oudere Europese gedachtegoed uit de Gestaltpsychologie en de Wurzburgse school - werd meer en meer invloedrijk in Europa in het algemeen en zeker ook in ons taalgebied. Van bijzondere betekenis is in dit verband de organisatie geweest te Amsterdam in juni 1977 van de NATO International Conference on Cognitive Psychology and Instruction (onder de dynamische leiding van Robert Glaser), waaraan deelgenomen werd door 150 onderzoekers uit 15 verschillende landen (Lesgold, Pellegrino, Fokkema, & Glaser, 1978). Dit alles leidde er overigens toe dat zeker reeds in de tweede helft van de jaren zeventig in het Nederlandse taalgebied a.h.w. twee scholen naast elkaar bestonden, de ene geba-

seerd op de Russische (Vygotskiaan)se handelingstheorie en de andere op de Amerikaanse informatieverwerkingstheorie. Bij verschillende gelegenheden werden beide scholen overigens met elkaar vergeleken en geconfronteerd, zo o.m. tijdens de ORD '78 te Utrecht (Kanselaar & Peeck, 1978) en in een symposium dat ik zelf georganiseerd heb tijdens de ORD '81 te Maastricht (De Corte, 1982). Bij dergelijke gelegenheden is weliswaar herhaaldelijk op bepaalde overeenkomsten gewezen (zie bijvoorbeeld Lodewijks, 1982). Toch meen ik met Span (1982) dat de verschillen fundamenteeler waren/zijn dan de gelijkenissen. Zo waren beide scholen weliswaar primair geïnteresseerd in de procesaspecten van het leren, doch zowel de conceptuele als de methodologische benadering van het leerproces bleef grondig verschillen. Vanuit conceptueel standpunt werden leerlingen in de Amerikaanse visie opgevat als informatieverwerkers (hetgeen allicht mede verklaart waarom zeker in de aanvangsfase van deze stroming tekstverwerking in het onderzoek een zo belangrijke plaats innam); in de Russische school daarentegen staat niet de notie informatieverwerking centraal, maar wel het concept handelen opgevat als doelgerichte activiteit. Op methodologisch vlak werd de Amerikaanse research steeds gekenmerkt door een hoge graad van interne validiteit, iets waar Russische onderzoekers vaak nogal licht mee omsprongen, althans naar Westerse maatstaven. Een ander gelijkenis is dat beide stromingen uiteindelijk beoogden bij te dragen tot de optimalisering van leerprocessen, maar in de Russische school tenderde men hier veel meer naar een sturende aanpak in vergelijking met de Amerikaanse opvatting. Tot een integratie waar wel eens op is aangedrongen (zie Span, 1978), is het alleszins niet gekomen, wellicht omdat dit ook niet goed mogelijk was (Span, 1982). Overigens heeft de handelingstheoretische stroming in de jaren tachtig meer en meer aan invloed ingeboet; ook hier is nog eens bevestigd geworden dat met het verdwijnen van de pionier en bezieler van een bepaalde stroming, ook de betekenis en de invloed van het ideeëngoed aftakelt. Derhalve kan men stellen dat bij het begin van de jaren tachtig de studie van leren en instructie uitgegroeid was tot een van de belangrijkste subdomeinen van het onderwijsonderzoek in het Nederlandse taalgebied, dat dan weliswaar niet exclusief maar toch in sterke mate aanleunde bij de cognitief-psychologische benadering.

Tot op zekere hoogte weerspiegelt de hiervoor geschetste ontwikkeling zich ook in ons eigen onderzoek. In de tweede helft van de jaren zeventig sloten wij met het Leuvens spuurwerk over wiskundig denken bij leerlingen uit de onderbouw van het basisonderwijs ook nogal aan bij de handelingstheoretische benadering. Illustratief hiervoor zijn de kwalitatief-psychologische analyses van de handelingsstructuren van eerste- en tweedeklassertjes (6- tot 8-jarigen) bij het oplossen van directe en indirecte (puntsommen) opgaven over optellen en aftrekken (bijvoorbeeld: $-9=8$; $15=+8$) (De Corte & Verschaffel, 1980a). Op grond van de antwoorden van 176 leerlingen op een schriftelijke toets bestaande uit 24 items en de hardop-denkenprotocollen van 12 kinderen, werden de adequate en inadequate handelingsstructuren van de leerlingen geanalyseerd en er werd vastgesteld dat 78% van de 1633 verkeerde antwoorden te typeren waren als fouten in het denken (en niet in het technisch) rekenen, d.w.z. fouten die resulteren uit het kiezen en uitvoeren van de verkeerde rekenoperatie om een opgave op te lossen. In een daaropvolgende construerend onderzoek werd met succes een experimenteel programma ontwikkeld en toegepast waarin systematisch onderricht werd gegeven om aan de vastgestelde tekorten, m.n. een deficiënte begripsvorming en een gebrek aan probleemoplossingstechnieken, te verhelpen (De Corte & Verschaffel, 1980b).

Vanaf het begin van de jaren tachtig werd het Leuvens onderzoek evenwel ook meer en meer georiënteerd op de informatieverwerkingsbenadering, inz. op het werk van Greeno en zijn medewerkers aan het Learning Research & Development Center (LRDC) te Pittsburgh (zie o.m. Riley, Greeno, & Heller, 1983) en van Carpenter en Moser aan het Wisconsin Research and Development Center te Madison (zie o.m. Carpenter & Moser, 1982). Enkele bezoeken aan LRDC en daaruit ontstane nauwe contacten en uitwisselingen waren daaraan niet vreemd.

DE JAREN TACHTIG: CONSOLIDATIE EN INTERNATIONALISERING VAN HET ONDERZOEK NAAR LEREN EN INSTRUCTIE

Dat de jaren tachtig kunnen beschouwd worden als het decennium waarin het onderzoek naar leren en instructie geconsolideerd werd, blijkt duidelijk wanneer men de programma's van de opeenvolgende Onderwijsresearchdagen nader bekijkt. Van bij de oprichting van de z.g. Onderzoeksthemagroepen binnen de Vereniging voor Onderwijsresearch, is "Onderwijsleerprocessen" een van de pijlers van het onderwijsonderzoek geweest. In een lezing op het eerste symposium van de Onderzoeksthemagroep, gehouden te Veldhoven in juni 1980, heb ik destijds gepleit voor een - zowel inhoudelijk als methodologisch - brede invulling van het domein, opgevat als de studie van het verloop en de resultaten van onderwijs- en leerprocessen in didactische situaties, en van de factoren die deze processen beïnvloeden (De Corte, 1981). Naar inhoud bedoelde ik hiermee dat men aandacht moet hebben voor:

- affectieve en psychomotorische processen en factoren, naast de cognitieve aspecten;
- processen en variabelen bij de individuele leerlingen, maar ook bij de leergroep en de leerbegeleiders en onderwijsgeevenden, en eveneens voor interacties tussen leerlingen en leerbegeleiders en tussen leerlingkenmerken en aspecten van onderwijsleersituaties;
- algemene wetmatigheden, individuele verschillen en ontwikkelingspsychologische aspecten.

Op methodologisch gebied hield ik een pleidooi voor een z.g. breedspectrumvisie, d.w.z. het gebruik maken van diverse types van onderzoeksofzet en van alle methoden en technieken die relevante gegevens over leer- en onderwijsprocessen kunnen opleveren, een visie die o.m. door De Klerk werd bijgetreden in zijn openingsrede van de Onderwijsresearchdagen 1982 te Tilburg. In dit verband verklaarde ik me o.a. ook voorstander van het in combinatie toepassen van constaterende studies en systematische onderwijsexperimenten of construerende onderzoeken, overigens helemaal in de lijn van het hiervoor kort beschreven voorbeeld over het oplossen van optel- en aftreksommen.

Hierbij aansluitend stelde ik inzake keuze van de onderzoeksthema's het in samenhang hanteren voor van twee algemene criteria en dit in het perspectief van het streven naar onderzoek waarin theorievorming en praktijkrelevantie hand in hand gaan:

1. de thema's moeten aansluiten bij reële problemen die zich in de onderwijspraktijk in brede zin stellen;
2. de thema's moeten mogelijkheden bieden om tot theorievorming over onderwijs- en leerprocessen te komen.

Mijn persoonlijke evaluatie van de ontwikkelingen binnen het vakgebied in de jaren tachtig is gematigd positief. Immers, reeds van in het begin van het decennium werd inderdaad naar een brede inhoudelijke invulling van de onderwijspsychologie gestreefd. Zo werd een van de ORD-bundels 1982 gewijd aan *Sociale en motivationele aspecten van het leren* (Boekaerts & Van Lieshout, 1982). In het "Woord vooraf" schrijven de redacteuren:

"In de onderwijskundige en onderwijspsychologische literatuur wordt de vraag naar ecologisch valide onderzoeksdesigns steeds duidelijker geformuleerd. Daardoor gaat de belangstelling van onderzoekers zich meer richten op die aspecten van het onderwijsleerproces, die in de reële klassetting kunnen worden onderzocht. Dit brengt met zich mee dat het eenzijdig accent op het cognitief functioneren van de leerling verschuift naar de studie van de factoren die het leerproces beïnvloeden. We zien dan ook een groeiend aantal publicaties die de sociale en de motivationeel-affectieve aspecten van het leerproces in de klassetting onderzoeken." (p. 1)

Tijdens dezelfde ORD werd ook uitvoerig aandacht besteed aan een aspect van het z.g. "teacher thinking", m.n. de planning van het onderwijzen (zie Halkes & Nijhof, 1982).

Verder reflecteert een derde bundel van de ORD 82 *Methodologie van de onderwijsresearch* (Van der Kamp & Van der Kamp, 1982) een brede kijk op onderzoeksmethoden, met aandacht voor o.m. ethnografische technieken en de Göteborgse fenomenografie.

Tenslotte kan opgemerkt worden dat de aandacht ook in toenemende mate uitging naar schoolrelevante opgaven en problemen uit de leervakken van het curriculum, hetgeen aansluit bij de twee genoemde criteria voor de keuze van onderzoeksthema's.

Het eigen Leuvens onderzoek uit die jaren tachtig was vooral toegespitst op de gedetailleerde analyse van de probleemoplossingsprocessen van lagere schoolleerlingen bij enkelvoudige redactie-opgaven en daarnaast op de studie van de mogelijkheden om via het leren programmeren in LOGO de denkvaardigheden van leerlingen uit de bovenbouw van het basis-onderwijs te bevorderen.

Het onderzoek over het oplossen van redactie-opgaven vormt een illustratie van de vrij succesvolle toepassing van de eerder beschreven breedspectrumvisie op de onderzoeksmethodologie. Bij dit speurwerk werd inderdaad op een gecombineerde wijze gebruik gemaakt van diverse technieken voor dataverzameling, m.n. schriftelijke toetsen, hardop denken, doorvragen, retrospectie en gedragsobservatie tijdens individuele interviews, oogbewegingsregistratie, handboekanalyse en construerend onderzoek. Deze reeks studies (voor een overzicht zie Verschaffel & De Corte, 1993) die stevig ingebed was in de internationale literatuur over het betreffende thema, leidde tot een aantal belangrijke resultaten over diverse aspecten van oplossingsprocessen, m.n. de prestaties van de leerlingen uitgedrukt in de moeilijkheidsgraad van de opgaven, de aard en de diversiteit van de oplossingsstrategieën en de aard en de ontstaanswijze van de fouten die kinderen maken. Ik kan hier slechts een zeer summier en partieel overzicht geven van deze resultaten.

Ten eerste werd vastgesteld dat de moeilijkheidsgraad van enkelvoudige vraagstukjes over optellen en aftrekken in sterke mate bepaald wordt door de semantische structuur van de opgave. Dit kwam vooral duidelijk tot uiting in de bevinding dat opgaven die met dezelfde rekenbewerking kunnen opgelost worden (bijvoorbeeld door van het grootste gegeven getal het kleinste af te trekken), maar die een verschillende semantische structuur hebben (veranderings-, combinatie-, of vergelijkingsopgave) een zeer uiteenlopende moeilijkheidsgraad hebben bij eersteklassertjes.

Ten tweede constateerden we een grote diversiteit aan oplossingsstrategieën, waarvan er vele door de kinderen zelf "uitgevonden" werden. De meest interessante bevinding was evenwel, dat opgaven die met eenzelfde bewerking kunnen opgelost worden, maar een andere semantische structuur hebben, ook verschillende oplossingsstrategieën uitlokken bij de leerlingen, en dit zowel op het materiële, als op verbale en mentale niveau van oplossen (bijvoorbeeld: op materieel niveau, bijvoegen van objecten bij een veranderingsopgave tegenover vormen van paren objecten bij een vergelijkingsopgave).

Ten derde is gebleken dat de bron van fouten bij enkelvoudige vraagstukjes veelal niet - zoals in het verleden meestal beweerd werd - ligt in het kiezen van de geschikte bewerking, maar wel in een daaraan voorafgaande fase in het oplossingsproces, namelijk de opbouw van een correcte representatie van de opgave. Kwalitatieve analyses van de fouten van leerlingen toonden verder aan dat ze doorgaans rechtstreeks en zeer systematisch voortvloeden uit de verkeerde representatie die een leerling van de probleemsituatie heeft opgebouwd. Oorzaken van deze onjuiste probleemvoorstelling bleken te zijn, ofwel een gebrekkige kennis van de eerder genoemde semantische schema's (verandering, combinatie, vergelijking) die aan de vraagstukjes ten grondslag liggen, ofwel het niet of verkeerd begrijpen van een woord of een uitdrukking uit de opgavetekst (bijvoorbeeld: het woord "samen" wordt opgevat als "ieder" of "beiden").

Belangrijk is dat deze onderzoeksresultaten een bijdrage hebben geleverd tot de theorievorming over het oplossen van enkelvoudige redactie-opgaven. Zo bleek dat de bestaande computermodellen over het oplossingsproces van dergelijke vraagstukjes (Riley, Greeno, & Heller, 1983) verfijnd en aangevuld moeten worden om rekenschap te kunnen geven van onze empirische gegevens over de moeilijkheidsgraad van de diverse soorten opgaven en over de wijze waarop kinderen deze opgaven (juist of verkeerd) representeren. Daarnaast leverden onze bevindingen over de aard en de ontwikkeling van oplossingsstrategieën een aantal belang-

rijke aanvullingen en nuanceringen bij voorafgaand onderzoek daaromtrent (Carpenter & Moser, 1982).²

Dit is slechts een selectie uit het Leuvens onderzoek op het gebied van het wiskundig probleemoplossen. Daarnaast werden ook studies verricht over de invloed van andere opga-vekenmerken dan de semantische structuur op het oplossingsproces bij enkelvoudige redactie-opgaven, m.n. de mate waarin de semantische relaties tussen de bekende en de onbekende hoeveelheden in de opgavetekst uitdrukkelijk en ondubbelzinnig geëxpliciteerd worden enerzijds, en de volgorde van de bekende en onbekende kwantiteiten in deze opgavetekst anderzijds (Verschaffel & De Corte, 1993). Met behulp van de techniek van de oogbewegingsregistratie werd o.m. een bijdrage geleverd tot het verklaren van de moeilijkheidsgraad van bepaalde (z.g. inconsistente) vergelijkingsopgaven (Verschaffel, De Corte, & Pauwels, 1992). Met het oog op een verbreding van de theorievorming werd eveneens reeds in de jaren tachtig de aard van de bestudeerde rekenopgaven verruimd naar vermenigvuldigings- en delingsvraagstukken (De Corte, Verschaffel, & Van Coillie, 1988).

Een tweede Leuvense onderzoekslijn die vooral in de tweede helft van de jaren tachtig werd ontwikkeld in een project over "Computers en leren denken", had betrekking op de cognitieve effecten van het leren programmeren in LOGO. Het is vooral in het raam van dit project dat construerend onderzoek werd verricht; in deze context introduceerden we ook het concept "krachtige onderwijsleeromgevingen": zulke omgevingen "laten enerzijds voldoende ruimte voor het zelfstandig exploreren van leertaken en projecten in wisselwerking met medeleerlingen, maar bieden tegelijk voldoende systematische begeleiding rekening houdend met individuele mogelijkheden en behoeften van elke leerling." (De Corte & Verschaffel, 1987a, p. 370; zie ook De Corte, 1990).

Gedurende een jaar werd gewerkt aan de ontwikkeling van een krachtige LOGO-leeromgeving, die vervolgens gespreid over het ganse schooljaar 1987-1988 geïmplementeerd en geëvalueerd werd in drie zesde klassen (12-jarigen) gebruik makend van de voortoets-natoets opzet met controlegroep. Bij confrontatie met het inmiddels zeer bekend geworden "cognitieve apprenticeship" model, dat rond dezelfde tijd tot stand kwam, bleek dat onze onderwijsleeromgeving belangrijke overeenkomsten vertoonde met dit model van Collins, Brown en Newman (1989), zowel wat betreft de nagestreefde doelen (geïntegreerd verwerven van domeinspecifieke kennis en strategische denk- en leervaardigheden) als de toegepaste leer- en instructietechnieken (modelleren, coaching, "scaffolding", discussie en reflectie in klasverband en tijdens groepswork). Uit de resultaten van het onderzoek kwam naar voren dat de leeromgeving inderdaad effectief was wat betreft het verwerven door de leerlingen van de domeinspecifieke LOGO-kennis en de beoogde denkvaardigheden binnen de LOGO-omgeving; daarenboven werd een significant nabij transfereffect vastgesteld. Een belangrijke conclusie die overigens in de lijn ligt van de bevindingen van andere onderzoekers, betreft de noodzaak van een krachtig instructiesysteem waarin de verwerving en de transfer van denkvaardigheden bij de leerlingen bewust en systematisch worden begeleid en nagestreefd (De Corte, Verschaffel, Schrooten, Indemans, & Hoedemaekers, 1990; voor een bondige samenvatting van het onderzoek zie De Corte, 1990).

Ik heb gemakshalve het eigen Leuvens spuurwerk gebruikt in een poging om aan te tonen dat men kan spreken van een consolidatie van het onderzoek naar leren en instructie in de loop der jaren tachtig. Hetzelfde geldt evenwel voor andere researchteams in het Nederlandse taalgebied. Deze ontwikkeling ging ook gepaard met een duidelijke internationalisering van het onderzoeksdomein tijdens dit decennium. Een eerste uiting hiervan is het feit dat elk jaar weer een groot aantal Nederlandse onderzoekers de Atlantische Oceaan overstak om de Annual Meeting van AERA bij te wonen en er zelfs, met wisselend succes wat de belangstelling van Amerikaanse zijde betreft, symposia te organiseren. Maar een tweede en m.i. belangrijker

fenomeen was de oprichting te Leuven in juni 1985 van EARLI tijdens de First European Conference for Research on Learning and Instruction. Hierbij wil ik erop wijzen dat de oprichting van EARLI een Nederlands-Vlaams initiatief was. Immers, de organisatiecommissie bestond uit twee Nederlanders - Hans Lodewijks en Pieter Span - en twee Vlamingen - Roger Parmentier en ikzelf - naast de meest Europees gerichte Amerikaanse onderwijspsycholoog ooit, m.n. de betreunde Dick Snow. Het mag ook wel vermeld worden dat EARLI er niet of toch zeer moeilijk zou gekomen zijn zonder de aanzienlijke steun in de beginjaren van SVO. Het is u allen bekend dat EARLI inmiddels is uitgegroeid tot de belangrijkste en internationaal - ook in de U.S.A. - zeer gewaardeerde vereniging van onderwijsonderzoekers, waarvan de twejaarlijkse conferenties tot nog toe steeds als van kwalitatief hoogstaand wetenschappelijk niveau beschreven zijn geworden. Ook het EARLI tijdschrift *Learning and Instruction* dat ik in 1991 als eerste editor mocht opstarten, heeft intussen een internationale status verworven, zoals blijkt uit de opname ervan in de "Social Sciences Citation Index" en "Current Contents". Het belangrijkste effect van de oprichting van EARLI lijkt me evenwel te zijn dat er duurzame, productieve contacten en uitwisselingen tussen Europese onderzoekers tot stand gekomen zijn en nog ontstaan, waarbij in het bijzonder ook de jongere generatie haar voordeel kan doen.

Ik moge hier nog vermelden dat in 1988 de ORD voor de eerste maal in Vlaanderen en meer bepaald in Leuven georganiseerd werd.

DE JAREN NEGENTIG: NIEUWE UITDAGINGEN VOOR HET ONDERZOEK NAAR LEREN EN INSTRUCTIE OP DE DREMPEL VAN DE 21STE EEUW

In de voorafgaande paragraaf heb ik gesteld dat de ontwikkelingen in het onderzoek naar leren en instructie tijdens de jaren tachtig als gematigd positief kunnen beoordeeld worden. Dit suggereert reeds dat er toch ook nog minder gunstige aspecten waren. Zo heeft de cognitief-psychologische benadering - door haar interesse in de processen van informatieverwerking bij meer complexe taken - ertoe bijgedragen dat het onderzoek zich veel meer dan voorheen ging richten op schoolrelevante opgaven en problemen, waardoor er een duidelijke "leervakgerichtheid" tot stand kwam in de onderwijspsychologie. Doch het accent in dit onderzoek kwam zeer sterk te liggen op het analyseren van de kennis en de vaardigheden die aan (competente) prestaties ten grondslag liggen. Er werd evenwel - zeker in vergelijking met het behavioristische tijdperk - vrij weinig aandacht besteed aan de leerprocessen die nodig zijn om kennis en vaardigheden te verwerven (Glaser & Bassok, 1989). Zo is er overtuigende empirische evidentie dat metacognitieve kennis en vaardigheden belangrijke determinanten zijn van het leren en de leerprestaties op school, maar het onderzoek naar hoe leerlingen die kennis en vaardigheden het best kunnen verwerven staat nog in de kinderschoenen (zie Hacker, Dunlosky, & Graesser, 1998). Daarnaast werd in de cognitieve aanpak de kennis te eenzijdig opgevat als "an integral self-sufficient substance, theoretically independent of the situations in which it is learned and used" (Brown, Collins, & Duguid, 1989, p.32), en het denken te zeer beschouwd als een puur mentaal proces dat zich afspeelt in het menselijk brein (Greeno, 1989); m.a.w. er werd onvoldoende rekening gehouden met het situatiegebonden karakter van kennis en denken.

Een andere belangrijke lacune waar herhaaldelijk op gewezen werd, betreft de blijvende kloof tussen onderzoek en theorie enerzijds en de onderwijspraktijk anderzijds. In haar Presidential Address tijdens de AERA Annual Meeting in 1994 betoogde Ann Brown (1994) in dit verband het volgende:

"* Enormous advances have been made in this century in our understanding of learning and development.

* School practices in the main have not changed to reflect these advances." (p.4)

Ann Browns evaluatie van de situatie weerklinkt in het standpunt dat Franz Weinert en ikzelf (Weinert & De Corte, 1996) terzake hebben geformuleerd in de *International encyclopedia of developmental and instructional psychology*: "After 100 years of systematic research in the fields of education and educational psychology, there is, in the early 1990s, still no agreement about whether, how, and under what conditions research can improve educational practice. Although research and educational practice have changed substantially since the beginning of the twentieth century, the question of how science can actually contribute to the solution of real educational problems continues to be controversial." (p.43)

Aandacht voor het onderzoek van leerprocessen

Met het analyseren van voornoemde en andere uitdagende kritische bedenkingen en het zoeken naar passende antwoorden, is inmiddels reeds een aanvang gemaakt. Zo is er de laatste tien jaar bij onderwijspsychologen weer een sterke interesse gegroeid voor de studie van leerprocessen. In de U.S.A. verscheen reeds in 1990 een themanummer van de *Review of Educational Research* (Iran-Nejad, McKeachie, & Berliner, 1990) onder de titel "Toward a unified approach to learning as a multisource phenomenon". In Europa werd ook in het begin van de jaren negentig met steun van de European Science Foundation, een interdisciplinair programma voor onderzoek van het leren opgestart onder de titel "Learning in Humans and Machines" (Reimann & Spada, 1995). Dit onderzoek naar leren werd overigens verrijkt en bevrucht door bijdragen vanuit de research over het leren in niet-schoolse en informele leersituaties, zoals antropologische studies over "apprenticeship" en onderzoeken in de industriële en bedrijfswereld (Resnick, 1987).

Dit spuurwerk heeft inmiddels geleid tot het identificeren van een reeks samenhangende kenmerken van effectieve leerprocessen in educatieve situaties die de bouwstenen vormen voor wat Bereiter (1990) een "educational learning theory" genoemd heeft. Deze kenmerken waarover thans in de literatuur een vrij ruime consensus bestaat (zie bijv. Brown & Campione, 1994; De Corte, 1995; Shuell, 1992), kunnen samengevat worden in de volgende definitie van leren: het is een constructief, cumulatief, zelfgereguleerd, doelgericht, gesitueerd, collaboratief en individueel verschillend proces van kennisverwerving, betekenisgeving en vaardigheidsontwikkeling. Toch zijn we nog veraf van een goed uitgewerkte, empirisch onderbouwde theorie van wat ik in het Engels noem "a theory of learning from instruction" (het door instructie begeleid leren).

Zo is overtuigend aangetoond dat leren inderdaad een constructief proces is, zelfs in leeromgevingen die gebaseerd zijn op een kennisoverdrachtsmodel. Doch zoals aangetoond werd door Phillips (1995) in een artikel met de titel "The good, the bad, and the ugly: The many faces of constructivism", komen er in de literatuur veel uiteenlopende versies van het constructivisme voor, die onderling sterk verschillen wat de theoretische uitgangspunten en epistemologische grondslagen betreft. Zoals betoogd werd door Fischbein (1990) met betrekking tot het wiskunde-onderwijs, is er nood aan een "more specific definition of constructivism as a psychological model for mathematics education." (p.12) Dit is des te meer het geval omdat de huidige constructivistische benaderingen geen bruikbare principes aanreiken voor het ontwerpen van krachtige leeromgevingen (Greer, 1996; Janvier, 1996). Met andere woorden, voortgezet theoretisch en empirisch onderzoek moeten gericht zijn op het verwerven van een grondiger inzicht in en een meer verfijnde analyse van de aard van constructieve leerprocessen die kunnen leiden tot competentie in diverse vakgebieden; tevens dient dit spuurwerk licht te werpen op de aard en de rol van instructie in het uitlokken en op gang houden van dergelijke leerprocessen.

Op een gelijkaardige wijze kan men stellen dat tien jaar onderzoek interessante beschrijvende resultaten heeft opgeleverd omtrent de zelfregulatie van het leren. Zo werd aangetoond dat zelfregulatie sterk correleert met leerprestaties en de belangrijkste kenmerken van leerlingen met een goede zelfregulatie werden geïdentificeerd en beschreven (Zimmerman & Risem-

berg, 1997). Doch zoals betoogd door Zimmerman (1994), de verklarende kracht van het construct "zelfregulatie" is nog steeds erg beperkt en we hebben nog maar weinig inzicht in de processen die een essentiële rol spelen in het effectief reguleren van het leren op school, in de ontwikkeling van zelfregulatievaardigheden bij leerlingen en in de cruciale aspecten van de instructie die de overgang van externe regulatie naar zelfregulatie bevorderen.

Mede onder invloed van de Vygotskiaanse opvatting over leren en ontwikkeling (Vygotsky, 1978), die in de loop der jaren tachtig een groeiende aandacht kreeg bij Westerse onderwijs- en ontwikkelingspsychologen, heeft de visie op kennen, denken en leren als gesitueerde processen - in reactie tegen het eenzijdig mentalisme van de cognitieve psychologie - terecht het belang beklemtoond van contextuele, sociale en culturele factoren voor de cognitieve ontwikkeling en voor het verwerven en toepassen van kennis en denkvaardigheden. Bruner (1996) heeft hieraan recent een lezenswaardig boek gewijd onder de titel *The culture of education*. Maar zoals betoogd werd door Gruber, Law, Mandl, & Renkl (1995): "Situating cognition still remains a 'loosely coupled' school of thought." (p. 177), en is recent door Anderson, Reder, en Simon (1996) vanuit het cognitivistisch perspectief bekritiseerd geworden voor het naar voren brengen van inaccurate en eenzijdige beweringen waaruit misleidende implicaties voor de onderwijspraktijk kunnen resulteren. Dit heeft overigens aanleiding gegeven tot interessante discussies in de literatuur (zie o.m. het eerste nummer van de jaargang 1997 van de *Educational Researcher*). Derhalve is er ook hier een grote behoefte aan creatief theoretische en empirisch onderzoek, dat wellicht het best geleid kan worden door het streven naar een nieuwe synthese waarin de positieve aspecten van de verschillende stromingen - de cognitieve en de gesitueerde benadering, maar eventueel ook aspecten van het behaviorisme - geïntegreerd worden. In dit verband onderschrijf ik het standpunt van Stella Vosniadou (1996): "Bringing brain and culture back in the study of cognition is an important development in the direction of the scientific study of learning." (p.105)

Een poging in de richting van een dergelijke integratie werd overigens recent gepubliceerd door niemand minder dan Greeno (1998), een van de voortrekkers van het gesitueerd leren (zie ook Greeno, Collins, & Resnick, 1996; Sfard, 1998). Het streven om een eenzijdige "situatieve" visie te overstijgen ligt ook aan de basis van het recente boek *Situated cognition: Social, semiotic, and psychological perspectives* (Kirshner & Whitson, 1997)

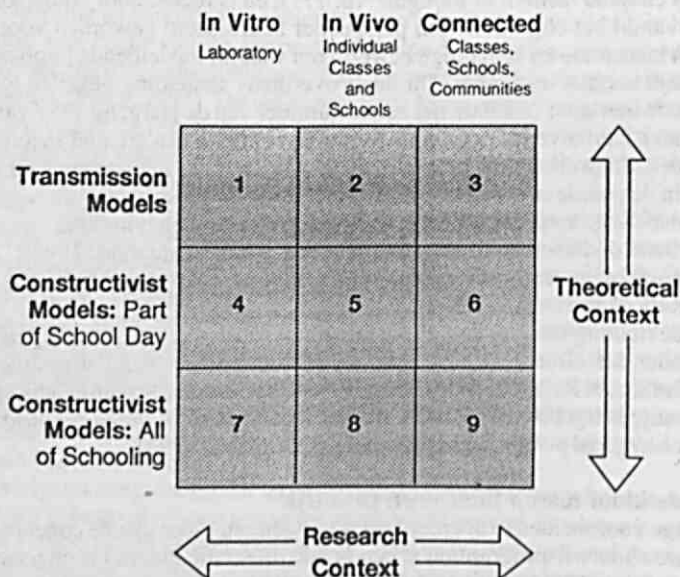
Aandacht voor de kloof tussen theorie en praktijk

Er zijn zeker aardige voorbeelden te noemen van pogingen om theoretische concepten te verwerken in bruikbare onderwijsprogramma's; een goede illustratie hiervan is ongetwijfeld het werk van Mettes & Pilot (1980) uit de late jaren zeventig over het leren probleemoplossen, gebaseerd op uitgangspunten van de Russische handelingspsychologie, inz. de theorie van Gal'perin. Toch is het onmiskenbaar dat de kloof tussen theorie en praktijk een aanhoudend pijnpunt blijft van het onderwijsonderzoek in het algemeen en het is ook de achillespees van de onderwijspsychologie. Elders heb ik, samen met Franz Weinert (Weinert & De Corte, 1996) betoogd dat een beloftevolle strategie voor het overbruggen van de bekende kloof bestaat "in the creation and evaluation in real classrooms of complex instructional interventions that embody our present (hypothetical) understanding of effective learning processes and powerful learning environments. In order to make a reasonable chance of success, such attempts at totally changing the classroom environment and culture should be undertaken in partnership between researchers and knowledgeable practitioners." (p. 48)

Een analoge visie - zij het meer uitgewerkt specifiek in verband met onderzoek over educatieve technologie - werd recent naar voren gebracht door de bekende Cognition and Technology Group at Vanderbilt (1996). Het z.g. "Looking at Technology in Context (LTC)" referentiekader relateert een dimensie die betrekking heeft op de educatieve context waarin onderzoek verricht wordt aan een tweede dimensie die verschillende leertheoretische standpunten representeert (zie Figuur 1). De dimensie "research context" gaat van onderzoek in

laboratoriumsituaties over studies in afzonderlijke, reële klassen naar speurwerk in bredere leergemeenschappen die tot stand komen door het onderling verbinden van klassen en scholen. De dimensie "theoretical context" gaat van onderzoek gebaseerd op een overdrachtsmodel van leren, naar studies waarin onderdelen van het curriculum aangepakt worden volgens een constructivistische benadering naar speurwerk dat erop gericht is het ganze schoolgebeuren te transformeren op constructivistische basis. De CTGV wijst erop dat het overgrote deel van het onderzoek over educatieve technologie uit het recente verleden thuishoort in de bovenste rij van het LTC-referentiekader en dat zulke vormen van technologiegebruik ook vrij gemakkelijk in te passen zijn in de traditionele onderwijspraktijk. De uitdaging bestaat erin het onderzoek meer en meer te verschuiven naar de tweede en de derde rij van het LTC-model; toepassingen van technologie die zich in de tweede en zeker in de derde rij situeren vergen evenwel fundamentele veranderingen in het klas- en schoolgebeuren.

Figuur 1: LTC (Looking at Technology in Context) Framework (CTGV, 1996)



Een aanpak in het onderzoek die beantwoordt aan het hiervoor ingenomen standpunt werd de laatste jaren door bepaalde onderzoekers naar voren gebracht onder de benaming "design experiments" (Brown, 1992; Collins, 1992) die gericht zijn op het uitbouwen van een "design science of education": "a design science must determine how different designs of learning environments contribute to learning, cooperation, motivation, etc." (Collins, 1992, p. 15)

Deze op interventie gebaseerde benadering, in nauwe samenwerking met praktijkmensen, heeft een tweevoudig doel dat overigens goed aansluit bij mijn eerder al vermelde criteria voor de keuze van onderzoeksthema's uit 1981 (De Corte, 1981), m.n. bijdragen zowel aan de theorievorming over door instructie begeleid leren als aan de (fundamentele) vernieuwing van de onderwijspraktijk (Brown, 1994; zie ook Greeno, Collins, & Resnick, 1996). De onderliggende idee is dat een doelmatige manier om tot een beter inzicht te komen in (constructieve) leerprocessen, erin bestaat onderwijsleeromgevingen te ontwerpen die zulke processen uitlokken en op gang houden.

Het lijkt erop dat deze aanpak zich langzaam aandient als een hoofdrichting in het onderzoek naar leren en instructie. Een in 1996 verschenen boek onder redactie van Schauble & Glaser (1996) vormt hiervan een illustratie: "Innovations in learning: New environments for

education documents the growth of a new kind of interdisciplinary teamwork that is evolving among practitioners, researchers, teacher educators, and community partners. The premise of this work is that the design of learning environments and the development of theory must proceed in a mutually supportive fashion." (p.xi)

Een van de meest representatieve voorbeelden van een onderzoeksprogramma waarin een krachtige leeromgeving ontworpen en geëvalueerd wordt is het "Fostering Communities of Learners" project van Brown en Campione (Brown, 1994; Brown & Campione, 1996). Uitgaande van een aantal leerprincipes (zoals actief leren, belang van metacognitie en zelfregulatie, dialogisch en collaboratief leren) wordt in dit project de leeromgeving in de klas fundamenteel gewijzigd, enerzijds op basis van vernieuwende onderwijsvormen, zoals wederkerig of rolwisselend onderwijs en de "jigsaw"-methode, en anderzijds door het creëren van een gans nieuwe klascultuur.

In een recent afgewerkt onderzoek hebben we in het Leuvens Centrum voor Instructiepsychologie en -Technologie ook een "ontwerpexperiment" uitgevoerd, waarin een krachtige leeromgeving voor het vaardig leren oplossen van contextgebonden wiskunde problemen bij leerlingen uit de bovenbouw van de basisschool werd ontwikkeld en vervolgens uitgeprobeerd in vier vijfde klassen (11-jarigen). De ontwikkeling van de leeromgeving gebeurde in nauwe samenwerking en overleg met de leerkrachten van de betreffende klassen en de implementatie ervan werd door deze leerkrachten zelf uitgevoerd (dit in tegenstelling tot het eerder vermelde LOGO-onderzoek uit de jaren tachtig). Centraal in de leeromgeving stond het verwerven door de leerkrachten van een algemene oplossingsstrategie met bijzondere aandacht voor een aantal waardevolle heuristieken en metacognitieve vaardigheden. De leeromgeving was gebaseerd op drie pijlers: 1. een gevarieerde verzameling uitdagende, realistische contextopgaven (in tegenstelling tot de stereotiepe klassieke schoolvraagstukken); 2. een combinatie van een stel krachtige instructietechnieken en werkvormen, zoals modelleren, groepswork en klassikale discussie waarbij veel aandacht besteed werd aan het expliciteren van en reflecteren over oplossingsstrategieën; 3. een nieuwe klascultuur gericht op het stimuleren van een positieve houding bij de leerlingen tegenover probleemoplossen en op het herdefiniëren van de rol van de leerkracht en de leerlingen in de wiskundeles. Het effect van de leeromgeving die bestond uit een 20 lessen gespreid over een periode van ongeveer drie maanden, werd nagegaan via een voortoets-natoets-retentietoetsopzet met controlegroep. De resultaten van het onderzoek tonen aan dat de ontworpen leeromgeving een betekenisvol positief effect had zowel op de prestaties en de methododische aanpak (bijvoorbeeld het gebruik van heuristieken) van de experimentele leerlingen, als op hun houding tegenover wiskunde problemen en het oplossen ervan. Belangrijk is op te merken dat deze leerwinst niet enkel geobserveerd werd bij de sterke en de middelmatige leerlingen, maar ook bij de zwakkeren. Daarenboven stelden we vast dat de extra aandacht voor de vaardigheid in en de houding tegenover probleemoplossen helemaal niet ten koste ging van leervorderingen in de andere onderdelen van het wiskundecurriculum. Tenslotte heeft dit project ook duidelijk het belang aangetoond van de nauwe samenwerking tussen onderzoekers en leerkrachten voor het welslagen van een ontwerpexperiment (voor een gedetailleerd verslag van dit onderzoek, zie Verschaffel, De Corte, Van Vaerenbergh, Lasure, Bogaerts, & Ratnckx, ter perse).

Het is echter wel van belang om hier nog een methodologische kanttekening aan toe te voegen. Immers, het werken aan theorievorming via interventie-onderzoek vergt de uitwerking van een passende methodologie voor het opzetten van ontwerpexperimenten in complexe educatieve settings op een zodanige wijze dat er toch theoretisch valide conclusies uit de empirische gegevens kunnen afgeleid worden. Dit is nog steeds geen eenvoudige zaak, omdat het streven naar ecologische validiteit ons er niet van ontslaat om ook de interne validiteit te waarborgen. Traditioneel heeft - zoals eerder al aangestipt - de Westerse "educational psychology" zich eerder bekommerd om de interne validiteit. Als gevolg daarvan was het onderzoek veelal gekenmerkt door wat Salomon (1996) genoemd heeft "isolation-by-reduction of phe-

nomena and processes" enerzijds en "study of the decontextualized individual" anderzijds. Het hoeft geen betoog dat deze karakteristieken van het traditionele onderzoek haaks staan op basisideeën van de hiervoor geschetste visie op het onderzoek van door instructie begeleid leren. Uitgaande van de opvatting dat de onderwijspsychologie gericht is op het ontwikkelen van educatieve omgevingen, betoogt Salomon (1996) - in de lijn van de aanpak via "design experiments" - dat het keurmerk van de discipline moet bestaan in het ontwerpen en bestuderen van "composites", d.w.z. complexe arrangementen zoals leeromgevingen waarin een groot aantal elementen en factoren opereren en interageren. Uiteraard vereist een dergelijke benadering de ontwikkeling van zowel passende onderzoekszetters en -strategieën, als van adequate technieken voor het verzamelen, verwerken en analyseren van empirische gegevens. Dit geldt overigens niet enkel voor de studie van de "composites" als Gestalten; immers, we moeten ook nog de samenstellende elementen van die composieten onder de loep nemen. Zo hebben we in het daareven vermelde Leuvense project over wiskundig probleemoplossen de holistische aanpak aangenomen, maar we hebben evenzeer geworsteld met het analyseren op basis van videobanden van de heuristisch strategieën en metacognitieve processen tijdens het oplossen van problemen in groepjes van twee leerlingen.

SLOTBESCHOUWING

Ik neem aan dat doorheen deze presentatie duidelijk geworden is dat het onderzoek naar leren en instructie in het Nederlandse taalgebied tijdens de voorbije vijftig jaar aanzienlijke veranderingen heeft ondergaan, die overigens grotendeels internationale evoluties weerspiegelen. Deze ontwikkelingen hebben zowel betrekking op de basisoriëntatie van het onderzoeksgebied, de epistemologische grondslagen, de theoretische kaders en de methodologische benadering van de onderzoeksthema's. Het meest comprehensieve overzicht van de huidige stand van zaken van het domein - zij het met een sterk Amerikaans accent - vindt men ongetwijfeld in het in 1996 verschenen *Handbook of educational psychology* onder redactie van Berliner & Calfee (1996a). In hun "Afterword" bij dat indrukwekkende volume van meer dan 1000 bladzijden geven deze bekende onderwijspsychologen een zeer positieve evaluatie van de "educational psychology": "Some observers have expressed concern about the decline in the research leadership of educational psychologists. We see a quite different picture reflected in the pages of this *Handbook*. First, it is clear that our field has been and continues to be highly productive and remarkably influential. ... We are "on the roll"; the preceding chapters exhibit an astounding freshness of ideas and enthusiasm for endeavours. Psychology as a discipline is in the midst of a paradigm shift, and educational psychology, as part of this discipline, is certainly at the forefront of these developments." (Berliner & Calfee, 1996b, p.1020)

En verder: "And so, as we approach the close of the twentieth century, we find this a time for educational psychologists to proudly acknowledge their achievements. They have influenced the way and means for studying education, providing evidence to demonstrate that the scientific study of education is possible and useful. ... We should celebrate the honor of serving as *the* major influence on the development of a rich and fractious research community for the scientific study of education." (Berliner & Calfee, 1996b, p.1020)

Allicht zal niet iedereen het enthousiasme van Berliner en Calfee in dezelfde mate delen. In elk geval hoop ik dat hun hoge graad van voldoening over het verleden niet zal leiden tot een soort zelfgenoegzaamheid onder de onderwijspsychologen, maar veeleer zal stimuleren tot het uitbouwen van theoretisch en methodologisch onderzoek van hoogstaande kwaliteit. Immers, zoals uit het laatste gedeelte van deze presentatie gebleken is staan onderzoekers naar leren en instructie voor belangrijke uitdagingen, inzonderheid ook wat betreft het overbruggen van de kloof tussen onderzoek en theorie enerzijds en praktijk en opleiding anderzijds. Over dit laatste zijn Berliner & Calfee (1996b) overigens erg wijzigzaam in hun "Afterword".

NOTEN

1. Correspondentie-adres: Erik De Corte, Center for Instructional Psychology and Technology (CIP&T), Katholieke Universiteit Leuven, Afdeling Didactiek, Vesaliusstraat 2, B-3000 Leuven, Belgium. Tel. (int)-32-16-32.62.48 - Fax. (int)-32-16-32.62.74. E-mail: Erik.DeCorte@ped.kuleuven.ac.be

2. Een artikel over dit onderzoek werd in 1987 door "The Research Advisory Committee of the National Council of Teachers of Mathematics" in de U.S.A. bekroond met de "1987 Award for Outstanding Article" in de *Journal of Research in Mathematics Education* (De Corte & Verschaffel, 1987b).

REFERENTIES

- Anderson, J.R., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher*, 25(4), 5-11.
- Bereiter, C. (1990). Aspects of an educational learning theory. *Review of Educational Research*, 60, 603-624.
- Berliner, D.C., & Calfee, R.C. (Eds.). (1996a). *Handbook of educational psychology*. New York: Macmillan.
- Berliner, D.C., & Calfee, R.C. (1996b). Afterword. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 1020-1022). New York, NY: Macmillan.
- Boekaerts, M., & van Lieshout, C.F.M. (1982). Woord vooraf. In M. Boekaerts & C.F.M. van Lieshout (Eds.), *Sociale en motivationele aspecten van leren. Bijdragen tot de Onderwijsresearchdagen 1982* (pp. 1-2). Lisse: Swets & Zeitlinger.
- Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2, 141-178.
- Brown, A.L. (1994). The advancement of learning. *Educational Researcher*, 23(8), 4-12.
- Brown, A.L., & Campione, J.C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp.229-270). Cambridge, MA: The MIT Press.
- Brown, A.L., & Campione, J.C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Bruner, J. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.
- Carpenter, T.P., & Moser, J.M. (1982). The development of addition and subtraction problem solving skills. In T.P. Carpenter, J.M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 2-24). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cognition and Technology Group at Vanderbilt. (1996). Looking at technology in context: A framework for understanding technology and education research. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 807-840). New York: Macmillan.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (Nato ASI Series F: Computers and Systems Sciences, Vol. 96, pp. 15-22). Berlin: Springer-Verlag.
- Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.
- De Corte, E. (1981). Beschouwingen met betrekking tot het onderzoek van onderwijsleerprocessen, anno 1980. *Tijdschrift voor Onderwijsresearch*, 6, 163-173.
- De Corte, E. (Ed.). (1982). *Onderzoek van onderwijsleerprocessen. Stromingen en actuele onderzoeksthema's. Bijdragen tot de Onderwijsresearchdagen 1981* (SVO-Reeks 53). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- De Corte, E. (1990). Ontwerpen van krachtige leeromgevingen. In M.J. Ippel & J.J. Elshout (Eds.), *Training van hogere-orde denkprocessen. Bijdragen aan de onderwijsresearch*, no. 32 (pp. 133-147). Amsterdam/Lisse: Swets & Zeitlinger.

- De Corte, E. (1995). Learning theory and instructional science. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines. Towards an interdisciplinary learning science* (pp. 97-108). Oxford, UK: Elsevier Science Ltd.
- De Corte, E., Geerligs, C.T., Lagerweij, N.A.J., Peters, J.J., & Vandenberghe, R. (1981). *Beknopte didactologie* (Vijfde volledig herziene druk). Groningen: Wolters-Noordhoff
- De Corte, E., & Span, P. (Eds.). (1983). *Studies over onderwijsleerprocessen. Bijdragen aan een symposium ter gelegenheid van tien jaar Belgisch-Nederlandse samenwerking*. Leuven: Helicon.
- De Corte, E., & Verschaffel, L. (1980a). Kwalitatief-psychologische analyse van het oplossen van aanvankelijke rekenopgaven bij 6- à 8-jarige basisschoolleerlingen. *Pedagogische Studiën*, 57, 383-396.
- De Corte, E., & Verschaffel, L. (1980b). Een exploratief onderwijsexperiment met aanvankelijke rekenopgaven bij 6- à 8-jarige kinderen. *Pedagogische Studiën*, 57, 433-448.
- De Corte, E., & Verschaffel, L. (1987a). Cognitieve effecten van leren programmeren. *Nederlands Tijdschrift voor de Psychologie*, 42, 364-372.
- De Corte, E., & Verschaffel, L. (1987b). The effect of semantic structure on first graders' solution strategies of elementary addition and subtraction word problems. *Journal for Research in Mathematics Education*, 18, 363-381.
- De Corte, E., Verschaffel, L., Schrooten, H., Indemans, R., & Hoedemaekers, E. (1990). *LOGO als springplank voor het verwerven van denkvaardigheden bij zesdeklassers. Eindverslag van het project "Computers en leren denken"* (Leuvense Pedagogische Bijdragen). Leuven: ACCO.
- De Corte, E., Verschaffel, L., & Van Coillie, V. (1988). Influence of number size, problem structure and response mode on children's solutions of multiplication word problems. *Journal of Mathematical Behavior*, 7, 197-216.
- de Klerk, L.F.W. (1982). Uitdaging en teleurstelling. In L. van der Kamp & M. van der Kamp (Eds.), *Methodologie van onderwijsresearch. Bijdragen tot de Onderwijsresearchdagen 1982* (pp. 13-21). Lisse: Swets & Zeitlinger.
- Fischbein, E. (1990). Introduction. In P. Neshier & J. Kilpatrick (Eds.), *Mathematics and cognition: A research synthesis by the International Group for the Psychology of Mathematics Education* (pp. 1-13). (ICMI Study Series). Cambridge, England: Cambridge University Press.
- Gagné, R.M., & Rohwer, W.D. Jr (1969). Instructional psychology. *Annual Review of Psychology*, 20, 381-418.
- Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, 40, 631-666.
- Greeno, J.G. (1989). A perspective on thinking. *American Psychologist*, 44, 134-141.
- Greeno, J.G., Collins, A.M., & Resnick, L.B. (1996). Cognition and learning. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York, NY: Macmillan.
- Greeno, J.W., & the Middle School Mathematics Through Applications Project Group (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53, 5-26.
- Greer, B. (1996). Theories of mathematics education: The role of cognitive analysis. In L.P. Steffe, P. Neshier, P. Cobb, G.A. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 179-196). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gruber, H. Law, L.C., Mandl, H., & Renkl, A. (1995). Situated learning and transfer. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines. Towards an interdisciplinary learning science* (pp. 168-188). Oxford, UK: Elsevier Science Ltd.
- Hacker, D.J., Dunlosky, J., & Graesser, A.C. (Eds.). (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Halkes, R., & Nijhof, W.J. (1982). *Planning van onderwijzen. Bijdragen aan de Onderwijsresearchdagen 1982*. Lisse: Swets & Zeitlinger.
- Iran-Nejad, A., McKeachie, W.J., & Berliner, D.C. (Eds.). (1990). Toward a unified approach to learning as a multisource phenomenon. *Review of Educational Research*, 60, 509-624.
- Janvier, C. (1996). Constructivism and its consequences for training teachers. In L.P. Steffe, P. Neshier, P. Cobb, G.A. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 449-463). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kanselaar, G., & Peeck, J. (Eds.). (1978). *Strategieën in leer- en onderwijsprocessen. Bijdragen aan de Onderwijsresearchdagen 1978*. (SYO-Reeks 12). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- Kirshner, D., & Whitson, J.A. (1997). *Situated cognition: Social, semiotic, and psychological perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lesgold, A.M., Pellegrino, J.W., Fokkema, S.D., & Glaser, R. (Eds.). (1978). *Cognitive psychology and instruction*. New York/London: Plenum Press.
- Lodewijks, J.G.L.C. (1982). Van behavioristische leerpsychologie, via instructietechnologie, naar psycholo-

- gie van de informatieverwerking: Enige kenmerken van de Amerikaanse onderwijspsychologie. In E. De Corte (Ed.), *Onderzoek van onderwijsleerprocessen. Stromingen en actuele onderzoeksthema's. Bijdragen tot de Onderwijsresearchdagen 1981* (pp. 36-51). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- Mettes, C.T.C.W., & Pilot, A. (1980). *Over het oplossen van natuurwetenschappelijke problemen. Een methode voor ontwikkeling en evaluatie van onderwijs toegepast op een cursus Thermodynamica*. (OC Rapport nr. 42.) Enschede: T.H. Twente.
- Phillips, D.C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Research*, 24(7), 5-12.
- Reimann, P., & Spada, H. (1995). *Learning in humans and machines. Towards an interdisciplinary learning science*. Oxford, UK: Elsevier Science Ltd.
- Resnick, L.B. (1987). Learning in school and out. *Educational Researcher*, 16(9), 13-20.
- Riley, M.S., Greeno, J.H., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburgh (Ed.), *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.
- Salomon, G. (1996). Unorthodox thoughts on the nature and mission of contemporary educational psychology. *Educational Psychology Review*, 8, 397-417.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4-13.
- Schauble, L., & Glaser, R. (Eds.). (1996). *Innovations in learning: New environments for education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Shuell, T.J. (1992). Designing instructional computing systems for meaningful learning. In M. Jones & P.H. Winne (Eds.), *Adaptive learning environments: Foundation and frontiers* (Nato ASI Series F: Computers and Systems, Vol. 85, pp. 19-54). Berlin: Springer-Verlag.
- Span, P. (1978). Strategieën in leer- en onderwijsprocessen, een poging tot evaluatie. In G. Kanselaar & J. Peeck (Eds.), *Strategieën in leer- en onderwijsprocessen. Bijdragen aan de Onderwijsresearchdagen 1978* (pp. 188-201). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- Span, P. (1982). Inleiding tot het symposium: Stromingen in het onderzoek van onderwijsleerprocessen. In E. De Corte (Ed.), *Onderzoek van onderwijsleerprocessen. Stromingen en enkele actuele onderzoeksthema's. Bijdragen tot de Onderwijsresearchdagen 1981* (pp. 11-14). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- van der Kamp, L., & van der Kamp, M. (1982). (Eds.). *Methodologie van onderwijsresearch. Bijdragen tot de Onderwijsresearchdagen 1982*. Lisse: Swets & Zeitlinger.
- Van Parreren, C.F. (1981). *Onderwijsproceeskunde*. Groningen: Wolters-Noordhoff.
- Van Parreren, C.F., & Carpay, J.A.M. (1972). *Sovjetpsychologen aan het woord*. Groningen: Wolters-Noordhoff.
- Verschaffel, L., & De Corte, E. (1993). A decade of research on word-problem solving in Leuven: Theoretical, methodological, and practical outcomes. *Educational Psychology Review*, 5, 239-256.
- Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye-movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, 84, 85-94.
- Verschaffel, L., De Corte, E., Van Vaerenbergh, G., Lasure, S., Bogaerts, H., & Ratineckx, E. (1998). *Leren oplossen van wiskundige contextproblemen. Een onderwijsexperiment in de bovenbouw van de basisschool*. (Studia Paedagogica). Leuven: Universitaire Pers Leuven.
- Vosniadou, S. (1996). Towards a revised cognitive psychology for advances in learning and instruction. *Learning and Instruction*, 6, 95-109.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weinert, F.E., & De Corte, E. (1996). Translating research into practice. In E. De Corte & F.E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp. 43-50). Oxford, UK: Elsevier Science Ltd.
- Zimmerman, B.J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D.H. Schunk & B.J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational application* (pp. 3-21). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Zimmerman, B.J., & Risemberg, R. (1997). Self-regulatory dimensions of academic learning and motivation. In G.D. Pyle (Ed.), *Handbook of academic learning: Construction of knowledge* (pp. 105-125). San Diego, CA: Academic Press.

**NOOT VAN DE REDACTIE
BIJ HET ARTIKEL VAN PROF. DR. E. DE CORTE.**

Het artikel van prof. dr. E. De Corte wijkt af van de bijdragen die gewoonlijk in TOR worden geplaatst. Het is persoonlijker dan wij gewend zijn. Dit komt omdat het de weergave is van de geschiedenis van een onderzoeker die zich nu al meer dan vijftig jaar heeft bezig gehouden met onderzoek van "leren en instructie". In zijn bijdrage beschrijft prof. De Corte de wording van het onderzoek in de jaren zeventig, de consolidatie in de jaren tachtig en de uitdagingen voor de 21ste eeuw. Daarmee is het volgens de redactie een interessant artikel voor een ieder die zich bezighoudt met onderzoek van leren en instructie en niet in de laatste plaats voor hen die zich op dit moment daarvoor hopen te kwalificeren als AIO of OIO. Het artikel is een bewerking van de rede die prof. De Corte als 'invited address' uitsprak tijdens de laatste ORD. Voor publicatie heeft het manuscript de gebruikelijk review procedure doorlopen.

Notities en Commentaren

NOOT VAN DE REDACTIE BIJ NOTITIES EN COMMENTAREN

In de rubriek notities en commentaren treft u een bijdrage aan van prof. dr. J. Dronkers getiteld 'Het Trouw-rapportcijfer van scholen in het voortgezet onderwijs; een analyse van de inspectiegegevens over de schooljaren 1995/96 en 1996/97'. Daarna treft u een bijdrage aan van dr. W. Pijls getiteld 'De berekeningsmethode in "schoolprestatie" met daarop een antwoord van Dronkers. De redactie hecht eraan mede te delen dat de bijdrage van Dronkers en Pijls afzonderlijk zijn binnengekomen. De bijdrage van Pijls is dus geen reactie op de bijdrage van Dronkers in dit nummer maar op publicaties in het dagblad Trouw.

HET TROUW-RAPPORTCIJFER VAN SCHOLEN IN HET VOORTGEZET ONDERWIJS; EEN ANALYSE VAN DE INSPECTIEGEGEVENS OVER DE SCHOOLJAREN 1995/96 EN 1996/97¹

J. Dronkers

SCO-Kohnstamm Instituut, Universiteit van Amsterdam, Wibautstraat 4, 1091 GM Amsterdam

1. Inleiding

Het dagblad *Trouw* heeft vorig jaar via de rechter de beschikking gekregen over de inspectiegegevens over het jaar 1995/96 van alle scholen in het Nederlandse voortgezette onderwijs. Die krant heeft vervolgens deze inspectiegegevens, die alleen betrekking hebben op het intern rendement van die scholen, gepubliceerd (Agerbeek, Hageman, Kreulen & Lakmaker, 1997). Ten behoeve van deze publicatie heb ik met behulp van deze gegevens de 'toegevoegde waarde' van scholen uitgerekend, en die uitgedrukt in een rapportcijfer van 1 tot 10, met een zesde als gemiddelde (Dronkers, 1997). Deze publicatie deed veel stof opwaaien, niet alleen in de dag- en weekbladpers (Agerbeek, 1997a, 1997b), maar ook in de wetenschappelijke tijdschriften (van Putten & van der Kamp, 1998; Veenstra, Dijkstra, Peschar & Snijders, 1998). Nu *Trouw* ook de inspectiegegevens over het jaar 1996/97 van dezelfde scholen heeft verkregen (Agerbeek, 1998b) kan ik beter ingaan op een aantal bezwaren rond deze eerdere berekening van 'toegevoegde waarde'. Daarbij zal ik een aantal nieuwe analyses van de inspectiegegevens over het jaar 1995/96 presenteren en ook een vergelijking maken met de inspectiegegevens over het schooljaar 1996/97. Voor de goede orde wijs ik erop dat het in deze bijdrage niet gaat om een wetenschappelijk onderzoek, maar om een maatschappelijke dienstverlening met behulp van wetenschappelijke inzichten. Zoals ik elders reeds betoogd heb (Dronkers, 1998a: 143) behoort een dergelijke dienstverlening tot de taak van wetenschappers.

Het berekenen van de 'toegevoegde waarde' ten behoeve van *Trouw* is nodig om twee bezwaren tegen het alleen maar publiceren van ruwe gegevens over het intern rendement van scholen op te vangen (Dronkers, 1998a). Het eerste bezwaar is dat scholen alleen geordend zouden worden op een smalle indicator voor rendement, bijvoorbeeld het gemiddelde eindexamencijfer. Met eindexamencijfers als smalle indicator belooft men bovengemiddelde selectie tijdens de school en aan de poort met een predikaat 'goede school'. Het tweede bezwaar is dat geen rekening zou worden gehouden met de verschillen in sociaal-culturele samenstelling van de leerlingpopulaties van scholen. In dat geval krijgt bijvoorbeeld een school met veel leerlingen met weinig cultureel kapitaal (bijvoorbeeld door het opleidingsniveau van hun ouders), die er toch in slaagt zijn leerlingen gemiddelde schoolprestaties te laten halen, een even goed predikaat als een school met veel leerlingen met veel cultureel kapitaal, die dezelfde gemid-

delde schoolprestaties behalen. Toch is de eerstgenoemde school duidelijk beter dan de tweede school, omdat hij meer waarde 'toevoegt' aan kennis en vaardigheden die leerlingen van huis uit meekrijgen. Als men geen rekening houdt met deze verschillen tussen de leerlingpopulaties van scholen, belooft men bovengemiddelde selectie aan de schoolpoort en ver-groot onnodig het verschil tussen sociaal-cultureel 'arme' en 'rijke' scholen.

Om beide bezwaren te ondervangen heb ik toen een berekening in twee stappen uitgevoerd: 1. de berekening van het totaal intern rendement van scholen; 2. een correctie voor de sociaal-culturele samenstelling van de leerlingpopulatie van scholen. Het daaruit voortvloeiend rapportcijfer voor scholen is door *Trouw* terecht naast de afzonderlijke indicatoren van intern rendement en sociaal-culturele samenstelling van de leerlingpopulatie afgedrukt (Agerbeek et al, 1997). Het *Trouw*-rapportcijfer is daardoor voor de krantenlezer een samenvatting van de afzonderlijke indicatoren, die hij zelf kan controleren².

De totaal meting van het intern rendement voorkwam dat ik mij moest beperken tot eind-examencijfers en bood mij bovendien de mogelijkheid dit te corrigeren voor de kenmerken van de leerlingen. Ik heb een goed totaal intern rendement arbitrair gedefinieerd als een school met hoge eindexamencijfers, veel geslaagden, weinig zittenblijvers en weinig uitstroom in alle bijna klassen³ en onder een laag totaal intern rendement heb ik het tegengestelde hiervan verstaan. Ik heb voor deze operationalisatie gekozen, omdat ik veronderstel dat ouders graag weten op welke school er weinig leerlingen blijven zitten, er weinig ongediplomeerd uitstromen, en er veel wordt geslaagd met hoge eindexamencijfers. Deze operationalisatie van intern rendement heb ik voor elk onderwijstype afzonderlijk moeten berekenen, omdat het aantal klassen in de verschillende onderwijstypen verschilt en de eindexamens in de onderscheiden onderwijstypen anders zijn ingericht. Ik paste daarbij factoranalyse (PCA) toe om te zien of een dergelijke dimensie voor de verschillende onderwijstypen überhaupt bestond. Dat bleek bij alle onderwijstypen het geval, òf als eerste òf als tweede factor. Dat betekent dat de door mij arbitrair gedefinieerde intern rendementsmaat een reële dimensie in de data was. Maar het is niet de enige factor of in enkele gevallen zelfs niet de belangrijkste. Daarom is het de moeite waard die andere dimensies aan een nader onderzoek te onderwerpen. Dat gebeurt ook in dit artikel (paragraaf 6).

Er bestaan geen duidelijke maatstaven voor het voldoende of onvoldoende functioneren van scholen op het punt van hun intern rendement. Ook zijn er nergens indicaties wat men belangrijker vindt bij het intern rendement van een school: zittenblijven, ongediplomeerd schoolverlaters, slagen of hoge eindexamencijfers. Daardoor zijn er geen duidelijke indicaties wat bij een totaal meting van het intern rendement het zwaarste zou behoren te wegen. Daarom heb ik in tweede instantie met behulp van dezelfde factoranalyse de gewichten van de verschillende indicatoren berekend, zodanig dat scholen op de door mij gekozen intern rendementsmaatstaf zo goed mogelijk onderscheiden zouden kunnen worden. Dit gebruik van empirische gewichten is geen principiële maar een pragmatische keuze. Beter zou het zijn de gewichten van de verschillende indicatoren op normatieve gronden te kiezen, maar dat leek mij niet mijn taak als wetenschapper. Een andere mogelijkheid bij afwezigheid van normatieve gewichten is de gewichten vast te stellen op grond van een aantal schooljaren: er zijn echter nu maar pas inspectiegegevens over twee schooljaren beschikbaar. Door deze pragmatische keuze voor empirische gewichten krijgen de verschillende indicatoren soms uiteenlopende gewichten. Het is interessant na te gaan of deze gewichten in tijd stabiel zijn: dat zal in dit artikel ook gebeuren (paragraaf 5).

Hetzelfde probleem ontstond bij het corrigeren van het intern rendement van scholen voor de sociaal-culturele samenstelling van hun leerlingpopulatie. Welk gewicht moest aan de beschikbare indicator van de samenstelling van de leerlingpopulatie worden gegeven? De inspectiegegevens bevatten geen indicatoren voor de samenstelling van de leerlingpopulatie. De enige thans bruikbare indicator, die voor alle scholen beschikbaar was, bleek het percentage cumi-leerlingen te zijn, dat door het ministerie van OC&W verzameld wordt in het kader

van de bekostiging. Dit percentage heb ik bij gebrek aan beter gebruikt⁴. Deze indicator heb ik daarom gebruikt om met behulp van regressie-analyse de betekenis van het percentage cumi-leerlingen voor het intern rendement te berekenen en die parameter te gebruiken als gewicht. In dit artikel zal ook nagegaan worden of die gewichten stabiel in de tijd zijn (paragraaf 5).

Een betwistbaar punt in mijn meting van totaal intern rendement is de opname van zittenblijven zonder uitstroom als indicator daarvan. Men kan van mening zijn dat zittenblijven zonder uitstroom geen onderdeel uitmaakt van het intern rendement: het gaat er immers om dat leerlingen het diploma halen en niet hoelang zij daarover doen. In dit artikel zal ik ingaan op de gevolgen van het al dan niet opnemen van zittenblijven in de meting van het totaal rendement (paragraaf 7).

Een gebrek aan de vorig jaar gepubliceerde inspectiegegevens was dat zij het rendement van één jaar betreffen. Het is immers mogelijk dat een school in een ander jaar een veel hoger of veel lager rapportcijfer krijgt. In de discussie rond de rapportcijfers wordt dit probleem echter overdreven, omdat het rapportcijfer niet gebaseerd is op één eindexamenklas, maar op alle klassen (met uitzondering van de brugklas) in dat jaar. Nu vergelijkbare inspectiegegevens over een tweede schooljaar beschikbaar zijn, kunnen wij de stabiliteit van het rapportcijfer nader analyseren, wat in dit artikel ook zal gebeuren (paragraaf 2, 3 & 4).

Voor een goed begrip van de hieronder volgende analyses is van belang zich te realiseren dat de inspectiegegevens over beide schooljaren onder verschillende omstandigheden verzameld zijn. De gegevens over het schooljaar 1995/96 werden nog door de scholen aan de inspectie geleverd in de veronderstelling dat zij alleen door de inspectie gebruikt zouden worden en dat zij nooit voor een vergelijking tussen scholen gebruikt zouden worden. Door een rechtelijke uitspraak zijn deze gegevens openbaar geworden, wat nooit de bedoeling van de scholen of van de inspectie is geweest. De scholen en de inspectie hebben deze gegevens dus nooit serieus op hun betrouwbaarheid gecontroleerd⁵. Na de publicatie van de inspectiegegevens over het schooljaar 1995/96 in de maand oktober door *Trouw* heeft de inspectie de reeds bij de school verzamelde gegevens over het schooljaar 1996/97 ter controle nogmaals voorgelegd aan de scholen en daarbij duidelijk vermeld dat deze gegevens na enige tijd publiek gemaakt zouden worden. Ook zijn sommige inspecteurs⁶ actief geweest bij het helpen van scholen bij de correctie van de reeds door hen geleverde gegevens. Maar tegelijkertijd hebben niet alle inspecteurs alle geleverde gegevens vervolgens gecontroleerd. Dit heeft in een aantal gevallen betekend dat bepaalde groepen uitzonderlijke leerlingen terecht uit de gegevens over het schooljaar 1996/97 zijn verwijderd (allochtone leerlingen die vanuit het buitenland zijn instromen in het voortgezet onderwijs en asielzoekers), die in het schooljaar 1995/96 nog ten onrechte werden meegeteld. Het is echter niet duidelijk of alle inspecteurs op dit punt even actief zijn geweest en of een en ander even systematisch gebeurd is. Tenslotte is de omschrijving van de begrippen zittenblijver en uitstromer binnen een scholengemeenschap veranderd, omdat de gegevens nu op (neven)vestigingsniveau verzameld zijn in plaats van op hoofdstigingsniveau (Agerbeek, 1998a).

Inspectiegegevens bevatten alleen op schoolniveau geaggregeerde gegevens over het intern rendement van scholen. Het is derhalve onmogelijk voor de berekening van de toegevoegde waarde een multi-level analyse toe te passen. Het is overigens de vraag of dit wenselijk zou zijn. Geaggregeerde gegevens kunnen de vraag beantwoorden "hoe hoog het intern rendement van een school is vergeleken met die van scholen met een vergelijkbare leerlingpopulatie", multilevel-analyse geeft antwoord op de vraag wat "de gemiddelde vooruitgang is van vergelijkbare leerlingen van verschillende scholen" (Bosker, 1998). Deze vragen zijn niet identiek en dus ook hun antwoorden niet (Dronkers, 1998a). Beide vragen lijken mij voor een breed publiek en voor beleidsmakers belangrijk. Het verschil in antwoorden van geaggregeerde analyse en multi-level analyse wordt groter naarmate de binnengroepregressie afwijkt van de tussengroepregressie. Deze afwijking treedt inderdaad op, maar is relatief niet groot

in verhouding tot de totale variantie (Bosker, 1998). Dus voor de meeste scholen en leerlingen wijkt de binnengroepprogressie niet belangrijk af van de tussengroepprogressie en is het verschil in antwoord op beide vragen niet groot.

Bij de hier gepresenteerde analyses heb ik alle inspectiegegevens gebruikt, behalve als van een school erg veel gegevens ontbraken. Ook scholen met minder dan 50 leerlingen in het betrokken onderwijstype zijn niet in de analyses betrokken. Missende waarden zijn in voorkomende gevallen vervangen door het gemiddelde.

2. De stabiliteit van het Trouw-rapportcijfer in de tijd

De eerste stap in deze analyse is de berekening van het rapportcijfer voor scholen op grond van de inspectiegegevens 1996/97. In deze eerste stap gebruik ik dezelfde gemiddelden en standaarddeviaties als ik gebruikte bij de berekening van het rapportcijfer over de gegevens 1995/96. Ook gebruik ik dezelfde gewichten voor de verschillende indicatoren van het intern rendement en voor de correctie van de sociaal-culturele achtergrond van de leerling populatie.⁷ Ook heb ik dezelfde regels toegepast bij het verwijderen van scholen met te weinig leerlingen of met te veel missende gegevens. Een verschil in rapportcijfer in de beide jaren is dus geheel terug te voeren op het verschil tussen de inspectiegegevens van beide jaren. In de volgende vier tabellen worden de samenhangen tussen de rapportcijfers van de verschillende onderwijstypen weergegeven.

Tabel 1: Samenhang tussen Trouw-rapportcijfer VWO 1995/96 en 1996/97

'95/96	2	3	4	5	6	7	8	9	Totaal
3			3	3				1	7
4	1	3	5	9	4	2			24
5	1		12	34	35	9	1		92
6	1	1	4	44	98	31	6	1	186
7			2	17	71	34	5		129
8				2	4	7	2		15
9						1		1	2
Totaal	3	4	26	109	212	84	14	3	455

38% van de VWO-scholen krijgt in beide schooljaren hetzelfde rapportcijfer, 48% verschilt slechts één punt en 9% twee punten. De samenhang tussen deze twee rapportcijfers voor de VWO-scholen is daarom redelijk maar niet erg hoog: .40 (Pearson's R). Dat komt vooral door een aantal uitbijters waarbij het verschil tussen beide jaren groter is dan 2 punten (2% meer dan 2 punten lager en 1% meer dan twee punten hoger). In een aantal gevallen zou dat laatste herleid kunnen worden tot het niet meer meerekenen van allochtone leerlingen die vanuit het buitenland zij-instromen in het voortgezet onderwijs. Dat blijkt uit de vijf VWO-scholen die 3 of meer punten stegen⁸. Het betreft hier bijna allemaal scholen met een groot aantal van deze allochtone zij-instromers. Indien wij deze zeven uitbijters niet meerekenen, wordt de samenhang tussen beide rapportcijfers hoger: .46. Interessant is nog op te merken dat het gemiddeld rapportcijfer voor het VWO niet is gestegen, maar juist significant⁹ is gedaald (1995/96: 6.0; 1996/97: 5.9). Dit wijst erop dat de mogelijkheden voor VWO-scholen hun gegevens op te schonen beperkt zijn geweest. Ook heeft blijkbaar het weglaten van bepaalde categorieën leerlingen bij VWO-scholen door hun leerlingpopulatie weinig gevolgen voor hun rapportcijfer gehad.

Tabel 2: Samenhang tussen Trouw-rapportcijfer HAVO 1995/96 en 1996/97

'95/'96	3	4	5	6	7	8	9	10	Totaal
1			1						1
2	1		1						2
3	1		1	3	1				6
4		2	6	9	1	1			19
5		4	14	38	18	2		1	77
6		1	15	98	77	7	1	1	200
7			3	29	67	9	3	1	112
8				1	9	1			11
9					1	1			2
10					1				1
Totaal	2	7	41	178	175	21	4	3	431

42% van de HAVO-scholen krijgt in beide schooljaren hetzelfde rapportcijfer, 44% verschilt slechts één punt en 10% twee punten. De samenhang tussen beide rapportcijfers van de HAVO-scholen is daarom redelijk maar niet erg hoog: .43. Opnieuw gaat het hier om een relatief klein aantal uitbijters, waarbij het verschil groter is dan 2 punten: 0,2% lager en 3% hoger. Gedeeltelijk kunnen ook deze laatstgenoemde uitbijters verklaard worden door het niet meer meerekenen van bepaalde categorieën uitzonderlijke leerlingen (allochtone leerlingen die vanuit het buitenland zij-instromen in het voortgezet onderwijs, asielzoekers, dropout-projecten). Dat blijkt uit de lijst van 14 HAVO-scholen die 3 of meer punten stegen¹⁰. Het betreft hier bijna allemaal scholen met een groot aantal van deze allochtone zij-instromers. De sprong van een 3 naar een 7 wordt gemaakt door het Da Vinci College te Leiden die in 1995/96 nog wel zijn jeugdige asielzoekers ten onrechte meetelde, maar dat in 1996/97 terecht niet meer deed (Agerbeek, 1998a). Als dat laatste reeds in 1995/96 was gebeurd, had de HAVO van het Da Vinci in Leiden toen een 8 gehaald (Dronkers, 1998a). Indien wij deze 14 uitbijters niet meetellen, wordt de samenhang tussen de beide rapportcijfers hoger: .50. Voor de HAVO scholen is wel een opwaartse tendens in de schoolcijfers te zien: het gemiddeld rapportcijfer gaat significant omhoog van 6,0 voor 1995/96 naar 6,4 voor 1996/97.

Tabel 3: Samenhang tussen Trouw-rapportcijfer MAVO 1995/96 en 1996/97

'96/'97	5	6	7	8	9	Totaal
1	1					1
3	1	3		1		5
4	3	10	6	2	1	22
5	5	26	51	6		88
6	1	42	140	39	1	223
7	1	8	78	33	4	124
8			8	2	5	15
Totaal	12	89	283	83	11	478

Slechts 30% van de MAVO-scholen krijgt in beide schooljaren hetzelfde rapportcijfer, 48% verschilt slechts één punt, 22% verschilt twee punten, 2% verschilt 3 punten en 1% vier of meer punten. Toch is ondanks deze schijnbaar grote verschuivingen de samenhang tussen de rapportcijfers voor MAVO-scholen ongeveer even hoog als bij de andere schooltypen: .39. De gedeeltelijke verklaring voor deze verschuiving is hetzelfde als hierboven: bepaalde categorieën leerlingen zijn door de scholen niet meer in de inspectiegegevens opgenomen. Dat blijkt uit de lijst van 21 MAVO-scholen die 3 of meer punten stegen. Het betreft hier bijna allemaal scholen met een groot aantal van deze allochtone zij-instromers, e.d.¹¹ Indien wij deze 21 uitbijters niet meetellen, stijgt de samenhang tussen beide rapportcijfers tot .47. De andere definitie heeft bij de MAVO-scholen ook duidelijk geleid tot een significante verhoging van het gemiddelde cijfer van 6,0 voor 1995/97 naar 7,0 voor 1996/97. Dalers met meer dan 2 punten komen daarom ook niet voor. Maar deze verhoging van het gemiddeld rapportcijfer is blijkbaar zo systematisch, dat de samenhang tussen de rapportcijfers voor MAVO-scholen niet anders uitvalt dan bij andere schooltypen.

Tabel 4: Samenhang tussen Trouw-rapportcijfer VBO 1995/96 en 1996/97

'95/'96	2	3	4	5	6	7	8	9	Totaal
1	1								1
2			1						1
3			1	3					4
4		1	2	4	3	2			12
5			3	12	24	6			45
6			1	17	104	46	3	1	172
7				2	34	49	3		88
8					2	4	2		8
Totaal	1	1	8	38	167	107	8	1	331

54% van de VBO-scholen heeft in beide schooljaren hetzelfde rapportcijfer, 29% verschilt slechts één punt, 6% twee punten en 1% drie of meer punten. De samenhang is bij VBO-scholen iets hoger, maar niet veel hoger dan bij de andere schooltypen (.54). De verschillen in rapportcijfers zijn wel minder groot: geen enkele school daalt meer dan 2 punten en maar drie scholen stijgen meer dan 2 punten¹². Indien wij deze drie uitbijters niet meetellen wordt de samenhang tussen beide rapportcijfers .57. Het gemiddeld rapportcijfer stijgt dan ook significant van 6,0 voor 1995/96 naar 6,2 voor 1996/97. Het is wel merkwaardig dat de niet meer opnemen van bepaalde categorieën leerlingen in de inspectiegegevens relatief zo weinig gevolgen heeft voor het VBO rapportcijfer in vergelijking met HAVO- en MAVO-scholen, terwijl die categorieën toch ook op het VBO relatief veel moeten voorkomen. Gaven VBO-scholen in het verleden hun inspectiegegevens al veel beter op, zodat de verandering minder groot was? Hebben VBO-scholen minder gebruik gemaakt van de mogelijkheden hun inspectiecijfers te corrigeren? Heeft de hulp van de inspecteurs zich vooral gericht op schooltypen aan de bovenkant van het onderwijsstelsel?

Uit deze eerste resultaten blijkt duidelijk dat een deel van de verschillen in intern rendement tussen de twee schooljaren terug te voeren is op verschillen in definities, gehanteerd door inspectie en scholen. Op zich is het buiten de berekening van intern rendement houden van specifieke groepen uitzonderlijke leerlingen (jeugdige asielzoekers, buitenlandse leerlingen die vanuit het buitenland zij-instromen in het voortgezet onderwijs, drop-out projecten) juist,

omdat dat het zicht op het intern rendement vertroebelt. Wel moet dit dan systematisch gebeuren, wat voor het schooljaar 1996/97 nog niet zeker is.

In de volgende paragrafen zullen wij de verschillen tussen de rapportcijfers en de inspectiegegevens van de beide schooljaren nader analyseren. Ik beperk mij daarom om redenen van ruimte tot de VWO- en HAVO scholen.

3. De stabiliteit van eindexamencijfers en slaag-, zittenblijf- en uitvalpercentages in de tijd

Een mogelijke verklaring voor de niet erg hoge stabiliteit van de rapportcijfers voor beide schooljaren is de gebruikte methode om dit rapportcijfer te berekenen. Indien deze methode een belangrijke bron van instabiliteit zou zijn zouden de samenhangen voor een meerderheid van de eindexamencijfers en slaag-, zittenblijf- en uitvalpercentages tussen beide schooljaren

Tabel 5: Correlaties tussen de eindexamencijfers en slaag-, zittenblijf- en uitvalpercentages van 1995/96 en 1996/97, ongecontroleerd en gecontroleerd voor percentage cumi-leerlingen (kolom 1 en 2 voor VWO; 5 en 6 voor HAVO), de correlaties tussen het verschil van de Trouw-rapportcijfers in beide schooljaren en de eindexamencijfers en slaag-, zittenblijf- en uitvalpercentages (kolom 3 voor VWO; 7 voor HAVO) en de correlaties tussen het verschil van de Trouw-rapportcijfers in beide schooljaren en de verschillen tussen de eind-examencijfers en slaag-, zittenblijf- en uitvalpercentages in beide schooljaren (kolom 4 voor VWO; 8 voor HAVO)

	VWO				HAVO			
	95-97 bi-variate	95-97 gecontroleerd % cumi	Vershil Rap.- 97	Vershil Rap.- Verschil 95-97	95-97 bi-variate	95-97 gecontroleerd % cumi	Vershil Rap.- 97	Vershil Rap.- Verschil 95-97
%cumi	.95*	n.v.t.	.34*	.34*	.95*	n.v.t.	.40*	.43*
Slaag %	.44*	.22*	.16*	.16*	.56*	.42*	-.03	-.03
Nederl.	.56*	.53*	.10*	.28*	.57*	.52*	-.05	.24*
Engels	.60*	.66*	-.01	.30*	.59*	.56*	-.05	.11*
Wisk.	.42*	.45*	.19*	.46*	.53*	.46*	-.06	.28*
2%tt ¹³	n.v.t.	n.v.t.	n.v.t.	n.v.t.	.00	.00	-.05	-.09
2%zb	-.02	-.01	-.06	-.06	-.01	-.01	.05	-.01
2%zbut	-.02	-.04	-.01	-.01	.04	.04	.15*	.09
3%tt	-.01	.00	.13*	.12*	-.01	-.01	-.02	-.03
3%zb	.01	.04	-.01	-.02	-.01	-.01	-.03	-.04
3%zbut	.06	.16*	-.06	-.07	.01	.01	-.04	-.05
4%tt	.28*	.24*	.10*	.10*	.09	.07	.01	-.01
4%zb	.19*	.22*	-.12*	-.12*	.38	.37*	-.07	-.07
4%zbut	.25*	.25*	-.06	-.06	.29*	.20*	-.01	-.01
5%tt ¹⁴	.10*	.10	-.06	-.06	n.v.t.	n.v.t.	n.v.t.	n.v.t.
5%zb	.19*	.18*	-.09	-.09	n.v.t.	n.v.t.	n.v.t.	n.v.t.
5%zbut	.27*	.35*	-.05	-.06	n.v.t.	n.v.t.	n.v.t.	n.v.t.

Legenda: *p<.05

hoger moeten zijn dan de rapportcijfers. In tabel 5 worden die samenhangen weergegeven, zowel voor de VWO-scholen als voor HAVO-scholen.

3.1. VWO-scholen

De eerste kolom laat zien dat de samenhangen tussen twee van de drie VWO eindexamencijfers voor het Centraal Schriftelijk Deel (Engels en Nederlands) tussen de beide schooljaren hoger zijn dan de samenhang tussen de rapportcijfers. De samenhangen tussen de eind-examencijfers voor Wiskunde A en de slaagpercentages liggen ongeveer op hetzelfde niveau als de samenhang tussen de rapportcijfers. Maar de samenhangen tussen uitstroom- en zittenblijfpercentages in het VWO van beide schooljaren liggen beduidend lager. Met name tussen de percentages voor de tweede en derde VWO-klassen bestaat geen significante samenhang voor de beide schooljaren, terwijl de significante samenhangen voor de hogere VWO-klassen niet imponerend genoemd mogen worden. De hoogste samenhang (.95) bestaat tussen het percentage cumi-leerlingen in beide schooljaren, wat duidelijk laat zien dat sociale segregatie tussen scholen het meest stabiele kenmerk van VWO-scholen is.

De tweede kolom laat dezelfde samenhangen voor het VWO zien, maar nu gecontroleerd voor het percentage cumi-leerlingen. Deze controle is nodig, omdat deze samenhangen heel goed voor een deel kunnen voortvloeien uit de grote stabiliteit van de samenstelling van de leerlingpopulatie in beide schooljaren. Toch is de forse reductie in de samenhang van het slagspercentage tussen beide jaren als gevolg van de controle opvallend: van .44 naar .22. De samenhangen in de drie VWO eindexamencijfers voor het Centraal Schriftelijk zijn daarentegen wel stabiel, ook bij controle voor het percentage cumi-leerlingen. Het enige verschil tussen de eindexamencijfers voor het Centraal Schriftelijk en het slaagpercentage per VWO-school zijn de cijfers voor het schoolonderzoek, dat het andere onderdeel vormt van het eind-examen. Cijfers voor het schoolonderzoek worden gemiddeld met de cijfers voor het Centraal Schriftelijk, en op basis van dit gemiddelde slaagt de leerling al dan niet voor het eind-examen. Blijkbaar verschillen VWO-scholen sterk in de becijfering voor dat schoolonderzoek, zijn deze slaagpercentages minder stabiel voor de verschillende schooljaren dan de cijfers voor het Centraal Schriftelijk. Maar deze instabiele cijfers voor het schoolonderzoek werken blijkbaar wel zodanig dat VWO-scholen een redelijk stabiel ongecorrigeerd slaagpercentage in verschillende schooljaren hebben. De stabiliteit in het slagspercentage van VWO-scholen vindt dus eerder zijn basis in de sociale segregatie tussen scholen en in de kwaliteit van de CITO-toetsen dan in een stabiele becijfering van het schoolonderzoek.

Na controle voor het percentage cumi-leerlingen zijn alleen nog maar de samenhangen tussen de VWO-eindexamencijfers iets hoger dan de samenhang tussen de rapportcijfers over beide jaren. De samenhangen van de andere indicatoren van intern rendement van VWO-scholen ligt beduidend lager dan de samenhang van de rapportcijfers.

Omdat de samenhang tussen eindexamencijfers (tussen .42 en .60) en slaag-, zittenblijf- en uitvalpercentages (tussen .01 en .44) van beide schooljaren even hoog of zelfs lager is dan de samenhang tussen de rapportcijfers van beide schooljaren (tussen .41 en .48) trek ik de conclusie dat de niet erg hoge stabiliteit in het rapportcijfer voor VWO-scholen niet te herleiden is tot een instabiliteit van de methode gebruikt om het rapportcijfer te berekenen.

3.2. HAVO-scholen

De vijfde kolom laat zien dat de samenhangen tussen de drie HAVO eindexamencijfers voor het Centraal Schriftelijk Deel en de slaagpercentages van beide schooljaren hoger liggen dan de samenhang tussen beide rapportcijfers. Evenals bij de VWO-scholen liggen op HAVO-scholen de samenhangen tussen de zittenblijf- en uitstroompercentages van beide schooljaren lager dan de samenhang tussen beide rapportcijfers. Alleen de samenhangen tussen de zittenblijfpercentages van de vierde klas HAVO van beide schooljaren zijn significant. Uiteraard is er opnieuw een zeer hoge samenhang tussen de percentages cumi-leerlingen van beide schooljaren (.95).

De zesde kolom laat dezelfde samenhangen zien, maar nu gecontroleerd voor het percentage cumi-leerlingen. Deze controle is nodig, omdat een deel van de gevonden stabiliteit in de afzonderlijke indicatoren voor het intern rendement kan voortvloeien uit de eerder gevonden stabiliteit in leerlingpopulatie. De samenhang wordt bij de HAVO-scholen door de controle niet wezenlijk veranderd, in tegenstelling tot de VWO-scholen. Ook treedt bij de HAVO-scholen door de controle niet zo'n sterke daling in de samenhang tussen de slaagpercentages van beide schooljaren op, zoals bij de VWO-scholen. Wel is ook bij HAVO-scholen de stabiliteit van de slaagpercentages tussen beide schooljaren lager dan de stabiliteit van de eindexamencijfers. Dat wijst opnieuw op de invloed van de becijfering van het schoolonderzoek die blijkbaar van jaar tot jaar op een school varieert, maar die door zijn instabiliteit tendeeft naar een stabiel slaagpercentage in de tijd. De relevante correlaties in tabel 5 suggereren echter dat deze instabiliteit van de becijfering van het schoolonderzoek bij VWO-scholen groter is dan bij HAVO-scholen. Staan VWO-scholen onder een grotere druk van hun over het algemeen mondiger clientèle dan HAVO-scholen?

Omdat de samenhangen tussen eindexamencijfers (tussen .53 en .59) en slaag-, zittenblijf- en uitval-percentages (tussen -.01 en .42) van beide schooljaren even hoog of zelfs lager is dan de samenhang tussen de rapportcijfers van beide schooljaren (tussen .42 en .50) trek ik de conclusie dat de niet erg hoge stabiliteit in het rapportcijfer voor HAVO-scholen niet te herleiden is tot een instabiliteit veroorzaakt door de methode gebruikt om het rapportcijfer te berekenen.

4. Kenmerken van scholen met een hoger of lager Trouw-rapportcijfer

Wat zijn de kenmerken van VWO- en HAVO-scholen, waar het rapportcijfer voor het jaar 1996/97 gedaald of gestegen ten opzichte van het schooljaar 1995/96?

4.1. VWO-scholen

Kolom 3 van tabel 5 geeft de samenhang tussen het verschil van beide rapportcijfers en de eindexamencijfers, de zittenblijf- en uitstroombpercentages en het percentage cumi-leerlingen in het schooljaar 1996/97. VWO-scholen met een hoog percentage cumi-leerlingen, met een hoog gemiddelde eindexamencijfer voor Wiskunde A en met een hoog slagingspercentage hebben een hoger rapportcijfer in 1996/97 in vergelijking met hun rapportcijfer over 1995/96 dan de andere VWO-scholen. Maar er bestaan nauwelijks significante samenhangen tussen veranderingen in het rapportcijfer en zittenblijf- en uitstroombpercentages, zelfs niet in de hogere klassen. Het feit dat VWO-scholen met veel cumi-leerlingen een hoger rapportcijfer in 1996/97 halen bevestigt het beeld uit paragraaf 2. Immers scholen met veel cumi-leerlingen zijn relatief ook scholen met veel allochtone leerlingen die vanuit het buitenland zij-instroommen in het voortgezet onderwijs.

Het resultaat dat VWO-scholen met hoge slaagpercentages en eindexamencijfers voor Wiskunde A ook een hoger rapportcijfer halen, laat zien dat het rapportcijfer voor- of achteruitgang in het intern rendement goed weergeeft. Dat geldt niet voor zittenblijf- en uitstroombpercentages in alle klassen van de VWO-scholen. Dat zou ook onwaarschijnlijk zijn gegeven hun geringe stabiliteit over de beide schooljaren heen, zoals wij reeds in de vorige paragraaf zagen.

De vierde kolom van tabel 5 laat zien dat met name VWO-scholen, die in 1996/97 hogere eindexamencijfers (zowel Engels, Nederlands en Wiskunde A) dan in 1995/96 halen of die in 1995/96 een hoger percentage slaggers hebben dan in 1996/97, een hoger rapportcijfer halen en andersom. Veranderingen in deze aspecten van het intern rendement werken dus door in het rapportcijfer, zoals ook de bedoeling is. Ook een toe- of afname van het percentage cumi-leerlingen leidt tot een hoger of lager rapportcijfer, overeenkomstig het principe van de berekening van 'toegevoegde waarde'. Veranderingen gedurende beide schooljaren in zittenblijf- en uitstroombpercentages hebben geen significante invloed op de hoogte van het rapportcijfer. Maar opnieuw, dat zou ook onwaar-

schijnlijk zijn gegeven hun geringe stabiliteit over de beide schooljaren heen.

In hier niet weergegeven analyses ben ik nagegaan of het verschil in rapportcijfer van VWO-scholen tussen beide schooljaren nog met andere kenmerken samenhangt. Dat is niet het geval voor het aantal VWO-leerlingen, de denominatie van de VWO-school, de provincie waarin de VWO-school staat of de breedte van de schoolgemeenschap waartoe de VWO-school behoort.

Dit alles laat zien dat het rapportcijfer voor het schooljaar 1996/97 vooral beïnvloed wordt door veranderingen in het al dan niet opnemen van bepaalde groepen leerlingen in de inspectiegegevens en door veranderingen in de eindexamenresultaten. Dit resultaat ondersteunt in mijn ogen de geldigheid van het rapportcijfer als bruikbare indicator van de 'toegevoegde waarde' van een VWO-school ten aanzien van zijn intern rendement.

4.2. HAVO-scholen

Kolom 7 van tabel 5 geeft de samenhang tussen het verschil in rapportcijfer tussen 1995/96 en 1996/97 en de eindexamencijfers, de zittenblijf- en uitstroompercentages en het percentage cumi-leerlingen in het schooljaar 1996/97. HAVO-scholen met een hoog percentage cumi-leerlingen hebben een hoger rapportcijfer in 1996/97 in vergelijking met hun rapportcijfer over 1995/96 dan de andere HAVO-scholen. Maar er bestaan nauwelijks significante samenhangen tussen veranderingen in het rapportcijfer en gemiddelde eindexamencijfers, het slaging-, de zittenblijf- en uitstroompercentages. Alleen HAVO-scholen met veel cumi-leerlingen hebben een hoger rapportcijfer gehaald. Deze uitkomst past goed in de reeds eerder getrokken conclusie dat een belangrijk deel van de verschillen in rapportcijfers tussen beide jaren terug te voeren is op een zuiverder opgave door scholen van hun gegevens aan de inspectie.

De laatste kolom van tabel 5 laat voor HAVO-scholen hetzelfde beeld zien als bij VWO-scholen. HAVO-scholen die in 1996/97 hogere eindexamencijfers (zowel Engels, Nederlands en Wiskunde A) dan in 1995/96 halen krijgen een hoger rapportcijfer en andersom. Veranderingen in deze aspecten van het intern rendement werken dus door in het rapportcijfer, zoals ook de bedoeling is. Ook een toe- of afname van het percentage cumi-leerlingen leidt tot een hoger of lager rapportcijfer, overeenkomstig het principe van de berekening van 'toegevoegde waarde'. Veranderingen gedurende beide schooljaren in slaag-, zittenblijf- en uitstroompercentages hebben geen significante invloed op de hoogte van het rapportcijfer.

In hier niet weergegeven analyses ben ik nagegaan of het verschil in rapportcijfer van HAVO-scholen tussen beide schooljaren nog met andere kenmerken samenhangt. Dat is wel het geval voor het aantal HAVO-leerlingen (hoe groter het aantal leerlingen, des te kleiner het verschil in rapportcijfer) en voor de denominatie van de HAVO-school (reformatrische scholen gaan het meest achteruit; openbare en algemeen-bijzondere het meest vooruit in rapportcijfer). Significante samenhangen zijn niet gevonden voor de provincie waarin de HAVO-school staat of de breedte van de schoolgemeenschap waartoe de HAVO-school behoort.

Dit alles laat zien dat het rapportcijfer voor het schooljaar 1996/97 vooral beïnvloed wordt door veranderingen in het al dan niet opnemen van bepaalde groepen leerlingen in de inspectiegegevens en door veranderingen in de eindexamenresultaten. Dit resultaat ondersteunt in mijn ogen de geldigheid van het rapportcijfer als bruikbare indicator van de 'toegevoegde waarde' van een HAVO-school ten aanzien van zijn intern rendement.

5. De stabiliteit van gewichten voor de schooljaren 1995/96 en 1996/97

Zoals ik reeds in de eerste paragraaf aangaf, heb ik factoranalyse gebruikt om de gewichten van de verschillende indicatoren van het intern rendement vast te stellen. In deze paragraaf vergelijk ik de gewichten van de relevante factoren die voor de beide jaren afzonderlijk zijn berekend, zowel voor VWO-scholen als voor HAVO-scholen¹⁵.

Tabel 6: De factorladingen van de verschillende indicatoren van intern rendement voor de schooljaren 1995/96¹⁶ en 1996/97 op de factor, waarin hoge slaagpercentages en eind-examencijfers samengaan met lage zittenblijf- en uitstroombpercentages in de verschillende klassen en het gestandaardiseerde effect van het percentage cumi-leerlingen op dat totaal rendement (kolom 1 & 2 VWO-scholen; kolom 3 & 4 HAVO-scholen)

Indicator	VWO		HAVO	
	1995/96	1996/97	1995/96	1996/1997
Slaag %	.74	.58	.64	.80
Nederlands	.41	.35	.42	.52
Wiskunde A	.55	.43	.50	.41
Engels	.40	.08	.20	.64
2%tt	n.v.t.	n.v.t.	-.30	-.07
2%zb	-.17	-.47	-.44	-.29
2%zbuit	-.17	-.44	-.46	-.24
3%tt	.04	-.19	-.36	-.24
3%zb	-.32	-.49	-.45	-.46
3%zbuit	-.27	-.56	-.59	-.56
4%tt	-.08	-.14	-.35	-.43
4%zb	-.43	-.41	-.43	-.36
4%zbuit	-.48	-.55	-.53	-.60
5%tt	-.24	-.29	n.v.t.	n.v.t.
5%zb	-.43	-.29	n.v.t.	n.v.t.
5%zbuit	-.40	-.29	n.v.t.	n.v.t.
%cumi	-.46	-.42	-.47	-.54

5.1. VWO-scholen

Hoewel beide factoren gezien de factorladingen in kolom 1 en 2 min of meer dezelfde betekenis hebben voor het intern rendement van VWO-scholen (scholen met hoge eindexamencijfers en slaagpercentages en lage zittenblijf- en uitstroombpercentages enerzijds en scholen met lage eindexamencijfers en slaagpercentages en hoge zittenblijf- en uitstroombpercentages anderzijds) verschillen de factorladingen duidelijk tussen beide jaren. Het gestandaardiseerde effect van het percentage cumi-leerlingen op het intern rendement van VWO-scholen is in beide schooljaren min of meer gelijk. Het verschil in factorladingen heeft tot gevolg dat het VWO-rapportcijfer voor 1996/97 berekend met behulp van de gewichten van 1995/96 niet gelijk is aan het VWO-rapportcijfer voor 1996/97 berekend met behulp van de gewichten van 1996/97, hoewel de samenhang daartussen hoog is (.66). De samenhang tussen het VWO-rapportcijfer voor 1995/96 berekend met zijn eigen gewichten en het VWO-rapportcijfer voor 1996/97 berekend met zijn eigen gewichten is laag (.21). Beide correlaties betekenen dat een empirische keuze van gewichten op grond van de structuur van de data van de afzonderlijke schooljaren tot verschillen in rapportcijfers voor VWO-scholen tussen beide jaren leidt.

5.2. HAVO-scholen

Ook de factorladingen in kolom 4 en 5 hebben min of meer dezelfde betekenis voor het intern rendement van HAVO-scholen (scholen met hoge eindexamencijfers en slaagper-

centages en lage zittenblijf- en uitstroompercentages enerzijds en scholen met lage eind-examencijfers en slaagpercentages en hoge zittenblijf- en uitstroompercentages anderzijds), maar ook zij verschillen duidelijk tussen beide jaren. Het gestandaardiseerde effect van het percentage cumi-leerlingen op het intern rendement van HAVO-scholen is in beide schooljaren min of meer gelijk. Het verschil in factorladingen heeft tot gevolg dat het HAVO-rapportcijfer voor 1996/97 berekend met behulp van de gewichten van 1995/96 niet gelijk is aan het HAVO-rapportcijfer voor 1996/97 berekend met behulp van de gewichten van 1996/97, hoewel de samenhang daartussen hoog is (.67). De samenhang tussen het HAVO-rapportcijfer voor 1995/96 berekend met zijn eigen gewichten en het HAVO-rapportcijfer voor 1996/97 berekend met zijn eigen gewichten is laag (.19). Dit betekent dat een empirische keuze van gewichten op grond van de structuur van de data van elk afzonderlijk schooljaar tot verschillen in rapportcijfers voor HAVO-scholen tussen beide jaren leidt.

6. Het belang van andere dimensies voor het intern rendement van scholen

Naast de door mij voor het rapportcijfer gebruikte dimensie, bestaan er volgens de toegepaste factor analyse ook andere significante dimensies. Om de waarde van de door mij gebruikte dimensie beter te kunnen beoordelen, behandel ik hier de twee relevante dimensies van het intern rendement van VWO- en HAVO-scholen, zoals die uit de factor analyses met de inspectiegegevens over het schooljaar 1995/96 naar voren komen.

Tabel 7: De factorladingen voor de twee relevante factoren voor VWO- en HAVO-scholen, gebaseerd op de inspectiegegevens van 1995/96, berekend met PCA en Equamax-rotatie¹⁷

Indicator	VWO-school		HAVO-school	
	1	2	3	4
Slaag %	.04	.79	.83	.07
Nederlands	.17	.48	.64	.05
Engels	.46	.56	.56	.29
Wiskunde	-.11	.61	.59	-.11
2%tt	n.v.t.	n.v.t.	.05	.49
2%zb	.71	.04	.06	.69
2%zbuit	.76	.06	.00	.66
3%tt	.10	.07	-.02	.49
3%zb	.60	-.15	-.12	.53
3%zbuit	.73	-.06	-.29	.54
4%tt	-.18	-.14	-.16	.34
4%zb	.25	-.38	-.36	.25
4%zbuit	.46	-.35	-.52	.23
5%tt	.01	-.25	n.v.t.	n.v.t.
5%zb	.10	-.43	n.v.t.	n.v.t.
5%zbuit	.04	-.41	n.v.t.	n.v.t.

6.1. VWO-scholen

De factor analyse geeft dat er vier factoren zijn met een eigenvalue groter dan 1.00¹⁸. Op grond van de verschillen tussen deze eigenvalues en de interpreteerbaarheid van de factoren

meen ik dat alleen de eerste twee factoren bruikbaar zijn. Hun factorladingen zijn weergegeven in de eerste twee kolommen van tabel 7.

Op de eerste dimensie scoren die VWO-scholen hoog die in de tweede, derde en vierde klas hoge zittenblijf- en gemiddelde uitstroombpercentages hebben, wat merkwaardigerwijze samen gaat met een hoog eindexamencijfer voor Engels. Deze VWO-scholen hebben echter een gemiddeld slaagpercentage en onderscheiden zich ook niet in hun gemiddelden op de twee andere eindexamenvakken. Inspectie van de namen van de scholen die hoog op deze dimensie scoren laat zien dat het hier vooral om gymnasia gaat. De onderkant van deze dimensie wordt vooral gevormd door brede scholengemeenschappen of door reformatoische scholen. Een gedeeltelijke verklaring van deze eerste dimensie ligt in de 'beleidsrijke' verwerking van zittenblijfgegevens door de inspectie. Zittenblijvers die vervolgens door stromen naar een hogere klas van ander onderwijstype (bijvoorbeeld van 2VWO naar 3HAVO) tellen binnen categoriale scholen (zoals gymnasia) als zittenblijvers met uitstroom, terwijl zij in scholengemeenschappen niet als zittenblijvers geteld worden (zolang de overgang maar binnen de betrokken scholengemeenschap plaats vindt). Toch is de beleidsrijke invulling niet de enige verklaring van deze eerste dimensie. Ook zittenblijven zonder uitstroom heeft op deze eerste dimensie hoge factorladingen en dat kan niet verklaard worden door een 'beleidsrijke' verwerking van zittenblijfgegevens. Op categoriale scholen komt zittenblijven in dezelfde klas blijkbaar meer voor als in brede scholengemeenschappen, waarin gemakkelijker de voorkeur gegeven wordt aan doorstroom naar een ander, meestal lager, schooltype. Gezien deze interpretatie van deze eerste dimensie is het niet merkwaardig dat een hoge score op deze eerste dimensie negatief samenhangt met het rapportcijfer (-.32).

De tweede dimensie komt sterk overeen (pearson $r=.90$) met de dimensie die ik in de niet-geroteerde versie heb gebruikt voor de berekening van het rapportcijfer over de inspectiegegevens 1995/96. De geroteerde factorladingen zijn goed vergelijkbaar met de ongeroteerde uit de eerste kolom van tabel 6. Deze overeenkomst laat opnieuw zien dat het rapportcijfer voor VWO-scholen verwijst naar een reële dimensie van hun totale intern rendement.

6.2. HAVO-scholen

De factor analyse geeft dat er vijf factoren zijn met een eigenvalue groter dan 1.00¹⁹. Op grond van de verschillen tussen deze eigenvalues en de interpreteerbaarheid van de factoren meen ik dat alleen de eerste twee factoren bruikbaar zijn. Hun factorladingen zijn weergegeven in de laatste twee kolommen van tabel 7.

De eerste dimensie komt redelijk overeen (pearson $r=.60$) met de dimensie die ik in de niet-geroteerde versie heb gebruikt voor de berekening van het rapportcijfer over de inspectiegegevens 1995/96. De geroteerde factorladingen zijn redelijk vergelijkbaar met de ongeroteerde uit de derde kolom van tabel 6. Wel zijn door de rotatie de factorladingen van de zittenblijf- en uitstroombpercentages lager dan in de ongeroteerde versie en zijn de factorladingen van de eindexamencijfers en het slaagpercentage hoger dan in de ongeroteerde versie. Deze overeenkomst laat opnieuw zien dat het rapportcijfer voor HAVO-scholen verwijst naar een reële dimensie van hun totale intern rendement.

Op de tweede dimensie scoren vooral HAVO-scholen met hoge zittenblijf- en uitstroombpercentages in hun tweede, derde en vierde klassen, maar zij hebben slechts gemiddelde eindexamencijfers en slaagpercentages. De lijst van scholen met hoge scores op deze dimensie is erg gevarieerd. Dat wijst erop dat hoge zittenblijf- en uitstroombpercentages in HAVO-scholen erg verschillende achtergronden hebben, die om een nadere, hier niet uitgevoerde analyse vragen. De hoge negatieve samenhang met het rapportcijfer laat zien dat deze tweede dimensie een belangrijk onderdeel uitmaakt van het rapportcijfer voor HAVO-scholen, dat op een ééndimensionale ongeroteerde factoranalyse was gebaseerd.

7. Intern rendement en zittenblijven zonder uitstroom

In het rapportcijfer is zittenblijven zonder uitstroom opgenomen als indicator van het intern rendement van scholen. Daartegen is het bezwaar ingebracht dat het intern rendement alleen

zou moeten slaan op de kans het diploma te behalen en niet op de snelheid waarmee het diploma gehaald wordt. Daarom pas ik in deze paragraaf dezelfde procedure toe als bij de berekening van het rapportcijfer, maar nu alleen met slaag- en uitstroombpercentages en eind-examen-cijfers. Ook hierbij gebruik ik de inspectiegegevens over het schooljaar 1995/96.

Tabel 8: De factorladingen van de twee relevante factoren voor VWO-scholen en de ene relevante factor voor HAVO-scholen over slaag- en uitstroombpercentages en eind-examen-cijfers, alsmede de daarbij behorende gestandaardiseerde regressiecoëfficiënt voor de berekening van de toegevoegde waarde.

Indicator	VWO-school		HAVO-school
	1	2	1
Slaag %	.70	.40	.78
Nederlands	.49	.14	.53
Engels	.72	-.04	.40
Wiskunde A	.62	.27	.66
2%tt	n.v.t.	n.v.t.	-.19
2%zbuut	.46	-.65	-.29
3%tt	.12	.01	-.27
3%zbuut	.36	-.76	-.50
4%tt	-.21	.17	-.31
4%zbuut	-.07	-.71	-.57
5%tt	-.14	.02	n.v.t.
5%zbuut	-.34	-.24	n.v.t.
%cumi	-.41	-.34	-.54

7.1. VWO-scholen

Het weglaten van de zittenblijf indicatoren leidt bij VWO-scholen tot een verandering van de factorstructuur. De eerste factor²⁰ combineert nu een hoog slaagpercentage, hoge eind-examen-cijfers, hoge uitstroombpercentages in de tweede en derde klas en lage uitstroombpercentages in de vierde en vijfde klas. VWO-scholen die hoog scoren op deze dimensie selecteren dus in de onderbouw, maar houden leerlingen die de bovenbouw hebben weten te bereiken beter vast en halen goede eind-examenresultaten. Kortom, scholen die het hordeloop- met het expeditiemodel (of contest mobility en sponsored mobility) combineren. Maar door de grotere selectie in de onderbouw hebben deze VWO-scholen niet het hoogste interne rendement. Hetzelfde geldt overigens voor die VWO-scholen die laag scoren op deze dimensie: in de onderbouw houden ze leerlingen goed vast, maar verliezen veel leerlingen in de bovenbouw en halen ook geen goede eind-examenresultaten. De eerste dimensie komt nog het meest overeen met de dimensie waarop het rapportcijfer gebaseerd is (vergelijk eerste kolom van tabel 6 met de eerste kolom van tabel 8), maar hangt door de hoge uitstroom in de tweede en derde klas niet erg hoog samen met het rapportcijfer ($r = .50$). Voor VWO-scholen betekent dat het weglaten van de zittenblijf indicatoren tot een minder goede meting van het totale interne rendement leidt.

De tweede factor²¹ groepeer VWO-scholen waar veel leerlingen na zittenblijven vertrekken en waar vrij veel leerlingen slagen voor hun eind-examen maar wel met slechts gemiddelde eind-examen-cijfers. Evenals in 6.1. lijkt het als het hier in hoofdzaak handelt om de categoriale gymnasia. De samenhang tussen beide factoren is dan ook hoog (.72).

7.2. HAVO-school

De eerste dimensie van de factor analyse zonder de zittenblijf indicatoren verschilt weinig van de dimensie met zittenblijf indicatoren, waarop het rapportcijfer gebaseerd is. HAVO-scholen die hoog op deze eerste dimensie scoren hebben hoge slaagpercentages en eindexamencijfers en lage uitstroompercentages. De samenhang tussen scores op deze eerste dimensie en het rapportcijfer is daarom ook zeer hoog (.85). Bij HAVO-scholen maakt het dus voor de meting van hun intern rendement niet veel uit of zittenblijven al dan niet in de berekening wordt betrokken.

8. Conclusie

De verschillen tussen de inspectiegegevens van de schooljaren 1995/96 en 1996/97 zijn voor een belangrijke mate verantwoordelijk voor de niet hoge samenhang tussen de daarop gebaseerde rapportcijfers voor scholen over hun 'toegevoegde waarde' met betrekking tot hun totaal intern rendement. Deze verschillen tussen de inspectiegegevens vloeiden vooral voort uit de begrijpelijke verwijdering van bepaalde groepen uitzonderlijke leerlingen uit de inspectiegegevens (allochtone leerlingen die vanuit het buitenland zijn instromen in het voortgezet onderwijs, asielzoekers, drop-out projecten). Er zijn nauwelijks tekenen dat scholen hun resultaten door de extra controle ronde ten onrechte hebben opgepoetst. Wel is het mogelijk dat VBO-scholen van deze correctie ronde minder gebruik hebben gemaakt dan MAVO- en HAVO-scholen. Indien wij rekening houden met deze procedurele verschillen tussen de inspectiegegevens van de schooljaren 1995/96 en 1996/97, zijn de verschillen in rapportcijfer tussen beide jaren niet erg groot. Het vaak naar voren gebrachte bezwaar dat de rapportcijfers, die vorig jaar in *Trouw* werden gepubliceerd, slechts een onbetrouwbare momentopname van het intern rendement van scholen zouden laten zien is dus niet juist. De stabiliteit van de belangrijkste indicatoren van het intern rendement van scholen (samenstelling van leerlingpopulatie; eindexamencijfers voor Centraal Schriftelijk Deel; slaagpercentage dankzij de instabiele becijfering van het schoolonderzoek) is redelijk tot hoog en ook het daarop gebaseerde rapportcijfer. Opvallend is de geringe stabiliteit van uitstroom- en zittenblijfpercentages over de beide schooljaren, met name in de onderbouw, maar dat is geen gebrek van de berekeningswijze van het rapportcijfer. Het zou daarom beter zijn de gemiddelde uitstroom- en zittenblijfpercentages per school te gebruiken in plaats van die voor elke klas afzonderlijk (Agerbeek, 1998c; Dronkers, 1998b).

De gewichten die men toekent aan de verschillende indicatoren van het intern rendement zijn gedeeltelijk ook verantwoordelijk voor de niet hoge samenhang tussen de rapportcijfers. Blijkbaar verschilt de betekenis van de verschillende indicatoren van schooljaar tot schooljaar, bijvoorbeeld door de wisselende kwaliteit van het Centraal Schriftelijk Examen. Dit pleit ervoor de gewichten van de verschillende indicatoren op normatieve gronden te kiezen in plaats van op empirische overwegingen. Wel is de regressie-coëfficiënt van de leerlingpopulatie in beide schooljaren erg stabiel, evenals de samenhang tussen de samenstelling van de leerlingpopulatie, wat duidelijk laat zien dat sociale segregatie tussen scholen het meest stabiele kenmerk van VWO- en HAVO-scholen is.

Het totale interne rendement van scholen, zoals ik dat in het rapportcijfer heb uitgedrukt (een school met hoge eindexamencijfers, veel geslaagden, weinig zittenblijvers en weinig uitstroom in alle bijna klassen), is niet de enige wijze om scholen zinvol te ordenen op grond van hun intern rendement. Ook andere ordeningen zijn denkbaar (bijvoorbeeld zonder zittenblijven), maar deze alternatieve ordeningen leiden niet tot een logischer ordening van scholen met betrekking tot hun intern rendement. De door mij gebruikte ordening is dus, gegeven de beschikbare inspectiegegevens over twee schooljaren, realistisch en kan niet gemakkelijk door een meer evidente ordening vervangen worden.

Dat zelfde geldt ook voor rendementsmetingen zonder zittenblijven: een meting van het

totaal intern rendement zonder zittenblijven verschilt weinig van een meting van het totaal intern rendement met zittenblijven òf het meet het totaal rendement van een school slechter.

Samenvattend: de door mij gevolgde methode die de 'toegevoegde waarde' van een school met betrekking tot zijn totaal intern rendement berekende werkt in beide schooljaren redelijk en levert voor beide schooljaren een redelijk stabiel rapportcijfer op, met uitzondering van die scholen waarop veel allochtone leerlingen die vanuit het buitenland zij-instroomden in het voortgezet onderwijs, asielzoekers e.d. zitten die scholen vorig jaar ten onrechte hadden toegevoegd aan hun inspectiegegevens. Een aantal van de geopperde bezwaren tegen de *Trouw*-rapportcijfers zijn derhalve niet voldoende empirisch gefundeerd en dus onvoldoende om op die technische gronden vergelijkingen tussen gecorrigeerde rendementen van met name genoemde scholen af te wijzen.

Literatuur

- Agerbeek, M. (1998a). Scholen slordig met cijfers. *Trouw*, 9 mei, ZENZ 5.
- Agerbeek, M. (1998b). Schooljaren wisselen net als wijnjaren. *Trouw*, 9 mei, ZENZ 4-Z5.
- Agerbeek, M. (1998c). Schoolprestaties 1998. *Trouw*, 13 juni, ZENZ.
- Agerbeek, M., Hagemen, E., Kreulen, E. & Lakmaker, H. (1997). *Trouw*-onderzoek schoolprestaties. *Trouw*, 25 oktober, ZENZ.
- Bosker, R. (1998). *Maatschappelijke implicaties van (verkeerde) indicatoren*. Enschede: Vakgroep Onderwijsorganisatie en -management/ OCTO, Universiteit Twente. Paper gepresenteerd op de Onderwijsresearchdagen 1998, Universiteit Twente.
- Dronkers, J. (1997). Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren. *Trouw*, 8 november, 12.
- Dronkers, J. (1998a). Het betere is de vijand van het goede. Een reactie op de commentaren over het *Trouw* rapportcijfer. *Pedagogische Studiën*, 75, 142-150.
- Dronkers, J. (1998b). *Verschildt de toegevoegde waarde van het intern rendement van scholen in het voortgezet onderwijs? Enkele kanttekeningen bij het Trouw-onderzoek Schoolprestaties 1998*. Amsterdam: SCO-Kohnstamm Instituut. Samenvatting van een openbaar college gegeven op 19 juni 1998.
- Putten, C. M. van & Kamp, L. J. T. van der (1998). Dronkers' rapportcijfer voor de middelbare scholen: nauwkeurigheid, geldigheid en betekenis. *Pedagogische Studiën*, 75, 135-141.
- Veenstra, D. R., Dijkstra, A. B., Peschar, J. L. & Snijders, T. A. B. (1998). Scholen op rapport. Een reactie op het *Trouw*-onderzoek naar schoolprestaties. *Pedagogische Studiën*, 75, 121-134.

Noten

- Deze bijdrage is gebaseerd op een openbaar college gegeven op 11 mei 1998 op de Universiteit van Amsterdam, ter gelegenheid van de publicatie door *Trouw* van de inspectiegegevens over het schooljaar 1996/97 (Agerbeek, 1998b). Op ondergeschikte puntjes wijken enkele tabellen af van die op het college zijn gepresenteerd, omdat enkele kleine codefouten er daarna nog uitgehaald zijn. Op 19 juni 1998 heb ik een tweede openbaar college gegeven (Dronkers, 1998b) dat uitsluitend handelde over de achtergrond van de berekening van de toegevoegde waarde met behulp van de meer gedetailleerde inspectiegegevens over het schooljaar 1996/97, zoals die op 13 juni door *Trouw* gepubliceerd werden (Agerbeek, 1998c). Deze meer gedetailleerde inspectiegegevens zijn niet vergelijkbaar met de meer globale inspectiegegevens over het schooljaar 1995/96.
- Bij de *Trouw* publicatie in 1997 (Agerbeek et al, 1997) is dit ideaal nog onvoldoende bereikt (zie Dronkers, 1997), maar krantenlezers van de *Trouw* publicatie uit 1998 (Agerbeek, 1998c) konden zelf het samenvattend oordeel afleiden uit de afzonderlijke indicatoren.
- Behalve de brugklas (omdat dat scholengemeenschappen onevenredig zou bevoordelen) en de eindexamensklas (omdat zittenblijven en uitstroom daar de tegenhanger bleek van slagen).
- Ik laat hier de discussie buiten beschouwing of het percentage cumi-leerlingen per school de meest geschikte indicator is, maar ik verwijs daarvoor naar Veenstra et al (1998) en Dronkers (1998a). Maar ter voorkoming van misverstanden, het percentage cumi-leerlingen is niet de meest geschikte indicator, maar waar niet is verliest de koning zijn recht.

5. Hiermee wil ik niet zeggen dat de inspectie of de scholen terecht handelden door de betrouwbaarheid van de door hun verstrekte gegevens te verwaarlozen (Dronkers, 1998a).

6. Bijvoorbeeld een Haagse inspecteur (Agerbeek, 1998a).

7. De formules voor de verschillende schooltypen, zoals ik die gebruikt heb voor de berekening van het rapportcijfer over de gegevens van 1995/96, zijn in *Trouw* gepubliceerd (Dronkers, 1997). De hier weergegeven formule voor de MAVO wijkt af van de toen in de *Trouw* gepubliceerde formule. De laatste genoemde was foutief afgedrukt.

$slaag\% = (slaag\% - 0,922) / 0,07685$; $zned = (((10 * nedd + 10 * nedc) / 2) - 64,022) / 3,4206$;

$zwis = (((10 * wisd + 10 * wisc) / 2) - 59,3732) / 5,0162$;

$zeng = (((10 * engd + 10 * engc) / 2) - 64,2844) / 4,7143$; $z2\%tt = (2\%tt - 0,02639) / 0,02754$;

$z2\%zb = (2\%zb - 0,06109) / 0,05457$; $z2\%zbut = (2\%zbut - 0,06561) / 0,05489$;

$z3\%tt = (3\%tt - 0,02582) / 0,02831$; $z3\%zb = (3\%zb - 0,08408) / 0,05227$;

$z3\%zbut = (3\%zbut - 0,04863) / 0,04319$.

Mavo rapportcijfer: $((0,327 * zslaag\% + 0,271 * zned + 0,265 * zwis + 0,2199 * zeng - 0,113 * z2\%tt - 0,057 * z2\%zb - 0,137 * z2\%zbut - 0,145 * z3\%tt - 0,078 * z3\%zb - 0,195 * z3\%zbut) + 0,059845 * cumi + 5,678) / \text{afgerond}$.

8. CSG Revius Deventer, Stellingwerf College Oosterwolde, Prisma College B Bakker Utrecht, Mozaïek College Arnhem, Olympus College Rotterdam.

9. Gepaarde t-toets.

10. RK SGM City College Rotterdam, SGM Slauerhoff Leeuwarden, OSG Hugo de Groot Rotterdam, Chr. SGM Overvoorde Den Haag, Vlaardingse SGM Vlaardingen, SG Het Zaanlandse Zaandam, Maerlant Lyceum Den Haag, Scholengroep Esloo Den Haag, Lindenholt College Nijmegen, Prisma College Utrecht, RSG Schravenlant, Esprit Scholengemeenschap Amsterdam, Da Vinci Leiden, Johan de Wit Den Haag.

11. Maaswaal College Wychen, RK SGM St. Montfort Rotterdam, Ivo Mavo't Heuveltje Aerdenhout, SG Erasmus Almelo, SG Nieuw Rotterdam Rotterdam, Schoonhovens College Schoonhoven, SGM Overvoorde 's-Gravenhage, Atlas College Rijswijk, Montessori College Maastricht, Zernike College Groningen, College Riederwaard Rotterdam, Spectrum College Utrecht, Dendron College Horst, Olympus College Rotterdam, SGM Calvyn Amsterdam; Johan de Witt College 's-Gravenhage, Prisma College, B. Bakker Utrecht, Mozaïek College Arnhem, SGM Het Amsterdams Lyceum Amsterdam, Esprit Scholengroep Amsterdam.

12. Esprit Scholengroep Amsterdam, Meerjant College Amsterdam, Johan de Witt Den Haag.

13. In de berekening van het rapportcijfer 1995/96 ontbrak om technische redenen deze indicator voor de VWO. In deze heranalyse hebben wij hem daar ook steeds weggelaten.

14. In de berekening van het rapportcijfer 1995/96 hebben wij steeds de uitstroom- en zittenblijpercentages van de eindexamenklas buiten beschouwing gelaten, gezien de hoge samenhang daarvan met het slaagpercentage. In deze heranalyse hebben wij dezelfde procedure gevolgd.

15. Om de analyse vergelijkbaar te houden pas ik dezelfde procedure bij de factor analyses in deze paragraaf toe: Principale Componenten Analyse zonder rotatie. Deze aanpak komt overeen met de werkwijze die ik bij de berekening van de rapportcijfers over 1995/96 toepaste, maar verschilt van die welke ik in paragraaf 6 en 7 gebruik.

16. Deze factorladingen voor het schooljaar 1995/96 zijn alleen direct te vergelijken met de gewichten uit Dronkers (1997), omdat deze gewichten de uit deze factorladingen afgeleide regressie scores zijn.

17. Ik heb Equamax rotatie gebruikt omdat ik alle rendementsindicatoren in de verschillende dimensies een zo groot mogelijke waarde wilde geven, omdat zij alle met evenveel recht als een aspect van het intern rendement beschouwd mogen worden.

18. De eigenvalue bedraagt resp. 2,6, 2,2, 1,5 en 1,1 met een verklaarde variantie van 17%, 15%, 10% en 7%.

19. De eigenvalue bedraagt resp. 2.6, 1.9, 1.3, 1.1 en 1.0 met een verklaarde variantie van 20%, 15%, 10%, 8% en 8%.

20. Eigenvalue 2.2; variantie 19.8%.

21. Eigenvalue 1.8; variantie 16,5%.

DE BEREKENINGSMETHODE IN "SCHOOLPRESTATIES"

Wim Pijs

Faculteit Economische Wetenschappen, Erasmus Universiteit Rotterdam

Abstract

Last year, the quality of the secondary schools was measured and expressed by one single value (Trouw 1997, Trouw 1998). This value is the outcome of a weighted sum, where the weights are equal to the loadings of a component, obtained by factor analysis. This paper shows that these weights are not appropriate.

1. Inleiding

Na de eerste geruchtmakende beoordeling van scholen van voortgezet onderwijs (Trouw 1997), gebaseerd op de gegevens over het schooljaar 95-96, heeft inmiddels een tweede beoordelingsronde plaatsgevonden, gebaseerd op de periode 96-97 (Trouw 1998). De methode die gevolgd is om tot een eindoordeel per school te komen, was niet dezelfde in elke ronde. Beide keren werd eerst een aantal deelscores per school berekend. De aard en het aantal van deze deelscores waren echter in de eerste ronde afwijkend van de tweede ronde. Vervolgens werd uit de deelscores één eindcijfer gedestilleerd. In beide gevallen werd voor de bepaling van het eindcijfer gebruik gemaakt van factoranalyse. Deze statistische techniek is echter ontworpen voor geheel andere doeleinden. In dit artikel hoop ik aan te tonen, dat deze techniek onge-schikt is als instrument voor kwaliteitsmeting.

2. De gevolgde methode

Zoals gezegd, is het eindcijfer bepaald aan de hand van scores voor een aantal kenmerken. De eerste keer werden liefst zestien kenmerken in ogenschouw genomen, de tweede keer zeven. Waarom voor deze wijziging gekozen is, is in Trouw niet toegelicht. We zullen de methode voor de berekening van het eindcijfer hier globaal beschrijven. In paragraaf 4 zullen we aangeven, welke modificaties en verfijningen daar door Dronkers op zijn aangebracht.

De berekening van het eindcijfer omvat de volgende stappen (vergelijk Dronkers 1997-1). Op de gegeven kenmerken wordt Principale Componenten Analyse (PCA) toegepast. Daartoe moeten de gegevens eerst gestandaardiseerd worden (standaarddeviatie = 1 en gemiddelde = 0). Vervolgens wordt de componentenmatrix berekend. De lading van elk kenmerk op de eerste component wordt gekozen als gewicht voor dat kenmerk. Het eindcijfer in eerste aanleg is de gewogen som van de gestandaardiseerde kenmerken met als gewicht de ladingen op de eerste PCA-component. Door herschaling aan het einde komt men in een bereik van 1 t/m 5 terecht.

Ter toelichting geven we een voorbeeld. Beschouw het slaagpercentage van het Havo in 1996. De variantie is gelijk aan 7,568 (bij een gemiddelde 84,4). De lading op de eerste component en dus ook het gewicht voor de gestandaardiseerde waarde bedraagt 0,63754. Het gewicht voor het ongestandaardiseerde slaagpercentage is dus $0,63754/7,568 = 0,08424$. (In de krant (Dronkers, 1997-1) werden andere gewichten voor de gestandaardiseerde scores gegeven; deze gewichten waren de ladingen op de eerste component gedeeld door de eigenwaarde; deze operatie is niet wezenlijk en het waarom ervan laten we buiten beschouwing.)

Een gewogen optelling heeft de volgende meetkundige interpretatie. De score-serie van elke school bepaalt een punt S in een n -dimensionale ruimte, waarbij n het aantal kenmerken voorstelt. Een gewichten-serie bepaalt een punt G in dezelfde ruimte. Beschouw de lijn m door de oorsprong O en door G . De projectie van S op m levert per school een tweede punt T . De lengte van het lijnstuk OT is gelijk (op een constante schaalfactor na) aan de uitkomst van de gewogen optelling. Omdat OT steeds langs m valt, zou men m derhalve de meetlat kunnen noemen. Merk op, dat een andere gewichten-serie een andere meetlat oplevert. Omgekeerd kan men uit een gegeven rechte (de meetlat) een gewichten-serie afleiden. Dronkers stelt, dat de rechte k , gedefinieerd als de rechte voortgebracht door de gewichten in de eerste PCA-component, de onderlinge verschillen tussen scholen het beste weergeeft (Dronkers 1997-1, Dronkers 1998). Deze uitspraak is een parafrase van de wiskundige stelling, die zegt, dat de variantie langs k maximaal is. Dit impliceert echter niet dat de rechte k ook een geschikte meetlat voor kwaliteit is. Het is in zijn algemeenheid niet zo, dat de gewichten behorend bij de rechte k waarden hebben, die redelijk of wenselijk zijn in relatie tot kwaliteit. Het blijkt dat deze gewichten een ongewenst hoge of ongewenst lage waarde kunnen aannemen of een waarde aan de verkeerde kant van nul. In de volgende paragraaf wordt dit theoretisch onderbouwd. Men raadplege enkele dagbladpublicaties (Pijls 1998-1, Pijls 1998-2, Pijls 1998-3) om voorbeelden van gewichten in de beide edities van Schoolprestaties te zien, die nogal vreemd aandoen.

3. De nadelen van de gebruikte methode

Indien twee kenmerken in sterke mate positief of negatief correleren, zullen bij toepassing van PCA deze kenmerken hoge ladingen krijgen met gelijk respectievelijk tegengesteld teken. Deze tekens (plus/min) kunnen anders zijn dan men op normatieve gronden zou wensen. Het probleem dient zich bij voorbeeld aan in de volgende situatie, die Trouw wil ontmaskeren, te weten, hoge eindexamenresultaten als gevolg van een strenge selectie. Laten we deze situatie in gesimplificeerde vorm eens bekijken. Stel dat van elke school het gemiddeld eindexamen-cijfer alsmede het gemiddeld zittenblijf-cijfer is gegeven. Zoals gezegd, men wil hoge eindexamenresultaten minder belonen, indien ze verkregen zijn ten koste van veel zittenblijvers. Stel het voor dit doel "ideale" geval, dat elk procent meer zittenblijvers een 0,1 hoger rapportcijfer oplevert. Dit vindt zijn neerslag in bij voorbeeld de volgende formule, geldend per school: $C=6+0,1*Z$, waarbij C het gemiddeld eindexamencijfer en Z het zittenblijf-percentage voorstelt. De PCA zou in deze situatie één component aangeven, die leidt tot een gewogen optelling met gewichten +1 en +1 voor de gestandaardiseerde, of +10 en +1 voor de ongestandaardiseerde waarden. Dit is volledig in strijd met de oorspronkelijke bedoeling het zittenblijven te bestraffen.

Een tweede voorbeeld met een iets algemener karakter is het volgende. Veronderstel dat goede eindexamenresultaten positief gecorreleerd zijn aan dure schoolreizen. Als men op dure schoolreizen een straf wil zetten, levert PCA de verkeerde gewichten. Ziet men dure (en verre) reizen als een pluspunt, dan biedt PCA wel de juiste tekens in de gewichten. We zien weer, dat de ladingen niet op voorhand verenigbaar zijn met de invloed, die men kenmerken uit normatieve overwegingen wil toekennen.

Het geschetste probleem had zich reëel voorgedaan, indien de culturele minderheden (cumi) in de PCA betrokken waren geweest. Gezien de negatieve correlatie tussen eindexamenresultaten en de cumi-score zou PCA deze kenmerken een tegengesteld gewicht geven. Dit zou inhouden, dat een eindcijfer lager uitvalt naarmate er meer alloctonen op school zitten. Mogelijk heeft Dronkers hier de onlogische werking van PCA opgemerkt en om die reden de cumi-score buiten de PCA gehouden.

We behandelen nu een voorbeeld dat aantoont dat belangrijke kenmerken een ongewenst laag gewicht kunnen krijgen. Zoals reeds boven opgemerkt, krijgen kenmerken die sterk correleren,

elk een hoog gewicht. Analoog krijgen kenmerken die daar niet mee correleren een laag gewicht. Zie Tabel 1, die denkbeeldige scores van scholen voor de vakken Nederlands, Engels en wiskunde bevat. Zoals te zien in de lijst, heeft elke school voor Nederlands hetzelfde resultaat als voor Engels. De eerste component heeft voor wiskunde een lading gelijk nul. Dit zou er toe leiden, dat het eindcijfer alleen afhangt van Nederlands en Engels. De tweede component bevat alleen wiskunde. Wiskunde staat dus letterlijk en figuurlijk haaks op de talen. Als we in de eerste rij van de tabel de 9 voor wiskunde in een 8 veranderen, zou dit kenmerk in de eerste component zelfs een klein negatief gewicht krijgen. Scholen zouden dan bestraft worden voor een goed wiskunde-resultaat. Analoog zou een 4 (in plaats van een 5) voor wiskunde in de laatste regel een licht positief gewicht voor dat vak opleveren. Beide mutaties resulteren dus in onbevredigende gewichten. Het extreem lage gewicht dat VBO-Engels heeft in de eerste editie van Schoolprestaties, is waarschijnlijk aan het hier beschreven effect te wijten.

4. Oplossingen voor de problemen

In deze paragraaf geven we aan, in hoeverre Dronkers van de in paragraaf 2 beschreven methode is afgeweken om enkele van de in paragraaf 3 beschreven problemen te verhelpen. De beoordelingsronden over de schooljaren 95-96 en 96-97 worden aangeduid met resp. Schoolprestaties-1 en Schoolprestaties-2.

De problemen hebben zich nadrukkelijk aangediend Schoolprestaties-1. De eerste component van het VWO heeft een positieve lading voor sommige zittenblijf-kenmerken in de hogere leerjaren. (Deze component is te interpreteren als de component van de "selecterende scholen", de scholen die goede eindexamen-resultaten behalen in combinatie met ongunstige rendementen in de bovenbouw). Dronkers heeft derhalve de eerste component genegeerd en de tweede component genomen (Dronkers 1997-2), die overigens voor de uitvallers op 3-VWO nog steeds een (klein) positief gewicht heeft.

Bij toepassing van een groot aantal kenmerken wordt er wellicht altijd wel één met een ongewenst teken gewogen. Door een aantal kenmerken te clusteren wordt de kans op dit ongemak verkleind. In Schoolprestaties-2 (Trouw 1998) zijn derhalve de zittenblijf- en uitval-kenmerken niet meer uitgesplitst naar leerjaren.

Met de cumi-scores houdt men altijd problemen. Zoals gezegd, zijn deze in de praktijk sterk positief gecorreleerd met zittenblijf- en uitval kenmerken. Dientengevolge zullen de PCA-ladingen een gewicht met een verkeerd teken opleveren. Dronkers heeft hier de volgende oplossing bedacht. Een eindcijfer is in eerste instantie berekend zonder cumi mee te tellen. Vervolgens is een zogeheten correctie op dit eindcijfer toegepast. Daartoe werd een regressie-analyse uitgevoerd met het voorlopige eindcijfer als afhankelijke variabele. De regressie-coëfficiënt wordt geïnterpreteerd als de mate waarin het eindcijfer door cumi wordt belast. Deze coëfficiënt wordt derhalve gekozen als het cumi-gewicht. De genoemde interpretatie is overigens slechts zinvol, indien sprake is van lineaire regressie. Het is niet duidelijk, in hoeverre dit het geval is.

Ofschoon het voor de hand gelegen had het kenmerk instroom (een kenmerk, dat in Schoolprestaties-2 meetelde) op dezelfde manier te behandelen als uitstroom, is dit kenmerk buiten de factoranalyse gehouden en, analoog aan cumi, via regressie-analyse in de berekening betrokken. De reden, waarom dit gebeurd is, moge de lezer duidelijk zijn na de uiteenzetting in paragraaf 3.

5. Slot

We hebben gezien, dat PCA-ladingen niet op voorhand geschikte gewichten zijn in het licht van kwaliteitsmeting. De gehanteerde methode is derhalve ongeschikt voor het bepalen van

één enkel eindcijfer. Mocht in de toekomst uitvoeriger en betrouwbaarder cijfermateriaal beschikbaar komen, dan nog is het zeer de vraag, of een passende manier gevonden kan worden, dit materiaal tot één enkel eindcijfer te reduceren. Nog altijd geldt, dat je appels niet bij peren kunt optellen.

Dankwoord

De auteur dankt zijn naaste collega R. Potharst, alsmede de medewerkers G. van Driel, L. Klomp en J. Geilenkirchen van de afdeling Economische Statistiek voor de vruchtbare en onderhoudende discussies.

Literatuur

- Dronkers, J. (1997-1), *Schoolcijfer zegt niet dat scholen met 5 onvoldoende functioneren*, *Trouw*, p.12, 8 november 1997.
- Dronkers, J. (1997-2), *Uitvoer SPSS met Factor Analyse* (ongepubliceerd).
- Dronkers, J. (1998), *Schoolcijfers arbitrair, maar betere zijn er niet*, *De Volkskrant*, p.9, 6 maart 1998.
- Pijls, W. (1998-1), *Dronkers maakt kwaliteitskeurmerk nog altijd niet waar*, *Trouw*, p.11, 3 januari 1998.
- Pijls, W. (1998-2), *Schoolcijfers van Dronkers slaan helemaal nergens op*, *De Volkskrant*, p.9, 19 februari 1998.
- Pijls, W. (1998-3), *Trouw vergelijkt appels met peren*, *De Volkskrant*, p.9, 16 juni 1998.
- Trouw (1997), *Schoolprestaties, bijlage bij dagblad Trouw*, 25 oktober 1997.
- Trouw (1998), *Schoolprestaties 1998, bijlage bij dagblad Trouw*, 13 juni 1998.

Tabel 1

Ned.	Eng.	Wisk.
8	8	9
8	8	5
7	7	9
7	7	5
6	6	9
6	6	5

DE BEREKENING VAN HET TOTAAL INTERN RENDEMENT VAN SCHOLEN; EEN ANTWOORD OP PIJLS' BEZWAREN

J. Dronkers

SCO-Kohnstamm Instituut, Universiteit van Amsterdam

Pijls wijst in zijn commentaar op mijn gebruik van factoranalyse bij de rekening van het totaal rendement van scholen ten behoeve van *Trouw* op een overbekend punt. Een analysetechniek, zoals factoranalyse, geeft niet automatisch valide uitkomsten. Als bij het gebruik van technieken en de interpretatie van de uitkomsten niet goed wordt nagedacht, kan het misgaan. Deze waarheid is geen nieuwe boodschap en de vraag is of Pijls' concrete bezwaar hout snijden.

Lezers van mijn andere bijdrage in deze TOR kunnen zien dat Pijls ten onrechte meent dat de factoranalyse mijn uitgangspunt bij de vaststelling van een totaal intern rendement van scholen was. Daar en elders (Dronkers, 1998a: 146) omschrijf ik mijn uitgangspunt echter anders: "ik heb a-priori een goed totaal intern rendement gedefinieerd als een school met hoge eindexamencijfers, veel geslaagden en weinig zittenblijvers en weinig uitstroom in alle klassen. Ik heb voor deze operationalisatie gekozen, omdat ik veronderstel dat ouders graag weten op welke school er weinig leerlingen blijven zitten, er weinig ongediplomeerd uitstromen, en er veel wordt geslaagd met hoge eindexamencijfers". Ik had op dat moment deze indicatoren bij elkaar op kunnen tellen tot één schaal. Dat heb ik niet gedaan, omdat ik ook wilde vaststellen dat deze schaal een empirische fundering had in de data. Dat er een valide schaal zou kun-

nen bestaan, was uit analyses van schoolloopbaancohorten reeds lang bekend. Zo concluderen Bosker, Guldemond, Hofman & Hofman (1989: 91) dat "scholen die het op de ene indicator¹ goed doen, ook vaker hogere scores op andere indicatoren: dus hoge slaagpercentages gaan samen met minder doublures en een hogere retentiviteit van de school. Aangezien wij op diverse bestanden steeds tot deze resultaten geraken, zelfs als we een en ander uitsplitsen naar diverse opleidingstypes, kunnen wij concluderen dat deze bevindingen redelijk betrouwbaar zijn". Op grond daarvan komen deze auteurs dan ook tot één kwaliteitsindicator: de combinatie van de genoemde loopbaanindicatoren (Bosker et al. 1989: 92). De factoranalyse heb ik dus in eerste instantie toegepast om te zien of de verschillende rendementsindicatoren een schaal vormen. Het antwoord daarop luidde positief: de veronderstelde intern rendementschaal was inderdaad aanwezig, hetzij als eerste, hetzij als tweede factor. Als deze dimensie niet reëel aanwezig zou zijn geweest in de data (en daarop slaan de hypothetische voorbeelden van Pijls), dan had ik inderdaad niet factoranalyse kunnen gebruiken. Maar de veronderstelde schaal was wel aanwezig in de data, vooral als uitstroom- en zittenblijfpercentages gemiddeld worden over alle klassen, zoals in de tweede *Trouw* publicatie gebeurde (Agerbeek, 1998; Dronkers, 1998b)². Dit vaststellen van het bestaan van een veronderstelde dimensie lijkt mij een legitiem gebruik van factoranalyse en derhalve zijn de concrete bezwaren van Pijls praktisch niet relevant.

Uiteraard levert een andere samenstelling van indicatoren een andere schaal op, zoals Pijls terecht opmerkt. Maar dat betekent vooral dat men vooraf goed moet nadenken welke indicatoren een goede inhoudelijke relatie met het te meten begrip hebben. In het geval van intern rendement zijn leerlingkenmerken als percentage cumi-leerlingen en percentage zij-instromers geen valide indicatoren: het zijn leerlingkenmerken die het rendement kunnen beïnvloeden. Daarom horen deze leerlingkenmerken niet thuis in de factoranalyse van indicatoren van intern rendement en dat heb ik dan ook niet gedaan. Het hypothetische bezwaar van Pijls op het punt van het al dan niet betrekken van het percentage cumi-leerlingen in de factoranalyse snijdt daarom geen hout: het opnemen zou inderdaad inhoudelijk fout zijn (vermenging van indicatoren die naar verschillende begrippen verwijzen) en daarom tot fouten kunnen leiden. Dat heb ik dan ook niet gedaan en Pijls' bezwaren tegen het gebruik van factoranalyse zijn in dit geval niet ter zake. Zij-instroom, die in de tweede *Trouw* publicatie naast het percentage cumi-leerlingen als correctie factor is meegenomen (Agerbeek, 1998; Dronkers, 1998b), beschouwt Pijls ten onrechte als rendementsindicator, terwijl het een leerlingkenmerk is. Want het gaat hier om opstroomers na behalen van een lager diploma, ongediplomeerde afstroomers uit een hoger onderwijstype, leerlingen die verhuizen of overgeplaatst worden naar andere scholen. Het zou dus inhoudelijk onjuist zijn het percentage zij-instromers in de factoranalyse op te nemen, zoals Pijls lijkt voor te stellen.

Het hoofdbezwaar van Pijls, dat in mijn factoranalyses appels met peren bij elkaar opgeteld zijn, is dus alleen ter zake als verschillende soorten indicatoren in de factoranalyses zouden zijn opgenomen of als ik mechanisch steeds de eerste factor genomen zou hebben. Omdat ik dat niet gedaan heb, is zijn hoofdbezwaar praktisch ongegrond en is mijn aanpak valide.

Literatuur

- Agerbeek, M. (1998). *Schoolprestaties 1998. Trouw*, 13 juni, ZENZ.
- Bosker, R. J., Guldemond, H., Hofman, R. H. & Hofman, W. H. A. (1989). *Kwaliteit in het voortgezet onderwijs*. Groningen: Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen.
- Dronkers, J. (1998a). *Het betere is de vijand van het goede*. Een reactie op de commentaren over het Trouw rapportcijfer. *Pedagogische Studiën*, 75, 142-150.
- Dronkers, J. (1998b). *Verschilt de toegevoegde waarde van het intern rendement van scholen in het voortgezet onderwijs? Enkele kanttekeningen bij het Trouw-onderzoek Schoolprestaties 1998*. Amsterdam: SCO-Kohnstamm Instituut. Samenvatting van een openbaar college gegeven op 19 juni 1998

Voetnoten

1. Bosker et al (1989) gebruikten als indicatoren slaagpercentages, afstroom, drop-out, doubleren, herexamens en schoononderzoeken.

2. Pijls heeft de twee redenen van deze middeling gemist. De eerste reden is de geringe stabiliteit van uitstroom- en zittenblijpercentages, met name in de onderbouw. Deze rechtvaardiging werd reeds eerder gegeven op het openbaar college van 11 mei 1998 (waarvan de bewerking in deze TOR opgenomen is), en dat aan de tweede Trouw publicatie op 13 juni 1998 vooraf ging. De tweede rechtvaardiging van deze middeling is in de Trouw van 13 juni genoemd (Agerbeek, 1998). Door de middeling pasten alle getallen in de krant, waardoor "lezers met behulp van de cijfers in de lijst het oordeel zelf kunnen narekenen".

Boekbesprekingen

Wessum, L van (1997)

De sectie als eenheid: Samenwerking en professionaliteitsopvattingen van docenten in het voortgezet onderwijs

Academisch proefschrift: Vakgroep Onderwijskunde/ISOR/Universiteit Utrecht

Promotoren: N.A.J. Lagerweij en Th. Wubbels - ISBN 90-6709-012-3

Samenwerking tussen docenten staat in toenemende mate in de belangstelling. Het wordt van belang geacht voor de professionele ontwikkeling van docenten, de ontwikkeling van scholen in de richting van de lerende organisatie, de implementatie van innovaties en het behalen van leerprestaties. Het proefschrift van Loes van Wessum biedt nader inzicht in deze materie door de aard en mate van samenwerking binnen scholen te onderzoeken en na te gaan welke factoren verantwoordelijk zijn voor verschillen in samenwerking. Het onderzoek richt zich daarbij op vaksecties in het voortgezet onderwijs en in dit opzicht vormt dit onderzoek een aanvulling op het terrein van onderzoek naar vaksecties. In het navolgende wordt eerst de inhoud van het proefschrift nader besproken, waarnaar vervolgens enige kanttekeningen bij het onderzoek worden geplaatst.

Inhoud

Het proefschrift bestaat uit zeven hoofdstukken. Na een inleidend hoofdstuk waarin de relevantie van samenwerking tussen docenten wordt besproken, wordt in het tweede hoofdstuk inzicht geboden in de vraag welke samenwerkingsvormen er theoretisch te onderscheiden zijn. Op basis hiervan komt de auteur tot een eigen indeling, waarbij zij, voortbouwend op het werk van Little (1990), een viertal samenwerkingsvormen onderscheidt. Het gaat hier om overleg, afspraken, hulp en assistentie en gezamenlijk werk. Verder is er sprake van een kritische bespreking van het onderzoek naar verschillen in samenwerking binnen scholen en vaksecties. Op basis hiervan wordt een voorlopig onderzoekskader ontwikkeld, waarin de opvattingen van docenten een belangrijke rol spelen. De onderzoekster stelt daarbij dat juist op dit punt tot nu toe verricht onderzoek tekort is geschoten.

In het derde hoofdstuk wordt deze constatering nader uitgewerkt door nader in te gaan op deze opvattingen. Deze opvattingen worden conceptueel verankerd door aan te sluiten bij het begrip professionaliteit zoals dat ideaaltypisch door Hoyle (1975) is ontwikkeld. In operationele zin betekent dit dat er drie verschillende vormen van opvattingen worden onderscheiden, te weten de onderwijskundige autonomie-opvatting van docenten, hun organisatorische autonomie-opvatting en hun kennisopvatting. De eerste opvatting heeft betrekking op de vraag of docenten hun autonomie willen delen met hun sectiegenoten, terwijl de tweede opvatting betrekking heeft op de bereidheid van docenten om hun autonomie te delen met andere collega's en de schoolleiding. Wat betreft de kennisopvatting blijkt dat het gaat om de vraag of docenten het kennisbestand van hun vak beschouwen als vaststaand (of afgebakend) of als zich ontwikkelend.

In het vierde hoofdstuk wordt vervolgens het definitieve onderzoekskader gepresenteerd. Hierin wordt verondersteld dat samenwerking tussen sectiegenoten verklaard kan worden uit de bestaande professionaliteitsopvattingen, schoolkenmerken (onder andere schoolgrootte en positie van vaksectie), werkplekcondities (onder andere de rol van de schoolleiding en beschikbare tijd om samen te werken) en sectiekenmerken (onder andere vak, aantal deeltijders en overeenkomst in doelgerichtheid). Verder wordt aangegeven dat de professionaliteitsopvattingen van docenten verklaard kunnen worden uit docentkenmerken zoals leeftijd, sexe, ervaring, opleidingsniveau, de aanstelling die zij hebben en de doelgroep waaraan de docent lesgeeft.

Tot nu toe is de lijn van het proefschrift gevolgd. Om redenen van duidelijkheid wordt in

het navolgende hiervan afgestapt en zal in eerste instantie met name aandacht geschonken worden aan de wijze waarop het onderzoek is uitgevoerd. Vervolgens wordt aandacht geschonken aan de resultaten en de conclusies.

In het onderzoek is sprake van twee typen onderzoek. Het eerste betreft een kwalitatief onderzoek. Hierbij is een drietal scholen betrokken en, verdeeld over deze scholen, een elftal secties. Binnen elke school gaat het om steeds om een aantal secties die in de optiek van de schoolleiding intensief samenwerken en om een aantal secties die matig samenwerken. De wijze waarop de data verzameld zijn is zoals te doen gebruikelijk in kwalitatief onderzoek (interview, documentanalyse), terwijl daarnaast een vragenlijst is afgenomen. Om de data te analyseren is een aantal coderingsschema's ontwikkeld.

Het tweede type onderzoek is kwantitatief van aard. Centraal daarbij staat een aantal voorbeeldig geconstrueerde vragenlijsten die respectievelijk inzicht geven in de aard en mate van samenwerking van docenten binnen secties en in de professionele opvattingen van docenten. Deze vragenlijsten zijn afgenomen bij docenten van de scholen die aan de gevalstudies deelgenomen hebben en, in verkorte vorm, bij docenten van een vierde school (N=202). Deze docenten hebben daarbij ook nog gegevens versterkt over hun persoons- en aanstellingskenmerken. Voor de analyse van de gegevens is gebruik gemaakt van een tweetal analysetechnieken. Met behulp van variantie-analyses is getracht verschillen in samenwerking van secties tussen en binnen scholen te verklaren. Met behulp van regressie-analyses worden relaties tussen de wijze waarop docenten samenwerken en kenmerken als de aanwezige kennisopvatting, het gegeven vak, het opleidingsniveau en zijn/haar aanstellingskenmerken onderzocht.

Op basis van beide type onderzoek blijkt dat samenwerking van docenten vooral gericht is op de vakinhoud en de toetsing en veel minder op didactische aspecten. Verder komt uit de resultaten het informele karakter van de samenwerking naar voren.

Wat betreft de professionaliteitsopvattingen van docenten blijkt uit het onderzoek dat docenten gekenmerkt worden door een gedeelde onderwijskundige en organisatorische autonomieopvatting, zij het dat zij meer bereid zijn hun autonomie te delen met sectiegenoten dan met andere leden van de school (schoolleiding en docenten in andere vakgebieden). Verder blijkt dat zij een ontwikkelingsgerichte kennisopvatting hebben ten aanzien van vakinhoud, didactiek en toetsing. Ten aanzien van deze opvattingen zijn er tenslotte nauwelijks (systematische) verschillen tussen docenten, tussen secties en binnen en tussen scholen.

Op basis van de resultaten aangaande verschillen in samenwerking blijkt dat vakkenmerken geen belangrijke rol spelen. Hoewel het wel zo is dat docenten met een ontwikkelingsgerichte kennisopvatting eerder bereid zijn samen te werken dan docenten met een statische kennisopvattingen, zijn deze kennisopvattingen echter niet, zoals in eerder onderzoek is gesuggereerd, gebonden aan het vak waarin wordt lesgegeven. Verder blijkt dat samenwerking vooral verklaard moet worden uit culturele aspecten. De mate van overeenstemming wat betreft doelen speelt daarbij een rol, maar vooral de kwaliteit van de sociale relaties: als deze goed en intensief zijn, is er de meeste kans op samenwerking.

Voor de hand ligt de opmerking dat deze resultaten tenslotte in het hoofdstuk over de conclusies nog eens naar voren komen. Verder formuleert de onderzoekster aan aantal meer overstijgende conclusies. Een van deze conclusies is dat vaksecties eigenlijk nauwelijks een eenheid vormen, op oppervlakkige wijze samenwerken door 'enkelstagleren' en 'het hier en nu' voor op te stellen, en op belangrijke punten niet bereid hun autonomie te delen anderen binnen de school. Terecht zijn dan vervolgens haar opmerkingen dat deze aspecten op tal van terreinen voor problemen zorgen, bijvoorbeeld bij de implementatie van onderwijsvernieuwingen en bij de ontwikkeling van de school in de richting van een lerende organisatie. In dit verband doet zij ook een aantal suggesties om de bestaande situatie te verbeteren.

Commentaar

Zoals elk proefschrift, kent dit proefschrift discutabele en goede kanten. Om met het eerste beginnen. Een vraag is bijvoorbeeld of alle vraagstellingen wel zo relevant zijn. Immers, over

de aard en mate van samenwerking binnen vaksecties is best een en ander bekend en dit roept de vraag op of het nodig is dit aspect nog eens op explorerende wijze te onderzoeken. Beter ware het geweest een tot nu toe onderbelichte vraagstelling nader uit te werken.

Voorzover er wel nieuwe vraagstellingen aan de orde komen (de relatie professionaliteitsopvattingen en samenwerking bijvoorbeeld) is vervolgens een vraag of er niet een opzet gekozen had moeten worden die een betere beantwoording mogelijk had gemaakt. Nu moet de onderzoekster schipperen met een beperkte data-set en daardoor komen deze vraagstellingen uiteindelijk niet goed uit de verf. In dit opzicht is de (vrijwillige) keuze voor een kleinschalig onderzoek niet gelukkig geweest. Daarbij is ook de legitimering voor dergelijk onderzoek niet erg sterk. De verwijzing naar andere onderzoeken waarbij de dataverzameling moeizaam verliep lijkt mij in ieder geval nauwelijks een voldoende reden om zich maar te beperken tot een klein aantal scholen. Temeer, daar er ook voorbeelden zijn van proefschriften waar, met hoeveel moeite ook, wel veel respondenten bij betrokken zijn geweest. Een ander discutabel punt is de wijze waarop het begrip kennisopvatting is geoperationaliseerd. De onderzoekster stelt dat dit begrip betrekking heeft op de mate waarin docenten hun vakgebied beschouwen als zijnde statisch dan wel zich ontwikkelend. Op basis hiervan zou men verwachten dat de geformuleerde items betrekking hebben op (gepercipieerde) vakkenmerken. Echter, als gekeken wordt naar de operationalisatie hebben veel items daar nauwelijks betrekking op. Zo zeggen items als 'het is belangrijk dat een docent zijn/haar vak vakkennis bijhoudt' en 'het is belangrijk dat een docent de vakliteratuur bijhoudt' weinig inzicht in de vraag of een vak al dan niet statisch is. Dit betekent ook dat de geformuleerde conclusies op dit punt in een ander licht gezien moeten worden.

Tenslotte is er een opmerking te plaatsen over de geformuleerde conclusies. Hoewel in de kern van de zaak haar conclusies juist zijn, sneeuwen uitspraken als 'de sectie is geen eenheid' een beetje het feit onder dat secties eerst een methode kiezen en vervolgens op basis hiervan formele afspraken maken over wanneer welke leerstof aangeboden wordt en hoe deze worden getoetst. Dit is niet alleen bekend uit ander onderzoek, maar ook uit gegevens uit het proefschrift zelf. Dat dan vervolgens allerlei informele samenwerkingsverbanden de boventoon voeren, lijkt mij, gezien de verdeling van docenten over verschillende klassen en schooltypen, logisch en misschien zelfs wel wenselijk. Waarom zou iedereen overal betrokken bij moeten worden? Met andere woorden, dat informele samenwerkingsvormen een belangrijke plaats vervullen is duidelijk, maar moeten ook in het licht gezien worden van eerder genomen besluiten en gemaakte afspraken op de het niveau van de sectie. In dit opzicht geven de eerder geciteerde zinnen een (iets) te ongenueanceerde kijk op het functioneren van vaksecties.

Naast deze discutabele punten zijn er uiteraard ook positieve punten. De wijze waarop de literatuur is verwerkt en de zorgvuldige manier waarop het kwalitatieve onderzoek is uitgevoerd zijn hiervan een voorbeeld. Een ander positief punt is de wijze waarop de onderzoeksinstrumenten voor het kwantitatieve gedeelte zijn ontwikkeld. Deze instrumenten wachten zij als het ware op verdere toepassing in een meer grootschalig onderzoek. Verder is een positief aspect de inzichtelijke wijze waarop de data worden gepresenteerd. Een en ander heeft, ondanks bovengenoemde kritiek, toch wel geleid tot een meer genuanceerde kijk op het functioneren van vaksecties dan tot nu toe in de literatuur naar voren is gekomen.

Literatuur

- Hoyle, E. (1975). Professionalism, professionalism and control in teaching. In V. Houghton, R. McHugh & C. Morgan (Eds.), *Management in education. The management of organization and individuals* (pp. 341-320). London: Ward Lock.
- Little, J. (1990). *The persistence of privacy. Autonomy and initiative in teachers' professional relations. Teacher College Record*, 91, 4, 509-36.

Bob Witziers,
(Universiteit Twente, Faculteit Toegepaste Onderwijskunde)

J.J. Chris Blom

Use-oriented courseware development for agricultural education: An ecological approach
 Proefschrift, Landbouwwuniversiteit Wageningen, 1997, ISBN 90-5485-705-6.

'Elke soort kan in beginsel overal voorkomen, het milieu selecteert.' Dat is de kern van de ecologische benadering. Vertaald naar het gebruik van educatieve computerprogrammatuur, oftewel courseware, houdt dat in dat vooral de context bepaalt welke kwaliteitsaspecten van courseware de doorslag geven bij de beslissing om daarvoor al dan niet gebruik te maken. Dit is het uitgangspunt dat ten grondslag ligt aan het proefschrift van Chris Blom. Het probleem dat centraal staat in zijn dissertatie, is het teleurstellende gebruik van educatieve programmatuur in het agrarisch onderwijs in Nederland. Door bij de ontwikkeling van courseware rekening te houden met de context waarin deze zal worden gebruikt, zou in deze situatie verbetering moeten komen, aldus de auteur. Het proefschrift is opgebouwd uit een theoretisch deel en het verslag van vier deelstudies, uitmondend in een lijst met aanbevelingen voor ontwikkelaars van educatieve programmatuur.

Onderzoeksopzet

Het onderzoek omvat de volgende deelstudies:

1. een schriftelijk vragenlijstonderzoek bij docenten;
2. vraaggesprekken met docenten;
3. een studie naar het proces van courseware-ontwikkeling;
4. een studie naar de organisatie van courseware-ontwikkeling.

Op basis van de resultaten van de eerste deelstudie is een eerste lijst met aanbevelingen geformuleerd die bij de ontwikkeling van courseware in acht zouden moeten worden genomen. Iedere volgende deelstudie leidt tot verfijning en uitbreiding van deze lijst.

Het schriftelijke vragenlijstonderzoek is tweemaal uitgevoerd, in 1990 en 1992. Hiertoe zijn docenten van twaalf scholen voor middelbaar agrarisch onderwijs aangeschreven. Bij de eerste afname heeft dit geleid tot 193 ingevulde vragenlijsten. De oogst bij de tweede afname (bij dezelfde scholen) bedroeg 164 enquêteformulieren. In de vragenlijst is onder meer ingegaan op factoren die van invloed zijn op computer- en coursewaregebruik in de les.

De vraaggesprekken met docenten zijn in 1995 gehouden. Aan deze fase van het onderzoek hebben vijf docenten meegewerkt: twee docenten die lesgeven in het lager agrarisch onderwijs en drie docenten die lesgeven in het middelbaar agrarisch onderwijs. Zij zijn geselecteerd op basis van het gegeven dat zij recent bepaalde coursewarepakketten hadden aangeschaft. De vraaggesprekken dienden meer inzicht te geven in computer- en coursewaregebruik in specifieke onderwijsleersituaties.

In het derde deelonderzoek zijn 32 courseware-ontwikkelprojecten onder de loep genomen. Daarbij ligt de nadruk op zes projecten die tijdens de dataverzameling in de ontwikkelingsfase verkeerden. In deze projecten is informatie vergaard door vraaggesprekken met leden van de projectteams te voeren, door in het kader van het project geproduceerde documenten te verzamelen en door vergaderingen van de projectteams bij te wonen. Bij de 26 overige, tijdens de uitvoering van de studie nog op te starten of reeds afgesloten, projecten beperkte de dataverzameling zich tot de genoemde documenten. Bij de analyse van de data is gebruik gemaakt van een uitgebreid categorieënschema, dat als houvast heeft gediend bij het schetsen van de gang van zaken bij de ontwikkelprojecten.

In het vierde deelonderzoek, tot slot, is de organisatie rond courseware-ontwikkeling ten behoeve van het agrarisch onderwijs in de periode 1989-1992 in kaart gebracht. Dit is gebeurd op basis van de analyse van een groot aantal documenten, het bijwonen van vergaderingen van werkgroepen en het voeren van 26 vraaggesprekken met sleutelpersonen.

Resultaten

Uit het vragenlijstonderzoek komt onder meer naar voren dat computergebruik door docenten vooral samenhangt met het vak waarin zij lesgeven. Ook de hoeveelheid tijd die beschikbaar is voor het voorbereiden van lessen, vormt daarbij een belangrijke factor.

Uit de vraaggesprekken bleek dat docenten met name de voorbereiding van leerlingen op de agrarische praktijk van belang vinden. Courseware die daaraan een bijdrage kan leveren, maakt meer kans om in de lespraktijk te worden ingezet. Wat in dit deel van het onderzoek opvalt, is de beperkte hoeveelheid tijd die docenten aangeven te willen (of kunnen) investeren in het leren omgaan met educatieve programmatuur.

De studie naar de ontwikkeling van coursewarepakketten laat onder andere zien dat de betrokkenheid van leraren in ontwikkelteams vaak tot problemen leidt, doordat de desbetreffende leraar onvoldoende tijd kan investeren in het project. De onderzochte projecten waren voornamelijk gericht op de productie van courseware en niet op de implementatie daarvan in het onderwijs. Dit uitte zich er bij voorbeeld in dat er onvoldoende aandacht werd besteed aan schriftelijk materiaal voor de docent, dat pakketten nauwelijks werden geëvalueerd voordat zij op de markt kwamen, en dat er weinig aandacht werd besteed aan informatievoorziening aan leraren over de manier waarop het pakket in de les kon worden ingezet.

Het vierde deelonderzoek heeft geleid tot een overzicht van organisaties die een rol spelen of speelden bij de ontwikkeling van courseware ten behoeve van het agrarisch onderwijs. Tevens is een blauwdruk gemaakt van een nieuwe opzet, die ertoe zou moeten leiden dat er meer gebruik van courseware wordt gemaakt in de lespraktijk.

De aanbevelingen

Het onderzoek leidt tot een uitgebreide lijst van aanbevelingen, 35 in totaal. Daardoor zijn tegenstellingen bijna onvermijdelijk. Voorbeelden daarvan zijn: 'courseware development should be school specific' versus 'courseware should be embedded in regular curriculum materials'. Pakketten moeten gemakkelijk te gebruiken zijn, maar zouden tevens de ervaren gebruiker de ruimte moeten bieden om het geheel aan te passen aan de eigen wensen. Een belangrijke aanbeveling is dat docenten meer betrokken moeten worden bij de ontwikkeling van educatieve programmatuur, maar het onderzoek laat tevens zien dat juist de deelname van docenten aan ontwikkelteams als knelpunt kan worden ervaren, doordat docenten vaak onvoldoende tijd hebben om de gewenste bijdrage te leveren. Wat toepassingen voor de lespraktijk betreft, wordt aanbevolen de hoeveelheid voorbereidingstijd voor leraren zo beperkt mogelijk te houden. Het gevaar bestaat echter dat dit leidt tot dermate kleinschalige toepassingen, dat men zich kan afvragen of docenten deze nog de moeite van het proberen waard vinden (vgl. Keursten, 1994).

Een andere kwestie is, dat het niet eenvoudig zal zijn om educatieve programmatuur te ontwikkelen die voldoet aan alle aanbevelingen. Hierdoor zal er behoefte bestaan aan een rangordening in de lijst of aan wegingsfactoren bij de aanbevelingen.

Bijdrage aan theorie en praktijk

Het gebruik van Informatie- en Communicatietechnologie (ICT) in het agrarisch onderwijs is tot nu toe onderbelicht gebleven in onderzoek. In deze studie wordt de genoemde onderwijssector onder de loep genomen, hetgeen een positief punt kan worden genoemd. In tegenstelling tot in dissertatieonderzoeken met betrekking tot de situatie in het basis- en voortgezet onderwijs (zie respectievelijk Janssen Reinen, 1996 en Ten Brummelhuis, 1995) wordt hier echter niet getracht een model te schetsen van de mate waarin en de wijze waarop verschillende factoren het computergebruik door docenten beïnvloeden. Er wordt volstaan met het geven van een overzicht van relevante factoren. Ook worden de resultaten van de herhaalde afname van de vragenlijst bij docenten niet met elkaar vergeleken wat de relevante factoren betreft. Een dergelijke vergelijking zou voor de hand liggen, aangezien uit de bovengenoemde onderzoeken in basis- en voortgezet onderwijs de conclusie is getrokken dat in verschil-

lende fasen van implementatie (deels) andere factoren een rol spelen. Bovendien is de steekproef te klein om de aanzienlijke verschillen in computergebruik te verklaren die tussen de twaalf in het onderzoek betrokken scholen optreden. Dit bemoeilijkt het trekken van conclusies over veranderingen die nodig zijn om het gebruik van courseware te stimuleren.

Het tweede deelonderzoek, waarin vraagggesprekken met docenten zijn gevoerd, is erg beperkt van opzet, zeker in vergelijking met het grote aantal gesprekken dat in het derde en vierde deelonderzoek is gevoerd. Het ging slechts om vijf docenten, waarvan er bovendien twee werkzaam waren in het lager agrarisch onderwijs (dat in het vragenlijstonderzoek buiten beschouwing is gebleven).

De beide deelonderzoeken die zijn gericht op de ontwikkeling van courseware en de organisaties die daarbij een rol spelen, leiden tot een uitgebreid overzicht van 'actoren en factoren' die hier kunnen worden onderscheiden. Ook worden daarbij optredende problemen in kaart gebracht, hetgeen zeker voor ontwikkelaars zinvolle informatie zal vormen. Het is jammer dat in de deelstudie die zich richtte op de ontwikkeling van courseware, slechts drie ontwikkeltrajecten tot het einde konden worden gevolgd. Bovendien komen in dit deelonderzoek niet de pakketten aan bod die werden ingezet door de docenten waarmee vraagggesprekken zijn gevoerd. Juist hier zou het interessant zijn geweest om verbanden te zoeken tussen kenmerken van de courseware waarvan het ontwikkeltraject is gevolgd en de opvattingen van docenten over die programmatuur.

Het geheel overziend kan worden gesteld dat een zwak punt van deze studie is, dat de verschillende deelonderzoeken onvoldoende op elkaar afgestemd zijn en dat de deelstudies bij docenten te kleinschalig van opzet zijn om een goed beeld te krijgen van de context waarin de programmatuur wordt ingezet. Nu ligt de nadruk meer op de ontwikkeling van programmatuur en de processen die daarbij een rol spelen, dan op de factoren die van invloed zijn op het gebruik van die programmatuur door docenten. Wat bovendien wordt gemist, is aandacht voor het verschijnsel 'krachtige leeromgeving'. Juist dergelijke leeromgevingen, waarin ICT een belangrijke functie kan vervullen, stellen specifieke eisen aan courseware. Zal de ecologische benadering daadwerkelijk leiden tot meer gebruik van educatieve programmatuur in het onderwijs? De praktijk moet het leren.

Ed Smeets

Instituut voor Toegepaste Sociale wetenschappen
Katholieke Universiteit Nijmegen

Referenties

- Brummelhuis, A.C.A. ten (1995). *Models of educational change: the introduction of computers in Dutch secondary education*. Academic Dissertation. Enschede: Universiteit Twente.
- Janssen Reinen, I.A.M. (1996). *Teachers and computer use: the process of integrating IT in the curriculum*. Academic Dissertation. Enschede: Universiteit Twente.
- Keursten, P. (1994). *Courseware-ontwikkeling met het oog op implementatie: de docent centraal*. Academisch proefschrift. Enschede: Universiteit Twente.

L. Sontag

Vormgeving en effecten van onderwijs aan vier- tot zevenjarige leerlingen

Thesis Publishers, Amsterdam, 1997, ISBN 90-9011158-1, f 35,-

Achtergronden

Hoewel de integratie van kleuter- en lagere school al ruim tien jaar een feit is, is het aantal studies dat gericht is op het reilen en zeilen in de kleutergroepen van de basisschool zeer beperkt gebleven. Voor een deel kan dat ongetwijfeld verklaard worden uit de traditionele opvatting dat voor kleuters spelen centraal dient te staan, en niet leren. Evaluatie van het kleuteronderwijs, en dan zeker als het zou gaan om cognitieve-effectenmetingen is voor menig- een nog taboe. Er is de laatste jaren echter een kentering ontstaan in het denken over het

onderwijs aan kleuters. Kleuteronderwijs, of ruimer: onderwijs aan jonge kinderen, wordt momenteel meer en meer in verband gebracht met de bestrijding van onderwijsachterstanden. De evaluatie van het Onderwijsvoorrrangsbeleid (OVV) heeft laten zien dat autochtone en vooral allochtone arbeiderskinderen doorgaans al met een flinke achterstand in groep 1 van het basisonderwijs starten. Tegelijkertijd is duidelijk geworden dat zij deze achterstand in hun gang door het basisonderwijs doorgaans niet meer inhalen (vgl. Mulder, 1996). Het besef is doorgedrongen dat, wil men de onderwijskansen van deze groepen verbeteren, er zo vroeg mogelijk actie zal moeten worden ondernomen. Dit heeft geleid tot een explosieve groei van allerlei voor- en vroegschoolse programma's (bv. Opstap) en tegelijkertijd tot meer aandacht voor de cognitieve ontwikkeling in de kleutergroepen. Een belangrijke stimulans daarbij is gegeven door de Commissie Evaluatie Basisonderwijs (1994) en de Commissie (Voor)schoolse Educatie (1994).

In het kielzog van beide commissies zijn onlangs twee studies gepubliceerd waarin het onderwijs aan jonge kinderen centraal staat. De eerste studie, van Driessen & Claassen (1996) (zie ook Driessen, 1997, in dit tijdschrift), heeft betrekking op de relatie tussen cognitieve en niet-cognitieve leerlingkenmerken en de pedagogisch-didactische aanpak van leerkrachten in groep 1 en 2. Het betreft een landelijk, grootschalig onderzoek waarbij gebruik is gemaakt van data van het cohort Primair Onderwijs ('PRIMA'). De gegevens hebben betrekking op één meetmoment. De tweede studie, die in deze recensie centraal staat, is het proefschrift van Sontag (1997). Beide studies vertonen veel gemeenschappelijke trekken; ik zal daar op het eind van deze bespreking kort aandacht aan besteden. Eerst richt ik me echter op het onderzoek van Sontag.

Theoretisch kader

In het eerste deel van haar dissertatie gaat Sontag in op de theoretische achtergrond van haar onderzoek. De algemene vraagstelling luidt: Wat is de invloed van kenmerken van de school- en klasse-organisatie op de prestaties en de ontwikkeling van leerlingen in de onderbouw van het basisonderwijs? Alvorens vanuit de empirie een antwoord te geven op deze vraag presenteert Sontag eerst een overzicht van (klassieke) theorieën over de ontwikkeling van taal en cognitie bij jonge kinderen (o.m. Piaget, Vygotsky). Vervolgens bespreekt zij enkele, wat meer recente, modellen van en visies op ontwikkeling en onderwijs (bv. Ervaringsgericht Onderwijs, 'Basisontwikkeling'). Van daaruit gaat ze kort in op de resultaten van enkele studies op dit gebied. Daarbij wordt tevens een relatie gelegd met het effectieve-scholenonderzoek. Op basis van het voorafgaande wordt de algemene vraagstelling uiteengelegd in een aantal specifiekere onderzoeksvragen.

Onderzoeksoptzet

De steekproef voor het onderzoek bestond uit een cohort van 480 leerlingen en 99 leerkrachten (uit groep 1 en 2) van 28 scholen in de vier grote steden. De ratio achter deze regio-keuze wordt helaas niet duidelijk gemaakt; evenmin wordt een poging ondernomen aan te geven in hoeverre de verzamelde gegevens generaliseerbaar zijn. Het onderzoek concentreert zich op vier etnische groepen, namelijk Nederlanders, Surinamers/Antillianen, Turken en Marokkanen; 38% van de leerlingen is allochtoon. Deze leerlingen zijn gevolgd vanaf het moment dat ze begonnen in groep 1 tot aan het eind van groep 3. Twee keer per jaar is de cognitieve ontwikkeling (i.c. verbale en nonverbale intelligentie) getest, en bovendien zijn in groep 3 toets-technisch en begrijpend lezen en rekenen afgenomen. De school- en klaskenmerken zijn in groep 1 en groep 2 verzameld met behulp van uiteenlopende methoden, te weten vragenlijsten, interviews, logboeken en observaties.

Beschrijvende analyses

Uit een eerste serie beschrijvende analyses van intelligentie en prestaties komen de inmiddels bekende verschillen naar sociaal milieu en etnische herkomst naar voren. In groep 1 en 2 zijn

nonverbale verschillen vooral milieu-gerelateerd, terwijl verbale verschillen met zowel milieu als etniciteit samenhangen. De gegevens wat betreft de passieve woordenschat laten zien dat de Marokkaanse en Turkse leerlingen op het eind van groep 2 een inhaalrace lopen; dit neemt echter niet weg dat hun prestaties dan nog ver onder die van de Nederlandse en Surinaamse leerlingen liggen. De prestaties qua technisch lezen en rekenen in groep 3 blijken sterker met milieu dan met etniciteit samen te hangen. Tegelijkertijd blijkt dat er wel een samenhang bestaat tussen de etnische herkomst en begrijpend lezen, een vaardigheid die minder van instructie afhankelijk is.

Scholen en leerkrachten verschillen sterk wat betreft hun effectiviteitskenmerken, zoals ouderparticipatie, monitoring van vorderingen, taalbeleid (op NT2-onderwijs en zijn overigens de meeste leerkrachten niet voorbereid); dergelijke verschillen treden echter niet op met betrekking tot onderwijskundig leiderschap, teamconsensus en prestatiegerichtheid. De verwerving van de basisvaardigheden (taal, voorbereidend lezen en rekenen) vindt in de kleutergroepen vooral spelenderwijs plaats; gerichte leeractiviteiten worden slechts spaarzaam ondernomen.

Sontag heeft geprobeerd de effectiviteitskenmerken samen te vatten tot een beperkt aantal factoren. Evenals veel andere onderzoekers voor haar (bv. Driessen & Claassen, 1996) moest ook zij echter concluderen dat dat eigenlijk niet mogelijk is; dergelijke kenmerken hangen doorgaans namelijk als los zand aan elkaar. Vandaar dat zij uiteindelijk de toevlucht heeft genomen tot een werkwijze waarbij de afzonderlijke kenmerken eerst zijn getrichotomiseerd (0 'ineffectieve leerkrachten'; 1 'redelijk effectieve leerkrachten'; 2 'effectieve leerkrachten') en deze vervolgens zijn gesommeerd tot twee dimensies, namelijk schoolorganisatie en pedagogisch-didactische aanpak. Sontag is nog nagegaan in hoeverre er sprake is van stabiliteit van de pedagogisch-didactische aanpak over twee leerjaren (groep 1 en 2) heen. Op de afzonderlijke kenmerken blijkt er wel enige stabiliteit te zijn, maar qua aanpak-als-geheel niet. Ten dele valt dit volgens Sontag te verklaren uit de gedachte dat de aanpak van de leerkracht zou moeten sporen met de ontwikkelingsfase van de leerling.

Effectenanalyses

De kern van Sontags dissertatie wordt gevormd door multilevel-analyses waarbij wordt nagegaan in hoeverre de cognitieve en linguïstische ontwikkeling van de leerlingen wordt beïnvloed door de school- en klaskenmerken. Deze analyses zijn modelmatig opgebouwd. In een eerste serie analyses is de afhankelijke variabele de score op de tweede testafname in het betreffende leerjaar. Achtereenvolgens worden ter verklaring daarvan drie blokken van kenmerken ingevoerd/toegevoegd: leerlingkenmerken (o.a. SES, herkomst, de score op de eerste testafname), leerkrachtkenmerken (aanpak en organisatie) en contextkenmerken (o.m. cultureel-etnische situatie, groepssamenstelling). De resultaten van deze analyses kunnen teleurstellend worden genoemd. Hoewel er verschillen zijn tussen leerkrachten in hun pedagogisch-didactische aanpak en schoolorganisatie, houden deze geen verband met de verbale en nonverbale vaardigheden van hun leerlingen. De meeste verklaringskracht heeft steeds het resultaat op de eerder afgenomen toets, het aanvangsniveau.

In een tweede analyse analyseert Sontag de ontwikkeling in de eerste twee leerjaren; dat doet zij met behulp van een soort groeiscoringen. Omdat ze in de eerdere analyses toch niets deden, zijn in deze analyse de aanpak en schoolorganisatie niet meer meegenomen. Hier blijkt dat er sprake is van individuele verschillen in ontwikkeling. De verbale ontwikkeling is vooral afhankelijk van de leeftijd (jongere leerlingen groeien sneller) en etniciteit (Turken en Marokkanen sneller); bij de nonverbale ontwikkeling is vooral de leeftijd en het geslacht van belang (meisjes sneller).

Waardering

Een vlot geschreven proefschrift, waarin zonder poespas - maar deskundig - wordt toegewerkt naar het beantwoorden van de onderzoeksvragen. Voor degenen die alleen geïnteresseerd zijn

in die antwoorden is dit het ideale en efficiënte voorbeeld; wellicht dat anderen echter graag nog wat meer relaties uitgediept zouden hebben gezien (bv. in de multilevel-analyses naast de twee samenvattende dimensies ook de afzonderlijke effectiviteitskenmerken). Jammer is het dat er her en der wat hinderlijk storende foutjes in de tekst en tabellen zijn blijven staan; zo zijn bijvoorbeeld in alle multilevel-tabellen verschuivingen in de kolommen opgetreden en kloppen de bijbehorende teksten ook niet altijd. Een pluspunt van het onderzoek is dat gebruik is gemaakt van verschillende dataverzamelmethode, waarmee een kritiek die soms op puur pen-and-paper onderzoek wordt geleverd, wordt gepareerd; een nadeel is dat alleen een relatief beperkt aantal leerlingen uit enkele grote steden in de Randstad zijn onderzocht. Hoewel de bevindingen met name wat betreft de relaties tussen de cognitieve effect-maten en de zogenaamde effectieve-schoolkenmerken zonder meer teleurstellend zijn, is het wel tekenend dat deze overeenkomen met die uit kwantitatief onderzoek waarbij alleen met toetsen en vragenlijsten is gewerkt.

Wat betreft een vergelijking van Sontags onderzoek met dat van Driessen & Claassen (1996) kan ik heel kort zijn. Hoewel het design afwijkt, de steekproeven verschillen, de dataverzamelmethode variëren en de geselecteerde kenmerken niet in alle opzichten dezelfde zijn, is in beide onderzoeken gekozen voor dezelfde analysetechnieken en zijn ook de bevindingen nagenoeg identiek: de effectieve-schoolkenmerken blijken dus niet effectief. Evenals Scheerens & Bosker (1997) dat in het voorwoord van hun boek over dit onderwerp doen, zullen velen zich daarom afvragen: 'is the idea of school effectiveness not a myth?'.
 11

Referenties

- Commissie Evaluatie Basisonderwijs (1994). *Onderwijs aan jonge kinderen. Deelrapport 3*. Den Haag: SDU.
- Commissie (Voor)schoolse Educatie (1994). *(Allochione) kleuters meer aandacht*. Rijswijk: Ministerie van WVC.
- Driessen, G. (1997). Voorbereidend lezen en rekenen in kleutergroepen: werkwijzen en resultaten. *Tijdschrift voor Onderwijsresearch*, 22, (3), 145-160.
- Driessen, G. & Claassen, A. (1966). *Voorbereidend lezen, rekenen en schrijven. Toetsresultaten en leerling-en leerkrachtkenmerken in de kleutergroepen van het basisonderwijs*. Ubbergen: Uitgeverij Tandem Felix.
- Mulder, L. (1996). *Meer voorrang, minder achterstand? Het Onderwijsvoorrrangsbeleid getoetst*. Nijmegen: ITS.
- Scheerens, J. & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Geert Driessen
ITS

Universiteit van Nijmegen

Grafische calculators in de wiskundeles: een onderzoek naar de invoering van de grafische calculator in het VWO

E. Harskamp (GION-RUG), C. Suhre (GION-RUG) en A. van Streun (IWI-RUG)

ABSTRACT

Recent plans for the introduction of the graphics calculator in the mathematics curriculum of Dutch high schools (age 16-18) have raised concerns about the consequences for teachers and pupils. The most important issue is whether the graphics calculator supports or hampers mathematical problem solving.

The University of Groningen has carried out an experiment with pupils age 16 from 12 high schools. The design was a pretest-posttest design with two experimental conditions using the graphics calculator and one control condition without the calculator. The results point towards a change in teacher instruction behavior and in pupils' problem solving approaches. Teachers using the calculator appear to make more use of graphics and tables during instruction than teachers of the control condition. Pupils in the experimental conditions use a wider range of problem solving approaches and solve more problems graphically than pupils in the control condition. Female students make less use of the calculator than male students, but score equally well on the posttest. In both experimental conditions low scoring pupils on the pretest show better results on the posttest than their equals in the control condition. These results correspond with outcomes of international research.

TECHNOLOGIE IN WISKUNDE TWEDE FASE HAVO-VWO

Met de invoering in 1998 van het studiehuis in de Tweede Fase HAVO/VWO worden ook nieuwe leerplannen en examenprogramma's geïntroduceerd. Voor bijna alle schoolvakken is er sprake van een toename in het gebruik van moderne technologie, zoals computers. Dat geldt ook voor de nieuwe wiskunde-examenprogramma's, waarin rekening wordt gehouden met het gebruik van een grafische rekenmachine (GRM). Een GRM is een calculator die behalve rekenmogelijkheden de mogelijkheid biedt om grafieken van functies weer te geven in een venster en deze nader te onderzoeken.

Voor het vak wiskunde kan de introductie van grafische rekenmachines om verschillende redenen een bijdrage leveren aan de verbetering van de onderwijskwaliteit. Ten eerste nemen grafische calculators leerlingen bewerkelijke handelingen, zoals het tekenen van assenstels, het uitrekenen van punten en het trekken van lijnen en krommen uit handen. Leerlingen moeten leren om passende wiskundekennis (parameters en functievoorschriften) te leveren waarmee de calculator rekent en tekent. In de grafieken kunnen leerlingen oplossingen exploreren en controleren. Het onderwijs kan zich hierdoor meer richten op het leren oplossen van wiskunde problemen. Een tweede voordeel van de grafische calculator is dat meer nadruk gelegd kan worden op aanschouwelijk onderwijs en op mogelijkheden om problemen langs grafische weg op te lossen. Dit kan het begrip bevorderen van de relatie tussen functies en hun grafieken (zie Van Streun, 1996). Een grafische probleemaanpak biedt zwakke leerlingen wellicht meer mogelijkheden om een probleem op te lossen. Een derde voordeel is het feit dat

alle leerlingen over een eigen calculator kunnen beschikken. Daardoor is de drempel voor gebruik laag. Vooral voor meisjes kan dat een voordeel zijn. Meisjes maken zoals bekend doorgaans minder gebruik van computers dan jongens (Whitley, 1997).

Behalve voordelen kunnen er echter ook nadelen zijn verbonden aan het gebruik van de GRM. Een nadeel zou kunnen zijn dat leerlingen erg afhankelijk worden van de grafische calculator en klakkeloos gegevens van het venster overnemen (Tuska, 1993). Voordat de GRM op grote schaal gebruikt gaat worden, is het daarom belangrijk na te gaan hoe het apparaat in het wiskundeonderwijs kan worden geïntegreerd en welke uitwerking het apparaat vermoedelijk zal hebben op de leerresultaten van leerlingen.

In dit artikel wordt verslag gedaan van onderzoek naar de gevolgen van de integratie van de GRM in het wiskunde curriculum voor leerlingen in de bovenbouw van het VWO. Daarbij is uitgegaan van een curriculum dat veel toepassingsproblemen biedt met de GRM en dat daarbij grafische oplossingen stimuleert. Het curriculum biedt echter ook oefeningen voor algoritmische oplossingsprocedures. Om een realistische indruk te krijgen van de gevolgen dit curriculum is het uitgevoerd door docenten zonder veel ervaring in gebruik van de GRM. Onderzocht is of leerlingen door de GRM en door de veranderingen die de GRM meebrengt voor het onderwijs anders en beter problemen leren oplossen. Tevens is onderzocht of meisjes de GRM even vaak gebruiken als jongens en of eventuele verschillen in gebruik gevolg hebben voor de leerresultaten.

VERANDERINGEN IN HET LESGEVEN EN MOGELIJKE EFFECTEN OP LEERPRESTATIES

In de internationale literatuur zijn reeds tientallen artikelen verschenen over de grafische calculator in het wiskundeonderwijs. De meeste artikelen doen verslag van praktijkervaringen of geven argumenten voor de invoering van de grafische calculator. Slechts een beperkt aantal artikelen gaat over empirische studies naar de effecten van de grafische calculator op de manier van lesgeven en de leerprestaties (zie voor een overzicht Penglase & Arnold, 1996).

Veranderingen in de manier van lesgeven

Uit studies blijkt dat invoering van de GRM gepaard kan gaan met invoering van probleemgericht onderwijs. Een studie van Army (1992) toont dat integratie van de GRM in het wiskundeonderwijs leidt tot versnelling in de behandeling van de belangrijkste eigenschappen van grafieken en functies. Daardoor kan de docent meer tijd vrijmaken voor probleemgericht onderwijs waarbij leerlingen grafieken exploreren en gesprekken voeren over het oplossen van problemen. Onderzoek van Rich (1991) laat zien dat de GRM leerlingen stimuleert meer kritische vragen te stellen. Mede hierdoor zijn leraren meer geneigd om verschillende oplossingen van de leerlingen te bespreken en om leerlingen zelfstandig aan een probleem te laten werken. Dat blijkt ook uit het onderzoek van Heid (1988). Heid liet twee groepen studenten werken met grafische programma's op de computer om diverse grafieken (polynomen en rationale functies) te onderzoeken. Ze moesten overeenkomsten, verschillen en eigenaardigheden van de grafieken van deze functies achterhalen en beredeneren. Leerlingen begonnen na aanmoediging grafieken van diverse functies te exploreren en diverse oplossingswijzen te evalueren. Volgens Heid (1988) en Army (1992) bevordert integratie van de GRM in het wiskunde curriculum dat in het onderwijs meer nadruk komt te liggen op uitleg via een tabel of grafiek en minder nadruk op algebraïsche oplossingsmethoden.

Effecten op de leerprestaties

Naar de effectiviteit van de grafische calculator in de wiskunde van het voortgezet onderwijs zijn in het buitenland een aantal veldexperimenten uitgevoerd. Zo heeft Ruthven (1990) onderzoek uitgevoerd in vier Engelse scholen met leerlingen in de leeftijd van 16-19 jaar. De leer-

stof in dit experiment ging over het gebruik van grafieken en het differentiëren van functies. De leerlingen van de experimentele conditie gebruikten in dit experiment een jaar lang de grafische calculator. Ze gebruikten de grafische calculator voornamelijk voor het controleren van formules en functievoorschriften, door de gegevens van de grafieken op de calculator te vergelijken met de uitgerekende gegevens. De leerlingen van de controleconditie kregen dezelfde leerstof zonder grafische calculator. Uit de onderzoeksresultaten blijkt dat de leerlingen van de experimentele conditie op het natoetsonderdeel waarin ze een functievoorschrift bij een grafiek moesten geven, beter scoorden dan de controlegroep. Op het onderdeel waarbij leerlingen gegevens uit grafieken moeten interpreteren (en waarbij de grafische rekenmachine niet wordt gebruikt) was er echter geen verschil met de controlegroep. Opvallend was dat de meisjes in de experimentele groep op het onderdeel waar functievoorschriften bij grafieken moesten worden gezocht, beter scoorden dan de jongens.

Shoaf-Grubbs (1994) voerde onderzoek uit bij vrouwelijke leerlingen van een college in de Verenigde Staten. De leerlingen waren ingedeeld in twee condities, een experimentele en een controleconditie. In de experimentele conditie gebruikten de leerlingen de grafische calculator, in de controleconditie niet. In beide groepen was het lesmateriaal hetzelfde. Na afloop van het experiment werden eindtoetsen afgenomen, waarbij de leerlingen de grafische calculator niet mochten gebruiken. Opvallend was dat er op een algebratoets geen verschil bestond in scores tussen beide groepen. Op drie toetsen die het inzicht in de relatie tussen formules en grafieken testen, deed de experimentele groep het significant beter. Vooral de zwakke leerlingen uit de experimentele groep bleken te hebben geprofiteerd van het gebruik van de grafische calculator.

Maxwell en Quesada (1994) gebruikten voor hun experiment met collegestudenten in de Verenigde Staten een nieuw leerboek over functies en grafieken, waarin het toepassen van de grafische calculator is geïntegreerd in de leerstof. De experimentele groep volgde dit nieuwe leerboek. De controlegroep volgde het gewone precalculus leerboek zonder grafische calculator. De experimentele leerlingen scoorden hoger op open vragen over functies, grafieken en vergelijkingen dan de leerlingen uit de controlegroep. Er was geen verschil op meerkeuzevragen over deze onderwerpen tussen de onderzoeksgroepen. De open vragen leenden zich volgens de onderzoekers, goed voor het grafisch exploreren en controleren. De meerkeuzevragen meten vooral de kennis van feiten en procedures. De onderzoekers concluderen dat de grafische calculator de begripsvorming en het controleren van oplossingen extra bevordert en dat de leerlingen met de grafische calculator in de lessen geen achterstand in kennis van feiten en procedures oplopen.

Adams (1994) deed in acht klassen op dezelfde universiteit onderzoek naar het effect van het gebruik van de grafische calculator op het kiezen van waarden voor het domein, het bereik en de schaal voor grafieken van functies, en naar het effect van calculatorgebruik op het identificeren en opstellen van passende functievoorschriften bij verschillende grafieken. De studenten gebruikten allen hetzelfde leerboek. De studenten in het onderzoek werden per twee klassen verdeeld over vier onderzoekscondities: een groep met de grafische calculator en extra opdrachten, een groep met alleen de grafische calculator, een groep met alleen extra opdrachten en een controlegroep. De resultaten laten zien dat zowel het gebruik van de grafische calculator als de extra opdrachten leiden tot beter inzicht in de kenmerken van de grafiek van een bepaalde functie, maar ook tot het beter analyseren van het functievoorschrift dat bij een gegeven grafiek past. Opvallend is dat in de experimentele condities vooral de minder goede studenten profiteren van het gebruik van de grafische calculator.

In de hier weergegeven onderzoeken zijn door de docenten vrijwel geen knelpunten in het gebruik van de grafische calculator in de klas gesignaleerd. Dit hangt mogelijk samen met het gegeven dat de onderzoekers vaak het gebruik ervan grondig voorbereid hadden en het onderwijs vervolgens zelf gaven. Of dat ook geldt wanneer een doorsnee docent zonder specifieke voorbereiding de grafische calculator gaat gebruiken, komt in ons onderzoek aan de orde.

EXPLORATIEF ONDERZOEK NAAR DE EFFECTEN VAN GEBRUIK VAN DE GRM

De grafische calculator moet binnen afzienbare tijd worden geïntegreerd in het wiskundeonderwijs in de Tweede Fase van het voortgezet onderwijs. Onduidelijk is in hoeverre docenten hun lesgedrag zullen aanpassen en of de grafische calculator het oplossen van wiskunde problemen ondersteunt. Het is daarbij van belang om na te gaan hoe frequent leerlingen gebruik maken van de grafische calculator en of zich in het gebruik verschillen voordoen tussen jongens en meisjes. Voor invoering van de GRM op grote schaal is informatie over deze aspecten van het gebruik door docenten en leerlingen en over de leerresultaten van belang. Het onderzoek richt zich op drie vragen.

Onderzoeksvraag 1:

In hoeverre komt de manier van lesgeven overeen met de bedoelingen van het nieuwe wiskunde curriculum en welke problemen doen zich voor?

Onderzoeksvraag 2:

Wordt de GRM door jongens vaker gebruikt dan door meisjes?

Onderzoeksvraag 3:

Welke effecten heeft het gebruik van de grafische calculator op de oplossingsmethoden die leerlingen gebruiken en op de leerresultaten?

Verwachtingen

De grafische calculator leent zich er bij uitstek voor leerlingen opgaven te laten exploreren, functies te laten schetsen, uitleg te laten geven over de gevonden oplossing en voor het hardop denkend oplossen van een probleem door de docent. Bij *vraag 1* bestaat de verwachting dat leraren die met de grafische calculator lesgeven, meer nadruk leggen op probleemgericht onderwijs. De nadruk in de nieuwe leerboeken waarin de GRM is opgenomen, ligt op het leren opstellen van functievoorschriften en vergelijkingen om een grafiek te kunnen plotten waaruit een oplossing is af te lezen, en op het leren kennen van de eigenschappen van verschillende soorten functies door hun grafieken te bestuderen. Verwacht wordt dat docenten vaker uitleg zullen geven aan de hand van tabellen (ter voorbereiding op grafieken en functies), de grafische calculator of van grafische weergave van functies op het bord.

Bij *vraag 2* is de verwachting dat jongens vaker gebruik maken van de GRM dan meisjes. Meisjes maken doorgaans minder gebruik van technologische hulpmiddelen dan jongens (Whitley, 1997). Omdat leerlingen persoonlijk de beschikking hebben over een GRM is de verwachting dat het verschil in de mate van gebruik tussen jongens en meisjes beperkt zal zijn.

Wat *vraag 3* betreft, luidt de verwachting dat leerlingen die met de GRM les krijgen, vaker een grafische oplossingswijze hanteren dan leerlingen zonder GRM. Ze komen immers veelvuldig in aanraking met het tekenen (plotten) van de grafiek van een functie en het vinden van oplossingen van wiskunde problemen via grafieken. Of de verschuiving in oplossingswijzen ook gevolgen zal hebben voor de wiskundige prestaties staat nog te bezien. Het vermoeden bestaat dat vooral de wat zwakkere leerlingen profijt zullen hebben van de grafische calculator (zie vorige paragraaf).

OPZET VAN HET ONDERZOEK

Uitgangspunt bij de inrichting van het experiment is vergelijkbaarheid van de condities wat betreft leerlingpopulatie en de te behandelen leerstof. Omdat de nieuwe versie van de methode Wiskundelijk de GRM heeft geïntegreerd in leerstof die vrijwel identiek is met de voorgaande versie, is het onderzoek gehouden op scholen die de methode Wiskundelijk gebruiken. De doelgroep bestaat uit leerlingen uit klassen 4-VWO. De leerstof voor VWO-4 wordt behan-

deld in vier hoofdstukken. Hierin komen de volgende vier onderwerpen aan de orde: Functies en Grafieken, Veranderingen, Exponentiële Functies en Goniometrische Functies.

Onderzoeksgroep en experimenteel ontwerp

Het onderzoek is uitgevoerd in het schooljaar '95-'96 als een quasi-experiment. Er zijn twee experimentele condities onderscheiden en een controleconditie. In de experimentele condities gebruikten de leerlingen allemaal een grafische calculator van Texas Instruments, de TI-81. Bij deze calculator kregen de leerlingen, behalve een handleiding ook een handige instructiekaart met een samenvatting van veel voorkomende gebruikssituaties en bijbehorende knoppencombinaties.

De eerste experimentele conditie (**Exp 1**) bestond uit het gebruik van de grafische calculator gedurende het hele schooljaar. De klassen zijn verdeeld over twee scholen, waarin men er voor gekozen had, om vooruitlopend op het nieuwe wiskundeprogramma, de GRM geheel in te voeren. In deze conditie werkten de leerlingen met de nieuwe versie van de methode Wiskundelij die bij een deel van de opgaven aanwijzingen biedt voor gebruik van de GRM. De tweede experimentele conditie (**Exp 2**) werd uitgevoerd in vijf klassen, afkomstig van vier scholen die voorafgaande aan de integrale invoering van de GRM alvast wilden proefdraaien met de GRM. In deze conditie werkten de leerlingen slechts één hoofdstuk met de nieuwe versie van de methode Wiskundelij. Alleen bij dat hoofdstuk gebruikten ze de grafische calculator. Bij de overige hoofdstukken werkten ze met de oude versie van de methode. Vanwege de beperkte hoeveelheid GRM's gebruikte elke klas in deze conditie gebruikte GRM bij een ander hoofdstuk. Dit is gedaan om twee redenen. Enerzijds kon het onderzoek naar de implementatie uitgevoerd worden in meer klassen dan wanneer het hoofdstuk beperkt zou zijn gebleven tot één hoofdstuk. Anderzijds kon worden nagaan of het gebruik van de GRM zijn gebleven tot één hoofdstuk wordt behandeld, van invloed is op de prestaties. Hoewel in elk hoofdstuk een ander type functie aan de orde komt, zijn de kenmerken van het functie-onderzoek namelijk wel vergelijkbaar.

De klassen zijn niet volledig aselekt aan de experimentele condities toegewezen. De eerste experimentele conditie vereist namelijk dat de leerlingen het hele jaar de beschikking hebben over de grafische calculator. Voor deze conditie is gebruik gemaakt van klassen waar de school de GRM verstrekt. De overige klassen zijn aselekt toegewezen aan de tweede experimentele conditie en de controleconditie. Vier klassen fungeerden als controlegroep. Ze kregen onderwijs uit de oude versie van Wiskunde Lijn 4 vwo. Ze kregen dezelfde wiskunde als de andere klassen maar geen leerstof met de grafische calculator. De leerlingen in de controleconditie gebruikten alleen hun zakrekenmachine als hulpmiddel bij het oplossen van opgaven.

In tabel 1 staat een overzicht van het (quasi-)experimentele onderzoeksontwerp.

Tabel 1: Experimenteel onderzoeksontwerp

onderzoekconditie	aantal klassen	aantal leerlingen	typering van de conditie
Experimenteel 1	3	73	Nieuwe versie van de methode en grafische rekenmachine bij alle vier hoofdstukken
Experimenteel 2	5	117	Nieuwe versie van de methode en grafische calculator bij één hoofdstuk en verder de oude versie zonder de calculator
Controle	4	101	Oude versie van de methode bij alle vier hoofdstukken zonder grafische calculator

Aan het onderzoek deden iets meer meisjes (52 %) mee dan jongens. Het minimum percentage meisjes in een klas was 30 % en het maximum 79 %. De leerlingen in de drie condities maakten aan het begin van het schooljaar dezelfde voortoets. De voortoets toetst de aanwezige kennis over functies, grafieken en vergelijkingen. Bij alle klassen zijn bij één hoofdstuk lesobservaties uitgevoerd en leerlingen en docenten geïnterviewd. Alle leerlingen maakten aan het eind van het schooljaar dezelfde eindtoets met stof over de vier onderwerpen. Tijdens het schooljaar hebben de leerlingen uit de tweede experimentele conditie een afsluitende toets gemaakt over het hoofdstuk dat ze hebben gedaan. De klassen in de eerste experimentele conditie hebben alle hoofdstukken van de nieuwe versie van Wiskundelijk gedaan. Om de omstandigheden in de condities gelijk te houden hebben de leerlingen van de eerste experimentele conditie en van de controleconditie ook elk een toets over één hoofdstuk gedaan. De resultaten van de toetsen per hoofdstuk zijn elders beschreven (Harskamp, Van Streun en Suhre, 1996). Daarbij bleek dat de leerlingen die met de nieuwe versie van Wiskundelijk werkten, iets beter presteerden op de toetsen.

Bij de natoets hadden de leerlingen uit de experimentele condities de beschikking over de GRM. Ze mochten deze bij alle opgaven gebruiken. De leerlingen in de controleconditie mochten alleen een standaard rekenmachine gebruiken. Aangezien voor veel leerlingen in de tweede experimentele conditie het meer dan drie maanden geleden was dat ze de GRM gebruikt hadden is de volgende voorbereiding op de toets gevolgd. Veertien dagen voor de eindtoets kregen de leerlingen uit alle drie condities een vijftal opgaven met aanwijzingen om zich op de eindtoets voor te bereiden. De leerlingen in de experimentele condities kregen een aantal tips voor het gebruik van de grafische calculator. De leerlingen in de controleconditie kregen tips om de opgaven handmatig of met de gewone rekenmachine op te lossen. In de eindtoets zijn alleen opgaven opgenomen, die passen bij de gewone leerstof uit 4 vwo. Opgaven die zich heel goed lenen voor een grafische aanpak, maar zonder een grafiekenprogramma slecht zijn uit te voeren (bijvoorbeeld over families van grafieken), zijn weggelaten. In feite zijn de extra mogelijkheden, die de grafische calculator biedt, niet in de toetsen terug te vinden. Dit is gedaan om bias in de toetsen in het voordeel van de experimentele groepen, te voorkomen.

Het experimentele lesmateriaal

De grafische calculator is in de nieuwe versie van de methode Wiskundelijk voorzichtig en terughoudend geïntegreerd, zodat de leerstof vergelijkbaar is met die van de oude versie van de methode (zie Harskamp, Van Streun en Suhre, 1996). In de nieuwe versie staat het leren van de wiskundige begrippen en oplossingswijzen centraal en niet de instructie van alle mogelijkheden van een bepaald type rekenmachine. Bijvoorbeeld de knop voor het laten berekenen van de afgeleide (NDERIV) is niet behandeld. De oplossing van een vergelijking met een afgeleide wordt algebraïsch berekend of grafisch benaderd met de grafische rekenmachine. In de nieuwe versie is slechts een deel van de opgaven gericht op het toepassen van de grafische calculator. Deze opgaven zijn voorzien van een symbool dat aangeeft dat de grafische calculator mag worden gebruikt. De calculator kan overigens ook alle wetenschappelijke berekeningen aan die met een zakrekenmachine worden gedaan.

In het hoofdstuk Functies en Grafieken dient ongeveer de helft van de opgaven gemaakt te worden met de grafische calculator. In het hoofdstuk Veranderingen is dat 33 %. In het hoofdstuk Exponentiële Functies zijn 25 % van de opgaven bedoeld voor de grafische calculator en in het hoofdstuk Periodieke Functies is dat 35%.

Instrumenten en verwerking van gegevens

Bij alle docenten in het onderzoek zijn twee *lesobservaties* uitgevoerd van instructielessen. Tijdens deze lessen is per minuut gescoord of lesgedragingen voorkwamen die zijn te kwalificeren als:

1. probleemgericht onderwijs (laten exploreren, eigen oplossingen laten vertellen of functie laten schetsen),

2. het gebruik van grafieken en tabellen (uitleg geven van een opgave aan de hand van een tabel of grafiek op het bord, of door gebruik van de GRM).

De observaties zijn uitgevoerd door studenten wiskunde die zijn getraind in het scoren van de hier onderscheiden observatiecategoriën. Per docent is vervolgens het percentage lestijd berekend waarin sprake was van probleemgericht onderwijs en van het gebruik van grafieken en tabellen. Per les is steeds een korte descriptie gemaakt van de uitleg die de docent op het bord heeft gegeven en, bij de docenten van de experimentele groep, van de wijze waarop de GRM wordt ingezet in de les. Na afloop van de tweede lesobservatie is een gestructureerd interview gehouden over knelpunten in het lesmateriaal, organisatorische problemen en de tevredenheid over het GRM-gebruik.

Bij alle leerlingen zijn een voor- en eindtoets afgenomen. De voortoets bestaat uit zes opgaven over de stof van het voorafgaande leerjaar (vwo 3). De eindtoets bestaat uit tien opgaven met in totaal negentien vragen. Bij dertien van de negentien vragen moet een leerling niet alleen een antwoord geven, maar ook expliciet aangeven op welke wijze het antwoord is verkregen. Voor deze dertien vragen is de oplossingswijze gescoord. De oplossingswijzen zijn ingedeeld in vier categoriën: algoritmische oplossing, grafische oplossing, gebruik van heuristieken en onbekend of geen oplossing. Voor de dertien vragen is berekend welk percentage algoritmisch, grafisch, heuristisch of niet is opgelost. Voor de negentien vragen is het percentage goed berekend. De betrouwbaarheid van de voortoets, uitgedrukt in Cronbach's alfa, is .52. De betrouwbaarheid van de natoets was .81.

Analyse

Door middel van variantie-analyse is nagegaan of de drie onderzoekscondities verschillen in het (gemiddelde) percentage lestijd dat docenten hebben besteed aan probleemgericht onderwijs en aan het gebruik van grafieken en tabellen. Omdat het beperkte aantal docenten de kans op het waarnemen van een significant verschil (de statistische power) bemoeilijkt, wordt een significantieniveau van .10 gehanteerd.

Of de manier waarop leerlingen opgaven oplossen in de drie condities verschilt, is onderzocht door middel van variantie-analyses. De afhankelijke variabelen in deze analyses zijn het percentage opgaven dat algoritmisch, grafisch of heuristisch is opgelost. De onafhankelijke variabelen zijn de geneste factor klas en twee conditiecontrasten. Het eerste conditiecontrast had tot doel verschillen tussen controleconditie en beide experimentele condities te toetsen. Het tweede conditiecontrast toetste verschillen tussen beide experimentele condities.

Tot slot is door middel van een multiniveau-analyse is onderzocht of de onderzoekscondities verschillen in het prestatieniveau van leerlingen op de eindtoets, rekening houdend met verschillen in relevante klas- en leerlingvariabelen. Daarbij is ook onderzocht of er sprake is van verschillen tussen klassen in de relatie tussen voor- en eindtoets, en of deze verschillen het gevolg zijn van een wisselwerking (interactie) tussen de conditie en de prestaties op de voortoets. De multiniveau-analyse is uitgevoerd met het programma VARCL. In een drietal modellen is de bijdrage onderzocht van de volgende relevante factoren aan verschillen tussen leerlingen in prestaties op de natoets. In model 1 zijn drie controlefactoren gespecificeerd. Op leerlingniveau zijn dat de factoren voortoets en de sekse. Op klasniveau is dat het gemiddelde aantal exacte vakken (wiskunde A of B, schei- en natuurkunde) dat de leerlingen van een klas gekozen hebben. Het gemiddelde aantal exacte vakken is opgenomen als fixed factor om te controleren voor verschillen in niveau en belangstelling tussen de klassen. Voor de voortoets en de factor sekse is in model 1 behalve een 'fixed' effectcoëfficiënt ook een 'random' effectcoëfficiënt gespecificeerd. De 'fixed effect' coëfficiënten van de voortoets en de factor sekse geven aan wat het gemiddelde effect van deze factoren is in de diverse klassen. Door middel van de variantie in de random effectcoëfficiënten wordt vastgesteld of de klassen significant afwijken van deze gemiddelde effecten. Model 2 is een uitbreiding van model 1. In dit model zijn twee dummy-variabelen opgenomen om de verschillen van de experimentele condities met de controleconditie te evalueren. In model 3 is de interactie tussen beide

experimentele condities en de voortoets in de analyse opgenomen. Met behulp van dit model wordt bekeken of de aanwezige random variantie in de voortoets-natoets relatie toegeschreven kan worden aan interactie-effecten tussen voortoets en conditie. De vergelijking voor model 3 is:

$$y_{ij} = \beta_0 + u_j + \beta_1 x_{1ij} + \beta_2 x_{2ij} + v_{1j} x_{1ij} + v_{2j} x_{2ij} + \gamma_1 z_{0j} + \gamma_2 z_{2j} + \beta_3 z_{1j} x_{1ij} + \beta_4 z_{2j} x_{1ij} + \varepsilon_{ij} \text{ met:}$$

- y_{ij} : de score van leerling i uit klas j op de natoets;
 x_{1ij} : de score van leerling i uit klas j op de voortoets;
 x_{2ij} : de sekse van leerling i uit klas j (0 = jongen, 1 = meisje);
 z_{0j} : de controleconditie (wel = 1; niet = 0);
 z_{1j} : de eerste experimentele conditie (wel = 1; niet = 0);
 z_{2j} : de tweede experimentele conditie (wel = 1; niet = 0);

de fixed regressiecoëfficiënten zijn:

$$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2;$$

de varianties en covarianties in het model zijn:

$$\sigma_{\varepsilon}^2 = \text{var}(\varepsilon_{ij}) \quad \text{de variantie in leerlingprestaties op de natoets}$$

$$\sigma_u^2 = \text{var}(u_j) \quad \text{de variantie in intercepten}$$

$$\sigma_{v1}^2 = \text{var}(v_{1j}) \quad \text{de variantie in hellingcoëfficiënt } \beta_1$$

$$\sigma_{v2}^2 = \text{var}(v_{2j}) \quad \text{de variantie in hellingcoëfficiënt } \beta_2$$

$$\sigma_{uv1} = \text{cov}(u_j, v_{1j}) \quad \text{de covariantie tussen } u_j \text{ en } v_{1j}$$

$$\sigma_{uv2} = \text{cov}(u_j, v_{2j}) \quad \text{de covariantie tussen } u_j \text{ en } v_{2j}$$

De plausibiliteit van model 1 is getoetst aan de hand van het verschil in 'deviance'-waarde van dit model met een zogenoemd 'leeg model', met als vergelijking $y_{ij} = \beta_0 + u_j$. Analyse van het lege model is voornamelijk zinvol om te bekijken of de variantie in intercepten tussen scholen significant is. Het verschil in 'deviance' tussen beide modellen is chikwadraat verdeeld. De plausibiliteit van de uitbreiding met de genoemde factoren in de modellen 2 en 3 is vervolgens getoetst aan de hand van het verschil in 'deviance'-waarde tussen het minder factoren omvattende model.

ONDERZOEKSRESULTATEN

De docenten

Er is nagegaan of de docenten in beide experimentele condities meer probleemgericht onderwijs geven en bij hun uitleg vaker grafieken en tabellen gebruiken dan docenten in de controleconditie. Dit kan erop wijzen dat hun lesgedrag is aangepast aan de bedoelingen van het nieuwe curriculum. In tabel 2 zijn de resultaten weergegeven.

De docenten uit de experimentele condities lijken gemiddeld iets meer lestijd besteed te hebben aan probleemgericht onderwijs. Het verschil tussen de experimentele condities en de controleconditie is echter statistisch niet significant. Er is derhalve geen steun voor de verwachting dat de docenten in de experimentele condities meer probleemgericht onderwijs bieden dan docenten in de controleconditie. Uit de observaties komt wel naar voren dat de condities verschillen in het gebruik van grafische middelen ($F(2,9)=3.62$; $p=.07$). In de controleconditie

tonen de docenten zelden of nooit hoe een grafiek van een functie eruit ziet. Slechts één van de vier docenten in de controle conditie heeft enige tijd besteed aan het tekenen van een grafiek op het bord.

Tabel 2: Percentage van de tijd die de docent in de les besteedt aan probleemgericht onderwijs en aan het gebruik van de GRM en grafieken of tabellen. Het aantal docenten is tussen haakjes vermeld

condities	Probleemgericht onderwijs	Grm of tabellen/grafieken op het bord
experimenteel 1 (n = 3)	28 %	24 %
experimenteel 2 (n = 5)	30 %	18 %
controle (n = 4)	19 %	1 %

In de experimentele versie van Wiskundelij moet een kwart tot de helft van de opgaven met behulp van de GRM worden opgelost. In de experimentele condities hebben alle docenten enige tijd besteed aan uitleg met de GRM of aan grafieken op het bord. De docenten van beide experimentele groepen gebruiken de GRM of grafieken op het bord gemiddeld 18-24% van de lestijd. Sommige docenten uit de experimentele condities tekenen vaker grafieken op het bord dan dat zij in hun uitleg gebruik maken van de GRM. In de interviews en gesprekken geven de leraren van de experimentele klassen zelf aan dat zij anders lesgeven en meer grafisch werken dan voordien.

Tijdens de observaties is nagegaan waarvoor de GRM door de leerlingen veelal werd gebruikt. De GRM werd vooral ingezet voor het exploreren van de relatie tussen functies en hun grafieken en voor het controleren van uitwerkingen. De docenten van de experimentele condities volgen het boek. De grafische calculator wordt door hen alleen toegepast bij opgaven die daarvoor in het boek zijn aangemerkt. Wat betreft de problemen bij de implementatie van het nieuwe curriculum is het volgende vastgesteld. Zoals we al eerder zagen, krijgen de leerlingen in de experimentele groepen niet meer gelegenheid om problemen op te lossen en te bespreken dan de leerlingen uit de controlegroep. De docenten geven zelf aan dat het erg lastig is om problemen aan de orde te stellen en te bespreken binnen de beschikbare lestijd. Een algemene klacht van de docenten is de overladenheid van het 4 VWO-programma. Docenten vinden dat ze veel van de stof uit de wiskundemethode niet kunnen behandelen in twee of drie uren wiskundeles per week. In de meeste gevallen ontbrak een overhead projector om handelingen op de GRM aan alle leerlingen te tonen. Verder zijn knelpunten geconstateerd die te maken hadden met de duur van het experiment. Docenten van leerlingen die slechts één hoofdstuk met de GRM werken (tweede experimentele conditie), constateren dat veel leerlingen de GRM niet optimaal gebruiken voor exploratie van een opgave of voor controle van een antwoord. Dat komt omdat de leerlingen weinig tijd hebben gehad (hooguit drie weken) om goed te leren omgaan met de GRM. Docenten van Experimenteel 1 zijn hierover positiever. In hun klassen is de GRM het hele jaar gebruikt, zowel voor het exploreren van grafieken en het oplossen van opgaven als voor de controle van uitwerkingen. Ook gebruiken sommige leerlingen het programmeerdeelte van de GRM, voor onder meer de abc-formules bij kwadratische functies.

De docenten uit beide experimentele groepen zijn tevreden over het gebruikte lesmateriaal. Alle docenten geven aan dat er voldoende goede opgaven over het exploreren van nieuwe leerstof zijn opgenomen in de methode. Ook voor het controleren van uitkomsten met de GRM zijn voldoende opgaven beschikbaar volgens de docenten.

De docenten ervaren wel problemen met de prijs van de machine. Enkele docenten geven

aan dat ze de GRM niet willen aanschaffen omdat de kosten van dit apparaat te hoog zijn. Ook is nog niet duidelijk welk merk en model de leerlingen moeten aanschaffen. Er komen steeds betere apparaten op de markt met meer gebruiksmogelijkheden.

De leerlingen

De gebruikte oplossingswijzen

De oplossingswijzen zijn ingedeeld in vier categorieën. De eerste categorie betreft algoritmische oplossingen. Hieronder verstaan we al de oplossingen die een vaste procedure vereisen zoals ongelijkheid oplossen, een punt op een grafiek berekenen, afgeleide bepalen enz. De tweede categorie is de grafische oplossingswijze. De leerlingen kunnen een grafiek tekenen of deze plotten op de GRM en vervolgens de oplossing benaderen. De derde categorie is heuristische oplossingswijzen: het gebruik van een tabel of schema als hulpmiddel. De laatste categorie is 'geen oplossing', wanneer een opgave niet is gemaakt. Voor de dertien vragen is berekend welk percentage algoritmisch, grafisch of heuristisch is opgelost.

De verwachting was dat leerlingen uit de experimentele condities vaker een grafiek tekenen of plotten op de GRM (grafische oplossingswijze) dan leerlingen uit de controleconditie.

Tabel 3: Aantal keren dat op de natoets verschillende oplossingswijzen zijn gebruikt (gemiddeld percentage leerlingen per onderzoeksconditie)

condities	algoritmisch	grafisch	heuristisch	geen oplossing	totaal
experimenteel 1	41%	35%	10%	14%	100%
experimenteel 2	31%	37%	5%	27%	100%
controle	42%	11%	8%	39%	100%

In tabel 3 staat het overzicht van de gebruikte oplossingsmethoden gemiddeld over alle opgaven van de eindtoets. Wat opvalt is dat leerlingen uit beide experimentele condities duidelijk meer gebruik maken van grafische methoden dan de leerlingen uit de controlegroep. De leerlingen uit beide experimentele condities maken in ongeveer 35% van de gevallen gebruik van grafische oplossingswijzen tegen 11% van de gevallen in de controleconditie. Het verschil tussen de experimentele groepen en de controlegroep is significant ($F(1,9) = 6.2$; $p = .03$). Een tweede opvallend gegeven is dat de leerlingen uit de eerste experimentele conditie en de controlegroep in ongeveer 41% van de opgaven een algoritmische oplossingswijze gebruikt hebben. In de tweede experimentele conditie is het gebruik iets lager (31%). Het verschil tussen beide experimentele condities is echter niet significant ($F(1,9) = 1.71$; $p = .22$). Opvallend is dat leerlingen in beide experimentele condities vaker aan een oplossing zijn begonnen dan leerlingen van de controlegroep. Dit is af te lezen onder de kolom 'Geen oplossing'. In de experimentele groepen is in gemiddeld 14% en 27% van de gevallen geen oplossingspoging geconstateerd, terwijl dit percentage in de controlegroep 39% bedraagt.

Dat er in de controlegroep vaker geen oplossingspoging is geconstateerd heeft niet te maken met de tijd voor het maken van de toets. De tijd was in alle condities gelijk (twee uur) en moet ruim voldoende moet zijn voor het maken van de tien opgaven. Vermoedelijk kenden de leerlingen in de controleconditie bij bepaalde opgaven geen algoritmische aanpak en zagen ze geen kans deze opgaven heuristisch of grafisch op te lossen. Uit een aanvullende statistische analyse (zie Harskamp, Van Streun en Suhre, 1996, p. 116-120), waarin rekening is gehouden met de prestatieverschillen op de voortoets, blijkt dat leerlingen uit de eerste experimentele conditie significant meer gebruik maken van grafische methoden dan de leerlingen uit de controlegroep, zonder dat het aantal algoritmisch opgeloste opgaven verschilt. De hogere

toetsresultaten in deze conditie zijn te verklaren doordat de leerlingen bij meer opgaven een oplossingsmethode kunnen formuleren. Bij leerlingen die slechts één hoofdstuk met de grafische oplossingsmethode hebben gewerkt (de tweede experimentele conditie), zijn grafische oplossingen weliswaar vrij vaak gebruikt, maar is het gebruik van de algoritmische en heuristische oplossingen minder dan in beide andere condities.

Gebruik van de GRM door jongens en meisjes

Bij de 13 vragen waarbij naar de oplossingswijze is gekeken, is ook vastgesteld of de GRM is gebruikt. Er is in beide experimentele condities enig verschil tussen jongens en meisjes. Jongens hebben bij 5.6 opgaven de GRM gebruikt. De meisjes hebben gemiddeld bij 4.8 opgaven de GRM gebruikt. Het verschil is statistisch significant ($F(1,191) = 5.05$; $p = .03$). De meisjes en jongens in het onderzoek verschillen wat hun gebruik van de GRM betreft, niet van het gedrag van leerlingen uit andere onderzoeken. Bekend is dat meisjes minder vaak gebruik maken van computers dan jongens (Whitley, 1997). We zullen nog nagaan in hoeverre de factor sekse ook samenhangt met de leerprestaties.

De leerresultaten

Behalen leerlingen in de experimentele condities betere resultaten bij het oplossen van wiskunde problemen dan leerlingen in de controleconditie? Voor we op deze vraag ingaan, kijken we eerst naar relevante verschillen in de samenstelling van de onderzoeksgroep ten aanzien van de voortoets en het aantal exacte vakken (zie tabel 4). De relevantie van de factoren in tabel 4 wordt ondersteund door de correlaties met de natoets. De correlatie tussen voortoets en natoets is .50 (leerlingniveau). De correlatie van de factor aantal exacte vakken met de natoets op klasniveau is .38.

De beschrijvende gegevens wijzen uit dat alle leerlingen zeer laag scoorden op de voortoets. Alle leerlingen zijn kennelijk veel van de leerstof van leerjaar drie vlak na de zomervakantie vergeten. Het was echter in het kader van het onderzoek niet raadzaam te wachten tot leraren de kennis hadden opgefrist, omdat het experiment direct na de vakantie is gestart. Hoewel de leerlingen uit de twee experimentele condities hoger scoorden dan die uit de controlegroep, blijkt bij toetsing met klas als geneste factor dat de verschillen tussen de drie condities niet significant zijn ($F(2,9) = 2.26$; $p = .16$). Op de natoets verschillen de drie condities wel significant ($F(2,9) = 3.33$; $p = .08$). De leerlingen uit de twee experimentele condities scoorden hoger op de natoets dan de leerlingen uit de controlegroep. De samenstelling van de klassen in de drie onderzoeksgroepen verschilt evenmin. Voor de bepaling van de klassensamenstelling is gebruik gemaakt van de mate waarin leerlingen bij de overgang naar VWO-5 exacte vakken (wiskunde A of B, schei- en natuurkunde) kiezen. De verwachting is dat naast voortoetsresultaten van leerlingen ook de interesse van de klas voor exacte vakken van invloed zal zijn op de interesse van leerlingen voor computers en op de wiskunde prestaties. Tabel 4 laat de gegevens zien.

De vraag is nu of er sprake is van conditiesverschillen, wanneer rekening wordt gehouden met voortoetsverschillen en verschillen tussen klassen in voorkeur voor exacte vakken. Die vraag is met het programma VARCL in een aantal verklarende variantiemodellen onderzocht. De resultaten daarvan zijn samengevat in tabel 5. Uit de analyse van het lege model blijkt dat de schoolspecifieke intercepten significant van elkaar verschillen. We zullen in het vervolg van de bespreking zien of de te onderzoeken factoren daarop van invloed zijn. In model 1 zijn de fixed effects parameters voor drie controlefactoren opgenomen en de random effect parameters voor de variantie van de regressie van de eindtoets op de voortoets en van de factor sekse. Vergelijking met het lege model leert dat de deviantie significant afneemt. De voortoets is een belangrijke factor van betekenis. De (helling)variantie van de voortoets en de factor sekse zijn echter ook significant. Dit betekent dat er verschillen zijn tussen klassen in de relatie tussen voor- en natoets en in de verschillen tussen jongens en meisjes. Model 2 omvat

Tabel 4: Beschrijvende gegevens van de leerlingen in de drie onderzoekscondities; gemiddelden en standaarddeviaties op het aangeduide niveau (tussen haakjes)

	Experimenteel 1 (hele jaar GRM)	Experimenteel 2 (hoofdstuk GRM)	Controlegroep (geen GRM)
<i>leerlingniveau</i> voortoets (score 0 - 6) eindoets (score 0 - 19)	2.21 (1.47) 10.18 (2.68)	1.33 (1.35) 7.88 (3.78)	1.04 (1.07) 6.81 (3.84)
<i>klasniveau</i> gemiddeld aantal exacte vakken (score 0-3)	1.98 (1.00)	1.96 (0.20)	1.60 (0.64)

Tabel 5: Regressie van de eindtoetscore op de leerling- en klasmerken; coëfficiënten en - tussen haakjes - de standaard-schattingfouten

	Leeg model	Model 1	Model 2	Model 3
Grand Mean/Intercept	8.07	5.37	10.40	9.70
Voortoets		1.15 (0.24)	1.08 (0.23)	1.70 (0.30)
Sekse		-0.51 (0.43)	-0.44 (0.42)	-0.43 (0.39)
Aantal bètavakken		0.74 (0.97)	-0.65 (0.87)	-0.40 (0.88)
Controleconditie			-4.43 (1.29)	-3.98 (1.30)
Experimentele conditie 1			-2.29 (1.20)	-2.04 (1.22)
Interactie voortoets x Experimentele conditie 1				-1.09 (0.41)
Interactie voortoets x Experimentele conditie 2				-.69 (0.39)
variantie leerlingniveau	9.23	6.93	6.94	7.06
variantie klasniveau	5.00 (sign.)	3.46 (sign.)	3.93 (sign.)	2.82 (sign.)
random hellingvariantie voortoets	----	0.42	0.34	0.06 (n.s.)
random hellingvariantie sekse		0.91	0.82	0.58 (n.s.)
Deviance	1514.52	1446.91	1442.27	1437.93
Vershil in deviance met vorige model		67.61	4.64	4.34
Vershil aantal vrijheidsgraden		7	2	2
P-waarde		.00	.098	.11

tevens de parameterschattingen van de twee conditie-effecten. Uit de afname van de deviantie valt af te leiden dat er sprake is van een conditie-effect. De verschillen in de relatie tussen voor- en natoets en in de verschillen tussen jongens en meisjes zijn ook in model 2 significant. Gegeven het feit dat zich verschillen voordoen tussen klassen in de relatie tussen voor- en natoets is nu de vraag aan de orde of het conditie-effect voor goede en zwakke leerlingen even sterk is. Dit is nagegaan door opname van de interactie-effecten van beide condities met de voortoets (model 3).

Uit tabel 5 blijkt dat de interactie-effecten leiden tot een verdere daling van de deviantie. Het verschil met model 2 is $p = .11$. Merk op dat in model 3 de hellingvariantie voor de voortoets terug valt naar .06. Een belangrijk gevolg van de opname van de interactie-effecten is ook dat de random effects variantie voor de factor sekse niet significant meer is in model 3. We kunnen ervan uitgaan dat de interactie-effecten tussen beide experimentele condities en de voortoets de voornaamste verklaring zijn voor de in de eerste twee modellen aanwezige random hellingvariantie. Hoewel de variantie in schoolspecifieke intercepten is afgenomen is deze in model 3 nog wel significant.

Wat valt er af te leiden uit de multiniveau-analyse? De regressiecoëfficiënten van model 3 in tabel 5 laten zien dat er sprake is van een positief verschil tussen beide experimentele condities en de controleconditie. Dit positieve verschil beperkt zich echter tot de zwakke tot middelmatige leerlingen. Dat blijkt uit het feit dat de coëfficiënten betreffende het interactie-effect van de experimentele condities met de voortoets negatief zijn. Hoofd- en interactie-effecten tezamen wijzen erop dat alleen leerlingen met een lage score op de voortoets beter presteren. Naarmate leerlingen hogere scores hebben op de voortoets wordt het verschil tussen de experimentele condities en de controleconditie kleiner. De positieve effecten voor zwakke en middelmatige leerlingen zijn in de eerste experimentele conditie iets groter dan in de tweede experimentele conditie. De bevinding dat zwakke leerlingen relatief beter scoren komt overeen met buitenlandse onderzoeksbevindingen (Shoaf-Grubs, 1994; Adams, 1994).

CONCLUSIES EN DISCUSSIES

Vanaf 1998 wordt de grafische calculator ingevoerd in de bovenbouw van HAVO en VWO. Er zullen nieuwe leerboeken worden ontwikkeld om het gebruik van de grafische calculator te integreren in de wiskundeleerstof. In dit onderzoek is onderzocht welke veranderingen zijn te verwachten voor de inrichting van het onderwijs en voor de leerprestaties. Om een reëel beeld te krijgen van de te verwachten veranderingen hebben de docenten van de experimentele groepen bewust geen training ontvangen. Ze zijn slechts één keer bij elkaar geweest om met elkaar de ervaringen uit te wisselen. Zij kregen het lesmateriaal toegestuurd en zonodig grafische calculators. De meeste docenten van de experimentele groepen in het onderzoek hadden vooraf weinig of geen ervaring in het gebruik van de grafische calculator. De leerlingen evenmin, maar zij blijken geen moeite te hebben om het elementaire gebruik onder de knie te krijgen, mits zij zelf het hele jaar over zo'n machine kunnen beschikken. Incidenteel gebruik op school of alleen bij een enkel onderwerp leidt niet tot de vereiste vaardigheid die nodig is bij het oplossen van wiskundige problemen.

De verwachtingen bij de eerste onderzoeksvraag waren dat leraren met de GRM meer probleemgericht onderwijs geven en meer uitleg zullen geven met behulp van de GRM of een grafiek op het bord. Uit de observaties in de lessen, blijkt dat de docenten uit de experimentele klassen niet meer probleemgericht onderwijs geven maar wel meer aandacht besteden aan de grafische voorstellingen van functies. Omdat een overheadprojector met een vierer van het grafische calculator ontbrak, werden de grafieken op het bord getekend. In de venster van de grafische calculator ontbrak, werden de grafieken op het bord getekend. In de experimentele groepen is meer gebruik gemaakt van uitleg aan de hand van tabellen en grafieken dan in de controlegroep. De docenten kunnen blijkbaar zonder arbeidsintensieve voorschooling dan in de controlegroep. De docenten kunnen blijkbaar zonder arbeidsintensieve voorschooling redelijk uit de voeten met het nieuwe curriculum, mits er kwalitatief goed lesmateriaal beschikbaar is. Wel is duidelijk geworden dat een aangepaste didactiek van probleemgericht onderwijs niet vanzelf tot stand komt. Het is aan te bevelen om bij de invoering van de GRM in het wiskundecurriculum docenten vooraf na te scholen in de selectie van leerstof uit hun methoden en in het interactief lesgeven en bespreken van oplossingsmethoden van leerlingen.

De verwachting bij de tweede onderzoeksvraag was dat meisjes minder vaak de GRM zullen gebruiken dan jongens. Deze verwachting wordt ondersteund als we kijken naar het gebruik van de GRM op de eindtoets. Meisjes gebruiken de GRM iets minder vaak dan jongens (4,8 tegen 5,6 opgaven van in totaal 13 opgaven), maar de multiniveau-analyse wijst uit dat hun resultaten door de bank genomen niet slechter zijn dan die van jongens. Deze onderzoeksbevindingen komen niet geheel overeen met die van Ruthven (1990). Deze constateerde dat meisjes in de groep die de calculator gebruikte, betere prestaties behaalden dan de jongens. Wellicht gebruikten de meisjes de calculator in het onderzoek van Ruthven frequenter dan in ons onderzoek het geval is. De leerlingen in Ruthvens onderzoek gebruikten de calculator al voorafgaande aan het experiment. Het onderzoek van Vazquez (1991) bevestigt onze resultaten: ze vond geen verschil in prestaties tussen jongens en meisjes in de experimentele con-

ditie die met de GRM had gewerkt, maar er was wel een effect voor de experimentele groep als geheel door het gebruik van de calculator.

De verwachting ten aanzien van de *derde onderzoeksvraag* was dat in experimentele groepen vaker een grafische oplossingswijze wordt gevolgd door de leerlingen en dat vooral de zwakkere leerlingen profijt zullen hebben van de calculator. Uit dit onderzoek blijkt dat het gebruik van de grafische calculator gedurende een geheel schooljaar leidt tot een uitbreiding van het repertoire aan oplossingswijzen. Leerlingen gebruiken vaker een grafische oplossingswijze en hanteren daarnaast een algoritmische oplossingswijze. Leerlingen die incidenteel met de GRM werken, gebruiken vooral een grafische oplossingswijze. De algoritmische oplossingswijze wordt enigszins verdrongen. De verklaring hiervoor kan zijn dat leerlingen meestal wel goed uit de voeten kunnen met een algoritmische of algebraïsche oplossingswijze, terwijl de zwakke leerlingen dit minder goed kunnen en nu terug vallen op de grafische oplossingswijze. In de controlegroep zonder grafische calculator zien we dan ook dat zwakke leerlingen opgaven algoritmisch oplossen of er niet aan beginnen.

Onze conclusie is dat de implementatie van de grafische calculator goed uitvoerbaar is. Bij geschikt lesmateriaal kan het gebruik leiden tot een uitbreiding van het repertoire aan oplossingsmethoden van de leerlingen. Vooral minder algebraïsch vaardige leerlingen zullen daar baat bij hebben.

Een vraag voor vervolgonderzoek is of, als gevolg van het GRM-gebruik, leerlingen een betere koppeling tot stand kunnen brengen tussen de functievoorschriften aan de ene kant en de bijpassende grafische voorstellingen anderzijds. Dit zou namelijk een verklaring zijn voor de goede prestaties van leerlingen uit de eerste experimentele conditie. In deze groep worden zowel met algoritmische als met grafische oplossingswijzen goede resultaten behaald. Ook uit andere onderzoeken komt deze verklaring voor de extra leerwinst bij GRM-gebruik naar voren (zie Penglase & Arnold, 1966). Het is interessant om na te gaan of vooral meisjes zullen profiteren van de GRM als ze deze frequent gaan gebruiken.

LITERATUUR

- Adams, T. (1994) Graphing Function Problems in Teaching Algebra. *Florida Educational Research Council, Research Bulletin*, 25, p. 122-143.
- Army, P.D. (1992). *The concept of function: Misconceptions and remediation at the collegiate level*. (Doctoral Dissertation, Illinois State University, 1991). Dissertation Abstracts International, 52/08, 2850.
- Harskamp, E.G., Streun, A van., Suhre, C.J.M (1996). *Handwerk en technologie in wiskunde Tweede Fase Voortgezet Onderwijs*. Groningen: GION.
- Heid, M.K. (1988). *Resequencing skills and concepts in applied calculus using the computer as a tool*. Journal for research in mathematics education, 19,(1), 3-25.
- Maxwell, M.E. en Quesada, A.R. (1994). The effects of using graphing calculators to enhance college students' performance in precalculus. *Educational Studies in Mathematics*, 27, p.205-215.
- Penglase, M & Arnold, S. (1996) The Graphics Calculator In Mathematics Education: A critical Review of Recent Research. *Mathematics Education Research Journal*, 8, p. 58-91.
- Rich, B.S. (1991). *The effect of the use of graphing calculators on the learning of function concepts in pre-calculus mathematics*. (Doctoral Dissertation, University of Iowa, 1990). Dissertation Abstracts International, 52/03, 835.
- Ruthven, K. (1990). The influence of graphic calculator use on translation from graphic to symbolic forms. *Educational Studies in Mathematics*, 21, p.431-450.
- Shoaf-Grubbs, M.M. (1994). Research Results on the effect of the graphic calculator on female students' cognitive levels and visual thinking. L.Burton & B.Jaworski (Eds). *Technology in Mathematics Teaching - a bridge between teaching and learning*. p.213-230. Londen: Lawrence Erlbaum Associates.
- Streun, van A. (1996). Zelfstandig leren met een grafiekenprogramma bij de hand. *Euclides*, 72, p.25-28.
- Tuska, A. (1993). *Students' errors in graphing calculator-based precalculus-classes*. Dissertation Abstracts International, 53/08, 2725

- Vasquez, J.L. (1991). The effect of the Calculator on Student Achievement in graphing linear functions (Doctoral dissertation, University of Florida). *Dissertation Abstracts International*, 51/11,3660
- Whitley, B. E. (1997). Gender Differences in ComputerRelated Attitudes and Behavior: A MetaAnalysis. *Computers in Human Behavior*, 13, 1, 122.

Doeloriëntaties, Waargenomen Eigen Competentie en Studieresultaten

Anneke Vrugt, Johan Hoogstraten & Frans J. Oort
Faculteit der Maatschappij- en Gedragwetenschappen, Afdeling Psychologie, Universiteit van Amsterdam, Roetersstraat 15, 1018 WB Amsterdam

SAMENVATTING

De centrale vraagstelling van dit onderzoek is op welke wijze taak- en ego-oriëntatie op het gebied van de studie bijdragen aan de studiemotivatie en prestaties. In verband daarmee is allereerst nagegaan of de structuur van de doeloriëntatie-items in ons onderzoek overeenkomt met die bij Duda en Nicholls (1992). Uit verschillende analyses blijkt dat de structuur van de doeloriëntatie-items het beste beschreven wordt door vijf componenten, waarbij aan de vier door Duda en Nicholls onderscheiden componenten (Ego-oriëntatie, Werkvermijding, Taakoriëntatie en Coöperatie), een component is toegevoegd die wij Hard werken hebben genoemd. Drie van de items die bij Duda en Nicholls hoog laden op de component Taakoriëntatie, laden in ons onderzoek hoog op de vijfde component Hard werken. In overeenstemming met de verwachtingen blijken taakoriëntatie en coöperatie bij te dragen aan de waargenomen eigen competentie ("perceived self-efficacy"). In tegenstelling tot de verwachtingen blijken de ego-oriëntatie en werkvermijding geen bijdrage te leveren aan de sociale vergelijking. In aansluiting bij de resultaten van eerder onderzoek (Vrugt, 1994) is ook gebleken dat de waargenomen eigen competentie en sociale vergelijking bijdragen aan de gevoelens over de eigen studievaardigheden en via die gevoelens aan de studieresultaten. In overeenstemming met de verwachtingen correleren taakoriëntatie, coöperatie en hard werken met de attributie van falen aan gebrek aan inzet en strategie-ontwikkeling.

ABSTRACT

This study deals with the way in which task and ego orientation in students' study contribute to their study motivation and achievements. In connection with this we first examined whether the structure of the goal orientation items in our study corresponds with the structure of the goal orientation items in Duda and Nicholls (1992). The results of several analyses showed that the structure of the goal orientation items is best described by five components. To the four components distinguished by Duda and Nicholls (Ego orientation, Work avoidance, Task orientation and Cooperation) we added a fifth component, namely Working hard. Three items which had high loadings on the component Taakorientation in Duda and Nicholls had high loadings on Working hard. Consistent with expectation task orientation and cooperation contributed to perceived self-efficacy. Contrary to expectation, ego orientation and work avoidance did not contribute to social comparison. Consistent with earlier research (Vrugt, 1994) it was also found that perceived self-efficacy and social comparison contributed to students' feelings about their study capabilities and via these feelings to their study results. In agreement with expectation, task orientation, cooperation, and working hard correlated with the attribution of failure to lack of application and strategy development.

Mensen houden er verschillende theorieën op na over hun functioneren. Bijvoorbeeld theorieën over de determinanten van de eigen prestaties. Dergelijke theorieën kunnen van invloed zijn op iemands motivatie en de activiteiten die hij of zij in een prestatiegerichte context, zoals studie of werk, ontplooit. In deze theorieën over de determinanten van de eigen prestaties zijn twee gezichtspunten te onderscheiden, die met verschillende termen worden aangeduid. Bandura (1989) heeft het over de opvattingen die mensen erop nahouden over intelligentie, als vaardigheid die ontwikkeld kan worden, of als stabiele entiteit die moeilijk te veranderen is. Elliot en Dweck (1988) onderscheiden twee doelen die individuen nastreven in prestatiegerichte situaties: leerdoelen en uitkomstdoelen. Als individuen er leerdoelen op nahouden streven zij ernaar de eigen vaardigheden te vergroten en nieuwe taken te beheersen. Individuen met uitkomstdoelen proberen een positief oordeel over de eigen capaciteiten te handhaven en negatieve evaluaties te vermijden, omdat uit die negatieve evaluaties kan blijken dat hun capaciteiten onvoldoende zijn. Duda en Nicholls (1992) spreken over doeloriëntaties en maken daarbij onderscheid tussen taak- en ego-oriëntatie. Ook al verschillen de termen die gebruikt worden, de begripsomschrijvingen komen duidelijk overeen. Omdat ons onderzoek aansluit bij dat van Duda en Nicholls (1992) zullen wij eveneens gebruik maken van de termen taak- en ego-oriëntatie.

Individuen die er een taakoriëntatie op nahouden gaan ervan uit dat zij goede prestaties kunnen behalen door het verwerven van kennis en inzicht, door hard werken en het begrijpen van de stof. Zij realiseren zich dat samenwerking met anderen en hulp van anderen ook kan bijdragen aan het tot stand komen van een goede taakuitvoering. Individuen met een ego-oriëntatie achten zichzelf succesvol als zij anderen verslaan en hun superioriteit kunnen demonstreren. Zij beschouwen veel inzet als indicatie van geringe capaciteiten. Capaciteiten blijken, vanuit dit gezichtspunt, vooral als iemand zich weet te onderscheiden zonder daarvoor hard te werken.

In een onderzoek bij scholieren hebben Duda en Nicholls vragen gesteld over doeloriëntaties op het gebied van de studie. Factoranalyse heeft uitgewezen dat vier factoren onderscheiden konden worden, namelijk de factoren Taakoriëntatie en Samenwerking die de taakoriëntatie representeren, en de factoren Ego-oriëntatie en Werkvermijding die de ego-oriëntatie representeren.

De centrale vraagstelling van dit onderzoek is op welke wijze taak- en ego-oriëntatie op het gebied van de studie bijdragen aan de studiemotivatie en prestaties. In verband daarmee is allereerst nagegaan of de structuur van de doeloriëntatie-items in ons onderzoek overeenkomt met die bij Duda en Nicholls (1992).

De doeloriëntaties op het gebied van studie hebben betrekking op globale opvattingen over de wijze waarop de eigen goede prestaties tot stand komen. Deze opvattingen zijn niet specifiek toegesned op de taak die men op een bepaald moment uitvoert. Daarom lijkt het aannemelijk dat de doeloriëntaties geen directe invloed uitoefenen op een bepaalde taakprestatie, maar de motivatie beïnvloeden doordat zij inwerken op variabelen die specifiek verband houden met een bepaalde taakprestatie en dus indirect de taakprestatie stimuleren.

Volgens Bandura (1986, 1989, 1997) spelen de opvattingen die iemand heeft over zijn of haar competentie op een bepaald terrein ("perceived self-efficacy") en de sociale vergelijking een belangrijke rol bij het tot stand komen van goede prestaties. Bandura (1986) omschrijft "perceived self-efficacy" als volgt: "... People's judgement of their capabilities to organize and execute courses of action required to attain designated types of performances" (p. 395). Uit onderzoek van Vrugt (1994) bij eerstejaarspsychologiestudenten is gebleken dat de opvattingen over de eigen competentie op het gebied van de studie en de vergelijking van de eigen studievaardigheden met die van een medestudent(e) een bijdrage leverden aan de gevoelens over de eigen vaardigheden. De gevoelens over de eigen studievaardigheden bleken op hun beurt bij te dragen aan de studieresultaten die behaald werden. Om na te gaan of de doeloriëntaties bijdragen aan de waargenomen eigen competentie en aan de sociale vergelijking is het onderzoek van Vrugt (1994) gerepliceerd.

Hofmann en Stricklund (1995, p. 497) bespreken een idee van Nicholls (1984) dat personen met een ego-oriëntatie en hoge beoordeelde competentie even goed presteren als personen met een taakoriëntatie. Volgens ons is dit idee vooral van toepassing op gevorderden op een bepaald gebied, maar niet op beginners. Gevorderden die hun vaardigheden ontwikkeld hebben en zichzelf op grond daarvan competent achten, kunnen er vooral op gericht zijn zichzelf van anderen te onderscheiden. Dit laatste hoeft geen belemmering te vormen voor goede prestaties omdat zij op grond van hun competentie in staat zijn goed te presteren. Onze onderzoeksgroep bestaat uit beginners, studenten die nog maar kort psychologie studeren en ook voordien geen ervaring hebben opgedaan met dit academische vak. Bij beginners op een bepaald terrein is een taakoriëntatie onontbeerlijk omdat beginners zich de vaardigheden die nodig zijn voor het behalen van goede prestaties nog eigen moeten maken. De onderstaande gedachtegang is dan ook toegespitst op beginners.

Individen met een taakoriëntatie voelen zich succesvol als zij hard werken en daardoor nieuwe vaardigheden leren. Een dergelijke inzet zal ertoe leiden dat men de eigen vaardigheden ook daadwerkelijk ontwikkelt. Als men de eigen vaardigheden ontwikkelt zal dit ook als zodanig ervaren worden, hetgeen resulteert in een relatief hoge waargenomen eigen competentie. We verwachten dus dat de taakoriëntatie zal bijdragen aan de waargenomen eigen competentie.

Individen met een ego-oriëntatie menen dat zij succesvol zijn als zij zich onderscheiden van anderen, beter zijn dan anderen en anderen verslaan. Ego-georiënteerde personen streven er dus naar de eigen superioriteit te tonen. Dit streven zal de richting van de sociale vergelijking beïnvloeden. Men kan zichzelf op het gebied van studievaardigheden vergelijken met een betere, slechtere, of gelijke medestudent(e). Vergelijking van de eigen studievaardigheden met die van een medestudent(e) die slechter af is (neerwaartse vergelijking), geeft aanleiding tot positieve gevoelens over de eigen studievaardigheden (Vrugt, 1994). Ego-georiënteerde personen kunnen aan neerwaartse vergelijking een gevoel van superioriteit ontlenu. Verwacht wordt daarom dat de ego-oriëntatie zal bijdragen aan neerwaartse vergelijking.

Op basis van de resultaten van Vrugt (1994) wordt tevens verwacht dat de waargenomen eigen competentie en de sociale vergelijking bijdragen aan de gevoelens over de eigen studievaardigheden en dat die gevoelens op hun beurt bijdragen aan de prestaties.

Bandura (1989) en Elliot en Dweck (1988) hebben gesuggereerd dat personen die leerdoelen nastreven anders reageren op falen dan personen die uitkomstdoelen nastreven. Personen die uitkomstdoelen nastreven houden zich bezig met de vraag: Zijn mijn capaciteiten voldoende? Falen wijst er voor hen op dat de eigen capaciteiten niet voldoende zijn. Omdat personen die uitkomstdoelen nastreven menen dat capaciteiten een statisch gegeven vormen dat niet eenvoudig te veranderen valt, zal falen aanleiding geven tot attributies aan gebrek aan talent en gevoelens van hulpeloosheid. Personen die leerdoelen nastreven houden zich vooral bezig met de vraag: Hoe kan ik mijn vaardigheden ontwikkelen? Falen verschaft hen informatie over de taakuitvoering, over de noodzaak om andere strategieën te ontwikkelen en de inzet te vergroten. Op grond van het voorafgaande wordt verwacht dat de taakoriëntatie verband houdt met de attributie van falen aan gebrek aan inzet en strategie-ontwikkeling, en dat de ego-oriëntatie gerelateerd is aan de attributie van falen aan gebrek aan talent en hulpeloosheid.

METHODE

Aan het onderzoek is meegewerkt door 177 eerstejaarsstudenten in de psychologie die in het kader van studieverplichtingen hebben deelgenomen aan een aantal collectieve testzittingen. Tijdens een van deze zittingen zijn de doeloriëntaties gemeten met behulp van vragen die ook gebruikt zijn door Duda en Nicholls (1992). Tijdens een volgende testzitting zijn de denkbeelden over de eigen competentie gemeten met behulp van een aantal vragen uit de Academic-Self-

Efficacy-vragenlijst, ontwikkeld door Wood en Locke (1987). Tevens zijn er vragen beantwoord over de richting van de sociale vergelijking, de gevoelens over de eigen studievaardigheden en over attributies. Enkele weken na deze testzittingen vond het relevante tentamen plaats.

Doeloriëntaties

De doeloriëntaties zijn gemeten met behulp van de 21 vragen die Duda en Nicholls (1992) gebruikt hebben om doeloriëntaties ten aanzien van de studie te meten. De deelnemers werden verzocht bij zichzelf na te gaan wanneer ze zich het meest succesvol voelen in de studie psychologie. De bewering "Ik voel mezelf succesvol als:..." werd gevolgd door 21 stellingen die elk voorzien waren van vijfpuntsschalen met als extremen: "duidelijk oneens" (1) en "duidelijk eens" (5). Een voorbeeld van een stelling die taakoriëntatie representeert, is: "...ik iets interessants leer". Een voorbeeld van een stelling die betrekking heeft op ego-oriëntatie is: "...ik anderen overtroef".

Waargenomen Eigen Competentie

In het onderhavige onderzoek is gebruik gemaakt van de vragen over Self-Efficacy-Strength (SES) uit de Academic-Self-Efficacy-vragenlijst van Wood en Locke (1987). Wood en Locke onderscheiden zeven taakdomeinen: begrijpen van de stof, onderscheiden van begrippen, uitleggen van begrippen, herinneren, concentratie tijdens de les, aantekeningen maken tijdens de les en concentratie tijdens het tentamen. Voor elk taakdomein zijn vier items met de volgende intervallen gebruikt. Bijvoorbeeld: "Ik begrijp 50 procent (70, 90 of 100 procent) van de begrippen die aangeboden worden." De deelnemers werden verzocht voor elk niveau aan te geven of zij dit konden realiseren (ja of nee). De deelnemers werden eveneens verzocht om bij elk niveau aan te geven hoe zeker zij waren dat zij dit niveau konden realiseren (op een schaal van 0 tot 10). Deze laatste vragen representeren de sterkte van de eigen competentie (Self-Efficacy-Strength: SES). De SES-vragen zijn gebruikt als maat voor de waargenomen eigen competentie (zie ook Silver, Mitchell & Gist, 1994). De vragen zijn gesteld over het studie-onderdeel waarvan de tentamenscores gebruikt zijn als maat voor de studieresultaten: Inleiding in de psychologie III. Omdat wij vooral geïnteresseerd zijn in de zekerheid over de competentie en niet in de zekerheid over de incompetentie zijn alleen de zekerheidsscores gebruikt over het niveau dat men meende te kunnen realiseren. Bij de analyses is gebruik gemaakt van de gemiddelde itemscore.

Sociale Vergelijking

De deelnemers werden verzocht te denken aan een medestudent(e) met wie zij zichzelf meestal vergelijken als het gaat om studievaardigheden en -resultaten. Op een zevenpuntsschaal, met als extremen "veel beter" (7) en "veel slechter" (1), kon aangegeven worden hoe goed deze medestudent(e) is in vergelijking met zichzelf.

Gevoelens

Er zijn zes vragen gesteld over de gevoelens die men koestert over de eigen studievaardigheden en -resultaten. Daarbij is gebruik gemaakt van bipolaire vijfpuntsschalen, bijvoorbeeld met trots (5) en schaamte (1) als extremen. Cronbach's alfa is .82. Bij de analyse is gebruik gemaakt van de gemiddelde itemscore.

Attributies

Er zijn vijf vragen gesteld die betrekking hebben op attributies aan gebrek aan talent en gevoelens van hulpeloosheid als men zou zakken voor het tentamen. Een voorbeeld is: "Als ik voor het tentamen Inleiding III zou zakken, betekent dit dat ik op het gebied van Inleiding III onvoldoende capaciteiten bezit". Elke stelling was voorzien van een zespuntsschaal met als extremen "duidelijk oneens" (1) en "duidelijk eens" (6). Cronbach's alfa voor deze schaal is .73. Bij de analyses is gebruik gemaakt van de gemiddelde itemscore.

Tevens zijn er vijf vragen gesteld die betrekking hebben op de inzet en strategie-ontwikkeling die ontplooid worden naar aanleiding van het zakken voor het tentamen. Een voorbeeld is: "Als ik voor het tentamen Inleiding III zou zakken, zal ik extra tijd en inzet besteden aan de herkansing". Ook in dit geval was elke stelling voorzien van een zespuntsschaal met als extremen "duidelijk oneens" (1) en "duidelijk eens" (6). Cronbach's alfa voor deze schaal is .60. Bij de analyses is gebruik gemaakt van de gemiddelde itemscore.

Studieresultaten

Als maat voor de studieresultaten is gebruik gemaakt van het cijfer dat behaald werd voor het studie-onderdeel Inleiding in de psychologie III. De keuze is gevallen op dit studie-onderdeel omdat de studenten ten tijde van de afname van de vragen al enige feedback hadden gekregen over hun competentie op dit gebied, in de vorm van twee tentamenuitslagen. Uit eerder onderzoek (Vrugt, Langereis & Hoogstraten, 1997) valt op te maken dat de waargenomen eigen competentie de studieresultaten beter voorspelt na zulke feedback.

RESULTATEN

Allereerst is nagegaan in hoeverre de structuur van de 21 doeloriëntatie-items in ons onderzoek overeenkomt met de structuur die door Duda & Nicholls (1992) is vastgesteld. In Duda en Nicholls worden de resultaten van een principale componentenanalyse (PCA) beschreven. Zij vonden vier componenten met een eigenwaarde groter dan één, kozen een oblique oplossing door te roteren naar het Oblimin criterium, en benoemden de gevonden componenten Ego-oriëntatie, Taakoriëntatie, Werkvermijding en Coöperatie. Deze vier componenten verklaren 61.9 procent van de totale variantie van de 21 items. In Tabel 1 staan de resultaten van Duda en Nicholls weergegeven.

Voor een PCA van de 21 doeloriëntatie-items zijn 170 proefpersonen beschikbaar. Er blijken vijf componenten te zijn met een eigenwaarde groter dan één. Met vier componenten wordt 61.8 procent van de totale variantie verklaard, vrijwel evenveel als bij de PCA van Duda en Nicholls. Met vijf componenten wordt 68.1 procent van de variantie verklaard.

Om de in een PCA getrokken componenten te kunnen interpreteren moet de originele matrix van componentladingen geroteerd worden naar een betekenisvolle oplossing. Kiezen we, net als Duda en Nicholls (1992), voor een Oblimin rotatie van de vier-componentenmatrix, dan zijn de eerste twee componenten goed te interpreteren als Ego-oriëntatie en Taakoriëntatie, lijkt de vierde component Coöperatie weer te geven, maar is de derde component moeilijk te duiden. Als we hetzelfde rotatiecriterium toepassen op de vijf-componentenoplossing dan krijgen we goed interpreteerbare componenten. Van deze oplossing zijn in Tabel 2 de componentladingen vermeld, evenals de verklaarde variantie per component, en de correlaties tussen de componenten.

Vier van de vijf componenten kunnen we dezelfde namen geven als Duda en Nicholls. De vijfde component noemen we Hard werken. De items die hoog op deze component laden (items 2, 9 en 12), behoren in de Amerikaanse vragenlijst tot de Taakoriëntatie-items. Een ander verschil met de Amerikaanse resultaten is dat item 21 niet op de Ego-oriëntatiecomponent laadt, maar wel op de Coöperatie-component. Dit kan verklaard worden doordat het item door een typefout van betekenis is veranderd. Alle overige items, 17 in getal, hebben hun hoogste lading op de component met dezelfde interpretatie als van Duda en Nicholls.

Voor verdere studie van de structuur van de Doeloriëntatie-items verwijzen we naar de Appendix. Daarin concluderen we dat de items van de Nederlandse vragenlijst grotendeels dezelfde structuur vertonen als de items in de Amerikaanse versie. Het voornaamste verschil betreft de Taakoriëntatie-schaal die in tweeën uiteenvalt. Op die manier ontstaat een nieuwe

Tabel 1: Resultaten Factoranalyse (Oblimin Rotatie) van Duda en Nicholls

	Factor 1 Ego- oriëntatie	Factor 2 Taak- oriëntatie	Factor 3 Vermijding	Factor 4 Coöperatie
Ik voel mezelf succesvol als:				
Ik meer weet dan anderen	.79			
Ik de beste tentamenresultaten behaal	.82			
Anderen dingen fout doen en ik niet	.69			
Ik de enige ben die vragen kan beantwoorden	.66			
Ik de slimste ben	.87			
Ik anderen overtroef	.75			
Ik betere resultaten behaal dan mijn vrienden	.58			
Anderen het niet zo goed doen als ik	.58			
Ik echt hard werk		.58		
Dat wat ik leer echt zinvol is		.77		
Ik een probleem oplos door hard te werken		.69		
Iets wat ik leer me stimuleert om over dingen na te denken		.78		
Ik een nieuw idee opdoe over hoe de dingen werken		.69		
Ik mijn uiterste best doe		.52		
Ik iets interessants leer		.92		
Ik iets leer waardoor ik meer over een onderwerp wil weten		.78		
Ik er makkelijk van af kom			.86	
Ik niet hard hoeft te studeren			.69	
Ik niets moeilijks hoeft te doen			.80	
Medestudenten en ik elkaar helpen om dingen uit te zoeken				.75
Medestudenten en ik elkaar helpen				.86
Eigenwaarde	5.82	4.40	1.75	1.02
% variantie	27.7	21.0	8.3	4.9
Correlaties tussen factoren				
factor 1				
factor 2	.21			
factor 3	.19	-.17		
factor 4	-.11	.56	-.21	

schaal die we Hard werken noemen. In de hierna volgende analyses worden somscores gebruikt voor ego-oriëntatie (7 items; Cronbach's alfa = 0.91), taakoriëntatie (5 items; alfa = 0.83), werkvermijding (3 items; alfa = 0.77), hard werken (3 items; alfa = 0.78) en coöperatie (2 items; $r = 0.80$). De product-momentcorrelaties tussen de somscores zijn in de bovendriehoek van de correlatiematrix in Tabel 2 vermeld, schuingedrukt. De correlaties tussen somscores blijken weinig te verschillen van de correlaties tussen de componenten na Oblimin rotatie. Uit de correlaties tussen de somscores blijkt dat ego-oriëntatie correleert met werkvermijding ($p < .011$). Taakoriëntatie correleert met coöperatie ($p < .011$), en met hard werken ($p < .011$), en ook coöperatie en hard werken correleren ($p < .05$). Deze resultaten zijn in overeenstemming met de theorie en sluiten aan bij die van Duda en Nicholls (1992).

Verwacht werd dat de taakoriëntatie zou bijdragen aan de waargenomen eigen competentie en dat de ego-oriëntatie zou bijdragen aan neerwaartse vergelijking. Eerst zijn er correlatie-

Tabel 2: Resultaten van principale componentenanalyse van de Doeloriëntatie-items: de vijf-componentenoplossing na Oblimin rotatie

Items (Ik voel mezelf succesvol als....)	EO	TO	Co	HW	WV	
1 Ik meer weet dan anderen	.76	.04	-.08	.28	-.08	
4 Ik de beste tentamenresultaten behaal	.65	-.03	-.06	.25	.14	
8 Anderen dingen fout doen en ik niet	.80	-.09	.04	-.02	.03	
11 Ik de enige ben die vragen kan beantwoorden	.79	.02	.12	-.04	.04	
14 Ik de slimste ben	.81	.15	.06	-.22	.08	
16 Ik anderen overtroef	.82	.05	.00	-.26	.02	
17 Ik betere resultaten behaal dan mijn vrienden	.86	-.12	.02	.06	-.00	
21 Anderen het niet zo goed doen als ik	-.05	-.01	.59	.17	.27	
2 Ik echt hard werk	.07	-.01	.02	.83	.06	
5 Dat wat ik leer echt zinvol is	.15	.54	.05	.16	.19	
9 Ik een probleem oplos door hard te werken	.00	.17	-.01	.71	-.02	
10 Iets wat ik leer me stimuleert om over dingen na te denken	-.06	.73	.07	.16	-.06	
18 Ik een nieuw idee opdoe over hoe de dingen werken	.10	.77	-.06	.11	-.11	
12 Ik mijn uiterste best doe	-.08	.10	.17	.72	-.00	
15 Ik iets interessants leer	-.05	.88	.04	-.16	.13	
20 Ik iets leer waardoor ik meer over een onderwerp wil weten	-.09	.79	.01	-.01	-.10	
19 Ik er makkelijk van af kom	.13	.07	-.03	-.14	.75	
6 Ik niet hard hoeft te studeren	-.01	-.04	.06	-.05	.85	
3 Ik niets moeilijks hoeft te doen	-.01	-.04	-.03	.21	.80	
7 Medestudenten en ik elkaar helpen om dingen uit te zoeken	.06	.01	.92	-.08	-.08	
13 Medestudenten en ik elkaar helpen	.04	.03	.93	.05	-.09	
Perc. verkl. variantie per component (totaal perc. verkl. variantie = 68.1)	21.1	13.8	10.2	10.4	10.2	
	Correlaties					
Afkortingen:	EO	TO	Co	HW	WV	
EO = Ego-oriëntatie	EO	1.00	.08	.15	.09	.44
TO = Taakoriëntatie	TO	.06	1.00	.30	.42	.05
CO = Coöperatie	Co	.11	.26	1.00	.25	.18
HW = Hardwerken	HW	.07	.28	.18	1.00	.14
WV = Werkvermijding	WV	.37	.02	.23	.10	1.00
<p>Noot: de steekproefgrootte is 170; componentladingen groter dan 0.40 zijn dik gedrukt; het rotatiecriterium is het genormaliseerde Directe Oblimin criterium (Harman, 1967, Paragraaf 15.5); in de correlatiematrix zijn boven de diagonaal, schuingedrukt, de product-momentcorrelaties tussen scores vermeld.</p>						

coëfficiënten berekend tussen de vijf doeloriëntatieschalen, waargenomen eigen competentie (SES), sociale vergelijking, gevoelens en studieresultaten. Alleen de taakoriëntatie-schaal correleert significant met SES, $r = .18$, $p < .05$. De taakoriëntatie-items representeren het begrijpen van de stof, dus hoe meer studenten menen dat het begrijpen van de stof bijdraagt aan hun goede prestaties hoe zekerder zij zijn over hun competentie.

Om te bepalen of de resultaten in overeenstemming zijn met de verwachtingen hebben we hiërarchische regressie-analyses uitgevoerd. In de eerste regressie-analyse fungeerden de vijf schalen die de doeloriëntaties representeren als predictoren. Omdat verwacht werd dat de taakoriëntatie zou bijdragen aan de waargenomen eigen competentie zijn eerst de schalen taakoriëntatie, hard werken en samenwerken, in deze volgorde, ingevoerd en daarna achter-eenvolgens de schalen ego-oriëntatie en werkvermijding. SES fungeerde als criterium. De resultaten staan in Tabel 3.

Tabel 3: Regressie van Waargenomen Eigen Competentie (SES) op Predictor Variabelen

Volgorde van invoer	R ²	Beta	t
1. Taakoriëntatie	.03	.18	2.35*
2. Hard werken	.03	-.03	-.36
3. Samenwerken	.07	-.20	-2.63*
4. Ego-oriëntatie	.07	-.06	-.81
5. Werkvermijding	.07	-.04	-.52
Totaal R ² = .07			E (7, 163) = 2.72
			p < .05

* p < .05

Tabel 4: Regressie van Sociale Vergelijking op Predictor Variabelen

Volgorde van invoer	R ²	Beta	t
1. Ego-oriëntatie	.004	-.06	-.88
2. Werkvermijding	.006	-.04	-.47
3. Taakoriëntatie	.006	.02	.28
4. Hard werken	.019	.12	1.44
5. Samenwerken	.026	.09	1.08
Totaal R ² = .03			E (7, 160) = .86
			p > .05

Tabel 5: Correlaties tussen Waargenomen Eigen Competentie (SES), Sociale Vergelijking, Gevoelens over de Eigen Vaardigheden en Studieresultaten

	M	SD	4	3	2
1. SES	8.27	.91	.03	.26**	-.12
2. Sociale vergelijking	4.34	1.10	-.18*	-.41**	
3. Gevoelens	3.52	.65	.22**		
4. Studieresultaten	6.67	2.12			

* p < .05
** p < .01

Tabel 6: Regressie van Gevoelens op Predictor Variabelen

Volgorde van invoer	R ²	Beta	t
1. SES	.10	.32	4.30**
2. Sociale vergelijking	.23	-.36	-5.20**
3. Taakoriëntatie	.25	.13	1.91
4. Hard werken	.25	.05	.71
5. Samenwerken	.26	.06	.77
6. Ego-oriëntatie	.26	.06	.56
7. Werkvermijding	.26	.07	.97
Totaal R ² = .26			E (7, 155) = 7.93
			p < .001

** p < .001

Uit Tabel 3 blijkt dat de taakoriëntatie-schaal een significante bijdrage heeft geleverd. Ook zien we dat de bijdrage van de schaal die betrekking heeft op samenwerken significant is. Deze variabelen verklaren tezamen 7% van de variantie in SES. Deze resultaten zijn in overeenstemming met de verwachtingen.

In de tweede regressie-analyse fungeerden de vijf schalen die de de doeloriëntaties representeren als predictoren, sociale vergelijking fungeerde als criterium. Omdat verwacht werd dat de ego-oriëntatie zou bijdragen aan de sociale vergelijking zijn eerst de schalen ego-oriëntatie en werkvermijding, in deze volgorde, ingevoerd en daarna achtereenvolgens de schalen taakoriëntatie, hard werken en samenwerken. De resultaten staan in Tabel 4. Uit deze tabel blijkt dat geen van de predictoren een significante bijdrage heeft geleverd.

Op basis van de resultaten van Vrugt (1994) werd tevens verwacht dat de waargenomen eigen competentie en de sociale vergelijking zouden bijdragen aan de gevoelens over de eigen studievaardigheden en dat die gevoelens op hun beurt zouden bijdragen aan de prestaties. In Tabel 5 staan de correlatiecoëfficiënten weergegeven voor de relatie tussen waargenomen eigen competentie (SES), sociale vergelijking, gevoelens over de eigen studievaardigheden en -resultaten (Gevoelens) en studieresultaten.

Uit deze tabel blijkt dat SES significant correleert met de gevoelens over de eigen studievaardigheden en -resultaten. Verder zien we een negatieve correlatie tussen de sociale vergelijking en de gevoelens over de eigen studievaardigheden en -resultaten. Neerwaartse vergelijking gaat dus gepaard met positieve gevoelens. Sociale vergelijking correleert eveneens negatief met studieresultaten: neerwaartse vergelijking gaat dus ook gepaard met goede studieresultaten. Ook de correlatie tussen gevoelens en studieresultaten is significant.

In de regressie-analyse fungeerde de variabele Gevoelens over de eigen studievaardigheden en -resultaten als criterium, als predictoren werden achtereenvolgens ingevoerd: SES, sociale vergelijking en de vijf schalen die de doeloriëntaties representeren. De resultaten van deze analyse staan in Tabel 6.

In Tabel 6 zien we dat SES en sociale vergelijking een significante bijdrage leveren aan de gevoelens over de eigen studievaardigheden en -resultaten. Ook blijkt neerwaartse vergelijking gepaard te gaan met positieve gevoelens. De resultaten zijn dus in overeenstemming met de verwachtingen.

In de volgende regressie-analyse fungeerde de variabele Studieresultaten als criterium. Als predictoren zijn achtereenvolgens ingevoerd: gevoelens over de eigen studievaardigheden en -resultaten, SES, sociale vergelijking en de vijf schalen die de doeloriëntaties representeren. De resultaten staan in Tabel 7.

Tabel 7: Regressie van Studieresultaten op Predictor Variabelen

Volgorde van invoer	R ²	Beta	t
1. Gevoelens	.05	.22	2.85**
2. SES	.05	.008	.10
3. Sociale vergelijking	.06	-.11	-1.39
4. Taakoriëntatie	.06	.07	-1.47
5. Hard werken	.07	-.08	1.03
6. Samenwerken	.07	-.07	.89
7. Ego-oriëntatie	.08	-.02	.32
8. Werkvermijding	.08	-.05	.65
Totaal R ² = .08		F (7, 153) = 1.64	p = .11

** p < .001

Uit Tabel 7 komt naar voren dat alleen de bijdrage van gevoelens over de eigen studievaardigheden en -resultaten significant is. Dit resultaat is eveneens in overeenstemming met de verwachting.

Verwacht werd ook dat de taakoriëntatie verband zou houden met de attributie van falen aan gebrek aan inzet en strategie-ontwikkeling, en dat ego-oriëntatie gerelateerd zou zijn aan de attributie van falen aan gebrek aan talent en hulpeloosheid. De correlaties tussen de schalen die de taakoriëntatie representeren en gebrek aan inzet en strategie-ontwikkeling zijn significant (taakoriëntatie: $r = .19$, $p < .05$; hard-werken: $r = .18$, $p < .05$; coöperatie: $r = .20$, $p < .05$). De correlaties tussen de ego-oriëntatie-schalen en de attributies van falen aan gebrek aan talent en hulpeloosheid zijn niet significant.

DISCUSSIE

De centrale vraagstelling van dit onderzoek is op welke wijze taak- en ego-oriëntatie op het gebied van de studie bijdragen aan de studiemotivatie en -prestatie. In verband daarmee is allereerst nagegaan of de structuur van doeloriëntatie-items in ons onderzoek overeenkomt met de structuur die door Duda en Nicholls (1992) is vastgesteld. Uit de analyses die beschreven staan in de resultatensectie en in de Appendix blijkt dat de structuur van de doeloriëntatie-items in ons onderzoek het beste beschreven wordt door vijf componenten, waarbij aan de vier door Duda en Nicholls onderscheiden componenten, een component is toegevoegd die wij Hard werken hebben genoemd. De items die bij Duda en Nicholls de hoogste lading hebben op Ego-oriëntatie, Werkvermijding of Coöperatie hebben ook in ons onderzoek de hoogste lading op die component. De items die bij Duda en Nicholls hoog laden op Taakoriëntatie blijken in ons onderzoek hoog te laden op twee verschillende componenten. De items die betrekking hebben het begrijpen van de stof, laden hoog op een component die we in navolging van Duda en Nicholls Taakoriëntatie hebben genoemd. De items die hoog laden op de vijfde component representeren Hard werken. In onderzoek bij tweedejaarspsychologie-studenten (Becker, Gillis & Jilma, 1995) bleek eveneens dat de taakoriëntatie-items van Duda en Nicholls hoog laden op twee verschillende componenten. Deze gegevens doen vermoeden dat hard werken en het begrijpen van de stof voor Amsterdamse psychologiestudenten twee verschillende manieren vormen om tot goede prestaties te komen. Op basis van onze resultaten kunnen daarover echter nog geen definitieve uitspraken worden gedaan.

Ook is gebleken dat de schalen die een taakoriëntatie representeren correleren; de schalen die een ego-oriëntatie representeren bleken eveneens te correleren, terwijl de taakoriëntatie-schalen niet correleerden met de ego-oriëntatie-schalen. Deze gegevens verschaffen een aanvullende onderbouwing van de theorie.

Verwacht werd dat de taakoriëntatie zou bijdragen aan de waargenomen eigen competentie en dat de ego-oriëntatie zou bijdragen aan neerwaartse vergelijking. De resultaten zijn ten dele in overeenstemming met de verwachtingen. Taakoriëntatie en coöperatie hebben een significante bijdrage geleverd aan de waargenomen eigen competentie. De items die taakoriëntatie meten, representeren het begrijpen van de stof. De opvatting dat het begrijpen van de stof een belangrijke determinant is van de eigen goede prestaties draagt dus bij aan de zekerheid over de eigen competentie. Coöperatie leverde een negatieve bijdrage: de opvatting dat samenwerken een belangrijke determinant is van de eigen goede prestaties gaat dus gepaard met relatieve onzekerheid over de eigen competentie. Samenwerking als determinant van goed prestaties geeft blijkbaar aanleiding om aan de eigen competentie te twifelen, misschien omdat men door samenwerking geen duidelijk beeld kan vormen over de eigen bijdrage.

Tevens werd verwacht dat de waargenomen eigen competentie en de sociale vergelijking zouden bijdragen aan de gevoelens over de eigen studievaardigheden en dat die gevoelens op hun beurt zouden bijdragen aan de prestaties over de eigen competentie. In overeenstemming

met de resultaten van eerder onderzoek (Vrugt, 1994) is gebleken dat de waargenomen eigen competentie en neerwaartse vergelijking een bijdrage hebben geleverd aan de gevoelens over de eigen studievaardigheden die op hun beurt hebben bijgedragen aan de studieprestaties. Op grond hiervan kunnen we concluderen dat de opvattingen over de determinanten van de eigen studieprestaties, als relatief globale opvattingen, de motivatie beïnvloeden doordat zij inwerken op de waargenomen eigen competentie, die direct betrekking heeft op de relevante taak. Taakoriëntatie en samenwerking verklaren tezamen 7% van de variantie in de waargenomen competentie. Er wordt dus maar een gering deel van de variantie verklaard. Wel moet bedacht worden dat het hier opvattingen betreft over de determinanten van goede prestaties op het gebied van de studie, die van algemene aard zijn. Aan dergelijke algemene opvattingen hoeven de betrokkenen geen consequenties te verbinden voor hun taakspecifieke motivatie en gedrag. Zo bezien kan gesteld worden dat een gering, maar niet te verwaarlozen deel van de variantie in de waargenomen eigen competentie verklaard wordt door taakoriëntatie en coöperatie.

Opvallend is dat alleen de taakoriëntatie, die betrekking heeft op het begrijpen van de stof, en samenwerking een bijdrage hebben geleverd aan de waargenomen eigen competentie. Deze resultaten ondersteunen de overwegingen die in de inleiding geformuleerd zijn over het belang van een taakoriëntatie voor de motivatie van beginners. In aansluiting hierbij is gebleken dat de ego-oriëntatie geen noemenswaardige bijdrage heeft geleverd aan de sociale vergelijking en dus ook niet via de sociale vergelijking heeft bijgedragen aan de studieresultaten. Ook deze gegevens suggereren dat een ego-oriëntatie bij beginners niet motiverend werkt.

Verder werd verwacht dat de taakoriëntatie verband zou houden met de attributie van falen aan gebrek aan inzet en strategie-ontwikkeling, en dat een ego-oriëntatie gerelateerd zou zijn aan de attributie van falen aan gebrek aan talent en hulpeloosheid. De schalen die de verschillende aspecten van de taakoriëntatie representeren: taakoriëntatie, samenwerking en hardwerken bleken te correleren met attributies aan gebrek aan inzet en strategie-ontwikkeling. Als men meent dat de eigen goede prestaties het gevolg zijn van het begrijpen van de stof, hard werken of samenwerken zal men bij falen geneigd zijn een verklaring te zoeken in termen van inzet en strategie-ontwikkeling. Op grond van deze resultaten kan eveneens geconcludeerd worden dat het begrippenkader dat ten grondslag ligt aan de doeloriëntaties zoals beschreven door Duda en Nicholls (1992) overeenstemt met gezichtspunten die op dit terrein geformuleerd zijn door Bandura (1989) en Elliot en Dweck (1988).

Onze resultaten verschaffen ook enig inzicht in het samenspel van doeloriëntaties, attributies en waargenomen eigen competentie bij beginners. Bij beginners op een bepaald terrein zal het optreden van falen niet altijd te voorkomen zijn. Vooral in de aanvangsfase als de opvattingen over de eigen competentie nog in ontwikkeling zijn zullen de opvattingen over de determinanten van de eigen prestaties bijdragen aan de wijze waarop falen verklaard wordt (Vrugt, Langereis & Hoogstraten, 1997). Personen met een taakoriëntatie zullen falen vooral toeschrijven aan gebrek aan inzet en strategie-ontwikkeling. Deze attributies zullen waarschijnlijk aanleiding geven tot een verhoogde inzet en tot nadenken over de stof en de eigen aanpak. Hierdoor komt de bekwaamheid tot ontwikkeling, en komt een positief beeld van de eigen competentie tot stand. Uit ons onderzoek blijkt dat een dergelijk beeld via de gevoelens over de eigen vaardigheden bijdraagt aan het realiseren van goede studieresultaten. Uit onderzoek van Silver, Mitchell en Gist (1994) kunnen we verder opmaken dat goede (studie)resultaten het positieve beeld over de eigen competentie versterken. Een taakoriëntatie kan dus de aanzet geven tot een positief proces van zichzelf-bevestigende voorspellingen.

Onze gegevens verschaffen geen inzicht in de psychologische processen bij personen met een ego-oriëntatie. We weten dus ook niet of de ego-oriëntatie eveneens de aanzet kan geven tot een proces van zichzelf-bevestigende voorspellingen. Verder onderzoek zal hierover uitsluitsel moeten verschaffen.

REFERENTIES

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, inc.
- Bandura, A. (1989). Self-regulation of motivation and action through internal standards and goal systems. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: W. H. Freeman and Company.
- Bekker, M., Gillis, M. & Jilma, J. (1995). *Factoren die de prestaties beïnvloeden in het onderzoeksdeel van het Onderzoekspracticum*. Intern Rapport Faculteit Psychologie, Universiteit van Amsterdam.
- Duda, J. L. & Nicholls, J. G. (1992). Dimensions of achievement motivation in school, work and sport. *Journal of Educational Psychology*, 84, 290-299.
- Elliot, E. S. & Dweck, C. S. (1988). Goals: an approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5-12.
- Hoffman, D. A. & Stricklund, O. J. (1995). Task performance and satisfaction: evidence for a task-by-ego-orientation interaction. *Journal of Applied Social Psychology*, 25, 495-511.
- Harman, H. H. (1967). *Modern factor analysis* (2nd edition). Chicago IL: The University of Chicago Press.
- Locke, E. A. & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs NJ: Prentice Hall.
- Silver, W. S., Mitchell, T. R. & Gist, M. E. (1995). Responses to successful and unsuccessful performance: the moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes*, 62, 286-299.
- Vrugt, A. (1994). Perceived self-efficacy, social comparison, affective reactions and academic performance. *British Journal of Educational Psychology*, 64, 465-472.
- Vrugt, A., Langereis, M. P. & Hoogstraten, J. (1997). Academic self-efficacy and malleability of relevant capabilities as predictors of exam performance. *Journal of Experimental Education*, 66, 61-74.
- Wood, R. E. & Locke, R. A. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measurement*, 47, 1013-1024.

APPENDIX PRINCIPALE COMPONENTENANALYSE VAN DE DOELORIËNTATIE-ITEMS

Bij principale componentenanalyse van de doeloriëntatie-items bleken vijf componenten een eigenwaarde groter dan één te hebben. De eigenwaarden van de eerste zeven factoren zijn 5.6, 3.9, 1.9, 1.6, 1.3, 0.9 en 0.7. Rotatie van vijf componenten volgens het Oblimin criterium leverde een goed interpreteerbare ladingenmatrix op. Op basis van deze ladingenmatrix, gerapporteerd in Tabel 2, zijn de doeloriëntatie-items in schalen ingedeeld. De somscores van deze schalen zijn gebruikt in de verdere analyses.

Echter, beantwoording van de vraag in hoeverre de structuur van de Nederlandse vragenlijst overeenkomt met het Amerikaanse origineel van Duda en Nicholls (1992) vergt nadere analyse. De gelijkheid tussen verschillende componentenoplossingen kan uitgedrukt worden in een congruentiemaat (toegeschreven aan verschillende auteurs; zie Harman, 1967, p. 269 e.v.). Deze congruentiemaat kan variëren van -1 tot +1 en heeft een zelfde interpretatie als een correlatiecoëfficiënt. Een probleem is dat Duda en Nicholls van hun oplossing niet de volledige ladingenmatrix hebben gepubliceerd. Zij vermelden alleen de ladingen met een grote waarde dan 0.40. Om de berekening van de congruentiemaat mogelijk te maken, hebben wij voor de niet gepubliceerde componentladingen de waarde nul ingevuld.

De Nederlandse vier- en vijf-componentenoplossingen zijn geroteerd volgens de bekende Varimax en Oblimin criteria (Harman, 1967). In Tabel A1 zijn voor elke oplossing de verklaarde varianties vermeld, de onderlinge correlaties, en de congruentiecoëfficiënten voor de samenhang met de componenten van Duda en Nicholls. Hoewel Duda en Nicholls een vier-componentenoplossing rapporteren, vertonen de vijf-componentenoplossingen een hogere congruentie van corresponderende componenten. Dit komt doordat in de Nederlandse vier-

Tabel A1: Verklaarde varianties, onderlinge correlaties, en congruentie-coëfficiënten van componenten van verschillende PCA's

Analyse	Comp. Naam	Verkl. var.	Correlaties tussen componenten				Congruenties met de componenten van D&N			
			DN/E	DN/T	DN/W	DN/C	DN/E	DN/T	DN/W	DN/C
DN/PCA4 Oblimin	DN/E	>4.20	1.00							
	DN/T	>4.22	.21	1.00						
	DN/W	>1.86	.19	-.17	1.00					
	DN/C	>1.30	-.11	.56	-.21	1.00				
PCA4 Varimax	EO	4.95	1.00				DN/E	DN/T	DN/W	DN/C
	TO	3.14	.00	1.00			.90	.04	.30	.01
	HW	2.35	.00	.00	1.00		.07	.89	-.20	.24
	Co	2.53	.00	.00	.00	1.00	.11	.63	.23	-.01
PCA4 Oblimin	EO	4.90	1.00				DN/E	DN/T	DN/W	DN/C
	TO	2.93	-.06	1.00			.90	.04	.23	-.05
	HW?	2.30	.16	.20	1.00		.06	.85	-.28	.19
	Co	2.46	.25	.10	.22	1.00	.03	.56	.21	-.08
PCA5 Varimax	EO	4.43	1.00				DN/E	DN/T	DN/W	DN/C
	TO	2.90	.00	1.00			.92	.04	.19	.05
	WV	2.14	.00	.00	1.00		.06	.92	-.00	.13
	Co	2.13	.00	.00	.00	1.00	.30	.06	.90	.03
	HW	2.18	.00	.00	.00	.00	.19	.20	.10	.86
PCA5 Oblimin	EO	4.43	1.00				DN/E	DN/T	DN/W	DN/C
	TO	2.90	.06	1.00			.93	.00	.04	.03
	WV	2.14	.37	.02	1.00		.01	.89	.00	.02
	Co	2.13	.11	.26	.23	1.00	.11	.03	.94	-.08
	HW	2.18	.07	.28	.10	.18	.13	.06	-.00	.89
PCA4 Patroon Oblique	EO	4.97	1.00				DN/E	DN/T	DN/W	DN/C
	TO	3.73	.03	1.00			.90	.09	.30	.04
	WV?	2.24	.02	-.27	1.00		.09	.95	-.05	.15
	Co	1.90	-.12	-.15	.44	1.00	.25	-.23	.64	.38
PCA5 Patroon Oblique	EO	4.56	1.00				DN/E	DN/T	DN/W	DN/C
	TO	2.76	-.06	1.00			.93	.05	.17	.13
	WV	1.79	-.18	.03	1.00		.04	.91	.01	.07
	Co	1.72	.15	-.18	-.27	1.00	.13	.02	.94	-.06
	HW	2.01	.02	-.18	-.08	-.14	.12	-.01	-.11	.90

Noot: de steekproefgrootte is 170; afkortingen: v.v. = verklaarde variantie, PCA4 = vier-componentenoplossing, PCA5 = vijf-componentenoplossing, DN/ = door Duda en Nicholls (1992) gerapporteerd, Oblimin = rotatie volgens het genormaliseerde Directe Oblimin criterium (Harman, 1967, Paragraaf 15.5), Patroon = rotatie naar een patroon van enen en nullen, EO = Ego-oriëntatie, TO = Taakoriëntatie, WV = Werkvermijding, Co = Coöperatie, HW = Hardwerken; een vraagteken geeft aan dat de interpretatie niet eenduidig is

componentenoplossingen geen component voorkomt die als Werkvermijding te interpreteren is. Het valt op dat de orthogonale Varimax en de oblique Oblimin rotaties vrijwel dezelfde resultaten opleveren. De in Tabel 2 gerapporteerde oplossing heeft congruentiecoëfficiënten van 0,93, 0,89, 0,89 en 0,94 voor de Ego-oriëntatie, Taakoriëntatie, Coöperatie en Werkvermijding. De nieuwe Hardwerken-component vertoont de meeste samenhang met de Amerikaanse Taakoriëntatie-component (0,50).

Teneinde te bezien of nog hogere congruenties bereikt kunnen worden, is geprobeerd de Nederlandse vier- en vijf-componentenoplossingen te roteren naar de gepubliceerde vier-componenten-oplossing van Duda en Nicholls. Met uiteenlopende criteria zijn zowel orthogonale als oblique rotaties uitgevoerd: rotatie naar alleen het gepubliceerde deel van de Amerikaanse ladingenmatrix, rotatie naar de hele Amerikaanse ladingenmatrix (waarbij nullen zijn ingevuld voor het niet gepubliceerde deel), rotatie naar een patroon van enen en nullen dat correspondeert met de Amerikaanse ladingenmatrix, en rotatie met als criterium de maximalisatie van de congruentie tussen corresponderende componenten.

Bij alle rotatiecriteria bleken vijfcomponenten-oplossingen beter dan vier-componenten-oplossingen, en oblique rotaties beter dan orthogonale rotaties. Echter, oblique rotaties naar alleen het gepubliceerde deel van de Amerikaanse ladingenmatrix resulteerden in ladingenmatrices en componenten-correlatiematrixes die slecht te interpreteren waren. Dit gold ook voor de rotaties met het maximale congruentiecriterium. De overige oplossingen leken sterk op elkaar. In Tabel A1 zijn daarom alleen de resultaten vermeld van de vier- en vijf-componentenoplossingen met oblique rotatie naar een patroon van enen en nullen, corresponderend met de Amerikaanse ladingenmatrix. Het blijkt dat de congruenties na deze patroonrotaties inderdaad nog wat hoger zijn. Bovendien lijkt er bij de vier-componentenoplossing na patroonrotatie wel een Werkvermijding-component naar voren te komen. In de vijf-componentenoplossing zijn de congruenties met Werkvermijding en Coöperatie echter veel hoger. Deze ladingenmatrix na patroonrotatie wijkt overigens nauwelijks af van de in Tabel 2 gerapporteerde ladingenmatrix na Oblimin rotatie. Het patroon van lage en hoge waarden in de twee ladingenmatrixes is identiek. Ook na de strengere patroonrotatie laadt Item 21 (oorspronkelijk Ego-oriëntatie) nog steeds het hoogst op de Coöperatie-component en nauwelijks op de andere componenten.

We concluderen dat de structuur van de Nederlandse doeloriëntatie-vragenlijst het best beschreven wordt door een vijf-componentenoplossing, waarbij aan de vier door Duda en Nicholls gevonden componenten een Hardwerken-component is toegevoegd. Uit Tabel A1 blijkt dat de Hardwerken-component in het algemeen zelfs wat meer variantie verklaart dan twee van de andere componenten.

Het didactisch handelen in het basisonderwijs

Dr W.J.C.M. van de Grift¹ en dr J.F. Lam²

1. INLEIDING

In dit artikel wordt verslag gedaan van de constructie van een observatie-instrument. Het instrument is gebaseerd op het Raschmodel en het is ontworpen om voor de Inspectie van het Onderwijs als middel te dienen voor het evalueren van de kwaliteit van het didactisch handelen van de onderwijsgeevenden die werken in groep drie en hoger in het basisonderwijs. Met behulp van dit instrument dienen de volgende evaluatievragen beantwoord te kunnen worden:

1. Wat is de kwaliteit van het didactisch handelen in het basisonderwijs?
2. Verschilt de kwaliteit van het didactisch handelen in de verschillende groepen in het basisonderwijs?
3. Gaat een hoge kwaliteit van het didactisch handelen van onderwijsgeevenden samen met een hoge schooloutput?

2. THEORETISCHE ACHTERGROND

De vraag: 'Hoe moet ik mijn onderwijs inrichten om zoveel mogelijk invloed op de resultaten van mijn leerlingen te hebben?' werd in de zeventiger jaren niet gesteld. Alleen het eerste deel van deze vraag 'Hoe moet ik mijn onderwijs inrichten?' werd gesteld en dat was op dat moment geen onderwerp van een empirische wetenschap. Wie op dat moment geïnteresseerd was in dit onderwerp ging te rade bij de (normatieve) pedagogiek en haalde aanvullende kennis bij de leerpsychologen. De pedagoog Langeveld (1959) benaderde het thema vanuit de fenomenologie. De Block (1965) maakte in zijn boek 'Algemene didactiek' leerpsychologische concepten toegankelijk voor pedagogen, maar kreeg om die reden wel het verwijt van eenzijdigheid van Van Gelder, Peters, Oudkerk Pool en Sixma (1972). De laatstgenoemden schreven een standaardwerk voor de opleiding van onderwijsgeevenden. Geleidelijk ontstond er, onder invloed van Van Parreren (1962, 1965) belangstelling voor het leerproces, dat leerlingen doormaken, maar daarmee was de relatie naar het handelen van de onderwijsgevende nog onvoldoende gelegd. Het werk van de onderwijsgevende stond in deze periode allermest centraal. Eerder was er de idee om zonder de onderwijsgevende het kind tot leren te brengen. Men raakte in Nederland in deze periode gecharmeerd van de ideeën van Skinner (1954, 1958) over geprogrammeerde instructie en teaching machines. Geprogrammeerde instructie zou elke leerling de gelegenheid gaan bieden om zelfstandig en in eigen tempo de leerstof door te werken. Dat er verschillen zijn in de manieren waarop leerlingen leren, werd vanuit deze inzichten niet onderkend. De onderwijsgevende was overbodig als leraar. De onderwijsgevende werd gezien als degene die het leerproces van de leerling organiseerde. Uiteraard zijn er uit deze ontwikkeling ook heel goede en belangrijke zaken naar voren gekomen. Dit geldt met name voor het denken in doelstellingen en hiërarchieën van doelen (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956, Mager, 1962).

Gagné (1965) legde volgens Walberg en Haertel (1992) de conceptuele basis voor de ontwerpers van instructieprocessen. Gagné legde een relatie tussen de vijf door hem onderscheiden leerdomeinen en de 'events of instruction'. Met dit laatste bedoelde hij onder meer: activiteiten om de aandacht van de leerling te trekken, het motiveren van de leerling, het ontwikkelen en gebruiken van een leerhiërarchie en het bieden van oefening en feedback op

basis van peilingen. Hiermee legde Gagné de relatie tussen leerpsychologische concepten en het handelen van onderwijsgeevenden.

Een ander belangrijk beginpunt van het empirisch onderzoek naar het didactisch handelen wordt door Walberg & Haertel (1992) gelegd bij Scriven (1967) en Cronbach (1963). Zij benadrukten het belang van formatieve evaluatie voor het opsporen van zwakheden in niet alleen nieuwe instructiematerialen maar ook in instructie-activiteiten. Daarmee kwam er aandacht voor empirisch onderzoek naar het didactisch handelen van onderwijsgeevenden.

In de tachtiger jaren ontstond bij een aantal Nederlandse onderwijsonderzoekers aandacht om de kwaliteit van het didactisch handelen van onderwijsgeevenden af te meten aan het effect ervan op de prestaties van de leerlingen. Een eerste mijlpaal vormde het eerste in Londen georganiseerde International Congress for School Effectiveness and Improvement (ICSEI).

In veel onderwijsonderzoek dat op dit moment naar de effectiviteit van het onderwijs wordt uitgevoerd, verwijst effectiviteit naar de mate waarin de beoogde doelstellingen worden behaald (Scheerens, 1989). De effectiviteit van een school kan worden opgevat als de mate waarin de school haar beoogde (leer)doelstellingen weet te realiseren. Hierbij wordt effectiviteit vaak uitgedrukt in scores behaald op een leerprestatietest. Voor de bepaling van de maat voor schooleffectiviteit wordt de volgende definitie gehanteerd (vgl. ook Bosker, 1991 en Brandsma, 1993): Een school is effectief als zij er, jaar in jaar uit, in slaagt aan het eind van acht jaar basisonderwijs hogere prestaties bij de leerlingen teweeg te brengen op schoolvorderingstoetsen, dan verwacht mag worden op basis van de instroomkenmerken (de aanleg en het milieu) van de leerlingen van die school.

Het is algemeen bekend dat scholen verschillen in de leerprestaties die zij bij hun leerlingen bereiken. Uit verschillende onderzoeken in het primair onderwijs blijkt dat de leerlingen op de ene school meer leren dan de leerlingen op een andere school. Deze verschillen blijken over de jaren heen redelijk stabiel te zijn (Van Batenburg, 1990; Bosker, 1991; Blok & Hoeksma, 1993).

In de onderzoeksliteratuur worden twee belangrijke groepen factoren genoemd die systematisch met de effectiviteit van scholen samenhangen: de kwantiteit en de kwaliteit van het onderwijs. Met de kwantiteit van het onderwijs wordt veelal verwezen naar termen zoals effectieve leertijd en gelegenheid tot leren. Tijd alleen is echter een onvoldoende verklaring voor de schoolprestaties. In die tijd moet ook kwalitatief goed onderwijs gegeven worden. Kwalitatief goed onderwijs blijkt het resultaat te zijn van een samenspel van verschillende aspecten. Het gaat zowel om de geschiktheid van het leermateriaal, de manier waarop de leerlingen worden gegroepeerd (zoals heterogene groepen of niveaugroepen), als om het gedrag van onderwijsgeevenden (Creemers, 1991; Slavin, 1994). De onderwijsgevende is op school de aangewezen persoon om de leerlingen die kennis en vaardigheden bij te brengen die leiden tot het bereiken van de gewenste onderwijsleerdoelen. De onderwijsgevende zal dit moeten doen met de vrijheden en beperkingen die het leermateriaal en het schoolbeleid met zich meebrengen.

Effectief onderwijs vraagt niet alleen om instructievaardigheden van de onderwijsgevende, maar ook om managementvaardigheden (zie o.a. Denham & Lieberman, 1980; Veenman et al, 1986; Veenman et al, 1992; Roelofs, 1993). Effectieve onderwijsgeevenden besteden hun tijd voornamelijk aan het lesgeven. Deze onderwijsgeevenden kunnen goed de orde in de groep handhaven. Ze verliezen weinig tijd met het ophalen en uitdelen van lesmaterialen. De tijd die met leswisselingen is gemoeid houden ze minimaal. Zij beginnen op tijd met hun lessen. De leerlingen hoeven niet lang te wachten op individuele hulp. De leerlingen begrijpen wat ze moeten doen. De instructie is goed afgestemd op het niveau van de leerlingen. Kortom de effectieve onderwijsgevende zorgt er voor dat de leerlingen hun tijd voornamelijk besteden aan leren, d.w.z. de leerlingen letten op tijdens de instructie en zijn taakgericht bezig tijdens de verwerking van de leerstof. Dit opletten en taakgericht bezig zijn, hangt nauw samen met de leerprestaties van de leerlingen (Anderson, 1984; Veenman, et al., 1986; Roelofs, 1993).

Onderzoekingen die verricht zijn naar de effectiviteit van leerkrachtgedrag hebben vele gedragingen opgeleverd die samenhangen met leerlingprestaties (en impliciet met de taakgerichtheid van leerlingen). Echter de vaardigheden afzonderlijk, leiden niet tot een verhoging van de effectiviteit. Het is eerder een combinatie van leerkrachtgedragingen die de instructie effectief maakt. De activiteiten van de onderwijsgevende zijn te onderscheiden in het geven van instructie, het bewaken van het zelfstandig werken van leerlingen (waaronder ook het verwerken van de leerstof) en een goed klassenmanagement.

Een instructiemodel waarin een combinatie van instructievaardigheden tot uitdrukking komt, is de directe instructie. Hierin vindt het onderwijsleerproces stapsgewijs plaats onder directe begeleiding van een onderwijsgevende (Rosenshine, 1987). Uit onderzoekingen in het basis-, speciaal en voortgezet onderwijs is gebleken dat dit een effectief instructiemodel is, met name ook voor kinderen uit achterstandssituaties (Carnine, 1989). Andere termen die naar dit instructiemodel verwijzen, zijn expliciet onderwijs, expliciete instructie en actief onderwijzen. In recente literatuur (Walberg, 1990; Creemers, 1991; Veenman, 1992; Veenman et al, 1992; Veenman et al, 1993; Roelofs, 1993) wordt vaak verwezen naar en gebruik gemaakt van het instructiemodel van Rosenshine en Stevens (1986). Directe instructie kent de volgende fasen: Dagelijkse terugblik en ophalen van voorgaande leerstof, presentatie en uitleg van (nieuwe) leerstof, inclusief het laten zien of voordoen, begeleidde inoefening, correctie en feedback, zelfstandig werken en wekelijks en maandelijks terugblik/evaluatie.

Samengevat kan geconcludeerd worden dat het onderwijs in de groep afhankelijk is van de onderwijsgevende in die groep. Hij of zij moet de leerlingen zover weten te brengen dat ze optimaal willen leren. Waarbij belangrijk is dat de onderwijsgevende over zowel goede instructie- als managementvaardigheden beschikt. Deze combinatie van vaardigheden zal in deze bijdrage gevat worden onder de term didactisch handelen.

3. STEEKPROEF

Door de Inspectie van het Onderwijs zijn in 1994 gegevens verzameld met als doel evaluatieve uitspraken te kunnen doen over de effectiviteit en het rendement van het didactisch handelen van onderwijsgevers in het basisonderwijs in Nederland. De steekproef is aselekt getrokken en bestaat uit 231 scholen. Deze steekproef is voldoende groot om proporties in de populatie te schatten met een nauwkeurigheid³ van 6.5%⁴. Gedurende de uitvoering van het project konden bij 4 scholen geen gegevens verzameld worden. Hier hebben steeds bijzondere redenen en oorzaken, zoals overmacht, een rol gespeeld. Het aantal geëvalueerde scholen komt daarmee te liggen op 227. De groep geëvalueerde scholen is op representativiteit voor de populatie gecontroleerd op de volgende kenmerken: denominatie, regionale spreiding, schoolgrootte en leerlingkenmerken. In tabel 1 wordt een overzicht geboden. Het blijkt dat de geëvalueerde scholen op deze punten niet noemenswaardig afwijken van de populatie van basisscholen.

Om de verbanden tussen het didactisch handelen en de schooloutput te kunnen berekenen zijn, na toestemming van de betreffende scholen, de gemiddelde Cito-eindtoetsresultaten over de jaren 1990, 1991, 1992, 1993 en 1994 bij het Cito opgevraagd. Het blijkt dat 63% van scholen uit de steekproef meedoet aan deze toets. Dit komt goed overeen met het landelijk percentage scholen van 68%. De steekproefnauwkeurigheid van deze groep van 146 'Citotoetssscholen' uit de steekproef is 8.2%, welke ruim voldoende is om nauwkeurige uitspraken te kunnen doen over de situatie in de populatie. Ook de groep 'Citotoetssscholen' is vergeleken met de populatie op bovengenoemde kenmerken en bleek redelijk representatief voor de populatie.

Tabel 1: *Overzicht van populatie- en steekproefkenmerken*

	populatie n = 8082	steekproef N = 231	citotoetscholen N = 146
<i>denominatie</i>	%	%	%
rooms-katholiek	29.0	31.3	36.5
protestant christelijk	30.2	35.2	35.9
openbaar	34.8	29.5	23.5
openbaar-bijzonder	6.0	3.9	4.1
<i>regionale spreiding</i>	%	%	%
Alkmaar	7.0	9.1	4.1
Arnhem	12.6	7.0	9.7
Breda	7.3	7.4	10.3
De Meern	8.0	8.3	5.5
Den Bosch	9.0	6.5	10.3
Groningen	9.0	10.9	11.7
Haarlem	6.6	8.3	9.7
Hengelo	5.6	4.8	6.9
Leeuwarden	6.9	4.8	6.2
Roermond	5.9	7.8	10.3
Rotterdam	7.7	7.0	5.5
Zoetermeer	9.0	11.3	7.6
Zwolle	5.3	7.0	2.1
<i>schoolgrootte</i>	%	%	%
≤ 150 leerlingen	44.9	43.0	44.1
151-249 leerlingen	35.8	39.6	40.7
≥ 250 leerlingen	19.2	17.4	15.2
<i>leerlinginstroomkenmerken</i>	%	%	%
gemiddeld % 1.00 leerlingen	55.1	53.4	54.4
gemiddeld % 1.25 leerlingen	33.7	34.6	36.0
gemiddeld % 1.40 leerlingen	.1	.1	.1
gemiddeld % 1.70 leerlingen	.3	.2	.3
gemiddeld % 1.90 leerlingen	10.9	11.7	9.2

4. PROCEDURE

De werkzaamheden in dit project zijn uitgevoerd door een coördinerend inspecteur, een stafinspecteur, een inspecteur basisonderwijs, een inspecteur speciaal onderwijs en een onderzoeker van de Universiteit Twente. Bij de uitvoering is de volgende procedure gevolgd. In elk van de scholen zijn door de Inspectie rekenen/wiskundelessen van elk ongeveer drie kwartier bezocht in de groepen 3, 5, 6 en 8. In totaal zijn 884 groepen bezocht. De inspecteur had van tevoren aangekondigd deze groepen te zullen bezoeken, zodat de onderwijsgevenden op het groepsbezoek voorbereid waren. Gedurende het groepsbezoek legde de inspecteur zijn of haar constatering en oordelen vast op een speciaal daarvoor ontworpen observatieformulier. Na afloop van elk groepsbezoek vond er een gesprek plaats met de onderwijsgevende waarin de inspecteur zijn of haar constatering toetste en indien daartoe aanleiding was naar

de achtergronden van het geobserveerde handelen vroeg. Na afloop van alle groepsbezoeken vond een gesprek met het schoolteam plaats over het thema 'effectiviteit en het rendement van het didactisch handelen'.

5. DIMENSIE VAN HET EVALUATIE-INSTRUMENT VOOR 'DIDACTISCH HANDELEN'

Het meetdoel van het te construeren evaluatie-instrument bestond in het vaststellen van de kwaliteit van het didactisch handelen bij de cognitieve onderdelen van het leer- en vormingsaanbod dat vanaf leerjaar drie in het basisonderwijs aangeboden wordt. De kwaliteit van het didactisch handelen is een dimensie die gekenmerkt wordt van de meest basale tot de meer complexe instructie- en managementvaardigheden van de onderwijsgevende.

De dimensie begint met instructie- en managementvaardigheden waarover een onderwijsgevende minimaal moet beschikken om überhaupt van lesgeven te kunnen spreken. Dit betekent dat er sprake is van helder en duidelijk gepresenteerde leerstof en dat er een duidelijke opbouw in de les zit. De les moet daarbij zodanig zijn voorbereid dat tijdens de les alles klaar ligt.

Een hoger niveau van het didactisch handelen wordt bereikt als de onderwijsgevende tijdens de uitleg van de leerstof nagaat of de leerlingen begrijpen wat er wordt uitgelegd. Belangrijk hierbij is ook dat veel leerlingen een beurt krijgen en dat de onderwijsgevende ervoor zorgt dat de leerlingen voordat ze zelfstandig aan het werk gaan, weten wat ze moeten doen. De les kan geïntensiveerd worden als de onderwijs gevende tijdens het zelfstandig werken ervoor zorgt dat de leerlingen goed en geconcentreerd doorwerken en dat iedereen doet wat hij moet doen. Het aantal ordemaatregelen zal ook laag zijn, omdat de leerlingen doorwerken.

De kwaliteit van de les kan verder geoptimaliseerd worden als de onderwijsgevende ook nog eens in staat is om de les af te stemmen op het niveau van de leerlingen.

Een echt hoog niveau in het didactisch handelen wordt bereikt als de onderwijsgevende in staat blijkt om de instructie aan te passen aan de zwakkere leerlingen en als er in de verwerking rekening gehouden wordt met de individuele verschillen in de groep.

6. INHOUDSVALIDITEIT

De constructie van de Raschschaal didactisch handelen vond plaats op basis van instrumenten die eerder door de Inspectie gebruikt zijn in evaluaties naar het didactisch handelen van onderwijsgevend in het basisonderwijs (Commissie Evaluatie Basisonderwijs, 1994; Inspectie van het onderwijs, 1994). Op basis van de indicatoren uit deze instrumenten is een grote pool met indicatoren aangelegd. Een panel van vijf inspecteurs heeft deze indicatorenpool beoordeeld op helderheid en eenduidigheid en gerangordend naar de mate waarin ze een indicator vormen voor de kwaliteit van het didactisch handelen van onderwijsgevend. Verder is een aantal formuleringen verbeterd. Met behulp van deze beoordeling is een domeinmatrijs van 67 indicatoren, die in totaal acht aspecten van het didactisch handelen van onderwijsgevend in het basisonderwijs beslaan, vervaardigd. In tabel 2 wordt een overzicht geboden.

Uit deze pool zijn de indicatoren voor het te construeren instrument geselecteerd. Aan deze selectie hebben de volgende overwegingen ten grondslag gelegen:

- * Het instrument moet gebruikt kunnen worden gedurende groepsobservaties.
- * Het instrument bevat tenminste alle 'basis' indicatoren waaraan een goede les dient te voldoen.
- * Het instrument bevat indicatoren uit tenminste elk van de aspecten van de domeinspecificatie, opdat een hoge voorspellende waarde bereikt kan worden voor elk van deze aspecten.

Tabel 2: Domeinmatrijs 'didactisch handelen'

	pool	indicatoren
helderheid instructie	7	2
directe instructie	15	7
gelegenheid bieden tot zelfstandig werken	6	6
organisatie van de les	13	5
afstemming op tempo en niveau van leerlingen	6	2
differentiatie in de instructie	4	1
differentiatie in de verwerking van de leerstof	3	1
evaluatie, registratie en feedback	13	0
totaal aantal indicatoren	67	24

- * Het instrument bevat geen indicatoren, waaraan onderwijsgeevenden redelijkerwijs (nog) niet kunnen voldoen.
- * Het instrument moet relatief snel afneembaar zijn en mag daarom in de definitieve versie niet veel meer dan ongeveer 20 indicatoren bevatten.

Van de indicatoren uit de domeinspecificatie zijn er slechts 54 vatbaar voor directe observatie tijdens schoolbezoek door inspecteurs. De overige 13, die overigens alle het aspect 'evaluatie, registratie en feedback aan leerlingen' betreffen, zijn indicatoren, waarvoor een gesprek met onderwijsgeevenden noodzakelijk is. Aangezien het de bedoeling is dat het nieuw te ontwerpen instrument een observatie-instrument wordt, is het aspect 'evaluatie, registratie en feedback' buiten beschouwing gebleven. De overige 54 indicatoren zijn benut voor verdere selectie. Hieruit zijn 24 indicatoren geselecteerd, waaruit het evaluatie-instrument 'didactisch handelen' is samengesteld. Tijdens de statistische analyses voor modelpassing bleek één indicator te 'sneuelen', zodat het instrument uiteindelijk uit 23 indicatoren bestaat.

7. CALIBRATIE VAN INDICATOREN

Elke indicator is gecalibreerd. Dat wil zeggen: Voor elke indicator uit de schaal is met behulp van het Raschmodel de plaats (de δ -waarde) op de dimensie 'didactisch handelen' berekend. De δ -waarden van de dimensie worden uitgedrukt in een logaritmische schaal, die in theorie kan uiteenlopen van $-\infty$ tot $+\infty$. De gemiddelde waarde van de δ van alle indicatoren wordt vastgezet op 0. Voor het calibreren van indicatoren met behulp van het Raschmodel kunnen diverse rekentechnieken gebruikt worden (cf. Fischer, 1974; Wright & Stone, 1979; Gustafsson, 1979; Verhelst, 1993). Hier is gekozen voor een techniek waarbij de vergelijkingen, die voor de calibratie van de indicatoren nodig zijn, worden opgelost met de 'simple method' waarbij de symmetrische functies worden berekend met de combinatiemethode (cf. Gustafsson, 1979). Deze techniek wordt gebruikt in de computerprogramma's van Fischer (1974), Gustafsson (1979) en Verhelst (1993). In tabel 3 worden de δ -waarden voor de indicatoren met hun bijbehorende stan daard error (s.e.) gepresenteerd.

In tabel 3 kan afgelezen worden dat de indicatorparameters (de δ -waarden) redelijk gespreid zijn over de dimensie. Ze lopen ze van -2.45 tot en met 4.39. Daarmee is de maatverdeling, waarmee het didactisch handelen gemeten kan worden, gereed. Wel moet geconstateerd worden dat de δ -waarde van een aantal indicatoren erg dicht bij elkaar ligt. Dit betekent, dat voor een efficiënte evaluatie met minder indicatoren kan worden volstaan. Hier staat tegenover, dat het verwijderen van indicatoren ook gevolgen zal kunnen hebben

voor de validiteit van het instrument. Daarom is vooralsnog besloten geen indicatoren te verwijderen.

Tabel 3: Eigenschappen van de indicatoren

indicatoren	log δ	s.e.	R_{it}	C
Leerlingen weten welke taken gemaakt moeten worden	-2.45	.37	.37	.96
Benodigde hulpmaterialen liggen op een goed bereikbare plaats	-1.49	.25	.65	.91
De onderwijsgevende behandelt leerstof die aansluit op het doel van de les	-1.49	.25	.43	.91
Leerlingen haken af door een te hoog lestempo*	-.83	.20	.51	.94
De les vertoont een duidelijke structuur	-.79	.20	.82	.86
De onderwijsgevende geeft uitleg in kleine opvolgende stappen	-.68	.19	.73	.91
De uitleg is helder	-.64	.19	.95	.87
Uit reacties van de leerlingen blijkt dat iedereen weet wat hij of zij moet doen	-.61	.18	.61	.89
De onderwijsgevende zorgt dat de leerlingen tijdens de verwerking taakgericht bezig zijn	-.51	.18	.89	.88
Leerlingen komen tot veel goede antwoorden op vragen van de onderwijsgevende tijdens de instructie	-.41	.17	.70	.83
Er is sprake van veel ordeverstoring in de klas*	-.35	.17	.63	.91
De onderwijsgevende zorgt ervoor dat leerlingen tijdens de verwerking van leerstof geconcentreerd zijn	-.35	.17	.87	.91
Het aantal ordemaatregelen tijdens de les blijft beneden de drie	-.21	.16	.65	.92
Er wordt voldoende gecontroleerd of de leerlingen de taken maken die ze moeten maken	-.21	.16	.76	.89
De onderwijsgevende geeft veel leerlingen een beurt	-.11	.16	.74	.86
De leerlingen werken goed door	.01	.15	.78	.90
De les is afgestemd op het niveau van de leerlingen	.26	.15	.63	.78
De les wordt vaak onderbroken*	.55	.14	.63	.89
Er gaat veel tijd verloren met wachten op individuele uitleg*.	.77	.13	.45	.89
Er is veel interactie tussen onderwijsgevende en leerlingen (er is dialoog tussen onderwijsgevende en leerling)	1.21	.12	.59	.77
De onderwijsgevende houdt in de verwerking rekening met de individuele verschillen in de groep	1.59	.12	.55	.83
De instructie wordt aangepast voor langzame leerlingen	2.36	.12	.61	.84
Aan het einde van de instructie geeft de onderwijsgevende een samenvatting van de lesstof	4.39	.16	.48	.82
<i>KR-20</i>			.76	
<i>gemiddelde consensus</i>				.88
<i>antwoordmodaliteiten: 0 = nee, 1 = ja</i>				
<i>schaalwaarden log θ</i>				
0 = -4.50 1 = -3.67 2 = -2.88 3 = -2.37 4 = -1.99 5 = -1.67 6 = -1.39 7 = -1.13 8 = -0.89 9 = -0.67				
10 = -0.45 11 = -0.23 12 = -0.01 13 = 0.21 14 = 0.44 15 = 0.69 16 = 0.95 17 = 1.24 18 = 1.57				
19 = 1.97 20 = 2.46 21 = 3.13 22 = 4.22 23 = 4.50				
*) indicatoren zijn omgescoord				

8. HOMOGENITEIT

Nagegaan is of het instrument voldoet aan de eisen die in de klassieke testtheorie aan een homogene schaal gesteld worden. In tabel 3 kan geconstateerd worden dat de KR-20 met .76 ruim boven de in veel onderzoek gehanteerde grens van .70 ligt. Ook de gemiddelde correlatie van de indicatoren met de somscore op de schaal (R_{it}) voldoet met ruim .65, royaal aan de eisen die in de klassieke testtheorie gesteld worden.

9. INTERBEOORDELAARSBETROUWBAARHEID

Het is belangrijk dat inspecteurs in gelijke situaties tot dezelfde constatering komen en dezelfde oordelen treffen. Om de consensus in constatering en oordelen te waarborgen vindt er bij ingebruikname van nieuwe instrumenten doorgaans scholing van de inspecteurs plaats. Dat is ook in dit project gebeurd. Tevens zijn er ten behoeve van de kwaliteitsborging zogenoemde duobezoeken uitgevoerd. Twee inspecteurs gaan samen op bezoek in dezelfde scholen en groepen en komen onafhankelijk van elkaar tot constatering en oordelen. Voor dit project hebben in totaal 37 duobezoeken plaatsgevonden waarbij in totaal 70 inspecteurs betrokken waren. Elk paar heeft per school, voor zover mogelijk, vier verschillende groepen geobserveerd. Het totaal aantal leerlinggroepen dat in duo's is bezocht, bedraagt 144. Op basis van deze gegevens zijn consensuscoëfficiënten per indicator berekend uitgedrukt in percentages overeenstemming⁶ tussen de paren inspecteurs. Uit tabel 3 blijkt dat er sprake is van goede consensus binnen de paren inspecteurs. De consensus (C) varieert van 77% tot 96%, terwijl de gemiddelde consensus over deze indicatoren 88% is. Dit geeft aan dat inspecteurs onafhankelijk van elkaar in hoge mate tot dezelfde constatering en oordelen komen.

10. DE FIT VAN HET RASCHMODEL

Voor de constructie van het instrumentarium is gebruik gemaakt van het Raschmodel. Daarom moet de verzameling indicatoren aan de hiervoor geldende assumpties van unidimensionaliteit en parallelle indicator karakteristieke curves voldoen.

De unidimensionaliteit van de schaal is in twee stappen nagegaan. Ten eerste moeten de indicatoren van deze schaal één onderliggende factor gemeenschappelijk hebben. Met behulp van het statistisch pakket LISREL is een analyse met één-factor model uitgevoerd op basis van een matrix met tetrachorische correlaties⁷. De factorladingen bleken uiteen te lopen van .23 tot .80 en de χ^2 bedroeg 176.98 bij $df = 230$, de bijbehorende p-waarde bedraagt .99. Dit geeft aan dat het model een goede fit heeft. Hieruit kan geconcludeerd worden dat er sprake is van een éénfactor model. Dit is een eerste aanwijzing voor de unidimensionaliteit van de schaal.

Een tweede belangrijk aspect van unidimensionaliteit is dat de maatverdeling zoals die door de δ -waarde van de indicatoren aangegeven wordt, niet beïnvloed wordt door andere zaken dan de eigenschap die men met de schaal beoogt te meten. In dit onderzoek is nagegaan of de maatverdeling van de dimensie 'didactisch handelen' beïnvloed wordt door kenmerken van de school of de groep. Voor elk van deze kenmerken zijn groepen respondenten onderscheiden. Vervolgens zijn voor elke groep respondenten de δ -waarden berekend. Deze indicatorparameters zijn vervolgens met behulp van de Andersentoets (een χ^2 -toets) getoetst op gelijkheid. De resultaten van deze berekeningen: de gevonden χ^2 -en de daarbij behorende vrijheidsgraden (df) met hun overschrijdingskans (p) staan in tabel 4. In deze tabel kan geconstateerd worden, dat op één na alle χ^2 -en, die ten behoeve van de unidimensionaliteit van de schaal berekend zijn, een overschrijdingskans hebben die groter is dan .01. Hieruit kan geconcludeerd worden, dat de unidimensionaliteit van de schaal niet in gevaar gebracht wordt door de onderzochte kenmerken van de school en groep.

Er is één uitzondering: De schaal blijkt gevoelig te reageren bij de beoordeling van het didactisch handelen van onderwijsgevendenden die in combinatiegroepen werken. Het is belangrijk om te onderzoeken of deze schending van de unidimensionaliteit van de schaal veroorzaakt wordt door slechts enkele indicatoren of dat de schaal als zodanig gevoelig is voor het gebruik in combinatiegroepen. Met behulp van de Andersentoets is daarom de bijdrage van elke indicator aan de χ^2 nader onderzocht. Geconstateerd kon worden dat zes moeilijkste indicatoren van de dimensie de schendingen veroorzaken. Het gaat hierbij om controle en orde, niveau-afstemming en differentiatie in de instructie, tijdverlies doordat leerlingen moeten wachten op individuele uitleg en de hoeveelheid interactie tussen onderwijsgevende en leerlingen. Op deze indicatoren is een vergelijking met dit instrument tussen enkelvoudige groepen en combinatiegroepen dus niet gerechtvaardigd bij de evaluatie van de kwaliteit van het didactisch handelen van onderwijsgevendenden. In combinatiegroepen heeft de onderwijsgevende immers niet altijd evenveel mogelijkheden om te controleren of de leerlingen de taken maken die ze moeten maken, als in de enkelvoudige groep. Immers als de onderwijsgevende de ene groep aan het werk heeft gezet moet hij of zij instructie geven aan de andere groep van de combinatie. Doordat de onderwijsgevende een combinatiegroep heeft, is de interactie tussen onderwijsgevende en leerling relatief minder dan in een enkelvoudige groep. De schaal kan wel gebruikt worden voor het beantwoorden van de vraag of het didactisch handelen in de combinatiegroep voldoet aan de basale eisen van goed onderwijs. Deze eisen maken immers deel uit van de indicatoren, die aan het begin van de dimensie staan.

De tweede eis die aan die aan een Raschschaal gesteld moet worden is de eis van de paralleliteit van de indicator karakteristieke curves (ICC's). Dat wil zeggen dat de indicatoren uit de schaal een gelijk discriminerend vermogen moeten hebben. Met andere woorden de maatverdeling op de dimensie moet gelijk blijven voor zowel zeer goede onderwijsgevendenden, als voor zwakke onderwijsgevendenden. Goede onderwijsgevendenden mogen met uitzondering van de invloed die hun grotere vaardigheid met zich mee brengt, geen grotere kans hebben om positief te scoren op de indicatoren dan zwakke onderwijsgevendenden. Wanneer niet aan deze eis voldaan is, fluctueert de plaats van de indicatoren op de schaalverdeling (de δ -waarde) afhankelijk van de kwaliteit van de aangetroffen eigenschap. Of de schaal hiervoor gevoelig is kan nagegaan worden met behulp van de Martin-Löftoets en een toepassingsvariant van de eerder genoemde Andersentoets. Beide toetsen zijn χ^2 -toetsen. Uit tabel 4 blijkt dat de χ^2 -en van de Martin-Löftoets en de Andersentoets hoge overschrijdingskansen hebben. Daarmee is de paralleliteit van de indicator karakteristieke curves aangetoond.

Tabel 4: Modelfit van de schaal

	χ^2	df	p
UNIDIMENSIONALITEIT VAN DE SCHAAL			
<i>Kenmerken van de school</i>			
schoolgrootte: ≤ 150 lln; 150-250 lln; ≥ 250 lln	74.01	44	.01
onderwijsvoorrranggebied: $\leq 25\%$ 1.90-lln; $> 25\%$ 1.90-lln.	31.79	22	.08
leerl. bevolk.: $\leq 25\%$ 1.00-lln; 25%-50% 1.00-lln; $\geq 50\%$ 1.00-lln.	43.89	44	.48
<i>Kenmerken van de groep</i>			
groep: 3,5,6,8	78.97	66	.13
combinatiegroep: ja/nee	67.67	22	.00
PARALLELLITEIT VAN DE ICC'S			
Martin-Löftoets	437.64	374	.01
Andersentoets	28,80	22	.15

Geconcludeerd kan worden dat de schaal voldoet aan de onderzochte assumpties van het Raschmodel.

11. BEGRIPSVALIDITEIT

Ten aanzien van de begripsvaliditeit zijn drie zaken onderzocht. Op de eerste plaats is onderzocht in hoeverre de Raschschaal voldoende samenhang vertoont met een zevental Likerschalen, die ontworpen zijn om aan de Raschschaal verwante aspecten te meten. Op de tweede plaats is onderzocht in hoeverre de Raschschaal een voorspellende waarde heeft voor de taakgerichtheid van de leerlingen. Ten derde is onderzocht of er een significant verband bestaat tussen Raschschaal en hoeveelheid van hoeveelheid tijd besteed aan klassenmanagement. Op deze drie aspecten wordt nu achtereenvolgens ingegaan.

In het project dat ten grondslag ligt aan de analyses waarover hier gerapporteerd wordt, is de kwaliteit van het didactisch handelen van onderwijsgevenden geëvalueerd met behulp van zeven Likerschalen⁸, die elk bedoeld zijn voor het meten van een aspect van dit didactisch handelen. De theoretische samenhang tussen de Raschschaal en deze zeven oorspronkelijke instrumenten biedt de mogelijkheid om de begripsvaliditeit van de Raschschaal te onderzoeken. Als aan beide instrumenten inderdaad dezelfde begrippen ten grondslag liggen, dan mogen we een sterke samenhang verwachten tussen de Raschschaal en de zeven Likerschalen. Om deze verwachting te kunnen onderzoeken, zijn de correlaties berekend tussen de Raschschaal en de zeven oorspronkelijke schalen. Deze correlaties zijn afgedrukt in tabel 5.

Tabel 5: Correlaties tussen diverse aspecten van didactisch handelen en de Raschschaal

	R	R _c
helderheid van de instructie	.72	.66
mate van directe instructie	.70	.54
leerlingen gelegenheid bieden tot zelfstandig werken	.77	-
organisatie van de les	.67	.57
afstemming op tempo en niveau van de leerlingen	.62	.55
mate van differentiatie in de instructie	.56	.55
mate van differentiatie in de verwerking	.40	.34

* voor alle correlaties geldt $p < .00$

In tabel 5 is steeds de correlatie (R) tussen de Raschschaal en de Likerschalen vermeld. Aangezien er door partiële overlap van indicatoren gedeeltelijk sprake is van autocorrelatie, kan het zijn dat de uitkomsten geflatteerd worden. Daarom is gecorrigeerd voor autocorrelatie (R_c). Aangezien alle indicatoren van de schaal 'leerlingen gelegenheid bieden tot zelfstandig werken' deel uitmaken van de Raschschaal is het niet mogelijk om deze correlatie te corrigeren voor autocorrelatie. Uit de resultaten kan geconcludeerd worden dat de Raschschaal globaal de helft van de variantie (R²) verklaart in de kwaliteit van de instructie, de mate van directe instructie, het bieden van de gelegenheid tot zelfstandig werken en de lesorganisatie. Dit is een zeer bevredigend resultaat. De Raschschaal heeft voor deze aspecten van het didactisch handelen dan ook een uitstekende voorspellende waarde. De correlaties met de drie Likerschalen die alle gericht zijn op aspecten van adaptief onderwijs zijn lager. Het gaat hierbij om afstemming op tempo en niveau van de leerlingen en om de differentiatie in de verwerking en in de instructie. Toch verklaart de Raschschaal nog ongeveer een kwart van de variantie in deze schalen. De Raschschaal heeft voor deze aspecten een redelijke voorspellende waarde. Als men echter specifiek de toestand van het onderwijs op het gebied van adap-

tief onderwijs wil evalueren, dan is het aangewezen om een meer specifiek hierop gericht instrument te gebruiken (Zie bijvoorbeeld Houtveen, 1994).

Een tweede aspect van begripsvaliditeit wordt gevormd door de mate waarin de score op de schaal voor didactisch handelen gepaard gaat met de taakgerichtheid van de leerlingen. In dezelfde groepen waarin het observatie-instrument is afgenomen hebben de inspecteurs het percentage leerlingen vastgesteld, dat tijdens de instructie niet zat op te letten of tijdens de verwerking van de leerstof niet taakgericht bezig was⁹. Tussen deze aspecten van niet-taakgerichtheid en de observatieschaal zijn correlaties berekend. Deze correlaties tussen zijn vermeld in tabel 6.

Tabel 6: *Correlatie tussen diverse kwantitatieve maten en de Raschschaal*

aspecten:	R
% leerlingen dat oplet tijdens instructie	.40
% leerlingen dat tijdens verwerking van leerstof taakgericht is	.38
% tijd voor klassenmanagement en orde handhaving	-.26
voor alle correlaties geldt $p < .00$	

Er blijkt een duidelijk en positief verband te zijn tussen de Raschschaal en het opletten tijdens de instructie en de taakgerichtheid tijdens de verwerking van de leerstof. Met andere woorden hoe beter de kwaliteit van het didactisch handelen des te meer de leerlingen opletten tijdens de instructie en des te taakgerichter de leerlingen zijn tijdens de verwerking van de leerstof. Immers in een kwalitatief goede les zullen de leerlingen beter opletten en taakgerichter bezig zijn. De onderwijsgevende zorgt ervoor dat de leerlingen opletten en dat zij zich niet met andere dingen bezighouden.

Een derde aspect van begripsvaliditeit is de mate waarin het didactisch handelen gepaard gaat met de hoeveelheid tijd besteed aan klassenmanagement in de groep. In dezelfde groepen waarin de inspecteur het observatie-instrument heeft afgenomen heeft de inspecteur ook de hoeveelheid tijd die aan klassenmanagement is besteed, bijgehouden¹⁰. Onder klassenmanagement wordt verstaan alle activiteiten die gericht zijn op het organiseren van de onderwijssituatie en op de ordehandhaving. Uit tabel 6 blijkt dat ook hier sprake is van een significant verband tussen de kwaliteit van het didactisch handelen en het percentage tijd besteed aan klassenmanagement (-.26). Dus hoe beter de kwaliteit van het didactisch handelen des te minder tijd er aan klassenmanagement en ordehandhaving verloren gaat. Dit mag ook verwacht worden. Een onderwijsgevende die constant bezig is met klassenmanagement zal niet veel tijd over houden om les te geven.

Geconcludeerd kan worden dat dit instrument een hoge voorspellende waarde heeft voor diverse belangrijke aspecten van het onderwijsproces. Daarmee heeft het instrument een goede begripsvaliditeit.

12. PREDICTIEVE VALIDITEIT

De Raschschaal is bedoeld voor het evalueren van het didactisch handelen van onderwijsgevendenden dat van invloed is op de output van hun school. De predictieve validiteit van de schaal hangt dus af van het bestaan van een significant verband tussen de score van de onderwijsgevendenden van een school op de schaal en de output van de school.

Een grondig opgezet onderzoek naar de effectiviteit van het didactisch handelen van onderwijsgevendenden vereist een experimenteel design met voor- en nametingen van de pres-

taties de leerlingen bij een experimentele groep en een controlegroep van onderwijsgeevenden. Dit onderzoek is echter niet opgezet als een experiment.

De veldsituatie van dit onderzoek was die van een schoolbezoek door de inspecteur, waarbij per school vier onderwijsgeevenden geobserveerd werden en waarbij de inspecteur inzage had in de resultaten op de Eindtoets Basisonderwijs van de betreffende school. Wanneer we een aantal assumpties accepteren, dan kan deze veldsituatie benut worden om de predictieve validiteit van de schaal voor het meten van de kwaliteit van het didactisch handelen te exploreren. Deze assumpties zijn de volgende:

De observaties van het didactisch handelen van een viertal willekeurig geselecteerde onderwijsgeevenden biedt een stabiel en representatief beeld van de kwaliteit van het didactisch handelen op de betreffende school.

De gemiddelde prestaties op de Eindtoets Basisonderwijs, zoals die jaarlijks door het Cito gemaakt wordt, bieden mits vastgesteld over een reeks van vijf jaar en mits gecorrigeerd voor milieu en etnische afkomst van leerlingen een betrouwbare en valide indicatie voor de opbrengst van een basisschool.

Om de predictieve validiteit te kunnen bepalen is ervoor gekozen om de output van scholen te baseren op de prestatiescores van de leerlingen op de Eindtoets Basisonderwijs over vijf schooljaren, na correctie voor milieu van afkomst van de leerlingen. Voor deze correctie is gebruik gemaakt van de gewichten die aan de leerlingen worden toegekend op grond van de opleiding en etniciteit van vader en moeder.

De berekeningen die nodig zijn voor de validiteitsbepaling 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 7. Voor de analyses zijn de data getransformeerd naar z-scores, zodat gevonden effecten geïnterpreteerd kunnen worden als gestandaardiseerde regressie-coëfficiënten (met tussen haakjes de bijbehorende standaard error).

Tabel 7. Verklaring voor verschillen in de output van basisscholen

	model 0	model 1	model 2
gestandaardiseerde regressiecoëfficiënten (standaard error)			
schoolomgeving		-.43 (.06)	-.42 (.06)
didactisch handelen			.16 (.07)
	% variantie	% onverklaarde variantie	t.o.v. model 0
tussen scholen variantie	62.94	42.99	40.56
binnen scholen variantie	37.06	36.56	36.58
deviantie	1147.75	1106.33	1101.21
aantal vrijheidsgraden (df)		1	1
verbetering van model (p)		<.001	<.025

Aan de hand van de resultaten in model 0 kan geconstateerd worden, dat de intra-klasse-correlatiecoëfficiënt $\rho = .63$ 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 stabiele verschillen in de gemiddelde Cito-eindtoetsresultaten van basisscholen.

De relatief kleine basisscholen in Nederland 'recruter' 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 leerlinggewicht in het betreffende schooljaar. Wanneer dit gemiddelde leerlinggewicht voor de vijf afzonderlijke schooljaren ingevoerd wordt in de regressievergelijking, dan blijkt uit model 1 dat 19.95% van de tussenscholenvariantie hierdoor verklaard te worden. (De binnenscholenvariantie verklaart 0.5%. Dat is te verwaarlozen.) Geconcludeerd kan worden dat de schoolomgeving relatief stabiel is.

Model 1 geeft tevens aan dat de gestandaardiseerde regressiecoëfficiënt voor de schoolomgeving significant ($\alpha = .05^{12}$) is. Het verschil in deviantie (=likelihood-ratio statistic) tussen model 0 en model 1 is 41.42, welke te vergelijken is met een χ^2 -verdeling met 1 vrijheidsgraad. Deze toets geeft aan dat dit significant is ($p < .001$) wat impliceert dat schoolomgeving als component opgenomen moet worden in het model. De schoolomgeving heeft dus een significant effect op de gemiddelde Eindtoetsresultaten van basisscholen over de jaren heen. Daarom moet voor dit effect gecorrigeerd worden voordat de variabele didactisch handelen wordt ingevoerd in de regressie-analyse.

Nu kan de vraag beantwoord worden of het didactisch handelen van onderwijsgeven in het basisonderwijs verband houdt met een stabiele schooloutput waarop de school ook daadwerkelijk invloed kan hebben. In model 2 kan geconstateerd worden, dat de variabele didactisch handelen een klein, maar significant effect heeft op de voor schoolomgeving gecorrigeerde stabiele output van basisscholen. Bijna 2.5% van de tussenscholenvariantie kan hiermee verklaard worden. Het verschil in deviantie tussen model 1 en model 2 is 5.12 met $df=1$. Deze blijkt significant te zijn met $p < .025$. Het model wordt door opname van de component didactisch handelen verbeterd. Hiermee is de predictieve validiteit van het instrument onderstrept.

13. DE KWALITEIT VAN HET DIDACTISCH HANDELEN IN HET BASISONDERWIJS

De Raschschaal voor het evalueren van de kwaliteit van het didactisch handelen is in 1997 opnieuw afgenomen bij de onderwijsgeven in de groepen 3, 5, 6 en 8 van een representatieve steekproef van 131 basisscholen. Dat biedt ons de mogelijkheid om een vergelijking te treffen tussen de situatie in 1994 en in 1997. In tabel 8 wordt een overzicht geboden van het gemiddelde en de spreiding (standaard deviatie) van de scores in groep 3, 5, 6 en 8 van het basisonderwijs in deze beide jaren.

Het blijkt dat er in feite geen belangrijke verschillen zijn tussen de onderwijsgeven in de verschillende jaargroepen. Evenmin zijn er verschillen tussen 1994 en 1997. Dit komt ook tot uitdrukking in afbeelding 1.

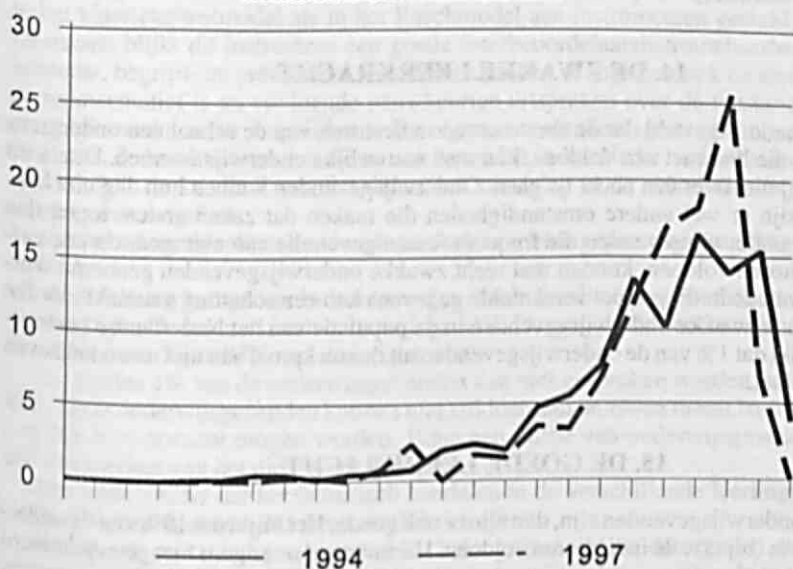
Op de x-as van afbeelding 1 is de frequentieverdeling van de vaardigheid in didactisch handelen in 1994 en in 1997 weergegeven. Deze vaardigheid loopt van -4.50 tot en met 4.50. Op de y-as is het percentage onderwijsgeven weergegeven. Hoewel de gemiddelde scores niet verschillen, blijkt wel dat de spreiding van de scores op de schaal in 1997 is afgenomen ten opzichte van 1994. In 1997 liggen de scores van de onderwijsgeven dichter bij het gemiddelde dan in 1994.

Tabel 8: Kwaliteit van het didactisch handelen in verschillende groepen van het basisonderwijs in 1994 en in 1997

	groep 3		groep 5		groep 6		groep 8	
	'94	'97	'94	'97	'94	'97	'94	'97
gemiddelde ruwe score (standaard deviatie)	18.7 (3.3)	19.0 (2.4)	18.7 (3.0)	19.0 (2.6)	18.7 (3.1)	18.5 (3.0)	18.9 (3.2)	19.1 (2.3)
gemiddelde vaardigheid (standaard deviatie)	2.3 (1.4)	2.2 (1.0)	2.3 (1.4)	2.2 (1.0)	2.3 (1.3)	2.1 (1.0)	2.4 (1.4)	2.2 (1.0)

Afbeelding 1

didactisch handelen



De gemiddelde onderwijsgevende voldoet in welke groep hij of zij ook werkt aan 19 van de 23 indicatoren.

Het bijzondere van een Raschschaal is dat niet alleen aangegeven kan worden aan hoeveel indicatoren de gemiddelde onderwijsgevende voldoet. Er kan binnen een zeker betrouwbaarheidsinterval ook aangegeven worden aan welke indicatoren die gemiddelde onderwijsgevende voldoet. Daarmee kan het didactisch handelen niet alleen kwantitatief (met een score), maar ook inhoudelijk geduid worden. Wanneer we weten dat de gemiddelde onderwijsgevende aan 19 van de indicatoren voldoet, dan kunnen we dit gedrag als volgt typeren:

Het didactisch handelen van de gemiddelde onderwijsgevende blijkt te voldoen aan de basale kenmerken waaraan een goede les moet voldoen: De leerlingen weten wat er van hun verwacht wordt, leerstof en lesdoel sluiten goed op elkaar aan, de leer- en hulpmaterialen liggen op goed bereikbare plaatsen, het lestempo is adequaat, de les is gestructureerd en de uitleg vindt plaats in kleine stappen en is helder.

Ook de intensiteit van de les en de taakgerichtheid van de leerlingen is op orde bij de gemiddelde leerkracht. De leerlingen krijgen tijdens de instructie de kans om tot veel goede ant-

woorden te komen en veel leerlingen krijgen een beurt. De leerlingen weten tijdens de verwerking van de leerstof wat ze moeten doen, de onderwijsgevende controleert in voldoende mate of de leerlingen de taken maken die ze moeten maken en de leerlingen zijn taakgericht en geconcentreerd bezig en werken goed door. Orde vormt geen probleem en lesonderbrekingen blijven achterwege. Er is veel op de inhoud van de les gerichte interactie tussen onderwijsgevende en leerlingen en er gaat maar weinig tijd verloren met bijvoorbeeld het wachten op individuele uitleg. De les is tevens afgestemd op het niveau van de leerlingen.

Bij de gemiddelde onderwijsgevende is er echter maar een geringe hoeveelheid interactie tussen hem of haar en de leerlingen. Er is te weinig dialoog tussen onderwijsgevende en leerling. Ook houdt de gemiddelde onderwijsgevende in de verwerking onvoldoende rekening met de individuele verschillen in de groep. De instructie wordt in onvoldoende mate aangepast voor langzame leerlingen. De meeste onderwijsgevers bieden aan het einde van de instructie geen samenvatting van de leerstof. Dit beeld is overigens geheel vergelijkbaar met het beeld dat naar voren kwam in de evaluatie van het basisonderwijs (Commissie Evaluatie Basisonderwijs, 1994abc).

14. DE ZWAKKE LEERKRACHT

In het voorafgaande is gesteld dat de eerste zeven indicatoren van de schaal een ondergrens vormen. Lessen die hier niet aan voldoen, kan men nauwelijks onderwijs noemen. Uiteraard moet men hierbij niet over één nacht ijs gaan. Onderwijsgevers kunnen hun dag niet hebben en anders zijn er wel andere omstandigheden die maken dat zaken anders lopen dan bedoeld. Echter onderwijsgevers die frequent lessen geven die aan niet meer dan de eerste zeven indicatoren voldoen, kunnen met recht zwakke onderwijsgevers genoemd worden. Op grond van de in dit project verzamelde gegevens kan een schatting gemaakt worden van het percentage zwakke onderwijsgevers in de populatie van het Nederlandse basisonderwijs. Het blijkt dat 1% van de onderwijsgevers uit de steekproef aan niet meer dan zeven indicatoren voldoet.

15. DE GOEDE LEERKRACHT

Als er zwakke onderwijsgevers zijn, dan zijn er ook goede. Het blijkt dat 19% van de onderwijsgevers aan (bijna) alle indicatoren voldoen. Hieruit moet overigens niet geconcludeerd worden dat deze onderwijsgevers 'onverbeterlijk' zijn. Uit diverse publikaties (Inspectie, 1994; Inspectie, 1995; Houtveen, 1995) blijkt dat de Achilleshiel van de onderwijsgevers in het basisonderwijs bestaat in het leveren van 'adaptief onderwijs'. Dit maakt, zoals eerder betoogd, maar heel partieel deel uit van het observatie-instrument. In adaptief onderwijs gaat het om het bieden van onderwijs, waarin de gebruikte onderwijsmethoden op het gebied van lezen, taal en rekenen voldoende specifiek zijn toegesneden op alle leerlingen, ook de leerlingen in risicosituaties. Daarbij is het belangrijk dat onderwijsgevers werken met een pedagogisch didactische aanpak die goed aansluit bij alle leerlingen, ook de leerlingen die een verhoogde kans lopen naar het speciaal onderwijs verwezen te worden. Uit recent Nederlands onderzoek (zie bijvoorbeeld: Houtveen & Osinga, 1994) is duidelijk geworden dat een pedagogisch didactische aanpak waarin diverse ingrediënten gecombineerd worden, beter helpt bij de zwakke leerling. Het gaat hierbij om:

- * het geven van 'directe instructie',
 - * het werken aan een zo groot mogelijke taakbetrokkenheid van de leerlingen bij de les en
 - * om het realiseren van de cyclus van adaptief onderwijs.
- (Zie ook het rapport 'Onderwijs op maat' van de Commissie Evaluatie Basisonderwijs.) De cyclus van adaptief onderwijs bestaat uit:

- * op groepsniveau de resultaten van de leerlingen, vooral bij lezen, taal en rekenen, regelmatig toetsen en
- * de resultaten ervan registreren,
- * diagnosticeren wat er met de leerling(en) die dreigen uit te vallen aan de hand is,
- * handelingsplannen maken voor individuele leerlingen of groepjes leerlingen die dreigen uit te vallen,
- * deze plannen uitvoeren en met behulp van toetsen weer evalueren of de doelen van de handelingsplannen bereikt zijn.

Zo niet dan weer opnieuw diagnosticeren, handelingsplannen maken en uitvoeren, enzovoorts.

16. CONCLUSIE

Het ontwikkelde observatie-instrument voor het evalueren van het didactisch handelen van onderwijsgeevenden in basisonderwijs bleek te voldoen aan de psychometrische eisen die zowel in het klassieke testmodel als in het Raschmodel aan instrumenten gesteld kunnen worden. Daarnaast blijkt dit instrument een goede interbeoordelaarsbetrouwbaarheid en een goede inhouds-, begrips- en predictieve validiteit te hebben. Aangezien ook de steekproef voldoende representatief is en voldoende nauwkeurige uitspraken over de toestand in de populatie mogelijk maakt, is het mogelijk om de volgende evaluatievragen te beantwoorden.

- 1) Wat is de kwaliteit van het didactisch handelen in het Nederlandse basisonderwijs?
- 2) Verschilt de kwaliteit van het didactisch handelen in de verschillende groepen van het basisonderwijs? en
- 3) Gaat een hoge kwaliteit van het didactisch handelen gepaard met een hoge leerlingprestaties?

Uit de resultaten kan geconcludeerd worden dat de kwaliteit van het didactisch onderwijsgeevenden in het Nederlands basisonderwijs goed is. De gemiddelde onderwijsgevende is in staat te voldoen aan bijna 83% van de indicatoren.

Bij slechts 1% van de onderwijsgeevenden kan niet gesproken worden van onderwijs in de klas. Deze onderwijsgeevenden kunnen niet voldoen aan de zeven meest basale kenmerken die van een les verwacht mogen worden. Bijna een vijfde van onderwijsgeevenden voldoet aan alle kenmerken van het didactisch handelen.

Het bleek verder dat het didactisch handelen in de verschillende leerlinggroepen van het basisonderwijs niet van elkaar verschilt in kwaliteit.

Ten slotte, en dat is misschien wel het belangrijkste, bleek dat scholen die jaar in jaar uit hoge leerlingprestaties hebben ook na correctie voor de omgeving waaruit de leerlingen 'gerecruiteerd' worden een hoge kwaliteit van het didactisch handelen laten zien.

LITERATUUR

- Anderson, L.W. (1984). Instruction and time-on-task: a review. In L.W. Anderson (Ed.), *Time and school learning: theory, research and practice*, (pp.143-163). London: Croon Helm.
- Batenburg, Th.A. van (1990). Variatie in schoolgemiddelden op de CITO-eindtoets basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 15, 362-369.
- Block, A. de (1965) *Algemene didactiek. Vooral in verband met de vorming van leerlingen van 10 tot 18 jaar*. Antwerpen: Standaard-Boekhandel.
- 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. *Tijdschrift voor Onderwijsresearch*, 18, 331-342.
- Bloom, B.S., M.D. Engelhart, E.J. Furst, W.H. Hill & D.R. Krathwohl, (1956). *Taxonomy of educational objectives: the classification of educational goals*. New York: Longmans Green.
- Blok, H. & M.H. Eiting (1988). De grootte van schooleffecten: hoe verschillend presteren leerlingen van verschillende scholen. *Tijdschrift voor Onderwijsresearch*, 13, 16-30.

- Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16, 208-216.
- Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs*. Groningen: RION.
- Brandsma, H.P. & J.W.M. Knuver (1989). *Basisschoolkenmerken als determinanten van het functioneren van leerlingen*. Groningen: RION.
- Bryk, A.S. & S.W. Raudenbush (1992). *Hierarchical linear models*. London: Sage.
- Carnine, D. (1989). Direct instruction model: special education. In T. Husen & T.N. Postlethwaite (eds.), *The International Encyclopedia of Education: Research and Studies (Supplementary Volume One)*, (pp. 251-257). Oxford: Pergamon Press.
- 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.
- Commissie Evaluatie Basisonderwijs (1994c). *Onderwijs-op-maat*. De Meern: Inspectie van het onderwijs.
- Creemers, B.P.M. (1991). *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas*. 's-Gravenhage: SVO.
- Cronbach, L.J. (1963). Course improvement through evaluation. In: *Teacher College Record*, vol 64, nr.8.
- 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.
- Denham, C. & Lieberman, A. (Eds.) (1980) *Time to learn: A review of the beginning teacher evaluation study*. Washington DC: National Institute of Education.
- Gagné, R.M. (1965). *The conditions of learning*. New York: Holt, Rinehart & Winston.
- Gelder, L. van, J.J. Peters, Th. Oudkerk Pool & J. Sixma (1972). *Didactische analyse*. Groningen: Wolters-Noordhoff.
- Grift, W. van de (1994). *Technisch rapport van het onderzoek onder 386 basis scholen ten behoeve van de evaluatie van het basisonderwijs*. De Meern: Inspectie van het Onderwijs.
- Gustafsson, J.-E. (1977). *The Raschmodel for dichotomous items: theory, applications and a computer program*. Göteborg: Institute of education, University of Göteborg.
- Gustafsson, J.-E. (1979). *Testing and obtaining fit of data to the Raschmodel*. Göteborg: Institute of education, University of Göteborg.
- Gustafsson, J.-E. (1979). *PML, a computer program for conditional estimation and testing in the Raschmodel for dichotomous items*. Göteborg: Institute of education, University of Göteborg.
- Houtveen, A.A.M. (1994). *Onderwijs op maat*. Utrecht: ISOR.
- Houtveen, A.A.M. & Osinga, N. (1994). *A case of school effectiveness: program and evaluation results of the Dutch National School Improvement Project*. Leeuwarden: GCO; Utrecht: ISOR.
- Langeveld, M.J. (1959). Algemene en speciale didactiek en het leren in verband der didactiek beschouwd. In: *Pedagogische Studiën* (36), 339-351.
- Langeveld, M.J. (1959). Het leerproces in het verband van de didactiek, I en II. In: *Pedagogische Studiën* (36), 385-393; 451-463.
- Mager, R.F. (1962). *Preparing instructional objectives*. Belmont, CA: Pitman Learning.
- Parreren, C.F. (1962). *Psychologie van het leren I en II*. Zeist/Arnhem: De Haan/Vanloghem Slaterus.
- Parreren, C.F. (1965). *Leren op school*. Groningen: Wolters-Noordhoff.
- Prosser, R., J. Rasbash & H. Goldstein (1991). *ML3, software for three-level analysis*. London: Institute of Education, University of London.
- Rosenshine, B.V. & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), *Third Handbook of Research on Teaching*. New York: MacMillan.
- Rosenshine, B.V. (1987). Direct instruction. In M.J. Dunkin (ed.), *The International Encyclopedia of Teaching and Teacher Education* (p 257-263). Oxford: Pergamon.
- Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*. 's-Gravenhage: SVO.
- Scriven, M. (1967). The methodology of evaluation. In: Tyler, R.W., R.M. Gagné, M. Scriven, *Perspectives of curriculum evaluation*. Chicago: Rand McNally.
- Skinner, B.F. (1954). The science of learning and the art of teaching. In: *Harvard Educational Review*. (24), 86-97.
- Skinner, B.F. (1958). Teaching machines. In: *Science*. (128), 969-977.
- Veenman, S. (1992). Effectieve instructie volgens het directe instructiemodel. *Pedagogische Studiën*, 69, 242-269.
- Veenman, S., Leenders, Y., Meyer, P. & Sanders, M. (1993). Leren lesgeven met het directe instructiemodel. *Pedagogische Studiën*, 70, 2-16.

- Veenman, S., Lem, P., Voeten, M., Winkelmolen, G. & Lassche, H. (1986). *Onderwijs in combinatieklassen*. Den Haag: SVO.
- Veenman, S., Lem, P. & Nijssen, F. (1988). *Omgaan met combinatieklassen: een programma voor schoolverbetering*. Den Haag: SVO.
- Veenman, S., Lem, P., Roelofs, E. & Nijssen, F. (1992). *Effectieve instructie en doelmatig klasmanagement: een schoolverbeteringsprogramma voor enkelvoudige en combinatieklassen*. Amsterdam/Lisse: Swets & Zeitlinger.
- Walberg, H.J. (1990). Productive teaching and instruction: assessing the knowledge base. *Phi Delta Kappan*, 470-478.
- Wright, B.D. & M.H. Stone (1979). *Best test design, Rasch measurement*. Chicago: Mesa Press.
- Verhelst, N.D. (1992). *Het eenparameter logistisch model (OPLM)*. Arnhem: Cito.
- Walberg, H.J. & G.D. Haertel (1992). Educational psychology's first century. In: *Journal of Educational Psychology*. 84 (1) 6-19.

NOTEN

1. Werkadres: Inspectie van het Onderwijs, Postbus 2725, 3500 GS Utrecht, tel. 030 6690707, fax 030 6667067, email: W.vandeGrift@owinsp.nl.
2. Werkadres: Universiteit Twente, Toegepaste Onderwijskunde, postbus 217, 7500 AE Enschede, tel. 053 4893660, fax 053 4893791, email: lam@edte.utwente.nl.
3. De formule voor steekproefnauwkeurigheid luidt: $n = N/[1 + N(e)^2]$, waarbij n = steekproefgrootte, N = populatie en e = steekproefnauwkeurigheid.
4. Een steekproefnauwkeurigheid van $e < 10\%$ is een veel gehanteerde grens om verantwoorde uitspraken over proporties in een populatie te kunnen doen.
5. De regionale spreiding is gebaseerd op de regionale indeling van 13 rijksinspectiekantoren van de Inspectie van het Onderwijs.
6. De formule die hiervoor is gebruikt luidt: $\Sigma o/\Sigma o_n$, waarbij o = overeenstemming per paar, i = indicator i , n = totaal aantal paren. Voor de consensus van de gehele schaal is de volgende formule gebruikt: $(\Sigma o/\Sigma o_n)k$, waarbij k = aantal indicatoren. De gebruikte observatieprocedure maakte het niet mogelijk om een bestaande consensuscoëfficiënt, zoals Cohens' Kappa, te gebruiken. Bij deze consensuscoëfficiënt wordt ervan uitgegaan dat de beoordeling van één bepaalde situatie door verschillende beoordelaars geschiedt. In dit project is echter elke groep eenmaal door een duo van inspecteurs bezocht.
7. Aangezien er sprake is van dichotome items zijn er tetrachorische correlaties berekend i.p.v. product-moment-correlaties.
8. Deze observatieschalen voldoen alle aan het criterium van een Cronbach's alpha van ten minste .70 en bij deze zeven observatieschalen is de interbeoordelaarsbetrouwbaarheid ten minste .80.
9. Voor beide instrumenten voor vaststellen van de taakgerichtheid van de leerlingen tijdens instructie en verwerking geldt, dat de interbeoordelaarsbetrouwbaarheid ten minste .80 bedraagt.
10. Voor dit instrument voor vaststellen van de hoeveelheid tijd die gemoeid is met klassenmanagement geldt, dat de interbeoordelaarsbetrouwbaarheid gemiddeld ten minste .80 bedraagt.
11. Afhankelijk van zijn/haar milieu van herkomst krijgt elke leerling een gewicht, dat een rol speelt bij de financiering van het onderwijs. Leerlingen van ouders uit allochtone gezinnen 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 ouders uit de groep van handarbeiders krijgen een gewicht van 1.25; en de overige leerlingen krijgen het gewicht 1.0
12. Een regressiecoëfficiënt is significant voor $\alpha = .05$ als deze 1.96 keer zo groot is als zijn standaard error.

Vaardig oplossen van contextgebonden wiskunde-problemen in de bovenbouw van de basisschool

L. Verschaffel, E. De Corte, S. Lasure, G. Van Vaerenbergh, H. Bogaerts
en E. Ratinckx¹

Centrum voor Instructiepsychologie en -Technologie (CIP&T), Katholieke Universiteit Leuven*

SAMENVATTING

Recent internationaal onderzoek heeft aangetoond dat veel leerlingen op het einde van de basisschool de verschillende aspecten die nodig zijn om wiskundige toepassingsproblemen op een efficiënte en succesvolle manier op te lossen onvoldoende beheersen. In dit onderzoek werd een 'krachtige leeromgeving' voor het leren oplossen van wiskundige toepassingsproblemen bij leerlingen van de bovenbouw van het basisonderwijs ontwikkeld en vervolgens in vier vijfde leerjaren (= groep 7) uitgeprobeerd. Daartoe werd aan de leerlingen uit de experimentele klassen een algemene strategie voor het vaardig oplossen van wiskundige toepassingsproblemen aangeleerd, met bijzondere aandacht voor een aantal waardevolle heuristische en metacognitieve strategieën. Intussen volgden de leerlingen van zeven controleklassen gewoon wiskundeonderwijs. Het effect van de leeromgeving werd geëvalueerd in een onderzoek met een voortoets-natoets-retentietoetsopzet. De resultaten leveren het bewijs dat het mogelijk is om via aangepast onderwijs een betekenisvolle invloed uit te oefenen op de ontwikkeling van de (wiskundige) probleemoplossingsvaardigheid van basisschoolleerlingen.

ABSTRACT

Recent research has shown that many upper elementary school children do not master the skill in solving mathematical application problems. In this design experiment a 'powerful learning environment' for teaching and learning modeling and problem-solving skills in upper elementary school mathematics was developed and tested in four fifth-grade classes. Pupils were taught a series of heuristic and cognitive strategies embedded in an overall metacognitive strategy for solving mathematical application problems. Meanwhile the pupils of seven control classes followed regular mathematics classes. The implementation and effectiveness of the experimental learning environment were tested in a study with a pretest-posttest-retention test design. The results indicate that the intervention had a positive effect on different aspects of pupils' mathematical modeling and problem-solving ability.

1. INLEIDING EN ACHTERGROND

In de recente vernieuwingsinitiatieven en -documenten met betrekking tot het wiskundeonderwijs, zoals het Concept Herziene kerndoelen basisonderwijs (SLO, 1996) in Nederland

* In deze bijdrage wordt verslag uitgebracht over het Onderwijskundig Beleids- en Praktijkgericht Wetenschappelijk Onderzoek nr. 94.07 in opdracht van het Ministerie van de Vlaamse Gemeenschap, Departement Onderwijs.

en de Eindtermen en Ontwikkelingsdoelen Basisonderwijs (1997) in Vlaanderen, staat het ontwikkelen van probleemoplossingsvaardigheden en attitudes en het toepassen van de verworven wiskundige kennis en vaardigheden in levensechte probleemsituaties centraal. Tal van onderzoeksgegevens hebben echter aangetoond dat veel leerlingen op het einde van de basisschool de verschillende aspecten die nodig zijn voor het ontwikkelen van een dispositie om wiskundige toepassingsproblemen op een efficiënte en succesvolle manier op te lossen onvoldoende beheersen (zie o.m. De Corte, Greer & Verschaffel, 1996; Lester, Garofalo & Kroll, 1989; Schoenfeld, 1992; Verschaffel, ter perse). Eerst en vooral blijken er tekorten te bestaan op vakinhoudelijk vlak; die hebben betrekking op allerlei kenniselementen (begrippen, formules, algoritmen, oplossingsschema's, ...) die moeten gebruikt worden om een probleem op te lossen. Daarnaast zijn er tekorten die eerder strategisch van aard zijn. Zo blijkt dat veel leerlingen vraagstukken op een oppervlakkige en ondoordachte manier aanpakken, omdat ze waardevolle (meta)cognitieve strategieën ontberen voor het aanpakken van échte problemen of de strategieën waarover ze wel beschikken niet spontaan toepassen. Verder vertonen heel wat leerlingen een sterke neiging om tijdens het modelleren en interpreteren van schoolvraagstukken hun ervaringskennis over de situatie die in de opgave ter sprake komt, te negeren. Ten slotte blijken nogal wat leerlingen er allerhande inadequate, negatieve of foutieve denkbeelden op na te houden over het leergebied wiskunde en over het leren en onderwijzen van vraagstukken. Deze onjuiste overtuigingen oefenen een negatieve invloed uit op de wijze waarop deze leerlingen wiskundige problemen aanpakken, op het soort van activiteiten dat zij kiezen wanneer ze met een probleem geconfronteerd worden en op de manier waarop zij hun eigen succes of mislukking evalueren. Hoewel de hiervoor gerapporteerde eerder negatieve resultaten deels gebaseerd zijn op de internationale onderzoeksliteratuur over het (leren) oplossen van wiskundige toepassingsproblemen, zijn er aanwijzingen dat het beeld dat uit dit overzicht van het internationaal onderzoek naar voren komt in grote lijnen ook opgaat voor de Vlaamse en Nederlandse situatie (zie bijv. Kuiper, Bos & Plomp, 1997; Verschaffel & De Corte, 1996).

De oorzaken van deze nogal teleurstellende resultaten van het vraagstukkenonderwijs op het vlak van strategische kennis en vaardigheden worden door de meeste auteurs vooral gezocht in de volgende drie kenmerken van het vigerend vraagstukkenonderwijs (voor meer informatie zie: De Corte e.a., 1996; Greer, 1997; Reusser & Stebler, 1997; Schoenfeld, 1992; Treffers, De Moor & Feijs, 1989; Verschaffel, ter perse).

- (1) De aard van de opgaven waarmee in de wiskundelessen op de basisschool wordt gewerkt. Leerlingen worden vooral geconfronteerd met zogenaamde standaardopgaven, waarin de relatie tussen de probleemsituatie en de vereiste bewerking(en) duidelijk is. Problematische opgaven die hen uitnodigen om zich te bezinnen over de ingewikkelde relatie tussen wiskunde en realiteit, en over de problemen die er kunnen rijzen bij het gebruik van wiskunde om reële probleemsituaties op te lossen, zijn eerder zeldzaam.
- (2) De manier waarop door de leerkracht en de leerlingen tijdens de vraagstukkenlessen met deze opgaven wordt omgesprongen. Vooral het individueel en schriftelijk (leren) oplossen van vraagstukken aan de hand van opgelegde, vaste standaardoplossingsschema's en -procedures wordt sterk beklemtoond. Daarbij wordt onvoldoende gebruik gemaakt van krachtige, interactieve instructietechnieken gericht op de ontwikkeling van waardevolle heuristische en metacognitieve vaardigheden (zoals modelleren, scaffolding, coaching, articuleren, reflecteren...) waarvan uit onderzoek gebleken is dat zij van cruciaal belang zijn voor vaardig wiskundig probleemoplossen.
- (3) De heersende klascultuur. Recent onderzoek, o.a. uit socio-constructivistische hoek (Cobb, Yackel & Wood, 1992; Gravemeijer, 1994), heeft aangetoond hoe allerlei subtiele en moeilijk zichtbare aspecten van het dagelijkse ritueel van het wiskundeonderwijs eveneens bijdragen tot de ontwikkeling van de eerder beschreven oppervlakkige aanpakstrategieën, inadequate denkbeelden en negatieve houdingen.

De laatste jaren hebben verscheidene onderzoekers onderwijsleeromgevingen ontworpen en ge-

evalueerd die tot doel hebben bij leerlingen uit de bovenbouw van het basisonderwijs of uit de onderbouw van het secundair onderwijs wiskundige probleemoplossingsvaardigheden te ontwikkelen. In vele van deze "design experiments" wordt min of meer expliciet uitgegaan van of aangesloten bij de principes van het door Collins, Brown en Newman (1989; zie ook Brown, 1992; De Corte e.a., 1996) ontwikkelde "cognitive apprenticeship"-model voor het leren en onderwijzen van complexe (meta)cognitieve vaardigheden. Hierna vermelden we kort drie recente construerende onderzoeken die rechtstreeks aan de basis lagen van onderhavige studie².

Lester e.a. (1989) voerden een construerend onderzoek uit naar de effecten op het probleemoplossingsgedrag van 12-13-jarige leerlingen van een onderwijsleeromgeving waarin leerlingen (1) oefenden in het gebruik van een aantal waardevolle heuristieken, (2) meer bewust werden gemaakt van hun probleemoplossingsstrategieën, en (3) geleerd werden om hun activiteiten tijdens het probleemoplossen te bewaken en bij te sturen. De ontwikkelde onderwijsleeromgeving bestond uit een set van geschikte problemen en een reeks lesplannen die o.a. de verschillende functies en activiteiten van de leerkracht beschreven. De instructie bestond uit een combinatie van probleemoplossen in kleine groepen, klassikale besprekingen en individuele opdrachten. Voor en na de instructie werd een schriftelijke toets afgenomen die bestond uit vijf onvertrouwde toepassingsproblemen. Uit de resultaten kwam een positief effect van de training op de probleemoplossingsvaardigheden van de leerlingen naar voren, maar dit globaal effect was minder groot dan verwacht; bovendien werden zeer grote inter-individuele verschillen in leerwinst vastgesteld.

Verschaffel en De Corte (1996) voerden een exploratief onderwijsexperiment uit waarin de mogelijkheid werd nagegaan om bij leerlingen uit de bovenbouw van de basisschool een tendens tot meer doordacht en realistisch wiskundig modelleren en oplossen te ontwikkelen. Terwijl twee controleklassen het normale wiskundeprogramma volgden, werkten de leerlingen van de experimentele klas een experimenteel programma af, bestaande uit vijf thematische leereenheden. Daarin werd telkens exemplarisch een bepaald soort van modeleringsprobleem onder de loep genomen (bijv. het interpreteren van de uitkomst van een niet-opgaande deling of het aanpakken van niet-lineaire verhoudingsproblemen). De vergelijking van de resultaten van de leerlingen van de experimentele klas en de controleklassen op de voor-, na- en retentietoets bracht aan het licht dat het programma de neiging van leerlingen tot oppervlakkig, niet-realistisch modelleren en interpreteren van wiskundige toepassingsproblemen aanzienlijk deed verminderen, hoewel het totaal aantal realistische reacties in de experimentele klas zowel op de natoets als op de retentietoets nog relatief laag was.

Van Haneghan, Barron, Young, Williams, Vye en Bransford (1992) maakten gebruik van videotecnologie om leerlingen te confronteren met rijke, authentieke en complexe probleemsituaties die ruimschoots de kans bieden voor het formuleren van een probleem, het opstellen van een model, het plannen hoe men het probleem gaat aanpakken en het interpreteren van de uitkomst, - steeds in relatie tot een reële context. Aan dit onderwijsexperiment werd deelgenomen door 10-11-jarigen met een niveau van wiskundevaardigheid dat boven het gemiddelde uitstak. Zowel aan de experimentele groep als aan de controlegroep werden drie onderwijssessies gegeven. In deze sessies werkte de experimentele groep onder leiding van een van de onderzoekers aan rijke problemen die op video werden aangeboden. Intussen werden de leerlingen van de controlegroep onderwezen in het oplossen van traditionele schoolvraagstukken. Voor en na de instructie werd zowel een klassieke vraagstuktoets afgenomen als een toets waarin gebruik werd gemaakt van op video aangeboden problemen. Op de klassieke vraagstuktoets werd geen verschil tussen beide groepen vastgesteld, maar op de andere toets, bestaande uit video-ondersteunde problemen, ging de experimentele groep aanzienlijk meer vooruit.

De beperkte positieve leereffecten behaald in deze drie onderzoeken zijn beloftevol. Ze laten immers zien dat het mogelijk is om de wiskundige probleemoplossingsvaardigheden van leerlingen te stimuleren via aangepast onderwijs. Toch kleven aan deze studies enkele tekorten. Ten eerste belichten zij elk slechts één bepaald aspect van dat onderwijs en veron-

achtzamen zij andere: het programma van Lester e.a. (1989) legt heel sterk de nadruk op het verwerven van een stel heuristische ingebed in een planmatige metacognitieve aanpakstrategie; het programma van Verschaffel en De Corte (1996) is helemaal gericht op het ontwikkelen van een dispositie tot meer realistisch wiskundig modelleren; en bij Van Haneghan e.a. (1992) staat het gebruik van videodiscs bij het uitwerken van authentieke, complexe, realistische contexten centraal, maar blijft de rol van de instructie die daarrond plaatsheeft sterk onderbelicht. Een tweede probleem rond de besproken studies is dat de instructie steeds verzorgd werd door de onderzoekers zelf in plaats van door de gewone leerkracht. Dit doet de vraag rijzen naar de implementeerbaarheid van deze programma's en naar de generaliseerbaarheid van de onderzoeksresultaten naar de reële onderwijspraktijk toe. Ten derde is de effectmeting in de drie studies vooral gebaseerd op een vergelijking van de prestaties van de leerlingen op een schriftelijke toets die voor en na de interventie werd afgenomen. Weliswaar werden ook andersoortige, meer procesmatige gegevens verzameld, maar dit materiaal werd over het algemeen minder systematisch geanalyseerd en is daardoor minder overtuigend. Ten slotte is in de besproken studies weinig of geen aandacht besteed aan de vraag of alle leerlingen in dezelfde mate van de interventie profiteerden, terwijl het met betrekking tot dergelijke interventies van cruciaal belang is te weten of ze niet enkel bij de sterkere maar ook bij de zwakkere leerlingen een positief effect sorteren.

In onderhavig artikel wordt verslag uitgebracht van een onderzoek waarin een 'krachtige leeromgeving' voor het leren oplossen van wiskundige toepassingsproblemen bij leerlingen uit de bovenbouw van het basisonderwijs ontwikkeld is en vervolgens in vier vijfde leerjaren is uitgetoetst. Bij dit 'design experiment' gingen we uit van de theoretische inzichten over krachtige leeromgevingen in het algemeen (Brown, 1992; Collins et al., 1989; De Corte e.a., 1996) en de bovenbeschreven construerende studies waarin (bepaalde van) deze inzichten reeds in min of meerdere mate toegepast werden op het leren oplossen van wiskundige toepassingsproblemen in het bijzonder. Daarbij is gepoogd om de sterke punten van de beschreven onderzoeken te combineren in één krachtige onderwijsleeromgeving, en de vermelde tekorten van de gehanteerde onderzoeksdesigns weg te werken.

2. BESCHRIJVING VAN DE LEEROMGEVING

2.1 Doel van de leeromgeving

Het doel van de leeromgeving bestaat erin leerlingen te helpen meer actieve, gemotiveerde, planmatige en bewuste oplossters van wiskundige toepassingsproblemen te worden. Dit wordt eerst en vooral nagestreefd door een algemene metacognitieve strategie voor het vaardig oplossen van wiskundige toepassingsproblemen aan te leren, die bestaat uit vijf stappen (zie Tabel 1). Bij de eerste en de tweede stap horen acht heuristieken waaraan we in de leeromgeving extra aandacht willen besteden gezien het belang ervan voor de vaardigheid in het oplossen van wiskundige toepassingsproblemen duidelijk is aangetoond (De Corte, 1995; De Corte e.a., 1996; Lester e.a., 1989; Schoenfeld, 1992; Verschaffel, ter perse).

Omwille van de nauwe verbondenheid van de (meta)cognitieve strategieën uit Tabel 1 met metacognitieve denkbeelden en houdingen, is een tweede belangrijke doelstelling de foutieve overtuigingen en negatieve attitudes die bij vele leerlingen rond het vak wiskunde en het oplossen van vraagstukken bestaan, in positieve zin te beïnvloeden. Voorbeelden daarvan zijn de opvattingen dat wiskundige problemen slechts één correct antwoord hebben, dat er slechts één enkele correcte manier is om een wiskundig probleem aan te pakken, dat de modale leerling niet eerder geziene vraagstukken onmogelijk zelfstandig kan oplossen, dat het kunnen oplossen van een vraagstuk slechts een kwestie van geluk is, dat de wiskunde die op school geleerd wordt weinig of geen uitstaans heeft met de echte wereld, ... (De Corte, 1995; Schoenfeld, 1992; Verschaffel, ter perse).

Tabel 1: Het vijfstappenmodel van vaardig oplossen van wiskundige toepassingsproblemen met bijhorende heuristieken

STAP 1:	IK STEL ME HET PROBLEEM VOOR
Heuristieken:	<ul style="list-style-type: none"> maak een tekening maak een tabel onderscheid noodzakelijke en overbodige gegevens gebruik je ervaringskennis
STAP 2:	IK BESLIS HOE IK HET PROBLEEM ZAL AANPAKKEN
Heuristieken:	<ul style="list-style-type: none"> maak een boomdiagram probeer verstandig uit zoek een patroon in de gegevens werk met eenvoudige getallen
STAP 3:	IK REKEN UIT
STAP 4:	IK INTERPRETEER MIJN UITKOMST EN FORMULEER MIJN ANTWOORD
STAP 5:	IK CONTROLEER

2.2 Pijlers van de leeromgeving

Een eerste pijler van de ontwikkelde leeromgeving is een gevarieerde set van complexe contextopgaven die uitdrukkelijk appèl doen op de beoogde heuristieken en metacognitieve vaardigheden. Bovendien wordt gewerkt met realistische, uitdagende opgaven, wat betekent dat zij zoveel mogelijk ingebed zijn in een context die afkomstig is uit het dagelijkse leven of uit de kinderlijke fantasiewereld en uitloopt op een zinvolle probleem- of vraagstelling voor kinderen van 10 à 11 jaar. Een ander kenmerk van de opgaven is hun open karakter: vele problemen zijn op meerdere manieren op te vatten, aan te pakken en op te lossen, wat aanleiding geeft tot overleg en discussie tussen leerlingen over de interpretatie, de aanpakwijze en het antwoord. Ten slotte is gestreefd naar grote variatie in de presentatiewijze van de problemen: puur tekstueel, in de vorm van een tabel, tekening, krantenknipsel of stripverhaal, ...

De tweede pijler betreft het aanwenden van een gevarieerd stel krachtige instructietechnieken en werkvormen. In de meeste lessen wordt een instructiemodel gevolgd dat ruwweg overeenkomt met dit uit het eerder vermelde exploratief onderzoek van Verschaffel en De Corte (1996). Dit model bestaat uit de volgende onderdelen: (1) een klassikale inleiding; (2) twee groepstaken die in vaste, heterogene groepen van drie à vier leerlingen worden aangepakt, telkens gevolgd door een bespreking met de ganse klasgroep, en (3) een individuele slotopdracht die eveneens achteraf klassikaal besproken wordt.

Een geschikte klascultuur, gericht op de beïnvloeding van de vakgebonden denkbeelden en houdingen van de leerlingen, vormt de derde pijler van de leeromgeving. Enkele typische kenmerken van deze cultuur zijn: (1) het creëren van ruimte voor het expliciteren van denkbeelden, misconcepties, oplossingsmanieren, gevoelens, ... van leerlingen in verband met het (leren) oplossen van wiskundige toepassingsproblemen; (2) het maken van afspraken over wat men verstaat onder een goed vraagstuk (bijv. 'sommige vraagstukken kunnen op meerdere manieren beantwoord worden'), een goede oplossing (bijv. 'soms is het niet zo zinvol om een heel precies antwoord te geven') of een goede oplossingsmethode (bijv. 'ook ervaren rekenaars tellen in geval van twijfel nog wel eens op de vingers'); en (3) het herdefiniëren van de rol van de leerkracht en van de leerlingen in de wiskundeles (bijv. 'jullie moeten niet verwachten dat ik als leerkracht alleen zal bepalen welke oplossing de beste is; dat is iets waarover we samen zullen overleggen').

2.3 Organisatie van de leeromgeving

De leeromgeving bestaat uit 20 lessen, verdeeld over drie grote delen:

- (1) Introductie op de vaardige oplossingsstrategie (les 1): in deze les leren de leerlingen onderscheid maken tussen een type- en een probleemopgave en wordt een globaal idee gegeven van het doel van de leeromgeving en de manier waarop dit zal worden nagestreefd.
- (2) Verwerven van de vaardige oplossingsstrategie (lessen 2 tot 16): in de lessen 2 tot en met 6 worden stap 1 en de vier daarbij horende heuristieken behandeld. De lessen 7 tot en met 11 behandelen stap 2 en de vier daarin vervatte heuristieken. In les 12 komt stap 3 aan bod. In de lessen 13 en 14 wordt dieper ingegaan op stap 4. Les 15 en les 16 handelen over de vijfde en laatste stap uit de vaardige oplossingsstrategie. Voor enkele voorbeelden van problemen uit dit deel van de lessenreeks verwijzen we naar de Figuren 1-3.

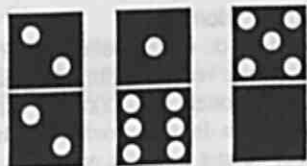
Figuur 1: Werkblad uit les 5 die handelt over de heuristiek 'gebruik je ervaringskennis'

In een Spaans dorpje bestaat de traditie dat elk jaar op 1 augustus de jonge mannen uit het dorp die tussen 20 en 30 jaar zijn een piramide bouwen door op elkaars schouders te gaan staan zoals op de tekening hieronder. Hoe hoog wordt die piramide ongeveer als de piramide 4 man hoog is?



Figuur 2: Werkblad uit les 9 die handelt over de heuristiek 'zoek een patroon in de gegevens'

Een dominospel bestaat uit allemaal verschillende dominostenen. Elke steen is verdeeld in twee helften en elke helft bevat 0, 1, 2, 3, 4, 5 of 6 stippen. Hieronder zie je enkele voorbeelden van stenen uit dit spel. Hoeveel stenen bevat dit spel in het totaal?



Figuur 3: Werkblad uit les 15 'fouten leren ontdekken en verbeteren'

Hieronder krijgen jullie een vraagstuk tesamen met drie foutieve oplossingen. De opdracht luidt: Ontdek waar de fout in elk van deze drie oplossingen zit en wat er precies fout is gelopen.

Een groothandelaar moet 200 doosjes met lampen verzenden en verpakt daarom de doosjes in een grote houten kist. In deze kist kunnen er 8 doosjes naast elkaar, 8 doosjes achter elkaar en 4 doosjes op elkaar staan. Zal deze man er in slagen om die 200 doosjes met lampen in deze kist te krijgen?

Oplossing 1:

$$8 + 8 + 4 = 20$$

Neem, deze man zal er niet in slagen om die 200 doosjes met lampen in deze kist te krijgen, want er kunnen maar 20 doosjes in deze kist.

Oplossing 2:

$$8 \times 8 \times 4 =$$

$$8 \times 8 = 64$$

$$\begin{array}{r} 64 \\ \times 4 \\ \hline 256 \end{array}$$

Neem, deze man zal er niet in slagen om die 200 doosjes met lampen in deze kist te krijgen, want er kunnen maar 116 doosjes in deze kist.

Oplossing 3:

$$8 \times 8 \times 4 =$$

$$8 \times 8 = 64$$

$$\begin{array}{r} 64 \\ \times 4 \\ \hline 256 \end{array}$$

Neem, want 256 is veel meer dan 200.

- (3) De vaardige oplossingsstrategie flexibel leren gebruiken in het kader van 'projecten' (les 17 tot 20): in deze lessen komt telkens één ingewikkeld probleem aan bod, waaraan gedurende een hele les gewerkt wordt en waarin de leerlingen al het geleerde uit de vorige lessen kunnen toepassen. Een voorbeeld van een probleem uit een van deze projectlessen wordt in Figuur 4 gegeven.

In de lessen speelt het zogenaamde E.H.B.O.-boekje (Eerste Hulp Bij het Oplossen van vraagstukken) een belangrijke rol. Dit bevat een globaal overzicht van de verschillende stappen en bladen met informatie over elke heuristiek en elke stap van de vaardige oplossingsstrategie. Telkens wordt een antwoord gegeven op de volgende vragen: (1) wat houdt deze heuristiek of deze stap in? (2) waar moet op gelet worden bij het uitvoeren ervan? en (3) wat is het belang ervan? Telkens wanneer er een bepaalde heuristiek of stap behandeld is, krijgen de leerlingen het bijhorend blad voor hun E.H.B.O.-boekje (voor een voorbeeld van een blad uit dit E.H.B.O.-boekje, zie Figuur 5).

2.4 Begeleiding vanwege het onderzoeksteam

Omdat de experimentele lessen niet door de onderzoekers maar door de leerkrachten gegeven werden, werd gezorgd voor een goede voorbereiding, begeleiding en ondersteuning van deze leerkrachten. Het model van professionele ontwikkeling van deze leerkrachten dat daarbij gevolgd werd, reflecteerde de visie van de onderzoekers op het (wiskunde)leren van kinderen, hetgeen impliceert dat er veel belang gehecht werd aan het creëren van een sociale context waarin de leerkrachten en onderzoekers van elkaar konden leren, in plaats van een model te hanteren waarbij de onderzoekers hun kennis en inzichten (proberen) overdragen op

Figuur 4: Voorbeeld van een projectles: het parking-probleem (les 17)

Wim en Margo hebben met kartonnen dozen een hele stad gebouwd, waarin ze hun speelgoedauto's kunnen laten rondrijden. In de stad is er nog geen parking. Het stuk grond tussen de kerk en het stadhuis lijkt daarvoor heel geschikt. Dit stuk grond heeft de vorm van een vierkant met een zijde van 50 cm. Wim heeft al een stuk karton van die grootte uitgeknipt.



Voor hoeveel speelgoedauto's is er plaats op deze parking?

1. Vul op de spandoek in hoeveel speelgoedauto's ten hoogste op de parking kunnen staan.
2. Teken op het karton hoe de parkeerplaatsen volgens jullie het best geschikt worden.
3. Leg uit hoe jullie tot deze oplossing zijn gekomen.

Figuur 5: Blad uit het E.H.B.O.-boekje dat hoort bij de heuristiek 'Onderscheid noodzakelijke en overbodige gegevens'

STAP 1: IK STEL ME HET PROBLEEM VOOR



ONDSCHIED NOODZAKELIJKE EN OVERBODIGE GEGEVENS

WAT?

ik ga na welke getallen uit de opgave ik nodig heb om het probleem op te lossen en welke getallen ik niet nodig heb

HOE?

- eerst bekijk ik aandachtig wat er gevraagd wordt
- dan bekijk ik alle gegevens en ga ik bij elk getal na of ik dit nodig heb of niet
- ten slotte omcirkel ik de getallen die ik nodig heb

WAAROM?

- zo dring ik door tot de kern van het probleem
- zo vermijd ik dat ik verkeerde getallen gebruik bij het berekenen van mijn antwoord

de leerkrachten (De Corte, 1998). Uitgaand van de veronderstellingen dat het verloop van wiskundeonderwijs te complex en te onvoorspelbaar is om helemaal vooraf helemaal te kunnen vastleggen en dat het onderwijsgedrag van leerkrachten gemedieerd wordt door hun kennis van en opvattingen over het vakgebied, lag de klemtoon van de voorbereiding en begeleiding van deze leerkrachten verder niet op het (leren) uitvoeren van bepaalde onderwijsgedragingen, maar op (leren) nemen van gefundeerde beslissingen (Carpenter & Fennema, 1992). Meer concreet hield de voorbereiding en begeleiding van de leerkrachten de volgende elementen in.

- (1) Een algemene handleiding met informatie over de achtergrond, de bedoeling en de globale opbouw van het programma, evenals een reeks algemene richtlijnen en aandachtspunten voor de aanpak van de lessen. Deze laatste hebben voornamelijk betrekking op de stappen die de leerkracht dient te zetten en de punten waarop hij/zij dient te letten bij het introduceren van een groepstaak of individuele opdracht, gedurende het groepsgevoel of individueel werken aan deze taak en bij de klassikale nabespreking.
- (2) Een specifieke handleiding per les waarin het concrete lesverloop wordt beschreven, voorbeelden van mogelijke juiste antwoorden en oplossingswijzen worden beschreven, enz.
- (3) Al het noodzakelijk lesmateriaal voor de leerlingen (werkbladen, objecten, ...).
- (4) Bijwonen van elke les door één lid van het onderzoeksteam, met gelegenheid tot overleg voor en na de les.
- (5) Regelmatige bijeenkomsten van twee uur waaraan deelgenomen werd door alle leden van het onderzoeksteam en door alle leerkrachten van de experimentele scholen die bij het onderzoek waren betrokken, evenals hun directies. Twee van deze bijeenkomsten vonden plaats voor de eigenlijke start van experimentele programma, drie in de loop ervan, en een achteraf. Tijdens de eerste twee bijeenkomsten werden de leerkrachten en directies uitgenodigd om commentaar te geven op ontwerpen van de algemene handleiding en van de specifieke handleidingen per les, die hen vooraf waren toegestuurd; voorstellen voor uitbreiding en aanpassing werden besproken met het hele team. Een van de voornaamste bedoelingen van deze bijeenkomsten was de leerkrachten 'co-ownership' geven van deze handleidingen. In de drie daaropvolgende bijeenkomsten werden ervaringen omtrent (de implementatie van) de voorbije lessen uitgewisseld, werden eventuele moeilijkheden besproken, en werden de komende lessen van naderbij bekeken. Tijdens de laatste bijeenkomst werden de leerkrachten geïnformeerd over de belangrijkste uitkomsten van het evaluatie- en implementatieonderzoek en werd samen gezocht naar een mogelijk verklaring voor bepaalde (deel)uitkomsten.

3. ONDERZOEKSMETHODE

3.1 Subjecten

Aan het onderzoek werd deelgenomen door vier willekeurig geselecteerde experimentele klassen (met respectievelijk 27, 19, 21 en 19 leerlingen) en zeven controleklassen (met respectievelijk 29, 22, 19, 21, 20, 17 en 18 leerlingen) van het vijfde leerjaar, die behoorden tot verschillende gemengde scholen uit de provincie Vlaams-Brabant³.

3.2 Globale onderzoeksopzet

Er werd gekozen voor een voortoets-natoets-retentietoetsopzet met een experimentele groep en een gelijkwaardige controlegroep, zoals in Tabel 2 wordt geschetst. In deze tabel staan ook de verschillende meetinstrumenten vermeld die gebruikt werden om de implementatie van de leeromgeving en het effect ervan op de leerlingen vast te stellen.

Tabel 2: Globale opzet van het onderzoek

	Experimentele groep (4 klassen)	Controlegroep (7 klassen)
Testmoment 1 (oktober '96)	Schriftelijke vraagstukken(voor)toets Vragenlijst 'houdingen en opvattingen (tegen)over wiskunde' Schoolvorderingentest Mondelinge vraagstukken(voor)toets	Schriftelijke vraagstukken(voor)toets Vragenlijst 'houdingen en opvattingen (tegen)over wiskunde' Schoolvorderingentest
Interventie (oktober '96 tot januari '97)	Krachtige leeromgeving (20 lessen) Video-opnames van lessen	Gewoon vraagstukkenonderwijs
Testmoment 2 (februari '97)	Schriftelijke vraagstukken(na)toets Vragenlijst 'houdingen en opvattingen (tegen)over wiskunde' Schoolvorderingentest Mondelinge vraagstukken(na)toets Interview met leerkrachten	Schriftelijke vraagstukken(na)toets Vragenlijst 'houdingen en opvattingen (tegen)over wiskunde' Schoolvorderingentest
Testmoment 3 (maart '97)	Schriftelijke vraagstukken(retentie)toets	Schriftelijke vraagstukken(retentie)toets

3.3 Instrumenten

Zowel in de experimentele klassen als in de controleklassen werden voor en na de interventie drie collectieve evaluatie-instrumenten afgenomen.

Ten eerste: een zelfgemaakte schriftelijke vraagstuktoets, bestaande uit 10 moeilijke toepassingsopgaven, die qua context en wiskundige structuur duidelijk afweken van de opgaven uit het experimenteel programma. Van deze toets werden er op de drie testmomenten telkens een parallel versie werd afgenomen. Tabel 3 bevat enkele items uit de vraagstuktoets die op het eerste moment werden gebruikt. Elk antwoord werd gescoord als 'correct antwoord' (CA), 'fout antwoord' (FA), 'rekentechnische fout' (TF) of 'geen antwoord' (GA). Dit classificatiesysteem had een zeer hoge betrouwbaarheid: drie onafhankelijke beoordelaars scoorden alle antwoorden van de leerlingen van twee willekeurig gekozen klassen op twee verschillende testmomenten en de interbeoordelaarsbetrouwbaarheid (in termen van de Kappa-coëfficiënt) bedroeg 0,99. Verder bedroeg de Cronbachs alfa-coëfficiënt voor de drie parallelversies respectievelijk 0,56, 0,75 en 0,79. De verklaring voor de lage interne consistentie van de voortoets is te vinden in de geringe variatie in de scores op deze toets, die over het algemeen uiterst laag waren (zie verder). Naast het scoren van de antwoorden in de vier bovengenoemde categorieën, werd gezocht naar aanwijzingen voor het gebruik van de aangeleerde heuristieken (zie Tabel 1). Concreter gezegd werd voor elk antwoord dat door een leerling uit de experimentele groep of de controlegroep tijdens de drie afnamemomenten op de schriftelijke vraagstuktoets gegeven werd, gescoord of er duidelijke sporen te vinden waren van het gebruik van een of meerdere van de acht aangeleerde heuristieken. Aanwezigheid van manifeste sporen van het gebruik van minstens een van deze acht heuristieken, werd gescoord met 1; afwezigheid met 0. Bij deze scoring werd uitgegaan van nauwkeurig omschreven operationalisering van de diverse heuristieken. Opmerkenswaard is verder dat de leerlingen op generlei wijze expliciet gevraagd of uitgedaagd werden om gebruik te maken van bepaalde heuristieken noch om de gebruikte heuristieken schriftelijk te rapporteren. Deze scores voor heuristiekgebruik kenden eveneens een hoge interbeoordelaarsbetrouwbaarheid: drie onaf-

hankelijke beoordelaars scoorden zowel de voor-, na- als retentietoets van één willekeurig gekozen experimentele klas op heuristiekgebruik, en bereikten daarbij een Kappacoëfficiënt van 0,96.

Tabel 3: Voorbeelden van voortoetsitems van de schriftelijke vraagstukentoets

- 1) Martha leest een boek. Plots merkt zij dat er bladzijden ontbreken, want onmiddellijk na bladzijde 135 komt bladzijde 173. Hoeveel bladzijden ontbreken er?
- 2) Lies heeft twee poppenhuizen. De vierkante vloer van het kleine poppenhuis heeft een zijde van 40 cm en bestaat uit 16 tegels. De vierkante vloer van het grote poppenhuis heeft een zijde die precies dubbel zo lang is als die van het kleine poppenhuis. Hoeveel tegels telt de vloer van het grote poppenhuis als daarvoor dezelfde tegels worden gebruikt?
- 3) Katrien bouwt huisjes met lucifers. Om 2 huisjes te bouwen, heeft ze 9 lucifers nodig. Om een rij van 5 huisjes te bouwen, heeft ze 21 lucifers nodig. Hoeveel lucifers zal ze nodig hebben om een rij van 10 huisjes te bouwen?



Ten tweede: een vragenlijst over de houdingen en opvattingen tegenover wiskunde, die aangeboden werd in de vorm van een Likertschaal met vijf antwoordalternatieven, gaande van 'helemaal niet akkoord' tot 'helemaal akkoord'. Voorbeelden van uitspraken uit deze vragenlijst zijn: 'Om de juiste oplossing voor een vraagstuk te vinden, moet je geluk hebben', 'Een goede leerkracht toont de leerlingen altijd hoe een bepaald soort vraagstukken moet opgelost worden alvorens de leerlingen zelf aan het werk te zetten', 'Als ik een vraagstuk niet direct kan oplossen, weet ik dat ik de oplossing nooit zal vinden'. Factoranalyse leverde twee interpreteerbare factoren op die samen 22% van de totale variantie verklaarden. De Cronbachs alfa-coëfficiënten voor beide factoren zijn respectievelijk 0,70 en 0,73, hetgeen bevredigend is. De eerste factor heeft betrekking op 'plezier en volharding bij het oplossen van vraagstukken' (zeven items); de tweede op 'een probleem-, proces- en realiteitsgerichte visie op het (leren) oplossen van vraagstukken' (14 items). Items uit de vragenlijst die op beide factoren een factorlading kenden die hoger was dan 0,20 werden toegewezen aan die factor waar zij inhoudelijk het best bij aansloten. De schaalscores waarmee hij de analyse gewerkt is, zijn gebaseerd op de som van de scores (maximumscore per item = 5) op de betreffende items. De factor 'plezier en volharding' kent aldus een maximumscore van 35, de factor 'visie op vraagstukken' een maximumscore van 70.

Ten derde: een schoolvorderingentest, bestaande uit 58 items verdeeld over acht subtests (logisch denken, verzamelingen, relaties, getal-kennis, bewerkingen, vraagstukken, metend rekenen en meetkunde), die een algemeen beeld geeft van het wiskundepeil van de leerlingen. De interne consistentie van de gehele test is hoog (gaande van 0,77 tot 0,89 van het derde leerjaar tot het zesde leerjaar) en de validiteit (uitgedrukt in een correlatie met de schooluitslag voor het vak wiskunde) blijkt eveneens voldoende hoog te zijn (gaande van 0,57 in het derde tot 0,77 in het zesde leerjaar) (Stinissen, Mermans, Tistaert & Vander Steene, 1985). Naast een totaalscore werden eveneens deelscores bepaald voor de verschillende rubrieken van deze test. Op de voortoets bedroeg de maximumscore 46 en op de natoets 58⁴.

Om zicht te krijgen op de eventuele kwalitatieve veranderingen in de oplossingsprocessen van de leerlingen als gevolg van de interventie, werden uit elke experimentele klas drie leerlingparen geselecteerd - een met twee sterke, een met twee middelmatige, en een met twee

zwakke leerlingen - die voor en na de interventie in het kader van een mondelinge, op video opgenomen vraagstuktoets vijf moeilijke toepassingsproblemen aangeboden kregen, die ze samen dienden op te lossen. De scoring gebeurde door twee 'blinde' beoordelaars die samen de videobeelden bekeken en in overleg bepaalden wanneer en hoe er gescoord diende te worden. In het analyseschema dat ze daarbij gebruikten, werden drie aspecten onderscheiden: (1) het resultaat van het oplossingsproces (CA, FA, TF of GA); (2) het gebruik van de aangeleerde heuristieken, en (3) het toepassen van een aantal typische metacognitieve activiteiten die kenmerkend zijn voor vaardige probleemoplossers, nl. oriëntering, planning, procesbewaking en reflectie-achteraf.

Bij alle leerkrachten van de experimentele klassen werd bovendien een steekproef van vier lessen op video opgenomen. De bedoeling van deze video-opnames was per leerkracht een 'implementatieprofiel' (Van den Akker & Voogt, 1994) op te stellen dat bestond uit tien componenten van leerkrachtgedrag die wij van essentieel belang achtten om van een goede implementatie van de leeromgeving te kunnen spreken (o.a. 'zorg voor een goede oriëntering op de taak', 'observeer en bied gepast hulp tijdens het groepswork', 'wijs erop dat er meerdere oplossingen en oplossingsmanieren bestaan tijdens de klassikale nabespreking').

Ten slotte werd van de vier leerkrachten van de experimentele klassen kort na het beëindigen van de lessenreeks (en alvorens zij informatie gekregen hadden over de resultaten van hun leerlingen op de toetsen) een interview afgenomen. De bedoeling daarvan was om van hen te vernemen welke indruk zij hadden van de leeromgeving, welke moeilijkheden zij hadden ondervonden bij de implementatie ervan en of zij bepaalde voorstellen hadden voor aanpassing van de leeromgeving.

4. ONDERZOEKSVRAGEN EN -HYPOTHESEN

De algemene hypothese die aan de basis lag van voorliggende studie was dat de krachtige leeromgeving een significante, positieve invloed zou hebben op de vaardigheid van de leerlingen in het oplossen van wiskundige toepassingsproblemen. We verwachtten derhalve dat de experimentele groep bij het tweede en het derde testmoment significant hoger zou scoren op de schriftelijke vraagstuktoets dan de controlegroep, terwijl dit niet het geval zou zijn op het eerste testmoment.

Bovendien verwachtten we dat de krachtige leeromgeving een gunstig effect zou hebben op de houdingen en overtuigingen (tegen)over dit soort van problemen. Deze hypothese werd getoetst door een vergelijking te maken tussen de antwoorden van alle leerlingen uit de experimentele klassen en de controleklassen op de twee schalen van de vragenlijst, die voor en na de interventie werd afgenomen. Onze predictie luidde dat de experimentele groep na de interventie op de beide schalen een significant hogere score zou bekomen dan de controlegroep, terwijl de scores van beide groepen niet zouden verschillen bij de voormeting.

Met betrekking tot de schoolvorderingstest werd geen predictie geformuleerd. Immers, aan de ene kant zou men kunnen argumenteren dat de aandacht die in de leeromgeving uitgaat naar het leren oplossen van toepassingsproblemen wel eens ten koste zou kunnen gaan van de vooruitgang op de andere onderdelen van het wiskundecurriculum. Doch aan de andere kant zou er ook een positief transfereffect kunnen optreden van hetgeen de leerlingen in de krachtige leeromgeving hebben geleerd naar de overige deelgebieden van het vakgebied toe.

Vervolgens verwachtten we ook een positief effect van de leeromgeving op de kwaliteit van het oplossingsproces van de leerlingen. Meer bepaald luidde onze hypothese dat de leerlingen uit de experimentele klassen ten gevolge van de krachtige leeromgeving meer (efficiënt) gebruik zouden maken van waardevolle heuristieken en metacognitieve vaardigheden tijdens het oplossen van wiskundige toepassingsopgaven.

Ten slotte werd bij elk van de afgenomen toetsen (de schriftelijke vraagstuktoets, de vragenlijst en de schoolvorderingstest) nagegaan of het verwachte effect zich in al de expe-

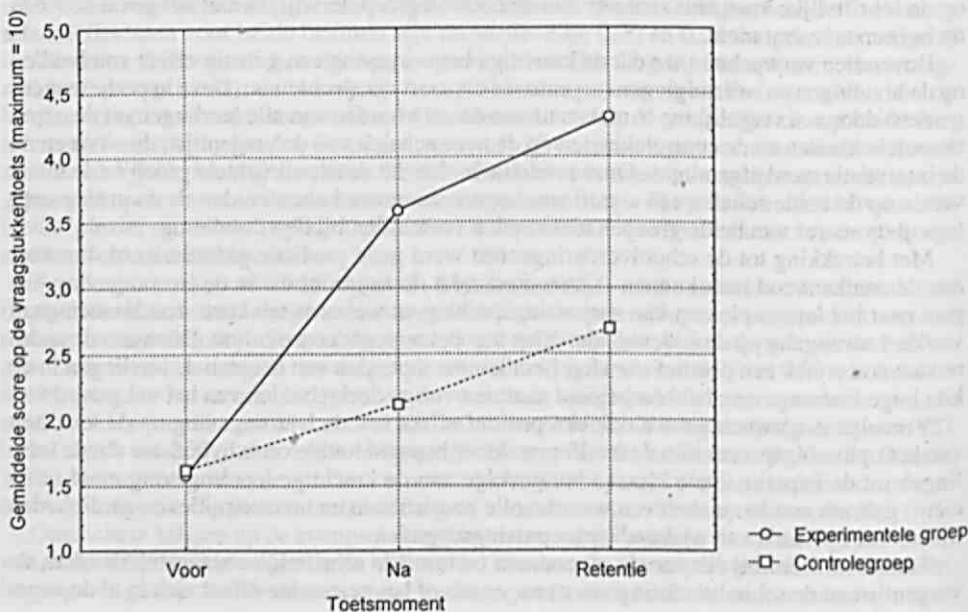
rimentele klassen in dezelfde mate voordeed en tevens of niet enkel de sterke, maar ook de middelmatige en de zwakke leerlingen van de leeromgeving profiteerden. Daartoe werden de leerlingen uit de experimentele groep en de controlegroep op basis van hun resultaten op de eerste afname van de schoolvorderingstest in drie niveaugroepen opgedeeld.

5. RESULTATEN

Om het effect van de leeromgeving op de resultaten van de drie collectieve toetsen (schriftelijke vraagstuktoets, vragenlijst en schoolvorderingstest) te achterhalen, werd een variantieanalyse uitgevoerd volgens een hiërarchisch factorieel design met drie herhaalde metingen (Testmoment 1, 2 en 3) voor de schriftelijke vraagstuktoets en twee herhaalde metingen (Testmoment 1 en 2) voor de vragenlijst en de schoolvorderingstest, en met de factoren Groep (experimentele groep versus controlegroep) en Klas, waarbij de factor Klas ingebed was in de variabele Groep. Hoofd- en interactie-effecten werden met behulp van a posteriori Tukey HSD toetsen verder uiteengehaald. Om de grootte van de gevonden effecten beter te kunnen inschatten, maakten we gebruik van Cohens' maat voor effectgrootte (Cohen, 1988).

Zoals voorspeld leverde de ANOVA op de resultaten op de drie schriftelijke vraagstuktoetsen een significante interactie op tussen Groep en Test [$F(2;396) = 30,60, p < 0,001$]. Figuur 6 geeft de gemiddelde score van de leerlingen uit de experimentele groep en de controlegroep op deze drie toetsen. Terwijl er voor de voortoets geen significant verschil was tussen de resultaten van de experimentele groep en de controlegroep was dit wel het geval voor de natoets en de retentietoets (volgens a posteriori Tukey HSD toetsen, $p < 0,05$). Dit gevonden effect is middelgroot (effectgrootte = 0,31) (Cohen, 1988). De enigszins verrassende toename van de gemiddelde score van de leerlingen uit de beide groepen tussen de natoets en de retentietoets is wellicht te wijten aan een leereffect veroorzaakt door het parallelle karakter van de toetsen gebruikt op de drie afnamemomenten.

Figuur 6: Gemiddelde scores van de experimentele groep en de controlegroep tijdens de drie afnames van de schriftelijke vraagstuktoets (voortoets, natoets en retentietoets)



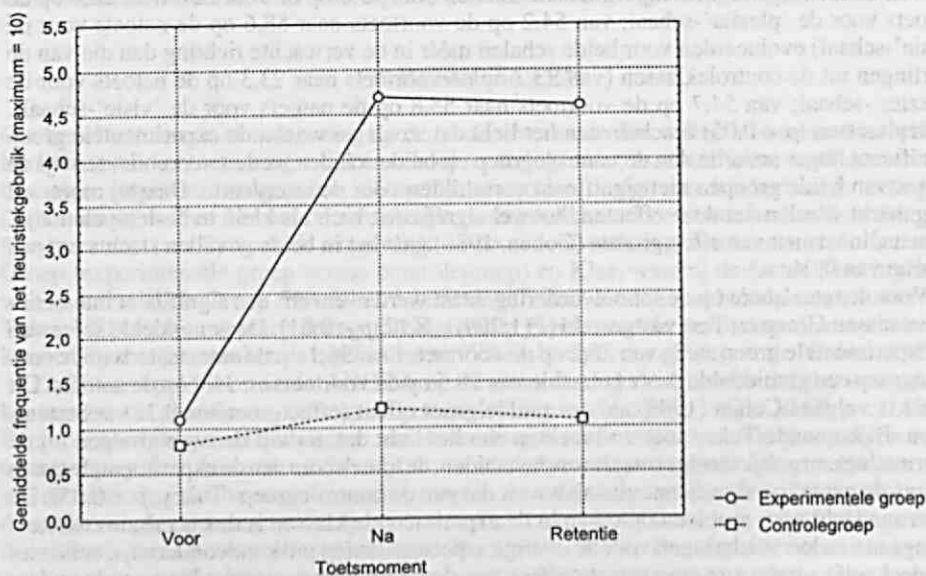
De interactie tussen Groep en Testmoment was ook voor beide schalen van de vragenlijst significant: $F(1;205) = 11,67, p < 0,001$ en $F(1;205) = 10,73, p < 0,01$ voor respectievelijk de 'plezier en volharding'-schaal en voor de 'visie op vraagstukken'-schaal. De gemiddelde scores van de leerlingen uit de experimentele klassen (van 23,6 op de voortoets naar 26,4 op de natoets voor de 'plezier'-schaal; van 54,2 op de voortoets naar 58,6 op de natoets voor de 'visie'-schaal) evolueerden voor beide schalen méér in de verwachte richting dan die van de leerlingen uit de controleklassen (van 23,2 op de voortoets naar 23,3 op de natoets voor de 'plezier'-schaal; van 54,7 op de voortoets naar 55,8 op de natoets voor de 'visie'-schaal). Tukey toetsen ($p < 0,05$) brachten aan het licht dat, zoals verwacht, de experimentele groep significant hoger scoorde dan de controlegroep op beide schalen na de interventie, terwijl de scores van beide groepen niet significant verschilden voor de interventie. Daarbij moet wel opgemerkt worden dat deze effecten, hoewel significant, toch als klein te bestempelen zijn. Immers, in termen van effectgrootte (Cohen, 1988) gaat het in beide gevallen slechts om een grootte van 0,04.

Voor de totaalscore op de schoolvorderingentest werd eveneens een significant interactie-effect tussen Groep en Test vastgesteld [$F(1;209) = 8,12, p < 0,01$]. De gemiddelde score van de experimentele groep steeg van 28,2 op de voortoets naar 36,1 op de natoets, terwijl de controlegroep een gemiddelde score behaalde van 28,5 op de voortoets en 34,4 op de natoets. Dit effect is volgens Cohen (1988) als een middelgroot effect (effectgrootte = 0,38) te bestempelen. Bijkomende Tukey toetsen brachten aan het licht dat, terwijl de twee groepen bij de voormeting een gelijkaardige totaalscore behaalden, de totaalscore van de experimentele groep tijdens de nameting significant uitstak boven die van de controlegroep (Tukey, $p < 0,05$). De extra aandacht voor probleemoplossen in de experimentele klassen is dus geenszins ten koste gegaan van leervorderingen voor de overige aspecten van het wiskundeonderwijs, er is integendeel zelfs sprake van enig transfereffect van de verworven vaardigheden naar de andere onderdelen van het wiskundecurriculum.

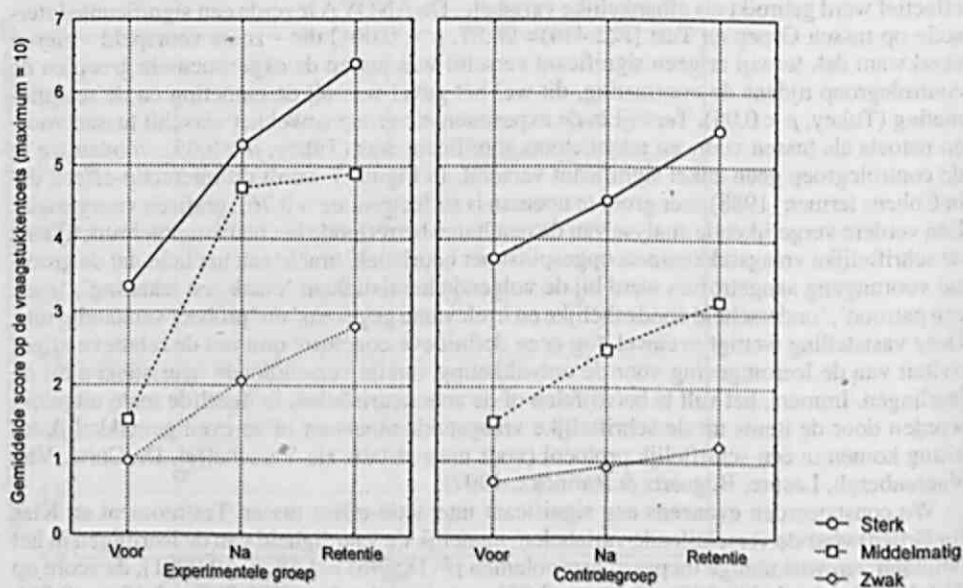
Om na te gaan of de leerlingen uit de experimentele groep meer gebruik maakten van de aangeleerde heuristieken dan de leerlingen uit de controlegroep, werd een univariate variatie-analyse uitgevoerd met Groep, Testmoment en Klas (in Groep) als onafhankelijke variabelen en het aantal vraagstukken waarin (ten minste) een van de acht aangeleerde heuristieken effectief werd gebruikt als afhankelijke variabele. De ANOVA leverde een significante interactie op tussen Groep en Test [$F(2;414) = 98,87, p < 0,001$] die - zoals voorspeld - hierop neerkwam dat, terwijl er geen significant verschil was tussen de experimentele groep en de controlegroep tijdens de voormeting, dit wel het geval was bij de nameting en de retentiemeting (Tukey, $p < 0,05$). Terwijl in de experimentele groep zowel het verschil tussen voor- en natoets als tussen voor- en retentietoets significant was (Tukey, $p < 0,05$), vonden we in de controlegroep geen enkel significant verschil. In Figuur 7 wordt dit interactie-effect, dat in Cohens termen (1988) zeer groot te noemen is (effectgrootte = 0,76), grafisch voorgesteld. Een verdere vergelijkende analyse van de resultaten betreffende het heuristiekgebruik tijdens de schriftelijke vraagstuktoets opgesplitst per heuristiek, bracht aan het licht dat de grootste vooruitgang aangetroffen werd bij de volgende heuristieken: 'maak een tekening', 'zoek een patroon', 'onderscheid noodzakelijke en irrelevante gegevens' en 'probeer verstandig uit'. Deze vaststelling wettigt evenwel nog geen definitieve conclusie omtrent de relatieve effectiviteit van de leeromgeving voor de ontwikkeling van de verschillende heuristieken bij de leerlingen. Immers, het valt te betwijfelen of de acht heuristieken in dezelfde mate uitgelokt worden door de items uit de schriftelijke vraagstuktoets en of ze even gemakkelijk tot uiting komen in een schriftelijk protocol (voor meer details zie Verschaffel, De Corte, Van Vaerenbergh, Lasure, Bogaerts & Ratinckx, 1997).

We constateerden eveneens een significant interactie-effect tussen Testmoment en Klas (in Groep) voor de verschillende variabelen, namelijk de vaardigheid van de leerlingen in het oplossen van wiskundige toepassingsproblemen [$F(18;396) = 4,58, p < 0,0001$], de score op de 'plezier'-schaal [$F(9;205) = 2,40, p < 0,05$] en de 'visie'-schaal [$F(9;205) = 2,18, p < 0,05$],

Figuur 7: Gemiddelde frequentie van het heuristiekgebruik in de experimentele groep en de controlegroep tijdens de drie afnames van de schriftelijke vraagstukkenoets (voortoets, natoets en retentietoets)



Figuur 8: Gemiddelde scores van de sterke, de middelmatige en de zwakke leerlingen uit de experimentele groep en de controlegroep tijdens de drie afnames van de schriftelijke vraagstukkenoets (voortoets, natoets en retentietoets)



en de schoolvorderingentest [$F(9;208) = 2,84, p < 0,01$]. Dit wil zeggen dat het eerder beschreven algemene effect van de krachtige leeromgeving niet in dezelfde mate teruggevonden werd in de vier experimentele klassen. In één experimentele klas, nl. klas 3, werd op al deze maten nauwelijks vooruitgang geconstateerd in vergelijking met de drie overige klassen (voor meer details zie Verschaffel, e.a., 1997). We merken hierbij nog op dat voor geen enkele van deze variabelen bij de voormeting significante verschillen tussen de vier experimentele klassen werden vastgesteld.

De trifactoriële variantieanalyse met Groep, Vaardigheidsniveau en Testmoment als onafhankelijke variabelen en het aantal correcte antwoorden op de schriftelijke vraagstuktoets als afhankelijke variabele, leverde geen significante drievoudige interactie op [$F(4;370) = 1,37, ns$]. Figuur 8 bevat de gemiddelde scores tijdens de drie afnames op de schriftelijke vraagstuktoets, uitgesplitst per niveaugroep. Uit deze resultaten kunnen we afleiden dat de drie niveaugroepen bijgedragen hebben tot het eerder gerapporteerde Groep x Test interactie-effect (zie Figuur 6). De leeromgeving heeft dus niet enkel geleid tot een betekenisvolle toename van de oplossingsvaardigheid van de sterke en middelmatige leerlingen, maar ook van die van de zwakke leerlingen.

Ook bij de overige afhankelijke variabelen, namelijk de score op de 'plezier'- en de 'visie'-schaal, de schoolvorderingentest en het heuristiekgebruik tijdens de schriftelijke vraagstuktoets bleek de interactie tussen Groep, Testmoment en Vaardigheidsniveau niet significant (respectievelijk $F(2;192) = 1,17, ns$; $F(2;189) = 0,72, ns$; $F(2;193) = 0,50, ns$ en $F(4;382) = 1,04, ns$). Dit wijst erop dat elk van de drie vaardigheidsniveaus op betekenisvolle wijze bijgedragen heeft tot de geconstateerde leerwinst van de experimentele groep.

Uit de resultaten op de mondelinge vraagstuktoets bleek dat de leerlingenparen uit de experimentele klassen tijdens de nameting globaal genomen bijna dubbel zo vaak spontaan gebruik maakten van heuristieken als tijdens de voormeting. Naast heuristiekgebruik gingen we bij deze leerlingenparen ook op zoek naar sporen van vier typische metacognitieve activiteiten, namelijk oriëntering, planning, procesbewaking en reflectie-achteraf. Het aantal gevallen waarin een van deze metacognitieve activiteiten voorkwam, steeg eveneens van voor naar natoets, maar hier was de toename minder groot (voor meer details over dit resultaat, evenals over de hiernavolgende resultaten, zie Verschaffel e.a., 1997).

Om voor elke leerkracht van de experimentele klassen tot een implementatieprofiel van de opgenomen lessen te komen, werd elk van de tien componenten gescoord op een vijfpunten-schaal van 0 (= volstrekt onvoldoende realisatie van de betreffende component) over 1 (= bevredigende implementatie) tot 2 (= een zeer goede implementatie). De gemiddelde implementatiescore voor de vier leerkrachten samen bedroeg 12,4 op 20, hetgeen de conclusie toelaat dat de realisatie van de krachtige leeromgeving globaal genomen bevredigend was. Uit de variantieanalyse bleek echter dat de vier leerkrachten nogal van mekaar verschilden in de mate waarin zij de tien componenten wisten te verwezenlijken [$F(3;12) = 10,227, p < 0,01$]; hun gemiddelde scores waren respectievelijk 10,1; 13,8; 10,9 en 14,6. De redelijk lage implementatiescore van leerkracht 3 kan een verklaring zijn voor de kleine vooruitgang van de leerlingen uit deze klas (zie hiervoor). Maar daartegenover staat dat de leerkracht van de klasgroep die op de meeste maten wel goed vooruitging (nl. klas 1) eveneens een lage implementatiescore behaalde.

Ten slotte werd in de interviews gepeild naar de beleving en de appreciatie van de leeromgeving door de leerkrachten. Daaruit bleek dat de vier leerkrachten zich schaarden achter de theoretische bereikbaarheid van het vaardig oplossingsmodel bij leerlingen van het vijfde leerjaar, dat zij positief stonden ten opzichte van de inhoud en de uitwerking van de lessenreeks, dat zij tevreden waren over de geboden begeleiding en tevreden terugkeken op hun deelname aan het onderzoek. Zij signaleerden echter ook diverse problemen. Ten eerste waren drie heuristieken uit het model volgens hen te moeilijk voor vijfdeklasser, nl. 'maak een boomdiagram', 'zoek een patroon' en 'vereenvoudig de getallen'. Ook vonden zij dat verscheidene toepassingsproblemen uit de leeromgeving om bijsturing vragen, vooral op het vlak van

het realiteitsgehalte. Wat het groepswork betreft, werden er door sommige leerkrachten vraagtekens geplaatst bij de vaste en/of heterogene samenstelling van de groepen. Tijdsgebrek werd volgens alle leerkrachten ervaren als het grootste probleem bij de implementatie van de leeromgeving. Ten slotte beklemtoonden twee van de vier leerkrachten dat zij graag op een nog directere en intensievere manier hadden geparticipeerd in de ontwikkelingsfase van de leeromgeving, al erkenden zij dat dit praktisch moeilijk haalbaar was geweest.

6. CONCLUSIES EN DISCUSSIE

Het doel van onderhavige studie bestond erin een krachtige leeromgeving te construeren, te implementeren en te evalueren, om zo bij leerlingen uit de bovenbouw van de basisschool een dispositie te ontwikkelen tot het vaardig oplossen van wiskundige toepassingsproblemen. Deze leeromgeving werd geïmplementeerd in vier klassen van het vijfde leerjaar, en het effect daarvan werd geëvalueerd in een onderzoek met een voortoets-natoets-retentietoetsopzet met een experimentele groep en een gelijkwaardige controlegroep waarin gewoon vraagstukken-onderwijs verstrekt werd.

Blijkens de resultaten op de schriftelijke voortoets en natoets met vraagstukken had de leeromgeving een betekenisvol positief effect op de ontwikkeling van de probleemoplossingsvaardigheid van de leerlingen. Uit de resultaten van de retentietoets kwam naar voren dat dit positief effect zeker niet verdween na stopzetting van de experimentele lessenreeks. De leeromgeving had tevens een gunstige invloed op de ontwikkeling van het plezier en de volharding van leerlingen bij het oplossen van vraagstukken, en van hun opvattingen en houdingen (tegen)over dit soort van opgaven. Uit de resultaten van de schoolvorderingentest bleek dat de extra aandacht in de experimentele klassen voor probleemoplossingsstrategieën en -vaardigheden en voor opvattingen en houdingen geenszins ten koste gegaan was van de leer-vorderingen voor de overige aspecten van het wiskundeonderwijs maar dat er integendeel zelfs een zeker transfereffect te constateren viel. Uit de vergelijking van de leerresultaten in de verschillende bekwaamheidsniveaus kwam tot uiting dat de genoemde vormen van leerwinst niet enkel voorkwamen bij de sterke en middelmatige leerlingen, maar ook bij de zwakkere. Als zodanig liggen de resultaten van onderhavige studie in de lijn van de bevindingen van andere studies waarin (meta)cognitieve strategieën werden aangeleerd, en met name van de drie construerende onderzoeken op het gebied van leren wiskundig probleemoplossen die we in het begin van dit artikel hebben besproken. Een belangrijk verschil met deze laatste studies is echter dat de ecologische validiteit van onze resultaten groter is, aangezien het experimenteel onderwijs niet door de onderzoeker(s) maar door de leerkrachten zelf verstrekt werd (Verschaffel, ter perse).

De positieve effecten van onderhavige studie zijn ons inziens toe te schrijven aan de volgende drie factoren: (1) het feit dat de leeromgeving niet eenzijdig gericht is op een specifiek aspect van de wiskundige dispositie, maar op het gelijktijdig en geïntegreerd verwerven van meerdere categorieën van kennis en vaardigheden; (2) de aandacht die er bij de uitwerking van de leeromgeving besteed is aan de kwaliteit van het opgavenaanbod, de kracht van de instructietechnieken en de gehele klascultuur, en (3) de zorg die in onderhavig onderzoek besteed is aan de begeleiding en ondersteuning van de leerkrachten.

De gehanteerde onderzoeksopzet laat echter niet toe om uitspraken te doen over het aandeel van de verschillende bovenvermelde factoren in het bereikte effect. Daartoe zijn verdere 'design experiments' vereist waarin bepaalde varianten van de ontwikkelde leeromgeving met elkaar worden vergeleken teneinde een beter zicht te krijgen op de sterke en zwakke punten ervan. Het ware ook wenselijk geweest dat de leeromgeving gerealiseerd en geëvalueerd werd in meer dan vier experimentele klassen. Dit had niet enkel geresulteerd in meer betrouwbare en generaliseerbare conclusies over het effect van de leergang maar had ons tevens in staat gesteld om de verschillen tussen leerkrachten in de implementatie van de ontwikkelde

leeromgeving én het verband daarvan met de leerwinst bij de leerlingen, op een meer systematische wijze te bestuderen. Daarnaast is het aannemelijk dat de resultaten beter zouden zijn geweest indien de instructietijd had kunnen opgevoerd worden tot meer dan 20 lestijden. Ten slotte kunnen vragen gesteld worden bij het gehanteerde onderzoeksinstrumentarium. Door het ontbreken van geschikte instrumenten voor de meeste variabelen waarin wij geïnteresseerd waren, dienden wij die noodzakelijkerwijs zelf te ontwikkelen. Maar wegens tijdsgebrek moest er dan wel gewerkt worden met voorlopige meetinstrumenten waarvan de objectiviteit, betrouwbaarheid en/of validiteit niet altijd voldoende kon worden onderzocht.

Naast deze methodologische kanttekeningen, hoeden de bekomen resultaten zelf ons voor overdreven optimisme. Zo herinneren we eraan dat - met uitzondering van de leerwinst op het vlak van het heuristiekgebruik tijdens de schriftelijke vraagstuktoets - geen van de vastgestelde verschillen in vooruitgang tussen de leerlingen van de experimentele groep en de controlegroep als 'groot' te bestempelen is; het gaat veeleer om 'middelmatige' of soms zelfs ronduit 'kleine' effecten. Daarnaast is er de vaststelling dat de resultaten in een van de vier experimentele klassen over de ganse lijn nogal tegenvielen. En ten slotte herinneren we eraan dat de leeromgeving de bestaande verschillen tussen zwakke en sterke leerlingen op vlak van prestaties niet deed afnemen.

Rekening houdend met deze relativerende bevindingen, blikken we tot slot van dit artikel nog even terug op enkele problematische aspecten van de ontwikkelde leeromgeving en/of de wijze waarop ze in onderhavige studie geïmplementeerd werd, die kunnen verklaren waarom de bereikte leereffecten niet (nog) groter waren. Daaraan gekoppeld formuleren we telkens suggesties voor verbetering van de leeromgeving en/of voor verder onderzoek.

Een eerste vraag is of de uitwerking van het model van vaardig probleemoplossen dat aan de basis ligt van de ontwikkelde leeromgeving met de acht daarin vervatte heuristieken (zie Tabel 1), wel de meest adequate is voor leerlingen van de basisschool. Voortgezet onderzoek moet uitwijzen op welke punten deze uitwerking verbeterd kan worden zodat het beter begrijpbaar en efficiënter hanteerbaar zou zijn.

Ten tweede: hoewel de ontwikkeling van adequate opvattingen en houdingen (tegen)over vraagstukken als een van de belangrijke doelen van de leeromgeving naar voren geschoven is en er ook op dat punt bij de leerlingen succes geboekt is, moeten we toegeven dat daaraan is en er ook op dat punt bij de leerlingen succes geboekt is, moeten we toegeven dat daaraan door de leerkrachten nog te weinig aandacht is besteed, mede omdat deze pijler in de ontwikkelde leeromgeving het minst uitgewerkt was. Daarnaast zijn ook de twee andere pijlers van onze omgeving (de opgaven en de instructietechnieken) voor verbetering vatbaar. In dit verband signaleren we vooral dat de kwaliteit van het groepswerk in een aantal gevallen duidelijk te wensen overliet.

Naast gebreken van de leeromgeving zelf, kunnen tekorten in de implementatie ervan door de leerkrachten mede verklaren waarom de bereikte effecten niet groter waren. Uit het onderzoek is duidelijk gebleken dat de effectieve realisatie van de leeromgeving zeer hoge eisen stelt aan de leerkrachten. Hoewel gepoogd is om de leerkrachten zo goed mogelijk te begeleiden bij de uitvoering van hun complexe en veeleisende taak, lag de klemtoon in onderhavige studie op de ontwikkeling en de evaluatie van de leeromgeving, en niet op het uittesten van een wetenschappelijk onderbouwd systeem van navorming en begeleiding van leerkrachten die deze leeromgeving in hun klaspraktijk willen realiseren. Zowel vanuit theoretisch als vanuit beleids- en praktijkgericht oogpunt is nader onderzoek rond dit laatste thema echter evenzeer gewenst.

Ten slotte is het zo dat de leergang zowel door de leerkrachten als door de leerlingen ervaren werd als een speciale lessenreeks die relatief los stond van de overige wiskundelessen. Het feit dat de realisatie van de leeromgeving plaatsvond in het kader van een welomlijnd wetenschappelijk onderzoek, droeg er wellicht toe bij dat de integratie van deze lessen in de gewone wiskundelessen moeilijk verliep. Het valt helemaal niet uit te sluiten dat deze gebrekige integratie de bereikte leereffecten heeft getemperd. We beklemtonen dan ook dat onze beslissing tot het uitwerken, implementeren en evalueren van een beperkte, welomlijnde les-

senreeks beruiste op praktische overwegingen en zeker niet op te vatten is als een pleidooi voor de invoering in het wiskundeonderwijs van dergelijke speciale lessenreeksen, los van de overige (wiskunde)lessen. De ontwikkeling van de wiskundige probleemoplossingsvaardigheid kan en mag niet gebeuren vanuit een specifiek curriculumonderdeel, maar dient ons inziens doorheen het hele wiskundeonderwijs en daarenboven doorheen het ganse onderwijs gestimuleerd en ondersteund te worden.

NOTEN

1. Correspondentie-adres: Lieven Verschaffel, Centrum voor Instructiepsychologie en -Technologie (CIP&T), Katholieke Universiteit Leuven, Afdeling Didactiek, Vesaliusstraat 2, B-3000 Leuven. Tel. 016-32.62.58. E-mail: Lieven.Verschaffel@ped.kuleuven.ac.be
2. De bespreking van deze drie "design experiments" zou kunnen worden aangevuld met heel wat andere recente studies met een zelfde of een analoge theoretische achtergrond en onderzoeksopzet, zoals die van Lampert (1990), Resnick en Nelson-LeGall (1987) of Silver, Smith, Lane, Salmon-Cox en Stein (1990). Bovendien is ook gedurende de voorbije decennia vanuit diverse theoretische paradigma's, zoals de denkpsychologie (bijv. Kohnstam), de handelingspsychologie (bijv. Landa, Van Parreren), de informatieverwerkingsbenadering (bijv. Simon) en de vakdidactische benadering (Polya, Freudenthal) veel onderzoek verricht naar de mogelijkheid om via aangepaste instructie de vaardigheid van leerlingen in het oplossen van wiskundige toepassingsproblemen te ontwikkelen.
3. De reden waarom het aantal klassen en leerlingen in de controlegroep groter was dan in de experimentele groep is dat we op die manier een zo representatief mogelijk beeld wilden verkrijgen van (de effecten van) het huidige vraagstukkenonderwijs.
4. Overeenkomstig de instructies van de constructeurs van de schoolvorderingentest, boden we tijdens het eerste afnamemoment enkel de eerste 46 items uit deze toets aan die betrekking hebben op de leerstof voor de eerste vier leerjaren van de lagere school, terwijl we bij het tweede afnamemoment ook de volgende 12 items uit deze toets gaven die betrekking hebben op de leerstof voor de derde graad.

REFERENTIES

- Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2, 141-178.
- Carpenter, T.P., & Fennema, E. (1992). Cognitively guided instruction. Building on the knowledge of students and teachers. *International Journal of Educational Research*, 17, 457-470.
- Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education*, 23, 2-33.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Collins, A., Brown, A., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L.B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honour of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.
- De Corte, E. (1995). Psychologie en reken/wiskunde-onderwijs. In L. Verschaffel & E. De Corte (Eds.), *Naar een nieuwe reken/wiskundendidactiek voor de basisschool en de basiseducatie. Deel 1: Achtergronden* (pp. 50-92). Brussel/Leuven: Studiecentrum Open Hoger Onderwijs (StOHO)/Acco.
- De Corte, E. (1998, August). *Marrying theory building and the improvement of school practice: A permanent concern for instructional psychology*. Presidential address of the Division of Educational, Instructional, and School Psychology presented at the 24th International Congress of Applied Psychology, San Francisco, CA.
- De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 491-549). New York: MacMillan.
- Gravemeijer, K.P.E. (1994). *Developing realistic mathematics education*. Utrecht: Freudenthal Institute, University of Utrecht.

- Greer, B. (1997). Modelling reality in mathematics classrooms: the case of word problems. *Learning and Instruction*, 7, 293-307.
- Instituut voor Leerplanontwikkeling (SLO) (1996). *Concept herziene kerndoelen voor het basisonderwijs*. Enschede: SLO.
- Kuiper, W.A.J.M., Bos, K.T., & Plomp, T. (1997). *Wiskunde en de natuurwetenschappelijke vakken in leerjaar 1 en 2 van het voortgezet onderwijs. Nederlands aandeel in TIMSS populatie 2*. Enschede: Universiteit Twente, OCTO.
- Lester, F., Garofalo, J., & Kroll, D. (1989). *The role of metacognition in mathematical problem solving. A study of two grade seven classes*. Final report to the National Science Foundation of NSF Project MDR 85-50346. Bloomington, IN: Mathematics Education Development Center, Indiana University.
- Ministerie van de Vlaamse Gemeenschap (1997). *Gewoon basisonderwijs: ontwikkelingsdoelen en eindtermen. Besluit van mei '97 en decreet van juli '97*. Brussel: Departement Onderwijs, Centrum voor Informatie en Documentatie.
- Resnick, L.B., & Nelson-LeGall, S. (1987, August). *Meaning construction in mathematical problem solving*. Paper presented at the Eleventh International Conference for the Psychology of Mathematics Education, Montreal, Canada.
- Reusser, K., & Stebler, R. (1997). Every word problem has a solution - The social rationality of mathematical modeling in schools. *Learning and Instruction*, 7, 309-327.
- 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.
- Silver, E.A., Smith, M., Lane, S., Salmon-Cox, L., & Stein, M. (1990). *QUASAR summary project*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.
- Stinissen, J., Mermans, M., Tistaert, G., & Vander Steene, G. (1985). *Leuvense Schoolvorderingstest Vernieuwde Wiskunde 2-6*. Brussel: C.S.B.O.
- Treffers, A., De Moor, E., & Feijs, E. (1989). *Proeve van een nationaal programma voor het reken/wiskunde-onderwijs op de basisschool. Deel 1. Overzicht einddoelen*. Tilburg: Zwijsen.
- Van den Akker, J., & Voogt, J. (1994). The use of innovation and practice profiles in the evaluation of curriculum implementation. *Studies in Educational Evaluation*, 20, 503-512.
- Van Haneghan, J.P., Barron, L., Young, M.F., Williams, S.M., Vye, N.J., & Bransford, J.D. (1992). The Jasper series: An experiment with new ways to enhance mathematical thinking. In D.F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 15-38). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Verschaffel, L. (ter perse). Mathematical modeling and problem-solving skills in upper elementary school children: analysis and improvement. In J. Hamers & J. Van Luit (Eds), *Thinking skills and teaching thinking*. Lisse: Swets & Zeitlinger.
- Verschaffel, L., & De Corte, E. (1996). Leren realistisch modelleren en interpreteren van vraagstukken. Een onderzoek bij leerlingen van de bovenbouw van de basisschool. *Pedagogische Studiën*, 73, 322-337.
- Verschaffel, L., De Corte, E., Van Vaerenbergh, G., Lasure, S., Bogaerts, H. & Ratinckx, E. (1997). *Vaardig oplossen van contextgebonden wiskunde problemen in de bovenbouw van de basisschool. Onderwijskundig Beleids- en Praktijgericht Wetenschappelijk Onderzoeksproject n° 94.07 in opdracht van het Ministerie van de Vlaamse Gemeenschap, Departement Onderwijs: eindrapport*. Faculteit Psychologie en Pedagogische Wetenschappen, Centrum voor Instructiepsychologie en -Technologie, Katholieke Universiteit Leuven.

Notities en Commentaren

INTERBEOORDELAARSOVEREENSTEMMING BIJ HET METEN VAN DE KWALITEIT VAN HET ONDERWIJZEN MET BEHULP VAN EEN BEOORDELINGSSCHAAL

Margo G.H. Jansen & Klaas J. Westerhof

Abstract

Quality of teaching is one of the key concepts in studies into educational effectiveness. In the project 'School en Instructie-Effectiviteit' a high inference observational instrument (KAS) measures quality of teaching. The instrument consists of 38 items, a priori grouped into five subscales. In this paper we focus on the interrater reliability of the KAS. The data were obtained in a training session where 11 observers applied the scale to the same video taped mathematics lesson. Due to missing data three items had to be excluded from further analysis. It was found that the reliability, obtained by computing Cronbach's α of the 35 remaining items was moderately high ($\alpha = .84$). The interrater agreement coefficient, which incorporates the differences in mean ratings between the observers was .82. From this we can infer that a mean rating of at least four independent observers will be necessary to ensure reasonably dependable scores with the KAS.

Inleiding

In het model 'leren op school' van Carroll (1963) wordt de kwaliteit van de instructie gezien als een variabele die de leerwinst beïnvloedt. Kwaliteit van de instructie wordt verondersteld positief samen te hangen met leerwinst (Walberg, 1986). De kwaliteit van de instructie wordt gedefinieerd op het niveau van het onderwijsleerproces. Het is een multidimensioneel begrip. Volgens Creemers (1994) wordt de kwaliteit van de instructie ten minste bepaald door de kwaliteit van het curriculum, de groeperingsvormen en het leerkrachtgedrag tijdens het onderwijsleerproces. Leerkrachtgedrag tijdens het onderwijsleerproces zullen we in het vervolg aanduiden als onderwijzen. Hier zullen we ons bezighouden met de vraag hoe de kwaliteit van het onderwijzen betrouwbaar is vast te stellen.

De kwaliteit van het onderwijzen kan worden vastgelegd met behulp van 'high-inference' observatie-instrumenten. Beoordelaars moeten vragen beantwoorden die verwijzen naar leerkrachtgedrag zoals: de leerkracht geeft de inhoud van de leerstof logisch weer, de leerkracht is enthousiast, de leerkracht verwijst naar voorafgaand leren. Hierbij wordt gebruik gemaakt van vier- of vijfpuntsresponsieschalen. De aldus verzamelde data worden verondersteld tenminste van ordinaal meetniveau te zijn. Bij het beoordelen van de betrouwbaarheid van een dergelijk instrument zijn twee aspecten van belang. Ten eerste de vraag of het instrument een schaal vormt en ten tweede of er sprake is van interbeoordelaarsbetrouwbaarheid. De eerste vraag heeft betrekking op de interne consistentie van het instrument. Bij de tweede vraag gaat het erom of de beoordelaars tot dezelfde score komen. Deze vragen zijn in het Project School-en Instructie-effectiviteit (SIE-project) beantwoord met gegevens die werden verkregen uit de training die de observatoren kregen voorafgaand aan het veldonderzoek.

In het SIE-project wordt het model voor leren op school van Carroll (1963) beproefd aan de hand van het vak wiskunde in VBO/MAVO en MAVO/HAVO brugklassen. De kwaliteit van het onderwijzen daarbij een van de onafhankelijke variabelen. Deze moest in het schooljaar 94/95 worden gemeten bij 56 leerkrachten in 28 scholen. De kwaliteit van het onderwijzen werd onder andere gemeten met behulp van een ratingscale. Het onderwijzen van de leer-

krachten is vier keer geobserveerd omdat een gemiddeld oordeel over meerdere observaties representatiever voor het onderwijzen van de leerkrachten wordt geacht dan een enkele observatie. Gezien de hoeveelheid observaties, de geografische spreiding ervan en de tijdsspanne waarbinnen ze moesten worden gedaan, is er gebruik worden gemaakt van meerdere observatoren. In het onderzoek werd gebruik gemaakt van 11 observatoren die van tevoren werden getraind in het gebruik van de ratingscale. De training bestond uit bestuderen en bespreken van de theoretische achtergrond van de ratingscale en vijf proefobservaties met behulp van op video opgenomen lessen. Na iedere les werden de afzonderlijke scores vergeleken en werden afspraken over toekomstige scoring gemaakt. De training werd afgesloten met de observatie van een les, die eveneens was opgenomen op videoband. Alle elf beoordelaars observeerden dezelfde wiskundeles met behulp van de ratingscale. Er is derhalve sprake van een tweedimensionele datastructuur, beoordelaars bij vragen, zonder replicaties, omdat er slechts sprake is van één object (een les).

De vraag is nu of er sprake is van een intern consistente schaal en of de observatoren overeenstemmen in de hoogte van de score. Alvorens we deze vragen behandelen, zullen we het observatie-instrument bespreken.

Het observatie-instrument

Uitgangspunten bij de constructie

Bij het construeren van de ratingscale is uitgegaan van de volgende uitgangspunten:

1. De schaal moet passen bij de operationele onderzoekspopulatie.
2. De schaal moet aansluiten bij onderzoeksliteratuur waarin als effectief beschreven vormen van onderwijzen wordt beschreven.

Ad 1: De onderzoekspopulatie.

De operationele populatie in het SIE-project bestaat uit de brugklassen VBO/MAVO/MAVO/HAVO. Van de leerlingen in deze populatie neemt men aan (Boekaerts e.a. 1995, p. 175) dat ze baat hebben bij directe instructie. Hiermee wordt een onderwijsmodel bedoeld dat sterk leerkrachtgestuurd is en dat uit een aantal fasen bestaat:

- 1) dagelijkse terugblik,
- 2) presentatie,
- 3) oefenen,
- 4) individueel verwerken,
- 5) terugkoppelen en periodieke terugblik.

Ad 2: Effectief onderwijzen

In de literatuur bestaat een brede consensus (Shulman, 1986, Westerhof 1989, Creemers, 1994, Boekaerts e.a., 1995, Lowyck e.a., 1995) over wat als een effectief onderwijzen moet worden beschouwd. Het is het proces waarin in een veilige omgeving de leerkracht de denkprocessen (cognities) van de leerlingen stuurt en de leerlingen actief bezig zijn zich leerinhouden eigen te maken. Uitgaande van deze consensus en het onderwijsmodel beschreven onder ad 1 is een observatie-instrument ontwikkeld voor het meten van de kwaliteit van het onderwijzen. Het observatie-instrument bestaat uit 8 categorieën. Categorie 1 heeft te maken met het activeren en herstructureren van voorkennis. Binnen deze categorie wordt geobserveerd of de leerkracht verwijst naar voorafgaand leren (terugblik). Bijvoorbeeld door de bedoeling van de les aan te geven, te verwijzen naar de vorige les of het bespreken van huiswerk. Categorie 2 betreft de wijze waarop nieuwe kennis wordt geïntroduceerd (wordt de nieuwe leerstof in een groter geheel aangeboden of in kleine hoeveelheden?), de helderheid van de presentatie en het gebruik van voorbeelden. Binnen categorie 3 wordt nagegaan of de leerkracht controleert of nieuwe kennis ook begrepen is. Bijvoorbeeld door het stellen van vragen en het laten verbaliseren van de gevolgde werkwijze. Aan de hand van categorie 4 wordt de kwaliteit van

het toepassen van en het oefenen met de leerstof nagegaan. Er wordt op gelet hoe de leerkracht de leerlingen instrueert hoe te oefenen en of deze daarop controle uitoefent. Categorie 5 'huiswerk' sluit daarop aan. Hier wordt geobserveerd of de leerkracht al dan niet duidelijk aangeeft hoe het huiswerk moet worden gemaakt. Met behulp van categorie 6 wordt geobserveerd of de leerkracht de denkprocessen al dan niet stuurt. Bijvoorbeeld door deze te vergelijken eigen denkprocessen of door het laten vergelijken van de eigen oplossingsmethoden met die van medeleerlingen. Binnen categorie 7 wordt nagegaan waaraan succes/falen wordt toegeschreven, of de leerkracht enthousiast is en meer of minder leertaakgericht is. Categorie 8 tenslotte gaat na hoe leertaakgerichtheid van de leerlingen wordt bevorderd. Bijvoorbeeld door het signaleren van overzicht, het voorkomen van wachttijden en het alert houden van de leerlingen.

Het antwoordformaat

De ratingscale bestaat uit een vierpuntsschaal waarbij de uiteinden van de schaal de situatie in de klas beschrijven en de observator moet aangeven welke situatie het beste de situatie in de klas beschrijft. Er is dus sprake van de volgende vier mogelijke situaties

- 1: situatie 1 beschrijft de klas duidelijk
- 2: situatie 1 beschrijft de klas het beste maar er is enig bewijs voor situatie 2
- 3: situatie 2 beschrijft de klas het beste maar er is enig bewijs voor situatie 1
- 4: situatie 2 beschrijft de klas duidelijk

Wanneer situatie 1 van toepassing is wordt een score van 1 toegekend etc. Er is voor de vorm gekozen op grond van ervaringen binnen het ISERP-project (Swint, 1993). De schaal bevat in totaal 38 items. De aantallen items per categorie 1 t/m 8, zijn respectievelijk 5, 6, 5, 3, 1, 4, 3 en 11. In bijlage 1 voor een voorbeeld van een item binnen elk der categorieën gegeven.

Een belangrijk probleem dat zich bij het doen van observaties in een veldsituatie voordoet is dat het van te voren niet is te voorzien of alle gedragingen die in de items van de ratingscale worden omschreven, ook daadwerkelijk geobserveerd kunnen worden. Daarom is het nodig een vijfde observatiecategorie 'niet te scoren' toe te voegen. Hiermee is gegeven dat er ontbrekende waarnemingen zullen zijn. Bij de verwerking van de gegevens zal daar rekening mee gehouden moeten worden.

Resultaten

Bij de eerste bewerking bleek dat er bij 3 van de 38 items te veel ontbrekende waarnemingen waren omdat de daarin beschreven gedragingen niet konden worden geobserveerd. Het betrof de items 20, 26 en 27. Bij de volgende analyses worden deze items buiten beschouwing gelaten. Categorie 5 'huiswerk' komt daardoor te vervallen. De overige missende data zijn per observator vervangen door het schaalgemiddelde.

Betrouwbaarheid als interne consistentie

De betrouwbaarheid van het meetinstrument is in termen van interne consistentie te beschrijven, als we de beoordelaars opvatten als items en de items als subjecten. Cronbachs alpha voor de gehele schaal bedraagt dan .84. De schaal is voldoende betrouwbaar. Met uitzondering van categorie 3, controle op begrijpen die een negatieve uitkomst vertoont, zijn de afzonderlijke categorieën als intern consistent te beschouwen. De alpha's van de overige categorieën lopen van .50 tot .95. Hierbij moet worden opgemerkt dat de categorieën die expliciet bedoeld zijn na te gaan hoe de leerkracht de cognities van de leerlingen beïnvloedt, begrijpen en metacognities, het slechtst scoren. Voor een overzicht, zie tabel 1.

De betrouwbaarheid in termen van interne consistentie zoals boven bepaald drukt uit dat de observatoren consistent beoordelen. Ze gaven of uniform hoge of uniform lage scores en zor-

gen daarmee voor een hoge gemiddelde correlatie tussen de 'items'. Dat zegt echter nog niets over de vraag of de items (beoordelaars) ook dezelfde gemiddelde score hebben.

Tabel 1: Overzicht van de categorieën van de ratingscale bedoeld om de kwaliteit van het onderwijs te meten, aantal items per categorie en de alpha-waarde van de schaal en de afzonderlijke categorieën.

Categorie	Schaal	Aantal items	Itemnummers	Alpha
		35		
1	Activeren voorkennis	5	1-5	.80
2	Nieuwe kennis	6	6-11	.87
3	Controle op begrijpen	5	12-16	negatief
4	Toepassen en oefenen	3	17-19	.95
6	Metacognities	4	21-24	.50
7	Klassenklimaat	1	25	
8	Taakgerichtheid	11	28-38	.67

Intraklassencorrelatiecoëfficiënten

Wanneer we de data beschouwen als gemeten op intervalniveau, dan zijn de maten voor beoordelaarsovereenstemming veelal gedefinieerd als ratio's van variantiecomponenten. Dergelijke ratio's staan bekend onder de naam van intraklassencorrelatiecoëfficiënten. In Eggen & Sanders (1993) wordt de meest voorkomende situatie behandeld, waarbij een aselechte steekproef beoordelaars een aselechte steekproef objecten beoordeelt (Eggen e.a., p. 457). In de in dit onderzoek gebruikte data-set spelen de items op de ratingscale de rol van object. Bij elf beoordelaars en 38 items ziet de datastructuur er als volgt uit.

Figuur 1: Datamatrix voor een gekruist design met twee factoren.

	1	2	.	b	.	k
1						
2						
.						
p				X_{pb}		X_p
.						
N				X_b		$X_{..}$

De gebruikte symbolen hebben de volgende betekenis:

- k = het aantal beoordelaars = 11,
- n = het aantal beoordeelde items = 38,
- p = de index voor items, met $p = 1, \dots, 38$,
- b = de index voor beoordelaars, met $b = 1, \dots, 11$,
- X_{pb} = de score voor item p van beoordelaar b,
- X_p = de somscore over beoordelaars voor item p,
- X_b = de somscore over items voor beoordelaar b,
- $X_{..}$ = de som van alle scores over items en beoordelaars.

De beoordeling (score) van item p door beoordelaar b wordt beschreven als:

$$X_{pb} = \mu + \alpha_p + \beta_b + \alpha\beta_{pb} + \epsilon_{pb} \quad 1$$

Het is een lineair model waarin naast een algemeen gemiddelde, een persoons-effect, een item-effect, een interactie-effect en een foutencomponent worden onderscheiden. Elk van deze effecten heeft een variantie die wordt aangeduid als variantiecomponent. Omdat er sprake is van slechts een replicatie per cel, is het niet mogelijk de foutencomponent en de interactiecomponent afzonderlijk te schatten. Het model wordt dan:

$$X_{pb} = \mu + \alpha_p + \beta_b + \epsilon_{pb} \quad (2)$$

met

$$\alpha_p = \mu_p + \mu$$

$$\beta_b = \mu_p + \mu$$

$$\epsilon_{pb} = X_{pb} - (\mu_p + \mu) - (\mu_p + \mu)$$

Wanneer beide factoren random zijn leidt dat niet tot 'bias' in de toetsing. Dit is wel het geval wanneer we een meer realistisch gemengd model zouden gebruiken met alleen de beoordelaars als 'random' factor en items als een vast effect (Kirk, 1968). De foutenterm en de interactie worden samengevat tot een component die we de residuele component zullen noemen.

De schatting van de variantiecomponenten kan als volgt uit de gemiddelde kwadratensommen worden berekend:

$$\sigma_\alpha^2 = (MS_p - MS_{res})/k$$

$$\sigma_\beta^2 = (MS_b - MS_{res})/n$$

$$\sigma_\epsilon^2 = MS_{res}$$

De beoordelaars overeenstemmingcoëfficiënt, ρ voor k beoordelaars is dan

$$\rho^2 = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + (\sigma_\beta^2 + \sigma_\epsilon^2)/k} \quad (3)$$

Formule (3) geeft aan in welke mate de gemiddelde beoordeling van k willekeurig gekozen beoordelaars overeenstemmen met de gemiddelde beoordeling van k andere, eveneens willekeurig gekozen beoordelaars. De invloed van het aantal beoordelaars op de mate van overeenkomst is te bepalen door k te variëren. Indien $k = 1$ dan is de coëfficiënt de overeenkomst tussen één willekeurig gekozen beoordelaar en andere eveneens willekeurig gekozen beoordelaars. Indien $k = 2$ dan geeft de coëfficiënt de gemiddelde overeenkomst aan tussen 2 willekeurig gekozen beoordelaars en 2 andere, eveneens willekeurig gekozen beoordelaars (Eggen & Sanders, 1993).

In het volgende zullen de resultaten van de analyse voor de matrix van 35 items * 11 beoordelaars worden weergegeven. Het algemeen gemiddelde bedraagt 2.26.

Tabel 3: Variantiebronnen van de KAS-ratingscale

Variantiebron	Kwadratensom	Vrijheidsgraden	Gemiddelde Kwadratensom
Tussen items (SS_p)	113.96	034	3.35
Binnen items	211.71	350	0.60
Tussen Beoordelaars (SS_b)	030.58	010	3.06
Residu (SS_{res})	181.66	340	0.53

Substitutie in formule (3) levert voor 11 beoordelaars een interbeoordelaarsovereenstemming op van .82. De formule geeft aan in welke mate de gemiddelde beoordeling van 11 willekeurig gekozen beoordelaars overeenstemt met de gemiddelde beoordeling van k andere, eveneens willekeurig gekozen beoordelaars. In onderwijskundig onderzoek is het echter zo dat er niet 11 willekeurig gekozen observatoren naar de ene en 11 willekeurig gekozen observatoren naar een school worden gestuurd om het onderwijzen te observeren. Tijd, geld en belasting van het veld laten doorgaans maar toe dat er één observator wordt gestuurd. Als deze behoort tot een groep observatoren, is de vraag interessant hoe vaak er door een willekeurig gekozen observator moet worden geobserveerd, wil de gemiddelde beoordeling door k observatoren voldoende betrouwbaar zijn. Uitgaande van de resultaten tijdens de training is het dus de vraag hoe vaak geobserveerd moet worden voordat het gecombineerde oordeel een redelijke betrouwbaarheid heeft. Dat is een andere formulering van de vraag in welke mate het aantal observatoren de mate van overeenstemming beïnvloedt (Eggen e.a. 1993, p. 460). In tabel 4 is de mate van overeenkomst tot 11 willekeurig gekozen observatoren weergegeven. In formule 3 varieert k van 1 tot 11.

Tabel 4: Mate van overeenkomst tussen aantallen willekeurig gekozen beoordelaars voor de gehele KAS-ratingscale

k =										
1	2	3	4	5	6	7	8	9	10	11
.30	.46	.56	.63	.68	.72	.75	.77	.79	.81	.82

Uit de tabel blijkt dat voor een redelijke gemiddelde overeenstemming tussen observatoren er minimaal 4 keer maar liever nog 5 of 6 keer door een willekeurig gekozen beoordelaar zou moeten worden geobserveerd.

Discussie

De betrouwbaarheid kan eveneens worden gedefinieerd als een ratio van variantiecomponenten, n.l.:

$$\rho^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2 / k} \quad (4)$$

Het verschil met (3) is dat de geschatte variantiecomponent voor de beoordelaars ontbreekt. Met andere woorden, er wordt geen rekening gehouden met de verschillen in gemiddelde, tussen de beoordelaars. Formule (4) is de formule waarmee Cronbachs alpha wordt berekend. Een hoge beoordelaarsbetrouwbaarheid is een nodige maar niet voldoende voorwaarde voor

een hoge beoordelaarsovereenstemming. Wanneer we Cronbachs alpha gebruiken laten we dus een belangrijk aspect van de beoordelaarsovereenstemming buiten beschouwing.

Als te beoordelen materiaal is in het onderzoek gebruik gemaakt van slechts één wiskundes die was opgenomen op videoband. Het is dus maar de vraag of de resultaten kunnen worden gegeneraliseerd naar andere lessen en andere observatieomstandigheden. Het zou aanbeveling verdienen om het onderzoek te herhalen voor een groter aantal lessen. Het probleem hiervan is de praktische uitvoerbaarheid, met name wanneer we alle observatoren alle lessen willen laten beoordelen.

Referenties

- Boekaerts, M., R.J. Simons (1995). *Leren en instructie. Psychologie van het leren en het leerproces*. Assen: Dekker & Van der Vegt.
- Brophy, J. and H.L. Goog (1986). Teacher Behavior and Student Achievement. In: M.C. Wittrock (ed.), *Handbook of Research on Teaching*. Third Edition. New York: Macmillan/Collier MacMillan.
- De Corte, E. (1992). Design and evaluation of powerful learning environments. In: B.P.M. Creemers, G.J. Reezigt, *Evaluation of Educational Effectiveness*. Groningen: ICO.
- Creemers, B.P.M. (1994). *The effective classroom*. London: Cassell.
- Eggen, T.J.H.H., P.F. Sanders (eds.) (1993). *Psychometrie in de praktijk*. Arnhem: CITO Instituut voor Toetsontwikkeling.
- Good, T. (1983). Research on Classroom Teaching. In: S. Shulman and G. Sykes (eds.), *Handbook of Teaching and Policy*. London: Longman.
- Kirk, R.E. (1968). *Experimental Design: Procedures for the Behavioral Sciences*. Belmont, California: Brooks/Cole Publishing Company, Inc.
- Lowyck, J., N. Verloop (eds.) (1995). *Onderwijskunde. Een kennisbasis voor professionals*. Groningen: Wolters-Noordhoff.
- Schaffer, E.C., P.S. Nesselrodt (1992). *The development of the Special Strategies Observation System*. Department of Curriculum and Instruction, University of Carolina at Charlotte. A paper presented at the American Educational Research Association Annual Meeting in San Francisco, April 20.
- Shulman, L.S. (1986). Paradigms and Research Programs in the Study of Teaching. A Contemporary Perspective. In: M.C. Wittrock (ed.), *Handbook of Research on Teaching*. Third Edition. Washington D.C.: American Educational Research Association.
- Swint, F. (1993). *Construction of a Classroom Observation Instrument*. Internal Memo. Groningen: GION.
- Walberg, H.J. (1986). Synthesis of Research on Teaching. In: M.C. Wittrock (ed.), *Handbook of Research on Teaching*. Third Edition. Washington D.C.: American Educational Research Association.
- Westerhof, K.J. (1989). *Effectiviteit van leerkrachtgedrag. Een empirische studie naar leerkrachtgedrag en de samenhang met leerwinst*. Ph.D. Dissertation.

INZICHTELIJKHEID VAN SCORINGSREGELS EN TOLERANTIE VOOR FOUTEN

A. Dirkwager (e-mail: aried@xs4all.nl)

Wanneer proefpersonen (studenten) een toets maken is het wenselijk dat de mogelijke consequenties van hun antwoorden op de items hen duidelijk zijn, alleen dan zijn zij in de gelegenheid hun antwoorden optimaal te geven (kiezen). Met name wanneer voor foutieve antwoorden punten worden afgetrokken, zoals bij "correctie voor gissen", "ja-?-nee" (van Til, 1998) of Multipele Evaluatie (Dirkwager 1981, 1998), is het van belang dat zij duidelijk weten welk risico zij lopen als zij antwoorden.. Wanneer dit het geval is is de scoringsregel "inzichtelijk" en mogen we aannemen dat er zodanig gerespondeerd wordt dat iedere geteste persoon -gegeven zijn capabiliteiten- zijn verwachte score optimaliseert. Deze laatste hypothese kan met een antwoordanalyse statistisch getoetst worden. Wanneer de hypothese dat de proefpersonen hun verwachte score optimaliseren verworpen moet we concluderen dat de scoringsregel niet "inzichtelijk" was.

Een scoringsregel kan op drie manieren "inzichtelijk" worden gemaakt. Ten eerste door mathematische analyse, ten tweede door een algemene toetsinstructie waarin gezegd wordt hoe men de vragen moet beantwoorden (deze instructie kan, gezien de resultaten van de mathematische analyse, misleidend zijn), en ten derde door het geven van de te verwachten score bij iedere antwoordmogelijkheid voor ieder item. Bij de laatste manier staat de scoringsregel bij de toetsafname weliswaar reeds vast, maar zij behoeft, evenmin als de mathematische achtergrond, voor de proefpersoon inzichtelijk te zijn - de proefpersoon krijgt volledige duidelijkheid over de consequenties van zijn mogelijke antwoorden en daar gaat het om.

Een mathematische analyse gaat doorgaans het begrip van proefpersonen (en van docenten die een toets afnemen?) te boven, maar zij biedt wel een solide basis voor een toetsinstructie die tot optimalisering van de verwachte score leidt. Dit sluit echter de mogelijkheid niet uit dat de toetsinstructie misleidend is of niet goed begrepen wordt omdat zij te ingewikkeld is of niet duidelijk genoeg is. Blijft dus de noodzaak van een goede mathematische analyse en een simpele manier om de consequenties van mogelijke antwoorden aan de proefpersonen duidelijk te maken.

Cita van Til (1998) hanteert drie antwoordopties, ja, ? en nee, met als scoringsregel nul punten als "?", plus een punt als juist en min een punt als onjuist. Drie dingen kunnen hier voor aangetoond worden: 1. De verwachte score wordt gemaximaliseerd wanneer nooit het vraagteken gekozen wordt, en 2. De meetfoutvariantie is een functie van het kennisniveau, van de feitelijke kans-op-juist, en maximaal als deze kans 50% is, 3. Studenten met het laagste kennisniveau profiteren het meest van deze meetfoutvariantie doordat zij een hoge kans hebben (ten onrechte) een voldoende te behalen door toevalsfluctuaties, studenten met een hoog kennisniveau lopen weinig risico daar de meetfoutvariantie voor hen klein is.

Het proefschrift van Cita van Til (1998) geeft empirische evidentie in hoeverre studenten zich overeenkomstig een optimale strategie gedragen en dus in hoeverre de gegeven toetsinstructie inzichtelijk en correct was. Het blijkt dat gemiddeld in 63,4% van de gevallen het vraagteken ("met zekerheid nul punten") gekozen is. Dit hoge percentage maakt het zeer onwaarschijnlijk dat de studenten een optimale strategie hebben gevolgd. Uitgesplitst naar algemeen kennisniveau rapporteert zij helaas het percentage vraagtekens niet, daaruit zou kunnen blijken dat alle studenten die niet voldoende zijn (terecht) op safe spelen en frequenter in geval van twijfel geen risico nemen en het vraagteken kiezen, in tegenstelling tot degenen met de minste kennis en degenen met ruim voldoende kennis die er baat bij hebben hun verwachte score te maximaliseren door nooit het vraagteken te kiezen. Gezien haar conclusie op pag. 151 ("Geconcludeerd kan worden dat deze antwoordinstructie tenminste aangescherpt moet worden, en bovendien de gehanteerde scoringsregel heroverwogen moet worden") is het echter ook zeer onwaarschijnlijk dat deze strategie gevolgd is.

Dirkzwager (1996) stelt een logaritmische scoringsregel voor, waarbij de studenten de likelihood rapporteren dat "ja" dan wel "nee" juist is, de persoonlijke inschatting van de kans-op-juist die vooraf gaat aan de keuze van "ja", "nee", of "?". Hiervoor kan aangetoond worden dat: 1. De verwachte score wordt gemaximaliseerd als deze persoonlijke schatting overeenkomt met de feitelijke kans-op-juist, als de student realistisch is wat betreft zijn kennis aangaande een item 2. De meetfoutvariantie een functie is van het kennisniveau en de gerapporteerde likelihood: voor realistische studenten is zij gelijk aan nul wanneer de feitelijke kans-op-juist 50%, niet hoger dan toeval, is, of wanneer deze kans 100% (perfecte kennis) is en maximaal bij onzekere "kennis", bij een kans-op-juist van ongeveer 80% en dus een kans-op-onjuist van 20%, 3. Studenten met een laag kennisniveau niet kunnen profiteren van meetfoutvariantie en ook bij voorgewende kennis vrijwel geen kans op een voldoende hebben.

Wat de door Dirkzwager (1996) voorgestelde scoringsregel betreft wijst van Til (terecht) op het bezwaar dat deze voor studenten weinig inzichtelijk is. Men zal in de instructie zich dan ook niet moeten concentreren op deze (mathematisch gefundeerde) regel, maar de studenten duidelijk laten zien hoeveel punten zij kunnen verwachten bij een goed, cq. een fout antwoord. In plaats van de door van Til gehanteerde zeer subjectieve schaal van 0%, 25%,

50%, 75% en 100% "zeker" (waarbij het voor de student geenszins duidelijk is wat deze percentages betekenen) zou een objectieve schaal van 50% tot 100% kans-op-juist gehanteerd moeten worden met opgave van het aantal punten dat men wint (cq. riskeert te verliezen) afhankelijk van de positie op deze schaal die men aankruist, overeenkomstig een goed verantwoorde scoringsregel. Dan zijn de consequenties van zijn antwoorden voor de student volstrekt "inzichtelijk", hoewel mogelijk voor de mathematisch niet geschoolde docent "onbegrijpelijk", en men is psychometrisch verantwoord bezig. Bij ja/?/nee vragen zou zo'n schaal er als het volgt uit kunnen zien (aangenomen dat bij de helft van de items "JA" juist is en bij de helft "NEE"):

Tabel 1: Relatie tussen antwoordopties en scores (Multipele Evaluatie).

	kans op juist	kans op onjuist	score als juist	score als onjuist
A. JA??	tussen 50% en 64%	tussen 36% en 50%	+ 1	- 1
NEE??	tussen 50% en 64%	tussen 36% en 50%	+ 1	- 1
B. JA?	tussen 64% en 87%	tussen 13% en 36%	+ 6	-10
NEE?	tussen 64% en 87%	tussen 13% en 36%	+ 6	-10
C. JA.	tussen 87% en 95%	tussen 5% en 13%	+ 9	-30
NEE.	tussen 87% en 95%	tussen 5% en 13%	+ 9	-30
D. JA!	tussen 95% en 98%	tussen 2% en 5%	+10	-50
NEE!	tussen 95% en 98%	tussen 2% en 5%	+10	-50
E. JA!!	100%	0%	+20	-500
NEE!!	100%	0%	+20	-500

Bij een toets van 100 vragen en een cut off score van 60 tussen "voldoende" en "onvoldoende" kan men op verschillende manieren een voldoende halen. Hoe een student een voldoende haalt hangt af van de wijze waarop hij gestudeerd heeft en van in hoeverre hij zijn eigen kennis juist weet in te schatten. Vier van de vele mogelijkheden (bij een test van 100 vragen) zijn:

1. Men weet niets zeker, kiest altijd "A", maar beantwoordde toch 80% juist.
2. Men weet van elke vraag "iets" (feitelijke kans-op-juist voor iedere vraag 67%), kiest altijd "B" en heeft van de honderd er inderdaad 67 "juist".
3. Men meent het antwoord op 28 vragen perfect te weten ("E") en heeft bij de overige vragen geen idee ("A", en de gemiddelde score is nul voor die vragen). In dat geval mag men zich een keer vergissen ("E" onjuist geeft -500 punten plus 28 maal 20 punten geeft een totaalscore van 60 punten).
4. Men weet drie vragen perfect ("E" juist) en heeft voor de overige 93 geen idee ("A" met een gemiddelde score van nul punten).

Uit "1" blijkt dat deze methode redelijk tolerant is voor de strategie-fout dat men een te lage optie kiest (dat men wat zijn kennis betreft "te bescheiden" is), uit "3" en "4" blijkt dat deze methode vrijwel intolerant is voor de strategie-fout dat men doet alsof men over betrouwbare kennis beschikt als dat niet zo is - dit in tegenstelling met de door van Til gehanteerde scoringsregel waar de beste strategie is nooit toe te geven dat men onkundig is.

Deze vier mogelijkheden komen overeen met vier studententypen.

"1" is een student die best wel het een en ander vrij goed weet (gemiddelde kans-op-juist is 80%) maar elke zelfkennis wat dit betreft mist. Het is waarschijnlijk een student die bij ja/?/nee hoe dan ook het vraagteken zou kiezen. Dit type student komt echter niet voor. Uit de data van van Til (1998) blijkt dat vraagtekenkeuze correspondeert met gemiddeld een kans-op-juist niet hoger dan toeval (54%) en dat bij een gemiddelde feitelijke kans-op-juist van 83% vrijwel altijd een antwoord gegeven wordt.

"2" is de generalist die van alles iets weet maar niets perfect.

"4" is de specialist die van heel weinig alles perfect weet en van al het overige niets.

"3" is ook een specialist, maar dan een waarvan wij als beoordelaar het acceptabel achten dat hij in een op de 29 gevallen volledig de fout ingaat. Wanneer we dit niet acceptabel achten (intolerant zijn voor het ten onrechte voorwenden van perfect betrouwbare kennis) kunnen we de score-als-onjuist voor "E" nog lager stellen dan -500, bijvoorbeeld -1850 voor het geval dat alle vragen op zijn specialisme betrekking hebben en we slechts één fout op de honderd willen tolereren: 99 maal 20 geeft 1910 punten minus 1850 geeft een score van 60 punten, net niet voldoende.

Tolerantie voor beoordelingsfouten van studenten

Naast vier typen studenten kunnen we ook twee soorten fouten onderscheiden die studenten kunnen maken.

Ten eerste kan de student te zeker zijn van zijn antwoord (hij voelt zich een "specialist" met een hoge kans-op-juist maar is in feite een "generalist" met een veel lagere kans-op-juist, hij doet zich "zekerder" voor dan gerechtvaardigd is).

Ten tweede kan een student (ongeacht zijn "zekerheid") een onjuist antwoord de hoogste kans op juist geven (een onjuist antwoord kiezen, ongeacht zijn "zekerheid").

Traditionele Multiple Choice scoring, waarvan ja/?/nee een speciaal geval is, houdt geen rekening met het eerste soort fout en is daar dus maximaal tolerant voor. De tolerantie voor het tweede soort fout is ook maximaal: zelfs als de student maximaal onzeker is moet hij -ongeacht zijn feitelijke kans-op-fout- altijd een antwoord geven om een zo hoog mogelijke score te krijgen. Een fout wordt steeds even zwaar "bestraft" met negatieve punten die met één goed antwoord op één andere vraag gecompenseerd kunnen worden.

Bij Multipele Evaluatie is dat ook het geval wanneer de student toegeeft dat hij het antwoord niet (voldoende) zeker weet, maar naarmate hij meer overtuigd zegt te zijn dat zijn (onjuiste) antwoord goed is wordt hij zwaarder "gestraft" met min-punten. Bij de scoring volgens tabel 1 zijn zelfs 25 perfecte antwoorden nodig om één fout als "specialist" te compenseren. Wanneer de vraag een praktijkgeval betreft waar een foute beslissing van een specialist ernstige gevolgen kan hebben is dit nog erg tolerant: zoals we zagen is 1 fout op de 28 (3.6% fout) toegestaan om een voldoende te krijgen. Een voorbeeld van een minder tolerante scoringsregel (1 fout op de honderd toegestaan) werd hierboven gegeven.

Tolerantie voor beoordelaarsfouten

Een docent-beoordelaar die zijn oordeel baseert op een objectieve toets maakt onvermijdelijk ook beoordelingsfouten. Als hij zich aan de toetsmethode en scoringsregel houdt vind deze beoordelingsfout zijn oorsprong in de aan toevalsfluctuaties toe te schrijven meetfout per item en student. We zagen dat deze meetfout afhankelijk is van het kennisniveau van de student in relatie tot het item - bij perfecte kennis is die meetfout altijd gelijk aan nul. Bij Multiple Choice (ja/?/nee) is zij maximaal als alle kennis ontbreekt en de kans-op-juist gelijk is aan 50%, in dat geval is zij bij Multipele Evaluatie ook gelijk aan nul. Bij Multipele Evaluatie is de meetfout maximaal bij imperfecte of onvolledige kennis, kans-op-juist ongeveer 80% en dus de kans dat de student met zijn bewering ongelijk heeft gelijk aan ongeveer 20%, in dat geval kunnen we met Multipele Evaluatie het kennisniveau niet nauwkeurig meten en zullen we dus vrij dikwijls de beoordelingsfout maken dat we -afhankelijk van de cut-off score- ten onrechte een onvoldoende toekennen, dan wel ten onrechte een voldoende geven. Bij extreme kennisniveaus is de kans op zo'n beoordelingsfout vrijwel gelijk aan nul. Bij Multiple Choice ligt dat anders omdat daarbij de meetfout bij volstrekte afwezigheid van kennis maximaal is - de kans dat ten onrechte een voldoende gegeven wordt is dan dus veel groter. De keuze welk risico op welk type beoordelingsfout hij wil lopen is uiteraard aan de docent-beoordelaar, maar het lijkt voor de hand te liggen het risico om -bij afwezigheid van kennis- ten onrechte een voldoende toe te kennen zwaarder te laten wegen dan het risico om bij imper-

fecte kennis een onvoldoende te geven. Dit pleit voor Multiple Evaluatie boven Multiple Choice.

Literatuur

Dirkzwager, A. (1981) *Multipel Evaluatie in plaats van Multiple Choice*. Tijdschr. v. Onderwijsresearch 1981,6,230-246

Dirkzwager, A. (1996a) *TestBet. Programma en Handleiding*. Zwijssen, Tilburg.

Dirkzwager, A. (1996b) *Scoringsregels en Toetsinstructie*. Tijdschr. v. Onderwijsresearch 1996,21,272-277

Dirkzwager, A. (1998) *A Bayesian Testing Paradigm: Multiple Evaluation, a feasible alternative for Multiple Choice*. Submitted to Educ. & Psych. Measmnt.

Van Til, C. (1998) *Voortgang in Voortgangstoetsing. Wageningen, Ponsen & Looijen*.

Boekbesprekingen

Biemans, H.J.A. (1997)

Fostering activation of prior knowledge and conceptual change

Academisch proefschrift Katholieke Universiteit Nijmegen

Promotor: prof. dr. P.R.J. Simons

Arnhem, ISBN 90 373 0361 7

Leren is per definitie een activiteit van de lerende, daar zal niet veel discussie over zijn. Echter over hoe we de lerende kunnen leren te leren valt nog veel te zeggen. Zo brengt de lerende zijn eigen ideeën ook wel preconcepties genoemd in het leerproces mee. Dit proefschrift brengt op nauwkeurige wijze de huidige stand van zaken op dit terrein in beeld. Tevens slaagt Biemans erin om het onderzoeksgebied verder te ontginnen. Aan de hand van een activiteitsmodel van preconcepties laat de promovendus zien dat leerlingen als ze op deze wijze leren tot betere leerprestaties komen. Het proefschrift is gebaseerd op lange termijn onderzoek op het terrein van het verhogen van het leervermogen. De empirische onderzoeken zijn zorgvuldig uitgevoerd en geanalyseerd.

In de eerste twee hoofdstukken bespreekt Biemans relevante constructivistische leertheorieën en onderzoeken. Biemans bakent de centrale begrippen uit de studie af. Voorkennis, preconcepties, informele kennis, naïeve theorieën, alternatieve concepties, alternatieve frameworks en originele ideeën zijn alle synoniem voor de wijze waarop de lerende naar een bepaald verschijnsel kijkt. De overige hoofdstukken gaan in op de onderzoeken die de promovendus heeft uitgevoerd. Het proefschrift bevat veel informatie die op heldere wijze uiteen wordt gezet.

Preconcepties en conceptuele verandering

We nemen als vanzelfsprekend aan dat leerlingen nieuwe informatie die we hen leren ook als zodanig opnemen. Leerlingen brengen in het leerproces echter hun eigen ideeën mee en vervangen deze niet automatisch door de nieuwe informatie (Ali, 1990). In veel onderzoek is aangetoond dat het belangrijk is dat leerlingen als ze nieuwe informatie verwerken, gebruik maken van hun voorkennis. Voorkennis helpt om concepten aan elkaar te relateren, het leermateriaal te structureren en kan leiden tot een rijkere en betere mentale representatie bij de lerende. Leerlingen maken echter niet altijd automatisch gebruik van hun voorkennis en relateren die dan ook niet aan de nieuwe informatie. In dat geval bestaat de mogelijkheid dat de nieuwe informatie als geïsoleerde kennis geleerd wordt en ervaren leerlingen moeilijkheden bij het toepassen van de nieuwe kennis in nieuwe situaties. Van belang is dan ook leerlingen tijdens het leren van nieuwe informatie te helpen om actief gebruik te maken van hun voorkennis.

De ideeën die leerlingen hebben kunnen incompleet zijn en soms zelfs onjuist. Deze ideeën worden niet zo eenvoudig aangepast en kunnen zelfs het proces van kennisopbouw belemmeren, met name wanneer de voorkennis ofwel de incomplete en/of incorrecte delen daarvan niet gecorrigeerd worden tijdens het leerproces. Leren is dan ook niet slechts het verwerven of construeren van nieuwe concepties, maar impliceert ook de reconstructie van bestaande concepties. Dat laatste wordt aangeduid met conceptuele verandering.

Instructie

Doel van instructie c.q. het leerproces is ervoor te zorgen dat de lerende zijn concepten dusdanig verandert dat die overeenkomen met de aangeboden informatie, in casu de wetenschappelijk geaccepteerde kennis. Over de vraag hoe dit proces van conceptuele verandering gefaciliteerd kan worden heeft een aantal onderzoekers zich gebogen. Belangrijke vraagstel-

ling hierbij is welke cognitieve processen een rol spelen bij de reconstructie van kennis en welke leeromgeving of welke instructiestrategieën dit proces kunnen faciliteren. Centraal in dit proefschrift staat dan ook de vraag welk instructiesysteem leerlingen kan helpen hun onvolledige of deels correcte/deels incorrecte voorkennis dusdanig te gebruiken dat het leerproces optimaal, zonder interferentie, verloopt en dat de beoogde leerdoelen worden bereikt. Meer expliciet welke instructiestrategieën en leermiddelen kunnen het proces van activering van voorkennis en reconstructie van kennis bevorderen.

CONTACT-strategie

Vier studies zijn uitgevoerd om de voorkennis te activeren en conceptuele verandering te bewerkstelligen. Het eerste onderzoek spitte zich toe op de CONTACT-strategie waarvan de effectiviteit door Ali (1990) was aangetoond. Deze strategie is gebaseerd op theorieën met betrekking tot conceptuele verandering en reconstructie van kennis (onder andere Nussbaum & Novick, 1982, Strike en Posner, 1985, Prawat, 1989). Betreffende theorieën geven vier condities aan die alle nodig zijn om het beoogde doel te bereiken: ontevredenheid met een bestaande preconceptie, het begrijpen van een nieuwe conceptie, het geloven in de plausibiliteit daarvan en geloven dat deze vruchtbaar is. Deze condities zijn vertaald in een CONTACT-strategie die uit vijf stappen bestaat (zie voor nadere bespreking hiervan: Ali, 1990):

1. zoeken naar eigen preconcepties;
2. vergelijken en contrasteren van deze preconcepties met de nieuwe informatie;
3. formuleren van nieuwe concepties op basis van de vorige stap;
4. toepassen van de nieuwe concepties;
5. evalueren van de nieuwe concepties op basis van de vorige stap.

Onderzoek 1

Vraagstelling in dit onderzoek was in welke mate de diverse stappen van de CONTACT-strategie bijdragen aan de effectiviteit van de strategie. Biemans doet dit door het ontmantelen van de strategie in de vorm van vier instructiecondities die verschilden met betrekking tot het aantal stappen. Een groep leerlingen doorliep alle stappen, een andere groep stopte bij het formuleren van een nieuw idee. Bij een derde groep werd alleen de preconcepties geactiveerd en bij de controle groep werd geen enkele stap van de strategie toegepast.

De volledige CONTACT-strategie leidde tot conceptuele verandering in tegenstelling tot de twee ontmantelde varianten van de strategie. Opvallend was dat leerlingen die alle stappen van de strategie doorliepen, beter presteerden bij testvragen die gerelateerd waren aan de centrale begrippen uit de teksten en minder bij testvragen die niet direct gerelateerd waren aan de teksten. Biemans verklaart dit met behulp van de selectieve-aandacht-hypothese. De strategie zou de aandacht richten op de centrale begrippen uit de tekst ten koste van andere informatie. Over het algemeen verschilden de condities weinig van elkaar voor wat betreft de leerprestaties. Aan de hand van de resultaten kan dan ook geen antwoord worden gegeven op de geformuleerde vraagstelling.

Bij de retentietest presteerden alle groepen behalve de CONTACT-groep lager. Dit impliceert een duurzaam leereffect voor de laatste groep. Biemans concludeert dan ook dat de CONTACT-strategie weliswaar effectief is maar verbetering behoeft.

De keus van Biemans om de strategie te ontmantelen is opmerkelijk te noemen, daar het activeringsmodel een sequentiële benadering impliceert, waarbij alle stappen noodzakelijk zijn. In de medische wereld wordt wel eens gezegd dat het niet eenvoudig is om bij een medicijn na te gaan welk specifiek onderdeel de genezende werking veroorzaakt, daar dit veelal een gevolg is van de combinatie. Naar mijn mening is juist de combinatie van de stappen cruciaal voor het activeringsmodel. Bovendien zijn de verschillen tussen de CONTACT-strategie en de conditie waarin de leerlingen tot en met stap 3 van de strategie doorliepen niet groot te noemen. Biemans loopt met zijn onderzoeksopzet daardoor een risico.

Onderzoek 2

In het tweede onderzoek vergelijkt Biemans een aangepaste CONTACT-strategie, de CONTACT-2-strategie met de oorspronkelijke CONTACT-strategie en een controle conditie zonder activering. Doel was de oorspronkelijke strategie te verbeteren ten aanzien van flexibiliteit, efficiëntie en effectiviteit. De CONTACT-2-strategie verschilt van de oorspronkelijke strategie, met name in de mate van hulpverlening aan de leerlingen, in de uitleg en de keuzemogelijkheden van de leerlingen. Daarnaast werden ook maatregelen genomen om de effecten van de selectieve-aandacht te minimaliseren. Voor nadere informatie over deze aanpassingen is het aan te bevelen het boek van Biemans te raadplegen.

De resultaten van dit onderzoek bevestigen dat de CONTACT-2-strategie effectiever was ten aanzien van de activering van voorkennis en conceptuele verandering dan de twee andere condities. Leerlingen construeerden juiste concepties, pasten deze toe in probleemsituaties en behaalden hogere leerprestaties. Opvallend was dat de effectiviteit van de CONTACT-2-strategie niet afhankelijk was van de mate van conceptuele overeenkomst tussen de toetsvragen en de centrale begrippen uit de teksten en het tijdstip van toetsing. Bovendien bleek dat de leerlingen minder leertijd nodig hadden. Daarmee bevestigt dit onderzoek dat de oorspronkelijke strategie indien voorzien van de door Biemans ingevoerde aanpassingen inderdaad effectief en efficiënt is.

Onderzoek 3

In het derde onderzoek richt Biemans zich op een ander aspect van de strategie, namelijk de hoge mate van externe sturing. Externe sturing is de ondersteuning die de leerling krijgt om zijn preconcepties te activeren en nieuwe juiste concepties te vormen. Zoals aangetoond leidt dit tot goede reconstructie van kennis bij leerlingen. Biemans stelt in dit onderzoek centraal de vraag hoe leerlingen zelf hun eigen preconcepties kunnen activeren en zelf processen die gericht zijn op conceptuele verandering kunnen aansturen. Hiermee betreft de promovendus het terrein van het leren leren; hoe leren we de leerlingen te leren. Meer concreet hoe kunnen we ze leren hun (gedeeltelijk) correcte of (gedeeltelijk) incorrecte voorkennis te gebruiken tijdens het leerproces, op dusdanige wijze dat zij niet daaraan vasthouden maar die aanpassen aan de wetenschappelijk geaccepteerde kennis. Biemans vergeleek twee manieren van terugtrekking van externe sturing. In het ene geval werd de ondersteuning geleidelijk teruggetrokken binnen elke instructiestap terwijl in de andere conditie het aantal stappen werd teruggebracht naarmate de training vorderde. De terugtrekking binnen elke stap bleek tot de beste conceptuele verandering te leiden. Biemans concludeert dat leren het best kan starten vanuit een leeromgeving die gekenmerkt wordt door een redelijk hoge mate van externe sturing. Naarmate de training vordert en de leerling beter presteert, c.q. in staat is de vereiste leeractiviteiten te initiëren en uit te voeren, kan de externe sturing worden afgebouwd.

Onderzoek 4

Zijn laatste onderzoek spitst Biemans toe op verschillen in cognitieve leeractiviteiten tussen leerlingen die succesvol zijn in het bestuderen van teksten terwijl ze met de CONTACT-2 strategie werken en leerlingen die minder succesvol zijn. Biemans maakt in zijn onderzoek gebruik van de hardop-denken methode om na te gaan welke leeractiviteiten de leerlingen gebruiken. De bevindingen zijn dat succesvolle leerlingen betere leerresultaten halen bij de verschillende CONTACT-2-stappen dan minder succesvolle leerlingen en dat ze de verschillende instructiestappen op een betere manier uitvoeren. Daarbij blijkt de wijze waarop leerlingen de tweede stap uit de strategie uitvoeren: vergelijken en contrasteren van preconcepties met de nieuwe informatie, cruciaal te zijn voor de effectiviteit van de strategie. De vraagstelling uit dit onderzoek lijkt op een nieuwe poging van Biemans om de strategie te ontmantelen. Naar mijn mening zijn er geen of onvoldoende argumenten en resultaten om daar opnieuw onderzoek naar te doen. Verder is gekozen om via de hardop-denkmethode na te gaan welke cognitieve activiteiten de leerlingen uitvoeren, terwijl ze met de CONTACT-2 strategie wer-

ken. Dit terwijl de CONTACT-2 strategie een "interne" leerstrategie is. De vraag is in hoeverre dit de (uitvoering van) activiteiten van de CONTACT-2 strategie heeft beïnvloed. Sommige stappen laten zich minder gemakkelijk, zonder externe hulp, hardop verwoorden. Daarbij geldt dat leerlingen afzonderlijk getraind moeten worden in hardop denken. Met andere woorden: ze moeten ook leren datgene wat ze in hun hoofd doen ook daadwerkelijk te verwoorden. In deze vaardigheid verschillen leerlingen. Mogelijkerwijs doen ze iets heel anders dan bij de strategie beoogd wordt en verwoorden slimme leerlingen goed wat de onderzoeker c.q. de strategie wil. Misschien verklaart dat ook het verschil in resultaten tussen onderzoek 1 en 4. In onderzoek 1 kon namelijk niet worden aangetoond welke stap(pen) van de strategie cruciaal is (zijn). Dit onderzoek plaatst het accent op het vergelijken en contrasteren.

Kanttekeningen

Biemans begint met het activeringsmodel, de CONTACT-strategie, waarvan de effectiviteit uit een eerder proefschrift is gebleken (Ali, 1990). Hij kiest echter niet voor het aanleren van deze strategie, maar voor het ontmantelen van de strategie. Op deze wijze probeert hij na te gaan welke stappen van de strategie tot de positieve leerresultaten leiden. Voor dit vraagstuk wordt echter geen antwoord gevonden en Biemans ziet zich genoodzaakt bij zijn tweede onderzoek verder te gaan met een aangepaste strategie. Je zou dan ook kunnen zeggen dat de keuze van de vraagstelling van het eerste onderzoek en de gevonden resultaten de keuze van de drie andere onderzoeken mede hebben bepaald. Zo kan helaas niet worden ingegaan op de generaliseerbaarheid van het activeringsmodel naar andere leersituaties, andere leergroepen en andere leerdomeinen. Het ware niet geheel onlogisch geweest om juist deze thematiek centraal te stellen in tegenstelling tot wat nu is gebeurd. Ook in het laatste onderzoek laat Biemans deze mogelijkheid voorbij gaan. Daarentegen zijn de inzichten in hoe leerlingen kunnen leren te leren in een bepaalde leersituatie aanzienlijk toegenomen, dankzij dit uitgebreide en nauwkeurig beschreven proefschrift.

Literatuur

- Ali, K.S. (1990). *Instructiestrategieën voor het activeren van preconcepties*. Academisch proefschrift. Helmond: Wibro.
- Prawat, R.S. (1989). *Promoting access to knowledge, strategy, and disposition in students: A research synthesis*. Review of Educational Research, 59, 1-41.
- Nussbaum, J. & Novick, S. (1982). *Alternative frameworks, conceptual conflict and accommodation: Towards a principled teaching strategy*. Instructional Science, 11, 183-200.
- Strike, K.A. & Posner, G.J. (1985). *A conceptual change view of learning and understanding*. In L.H.T. West & A.L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 211-231). Orlando: Academic press.

Ali, K.S.

Technische Universiteit Eindhoven

Marijke Kral

Instructie en leren in combinatieklassen. Een onderzoek naar effectiviteitsverschillen tussen combinatieklassen en enkelvoudige klassen in het basisonderwijs
(Dissertatie OU) Nijmegen: ITS - 1997 - ISBN: 90 5750 020 5

De relevantie van het probleem

In het verlengde van het vele ontwikkelings- en onderzoekswerk rond de combinatieklassenproblematiek dat het laatste decennium door Veenman c.s. is uitgevoerd, schreef Marijke Kral een dissertatie over dit onderwerp. Voor het eerst wordt diepgaand en grootschalig onderzocht of kinderen in combinatieklassen meer of minder opsteken van het onderwijs, en wat daarvan de eventuele oorzaken zouden kunnen zijn. Het onderwerp mag zich verheugen in

een grote belangstelling, deels gevoed door ongerustheid van ouders en practici, en deels gevoed door wetenschappelijke meningsverschillen.

Voor ouders en leerkrachten, opgegroeid in een tijd dat de formatietoekenning (met klas-sendelers) en de demografie ertoe leiden dat enkelvoudige klassen veruit overheersten, is de combinatieklas een gevreesde bron van onvoldoende aandacht voor het eigen kind respectievelijk een stressveroorzakende managementklus. Volgens sommige onderwijskundigen biedt de combinatieklas vele mogelijkheden: "peertutoring" wordt mogelijk, leeftijdsheterogene en/of vaardigheidshomogene groepen kunnen gevormd worden, het nodigt uit tot meer gedifferentieerd onderwijs, en de gevolgen van zittenblijven worden sterk gemitigeerd. Volgens andere onderwijskundigen daarentegen gaat er kostbare tijd verloren in combinatieklassen, waarvan de negatieve gevolgen alleen door middel van allerlei compensatiemechanismen binnen de perken zijn te houden. Zie daar het meningsverschil dat geen twijfel laat bestaan over de relevantie van het dissertatie-onderzoek.

De wetenschappelijke achtergrond

Veenman's review van het onderzoek naar de effecten van combinatieklassen (Veenman, 1995) leidde tot de samenvattende bevinding dat leerlingen in combinatieklassen cognitief niet onderdoen voor leerlingen in enkelvoudige klassen, en dat op het affectieve domein zich zelfs voordelen aftekenen. Het bleek voor de resultaten niet uit te maken of scholen (min of meer noodgedwongen) om organisatorische redenen combinatieklassen vormden of uit pedagogisch-didactische overwegingen. Kral bespreekt de kritiek die op deze bevinding geleverd is: de hypothese werd geopperd dat door selectiemechanismen en extra inspanning de negatieve effecten van combinatieklassen niet worden aangetroffen die er anders wel zouden zijn. Die selectiemechanismen betreffen de volgende: alleen oudere, ervaren leerkrachten krijgen combinatieklassen onder hun hoede, en bij het toewijzen van leerlingen aan enkelvoudige of combinatieklassen wordt rekening gehouden met de mate waarin ze zelfstandig kunnen werken. Tegenover deze compensatiehypothese stelde Veenman (1996) een onderbenuttingshypothese: combinatieklassen zouden gunstiger uitpakken voor leerlingen dan enkelvoudige klassen als leerkrachten de groeierings- en differentiatiemogelijkheden adequaat zouden oppakken.

Om zicht op deze controverse te krijgen wordt een verbinding gemaakt met het onderzoek naar effectieve instructie. Een gedegen effectonderzoek alléén zou immers geen licht op de opgeroepen vragen werpen. Door na te gaan of de taakgerichte lestijd, de doelmatigheid van de instructie, het klassenmanagement en de klassenorganisatie en de inrichting van het zelfstandig werken verschilt tussen enkelvoudige en combinatieklassen kan het (on)gelijk van de critici boven tafel komen.

Deze op zich heldere probleemstelling wordt in het proefschrift verbonden met de schaalvergrotingsoperatie die begin jaren negentig (na een heftig wetenschappelijk debat, waarvan we in het proefschrift niets terug vinden) in het basisonderwijs werd ingezet. Combinatieklassen zijn immers vooral te vinden op kleine scholen, waar volgens Claessen en Lagerweij (1990) onderwijskundig leiderschap, specifieke expertise, individualisering, zorgverbreding en onderwijs-evaluatie minder kans zouden krijgen dan op grote scholen.

Grote scholen zouden dus in sterkere mate kenmerken van effectieve scholen kunnen vertonen, kenmerken die op hun beurt effectieve instructie faciliteren en daarmee indirect hogere leeropbrengsten tot gevolg hebben. Kral werkt deze aandacht voor het schoolniveau conceptueel en onderzoeksmatig uit in de richting van schoolcultuur, en het beleidsvoerende en evaluerend vermogen van scholen.

De opzet van het onderzoek

Om licht te kunnen werpen op beide problemen (combinatieklassen en schaalvergroting) wordt een onderzoek opgezet waarbij 30 kleine scholen worden vergeleken met 15 grote scholen. De leerlingen uit de jaargroepen 4, 6 en 8 van beide typen scholen worden getest aan het begin

en aan het einde van het schooljaar ten aanzien van technisch lezen, spellen, begrijpend lezen en rekenen. Ook worden, vanwege de mogelijke selectieproblematiek, gegevens verzameld over intelligentie, geslacht, sociaal-economisch milieu van herkomst e.d. Via vragenlijsten en logboeken worden gegevens over de klassenorganisatie, de instructie, en de schoolorganisatie verzameld.

Kral gaat vervolgens uitgebreid in op de representativiteit van de onderzoeksgroep voor de Nederlandse populatie van basisscholen, en komt tot de conclusie dat scholen in de grote steden ondervetegenwoordigd zijn. Ook de psychometrische kwaliteiten van de diverse scholen komen aan bod, waarbij blijkt dat veel instructieschalen een zeer matige betrouwbaarheid hebben.

De resultaten

De onderzoeksresultaten worden door Kral over vijf hoofdstukken verdeeld gepresenteerd. De "proof of the pudding" is meteen aan de orde in het eerste van die vijf hoofdstukken. Aan de hand van multiniveau regressie-analyses wordt duidelijk gemaakt dat in combinatieklassen evenveel geleerd wordt als in enkelvoudige klassen. De mogelijke kritiek dat het maar om één schooljaar gaat, wordt door Kral handig ondervangen, doordat ze aantoont dat de jaargroep 8 leerlingen op de kleine scholen (die hun hele onderwijs carrière in combinatieklassen hebben doorgebracht) ook op de pretest (na controle voor IQ, SES, e.d.) niet onderdoen voor de andere leerlingen: dat is immers het cumulatief effect van zeven jaar onderwijs in combinatiegroepen.

Ook laten de analyses zien dat lager milieu-leerlingen in combinatieklassen niet meer achterblijven dan in enkelvoudige klassen. Tot zover stroken de uitkomsten met de review-resultaten van Veenman.

Het volgende hoofdstuk bespreekt de resultaten van het onderzoek naar de instructie-, klassenorganisatie- en klassenmanagementfactoren. Eén citaat maakt alles duidelijk: "Dat in combinatieklassen feitelijk net zo wordt gewerkt als in enkelvoudige groepen wordt nog eens bevestigd door het gegeven dat in de meeste combinatieklassen (80%) de leerlingen per jaargroep bij elkaar zitten, dat meestal alternerend instructie wordt gegeven aan de ene en de andere jaargroep(en) (door 87% vrijwel altijd) en dat de leerlingen die daar niet bij betrokken zijn dan meestal individueel werk doen (90% vaak of altijd)" (p. 132).

Dan komt een specifiek element in een apart hoofdstuk naar voren: tijdsbesteding. Hierover werd eerder separaat verslag gedaan in dit tijdschrift: combinatieklassen en enkelvoudige klassen verschillen niet qua tijdsbesteding aan rekenen, lezen en taal en ook niet wat betreft de tijd die de leerkracht uittrekt voor klassenmanagement. Wel komen er verschillen naar voren ten gunste van de enkelvoudige klassen met betrekking tot de tijd die de leerkracht uittrekt voor het begeleiden van oefeningen en de individuele hulp.

De drie volgende hoofdstukken in het proefschrift gaan nader in op de schaalvergrotingsvragen. Nadat Kral eerst heeft laten zien dat de kleine scholen in dit onderzoek meer achterstandsleerlingen herbergen, en dat het uitstroomniveau op deze scholen navenant lager ligt dan op de grote scholen, en en passant heeft opgemerkt dat 43% van die kleine scholen met wisselende combinatiegroepen werkt, kan ze vervolgens de beweringen van Claessen en Lagerweij beproeven. Noch qua schoolleiderschap noch qua schoolbeleid blijken de grote van de kleine scholen te verschillen. Hun evaluerend vermogen is daarentegen beter ontwikkeld, maar daar staat tegenover dat de leerkrachten minder bij de beleidsvorming zijn betrokken. Grote scholen overtreffen de kleine scholen qua prestatiegerichtheid, maar merkwaardigerwijze zijn de leerkrachten het op de grote scholen onderling minder eens over de doelgerichtheid van de school. Voor nog eens tien schoolcultuurvariabelen blijken geen verschillen op te treden tussen kleine en grote scholen. Wat nog wel blijkt is dat op kleine scholen de leerprestaties van de leerlingen significant vaker op schoolniveau worden geëvalueerd dan op grote scholen. De lezer van deze heuglijke tijding vraagt zich echter meteen af wat het dan te betekenen heeft dat die grotere scholen een groter evaluerend vermogen hebben.

Als proef op de som wordt vervolgens nagegaan of op kleine scholen de leerlingen minder vooruitgang boeken (quod non, waarmee ook het empirisch fundament van de schaalvergrotingsoperatie nogmaals een geduchte knauw krijgt) en of de verschillende schoolvariabelen met de leerwinst van leerlingen samenhangen (en wederom is het antwoord negatief).

Discussie

In de discussie beziet Kral de compensatie- versus onderbenuttingshypothese opnieuw vanuit het licht van de onderzoeksresultaten. Volgens haar is er weinig steun gevonden voor de compensatiehypothese: combinatieklassen herbergen niet de betere leerlingen en evenmin de oudere, ervarener leerkrachten. Ook de instructie in combinatieklassen onderscheidt zich niet van die in enkelvoudige klassen. Het eerder gegeven citaat wijst volgens Kral onmiskenbaar in de richting van de onderbenuttingshypothese: de differentiatiemogelijkheden die combinatieklassen bieden blijven onbenut.

Voordat ze op de schaalvergrotingsproblematiek ingaat, merkt Kral terzijde op dat ook dit schooleffectiviteitsonderzoek weer weinig heeft opgeleverd. En dan besluit ze met de conclusie dat schaalvergroting niet effectiever is.

Evaluatie

Het proefschrift roept waardering op, omdat een complexe problematiek grondig is onderzocht, daarbij gebruik makend van het vele eerdere onderzoekswerk naar combinatieklassen en de verworvenheden van het onderzoek naar effectieve instructie en effectieve scholen.

Maar wat de handigheid van het onderzoeksontwerp lijkt (de problematiek van de kleine scholen en de combinatieklassen komen in één onderzoek aan bod) is tevens ook de valkuil. Ik kan me niet aan de indruk onttrekken dat we nu juist door die twee problemen gecombineerd te onderzoeken van beide minder te weten zijn gekomen dan wanneer slechts één van de twee zou zijn onderzocht.

Omdat de kleine scholen en de combinatieklassen onlosmakelijk verbonden zijn in dit onderzoek doemen allerlei nieuwe compensatiehypotheseën op: de kleinschaligheid van de school compenseert de negatieve effecten van de combinatiegroep (of omgekeerd), of: omdat de klassen op de kleine scholen veel kleiner zijn dan op de grote scholen (20 versus 27) worden eventuele negatieve effecten van kleinschaligheid en/of combinatieklassen teniet gedaan door de kleine klassen. Het onderzoek roept ook technische vragen op: waarom 15 grote en 30 kleine scholen en niet evenveel van beide, en hoe zit het met het onderscheidingsvermogen met een dergelijke geringe steekproefomvang?

Waarom is, als die onderbenuttingshypothese serieus wordt genomen, niet overwogen op enkele scholen het veronderstelde positieve effect van combinatieklassen door middel van krachtige training zichtbaar te maken? En zou het wellicht niet een idee zijn dat in sommige combinatieklassen compensatie plaats vindt en in andere onderbenutting?

Ook vind ik de conclusie dat de effectieve scholen ingang "wederom" weinig heeft opgeleverd merkwaardig. Ten eerste omdat met onbetrouwbare schalen effecten nauwelijks te detecteren zijn (men zou het niet eens moeten proberen), ten tweede omdat de kans dat in de populatie aanwezige effecten ten onrechte niet gevonden worden vrij groot was gegeven het geringe aantal scholen, en ten derde omdat deze bevinding van Kral ook de door haar opgevoerde schoolcultuur en -organisatie variabelen geldt die geen plaats in het effectieve scholen onderzoek hebben. De conclusie dat schaalvergroting niet effectiever is kan niet automatisch leiden tot de conclusie dat het daarmee ongewenst is: grotere scholen zijn ook niet ineffactiever en ze zijn goedkoper. Iets anders is dat schoolgrootte-onderzoek niet hetzelfde is als onderzoek naar schaalvergroting, en dat we dus niet uit een vergelijking van kleine met grote scholen direct kunnen afleiden wat er gebeurt als die kleine scholen groter worden.

Toch levert het proefschrift, ondanks de hiervoor geventileerde kritiek, een belangrijke bijdrage aan de discussie over combinatieklassen. De bewijslast dat leerlingen in combinatieklassen niet achterblijven is nu toch weer enigszins uitgebouwd. Het is dan ook niet Kral te

verwijten, dat die onderbenuttingshypothese zelf onderbenut is door al die om adaptief en gedifferentieerd onderwijs roepende onderwijskundigen. Die hypothese kan alleen waar zijn als men dergelijk onderwijs onder normale condities succesvol én effectief weet te implementeren.

Literatuur

- Claessen, J.F.M. & Lagerweij, N.A.J. (1990). Naar een grote school. In: Projectgroep Schaalvergroting Basis-onderwijs, *Schaal en kwaliteit in het basisonderwijs*. Zoetermeer MOW.
- Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: a best evidence synthesis. *Review of Educational Research*, 65(4), 319-381.
- Veenman, S. (1996). Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research*, 66(3), 323-340.

R.J. Bosker
Universiteit Twente
Toegepaste Onderwijskunde
Postbus 217, 7500 AE Enschede

J. Prins (1997)

Studieuitval in het wetenschappelijk onderwijs. Studentkenmerken en opleidingskenmerken als verklaring voor studieuitval

Proefschrift Universiteit van Nijmegen. Nijmegen: Nijmegen University Press.
ISBN 90 5710 026 6.

Inleiding

In het Nederlandse hoger onderwijsbeleid vanaf 1949 stond aanvankelijk met name de lengte van de studieduur in de beleidsmatige belangstelling. Pas in de jaren tachtig is met de Wet Tweefasenstructuur (1981) en de introductie van het Plaatsen-Geld-Model (1982) de aandacht verschoven in de richting van het rendement en daarmee naar studieuitval (zie voor de complete historie Bijleveld, 1993).

Bijleveld (1993) geeft een uitgebreid overzicht van de verschillende benaderingen en modellen voor studieuitval in de angelsaksische literatuur. De interactionele benadering is het uitgangspunt geworden voor de hoofdstroom van het empirisch verklaringsonderzoek. De benadering gaat ervan uit dat student en omgeving met elkaar interacteren, en dat de ontwikkeling van de interactie van invloed is op de beslissing al dan niet de studie te staken. In de lijn van Tinto (1987) en Pascarella c.s. (Pascarella and Terenzine 1991; Pascarella, Terenzini and Wolfe, 1986; Stoecker, Pascarella, and Wolfe, 1988) staat met name de sociale integratie centraal in de verklaring; een mogelijke synthese van de verschillende modellen is te vinden bij Cabrera et al (1992).

In het Nederlandse onderzoeksproject "Verder Studeren" (Roeleveld, De Jong, Oosterbeek en Webbink, 1994; De Jong, Webbink en Roeleveld, 1995; De Jong, 1997) zijn elementen uit het model van Tinto gebruikt om het verloop van studieloopbanen in het Nederlands hoger onderwijs te analyseren. Dergelijk onderzoek sluit qua opzet en methodiek aan bij schoolloopbaanonderzoek in het kader van het thema "onderwijs en samenleving", in het bijzonder de ongelijkheid in onderwijskansen (zie Dronkers, Heijke, en Van der Velden, 1994a), waarin relaties worden onderzocht tussen afhankelijke variabelen als het bereikte onderwijsniveau (onderwijspositie), transities, keuze van en in scholen, seksespecifieke vakkenkeuze, en drop out, al dan niet van speciale groepen, en onafhankelijke variabelen als sociale en etnische herkomst, factoren op schoolniveau, het streefniveau van ouders, leerkrachten en leerlingen, overadvisering, steun van ouders, en verwachtingen van ouders, leerkrachten en leerlingen. In dit kader wordt soms ook de loopbaan op de arbeidsmarkt in de analyse betrokken (Dronkers, Heijke, en Van der Velden, 1994b; Roeleveld en Derriks, 1995; Hoeben, 1996) of wordt uit-

gegaan van het perspectief van de school- en instructie-effectiviteit (Creemers en Scheerens, 1991). Veelal worden methoden gebruikt als LISREL-analyse, logistische regressie, en multi-niveau-analyse.

Bijveld (1993) concludeert voor wat betreft het Nederlandse onderzoek aangaande studieuitval in het WO dat verschillende variabelen niet op het niveau van de studierichting zijn onderzocht, (inmiddels Jansen, 1996, 1997) multivariaat onderzoek schaars is, (bijvoorbeeld voor studentkenmerken: Meier, Bruinsma, en Geurts, 1991), de interactie tussen studentkenmerken en studierichtingkenmerken weinig aandacht heeft gekregen, en er sprake is van weinig theorievorming. Bijveld (1993) heeft derhalve zelf een interactioneel verklingsmodel ontwikkeld voor studieuitval. Het hier te bespreken proefschrift van Prins (1997) is er ook op gericht een interactioneel verklingsmodel te ontwikkelen en te toetsen.

Korte beschrijving van het proefschrift

De centrale probleemstelling van het proefschrift luidt: "Is vanuit de kenmerken van instromende studenten, de kenmerken van de opleidingen en de interactie van beide een adequate verklaring te vinden voor studieuitval?" Hierbij wordt 'studieuitval' opgevat als "het staken van een opleiding in het wetenschappelijk onderwijs". Dit betekent dat het omzwaaien naar een andere opleiding, zelfs binnen dezelfde faculteit, in het onderzoek wordt opgevat als studieuitval (zie noot 2, blz. 17).

In een inleidend hoofdstuk wordt de probleemstelling gesitueerd in een kort historisch overzicht van het hoger onderwijsbeleid. De invoering van de Wet Twee-Fasen-structuur in 1982 heeft er 15 jaar later wel toe geleid dat de gemiddelde studieduur is gedaald van meer dan 7 jaar tot ongeveer $5\frac{1}{2}$ jaar: echter, het numeriek rendement (op instituutniveau) is niet gestegen: 43% van de studenten heeft (nog) geen diploma gehaald, waarvan de ene helft struikelt in de propedeuse en de andere helft in de doctoraal fase. Het studierendement is te verhogen door de studieuitval te beperken waarbij drie soorten factoren een rol spelen: (a) stelsel- en systeemkenmerken, (b) studentkenmerken, en (c) opleidingskenmerken. Het proefschrift beperkt zich tot de twee laatste soorten factoren.

Desalniettemin wordt voor een goed begrip in hoofdstuk 2 ingegaan op de invloed van stelsel- en systeemfactoren op de studieuitval. Er wordt hierbij een onderscheid gemaakt tussen kenmerken van het voortgezet onderwijs, het hoger onderwijs, en de invloed van de overheid en instellingen op studeergedrag. Prins komt tot de conclusie dat alle beïnvloedingspogingen van overheid en instellingen vooralsnog niet hebben geleid tot een verhoging van het rendement. We zien wel grote verschillen in numerieke rendementen, gemiddelde studieduur van afgestudeerden, en tijdstippen van studieuitval als we opleidingen met elkaar vergelijken. Ook studenten verschillen sterk van elkaar op een groot aantal kenmerken. Studieuitval is daarmee de uitkomst van een interactie tussen studentkenmerken en opleidingskenmerken.

In hoofdstuk 3 wordt gerapporteerd over een literatuuronderzoek aangaande Nederlandse studies en Noordamerikaanse studies. In het Nederlandse onderzoek heeft men zich tot voor kort vooral beperkt tot enkele factoren en binnen een bepaalde opleiding. Het Noordamerikaanse onderzoek gaat uit van interactieve verklingsmodellen waarin de invloed van academische en sociale integratie op uitvalsbeslissingen centraal staat. De meest recente studies maken gebruik van multivariate statistische technieken waarmee zo'n 30% van de verklaarde variantie in studieuitval kon worden verklaard. Deze interactionele benadering wordt als uitgangspunt genomen voor het onderzoek.

In hoofdstuk 4 wordt op basis van recent Amerikaans en Nederlands onderzoek een conceptueel model opgebouwd dat studieuitval moet verklaren. Prins brengt een onderscheid aan tussen factoren op studentniveau en op opleidingsniveau. Factoren op studentniveau zijn: voorafgaande schoolloopbaan (type vooropleiding, samenstelling vakkenpakket, aansluiting VWO-WO); achtergrond (steun omgeving, herkomst, geslacht, leeftijd); aspiratie, motivatie en zelfvertrouwen; sociale integratie (deelname en betrokkenheid bij extra-curriculaire activiteiten); academische integratie (contacten met de staf, studievoortgang, tevredenheid met

de academische ontwikkeling en prestatie); en uitvallen of blijven. Factoren op programma-niveau zijn: studiebegeleiding, programmakenmerken (aantal contacturen, aantal parallelle vakken, gemiddeld aantal uren per vak); beoordelen en toetsen; en studentgerichtheid. De veronderstelling is dat de factoren op studentniveau een causaal model vormen. Het model geeft de causale relaties weer, en zijn geformuleerd in de vorm van 19 hoofdhypothesen.

Hoofdstuk 5 bespreekt de opzet van het onderzoek. Het onderzoek vond plaats op de KU Nijmegen. Jaarlijks stromen er zo'n 2500 studenten in en vallen er ruim 1000 uit. In begin 1995 is er een vragenlijst verzonden naar 925 uitvallers en 1000 aselekt gekozen niet-uitvallers. Het responsepercentage was 49% bij de uitvallers, en 66% bij de blijvers; in totaal waren er 1001 respondenten. De organisatiekenmerken van zes grote opleidingen, verdeeld over alpha-, bèta-, en gamma-faculteiten van de KU Nijmegen, zijn via interviews en documentatieanalyses verzameld en gedichotomiseerd. De analyse kent twee fasen: (a) een beschrijvende fase waarin ook het aantal items via factoranalyse en betrouwbaarheidsanalyse wordt beperkt, en (b) een toetsingsfase waarin het model - de factoren op studentenniveau - wordt getoetst en verbeterd met behulp van logistische regressie en LISREL. De invloed van de opleidingskenmerken op de causale relaties - in termen van hoeveelheid verklaarde variantie in de studieuitval - wordt nagegaan door elk opleidingskenmerk simultaan in te brengen bij de modeltoetsing.

Hoofdstuk 6 rapporteert de beschrijvende onderzoeksresultaten per item, in hoeverre de items verschillende factoren vormen, en in welke mate er verschillen zijn tussen de uitgevallen en gebleven studenten. In het geval factoren invariaat zijn of nauwelijks samenhang vertonen met andere factoren worden ze verwijderd uit het model. Alle achtergrondkenmerken, met uitzondering van 'steun ouders en vrienden' worden verwijderd. De belangrijkste (cor)relaties met uitval zijn: gemiddeld aantal studiepunten (.36), studiefase (.35), zelfvertrouwen (.35), tevredenheid met ontwikkeling en prestaties (.31), aspiraties (.26), en onderwijservaring als mentor of studentassistent (.23). De datareductie resulteert in een geoperationaliseerd verklaringsmodel van 21 factoren dat zal worden getoetst. Voor wat betreft de programmakenmerken wordt de hoeveelheid data gereduceerd door de zes opleidingen per kenmerk in twee groepen ingedeeld: een groep waarin het kenmerk niet of relatief weinig voorkomt, en een groep waarin het kenmerk relatief veel voorkomt.

In hoofdstuk 7 komen de resultaten van de logistische regressie en de LISREL-analyse aan de orde. De resultaten van de logistische regressieanalyse geven aan dat het model een significante verklaring geeft voor de studieuitval; het model verklaart 30% van de hoeveelheid variantie. Van ruim driekwart van de studenten kan correct worden voorspeld of ze zijn uitgevallen of niet. De eerste LISREL-run leidt tot een verwerping van het model. De analyse geeft echter ook suggesties aan voor verbeteringen in de modelspecificatie. Het verbeterde eindmodel verklaart 26% van de variantie in de studieuitval. De verschillende uitkomsten van de twee analyses worden als gering gezien en toegeschreven aan de specifieke voor- en nadelen van de betreffende technieken. De toetsing van de zes programmakenmerken leverde het volgende op: (a) twee kenmerken - aantal parallelle cursussen en de hoeveelheid contacttijd - beïnvloedden niet het model, maar wel de sterkte van de padrelaties; (b) vier kenmerken - studiebegeleiding, gemiddeld aantal studiepunten per vak, studentgerichtheid, en tentamenbeleid - veranderden de 'fit' van het model hetgeen wijst op een sterke invloed op de factoren en de relaties in het model. De totaal verklaarde variantie van uitval/blijven varieert tussen de 16% (opleidingen met veel studiebegeleiding) en 41% (opleidingen met een sterk tentamenbeleid).

Hoofdstuk 8 vat samen, formuleert conclusies en werpt enige discussievragen van theoretische, methodologische en beleidsmatige aard op. Het eindmodel verklaart 26% van de variantie in studieuitval/blijvers. Effect op de verklaring van de studieuitval hebben factoren op studentniveau als de intrinsieke gemotiveerdheid, aspiraties, hoog zelfvertrouwen, veel contacten met de staf, tevredenheid met de academische ontwikkeling en prestaties, studieprestaties, ouderejaars, betrokkenheid bij opleidingscommissies en studentenverenigingen, en

betrokkenheid als mentor of student-assistent. Academische integratie wordt daarbij gezien als belangrijker dan sociale integratie. Voorts bestaat er geen wederzijdse compenserende relatie tussen sociale en academische factoren; sociale integratie gaat vooraf aan academische integratie.

Zoals gezegd, ook alle factoren op opleidingsniveau hebben effect. De volgende factoren op studentniveau hebben geen effect: geslacht (per saldo), opleiding ouders, beroep ouders, steun omgeving, leeftijd, samenstelling VWO-pakket, aansluiting VWO-KU Nijmegen, type vooropleiding, en extrinsieke motivatie.

Commentaar

Prins heeft met zijn studie uitstekend aangesloten bij de empirische hoofdstroom aangaande het onderzoek naar studieuitval in het hoger onderwijs zoals geïnspireerd door Tinto. Hierbij gaat hij goed te werk door de studie breed op te zetten, kritisch de veelal Amerikaanse literatuur te verwerken, vraagstellingen en hypothesen precies uit te werken, netjes te operationaliseren, inventief variabelen reductie-methoden toe te passen, en goed te laten zien op welke gronden het model wordt bijgesteld op basis van de LISREL-resultaten. Het proefschrift is een hele mooie empirische voortzetting van de door Bijleveld (1993) ingezette modellering van de Nederlandse WO-situatie aangaande de verklaring van studieuitval.

Wel kent de studie ook de beperkingen waarvoor hij zelf waarschuwt (Prins geeft dat ook bij voorbaat toe): er is sprake van een ex post facto onderzoek in plaats van een longitudinaal onderzoek, en het vindt slechts plaats in één instelling. Tevens had ik af en toe wat problemen met het gebruik van de causale terminologie in de hoofdstukken over de opbouw en toetsing van het model. De tekst suggereert soms te veel dat de 'causale verbanden' meer zijn dan correlatieve verbanden, en dat een LISREL-benadering het mogelijk maakt dergelijke diepere verbanden op hun juistheid te toetsen. Gelukkig blijkt dan weer bij de uiteindelijke rapportage van de resultaten dat Prins één en ander wel degelijk ziet als correlatief, en dat de LISREL-analyse een exploratieve techniek is. Voorts zij opgemerkt dat het soms niet meekvalt in een studie als deze de hoofdlijn te blijven zien: het aantal variabelen is groot, en afkortingen daarvan gebruiken in de tekst verhogen niet de leesbaarheid. Ook het aantal hypothesen is groot waarbij de onderlinge relaties tussen de hypothesen door de complexiteit soms moeilijk te overzien zijn. Nog een kanttekening bij de definitie van het begrip 'studieuitval' (zie blz. 17 en 89): de omzwaaiers worden ook tot de uitvallers gerekend. Op zich is daar niets op tegen; echter, in de analyse was het interessant geweest de echte 'stoppers' te onderscheiden van de 'omzwaaiers'. Ik heb de indruk dat beide groepen toch wat andere overwegingen hebben om de studie te verlaten; het perspectief is geheel anders.

Is er nu vanuit de kenmerken van instromende studenten, de kenmerken van opleidingen - zoals aangegeven door Prins -, en de interactie van beide een adequate verklaring gevonden voor studieuitval? Mijn antwoord is ontkennend, omdat, door de bank genomen, slechts 30% van de variantie wordt verklaard. Eigenlijk was dit ook wel te verwachten, aangezien de basis is genomen in modellen die ook niet veel verder kwamen dan zo'n 30%. De modelspecificatie bevat waarschijnlijk omissies en/of fouten die moeten worden verbeterd. Voor de verbetering moet m.i. niet alleen longitudinaal onderzoek worden ondernomen, maar moet het onderzoek in eerste instantie ook kwalitatief van aard zijn - bijvoorbeeld diepte-interviews - om bij de betreffende studenten eens goed te achterhalen wat hen nu beweegt om de studie te staken. De aard en keuze van de variabelen moet weer opnieuw worden geïnspireerd door de te onderzoeken personen om tot valide verklaringen te komen. Op dit moment is m.i. het aggregatieniveau te hoog, en geven de gevonden statistische verbanden te weinig informatie over de processen die daadwerkelijk plaatsvinden. Voorts lijkt het van belang de kwantitatieve en kwalitatieve resultaten van het schoolloopbaanonderzoek in het basisonderwijs en voortgezet onderwijs bij het verdere onderzoek te betrekken.

Tot besluit nog een opmerking over de praktische bruikbaarheid van de resultaten. In het laatste hoofdstuk gaat Prins in op de beleidsmatige consequenties. In de eerste plaats valt op

dat hij geregeld terugvalt op mogelijke maatregelen op systeemniveau - studiefinanciering, het bekostigingsniveau - waarover het onderzoek juist niet gaat. In de tweede plaats kan ik het wel met hem eens zijn dat een studierichting zich moet richten op de middengroep, en dat in dit kader een mix van curriculaire maatregelen moet worden toegepast. Echter, de beleidsmatige hamvraag op het niveau van een studierichting is nu juist 'welke mix' en 'hoe' toe te passen op de middengroep? Ik vrees dat het type verklaringsmodellen, zoals nu opgesteld en getoetst door Prins, te weinig of geen antwoorden geven op deze vragen. De gevonden effecten van de opleidingskenmerken wijzen wellicht de weg: het opstellen van een model waarin meer gedetailleerd wordt ingezoomd op het interactieproces van een student met een onderwijsprogramma. Maar zoals reeds bepleit: laten we dat nu eens echt de studenten zelf vragen.

Literatuur

- Bijleveld, R.J. (1993). *Numeriek rendement en studiestaking*. Proefschrift Universiteit Twente. Enschede: Universiteit Twente / CSHOB.
- Cabrera, A.F., Castañeda, A.N. & Hengstler, D. (1992). The convergence between two theories of college persistence. *Journal of Higher Education*, 63, 2, 143-164.
- Creemers, B.P.M. & Scheerens, J. (red.) (1991). School- en instructieeffectiviteit. *Tijdschrift voor Onderwijsresearch*, 16, 4, 193-258 (themanummer).
- Dronkers, J., Heijke, J.A.M. & Velden, R.K.W. van der (red.) (1994). Ongelijkheid in onderwijskansen. *Tijdschrift voor Onderwijsresearch*, 19, 3, 189-259 (themanummer).
- Dronkers, J., Heijke, J.A.M. & Velden, R.K.W. van der (red.) (1994). Onderwijs en samenleving. *Pedagogische Studiën*, 71, 5, 323-366 (themanummer).
- Hoeben, W.Th.G.J. (red.) (1996). Rendement van het beroepsonderwijs. *Tijdschrift voor Onderwijsresearch*, 21, 3, 197-246 (themanummer).
- Jansen, E.P.W.A. (1996). *Curriculumorganisatie en studievoortgang. Een onderzoek onder zes studierichtingen aan de Rijksuniversiteit Groningen*. Proefschrift Rijksuniversiteit Groningen. Groningen: GION.
- Jansen, E.P.W.A. (1997). Curriculumorganisatie en studievoortgang. *Tijdschrift voor Onderwijsresearch*, 22, 2, 3-15.
- Jong, U. de, Webbink, H.D. & Roeleveld, J. (1995). *Verder studeren. De subjectieve factor: commitments en academische integratie in de studieloopbaan*. Den Haag: Ministerie van Onderwijs en Wetenschappen.
- Jong, U. de (1997). *Het Amsterdamse studieloopbaanmodel*. Paper gepresenteerd op de Onderwijsresearchdagen 1997 te Leuven.
- Meier, K.A., Bruinsma, G.J.N. & Geurts, P.A.Th.M. (1991). Snelheid van studeren bij universitaire studenten. Een verslag van een empirisch vervolgonderzoek naar individuele factoren. *Tijdschrift voor Hoger Onderwijs*, 9, 3, 181-194.
- Pascarella, E.T. & Terenzi, P.T. (1991). *How college affects students*. San Francisco: Jossey-Bass.
- Pascarella, E.T., Terenzini, P.T. & Wolfe, L.M. (1986). Orientation to college and freshmen year persistence/withdrawal decisions. *Journal of Higher Education*, 57, 2, 155-175.
- Roeleveld, J., Jong, U. de, Oosterbeek, H.D. & Webbink, H.D. (1994). *Verder studeren. Studievoortgang en onderwijskeuzen 1992*. Den Haag: Ministerie van Onderwijs en Wetenschappen.
- Roeleveld, J. & Derriks, M. (1995). Uitvallers uit de hoogste klassen van HAVO en VWO. *Pedagogische Studiën*, 72, 6, 437-455.
- Stoecker, J., Pascarella, E.T. & Wolfe, L.M. (1988). Persistence in higher education: a nineyear test of a theoretical model. *Journal of College Student Development*, 29, 3, 196-209.
- Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago / London: The University of Chicago Press.

Cees Terlouw
Universiteit Twente - Onderwijskundig Centrum
Postbus 217 - 7500 AE Enschede
C.TERLOUW@OC.UTWENTE.NL

Jaarverslag Vereniging voor Onderwijs Research 1997-1998

1. Samenstelling bestuur

Prof. Dr. N. Verloop	Voorzitter
Prof. Dr. H. van Hout	Vice-voorzitter, tevens Divisie Hoger Onderwijs
Dr. M. Mulder	Secretaris
Dr. Tj. Imbos	Penningmeester
Prof. Dr. J. van den Akker	Divisie Curriculum
Prof. Dr. J. Dronkers	Divisie Onderwijs en Samenleving
Prof. Dr. J. Pieters	Divisie Leren en Instructie
Prof. Dr. P. Sanders	Divisie Methodologie en Evaluatie
Dr. B. Versloot	Divisie Beroepsonderwijs, Bedrijfsopleidingen en Volwasseneneducatie
Prof. Dr. F. van Wieringen	Divisie Beleid en Organisatie
Prof. Dr. Th. Wubbels	Divisie Lerarenopleiding en Leraarsgedrag

2. Aantal leden

Ledenaantallen:

1996	586
1997	549
1998	577

Aantallen leden per Divisie:

Beroepsonderwijs, Bedrijfsopleidingen & Vowasseneneducatie	169
Leren & Instructie	143
Hoger Onderwijs	106
Lerarenopleiding & Leraarsgedrag	98
Methodologie & Evaluatie	85
Beleid & Organisatie	83
Curriculum	80
Onderwijs & Samenleving	57

3. Ledenwerving

Er heeft dit jaar een gerichte actie ten behoeve van het werven van leden plaats gehad. De leden van het VOR Promovendi-Overleg die nog geen lid van de VOR waren zijn door enkele bestuursleden per universiteit benaderd met het verzoek lid te worden van de VOR. De desbetreffende hoogleraren zijn benaderd met de suggestie in hun instelling te bevorderen dat promovendi de kosten van het lidmaatschap van de VOR mogen declareren bij hun werkzaamheid. Hierbij is verwezen naar de Faculteit der Toegepaste Onderwijskunde aan de Universiteit Twente waar dit reeds mogelijk is.

Verder is overleg geweest met de Nederlandse Vereniging voor Opvoedkundigen, Onderwijskundigen en Orthopedagogen, de NVO. Deze Vereniging is voorgesteld in navolging van de NVvO en de VULON combinatielidmaatschappen mogelijk te maken. NVO-leden zouden dat met reductie lid kunnen worden van de VOR, en omgekeerd. Alvorens hiertoe over te gaan zal het ledenbestand van de VOR worden vergeleken met dat van de NVO. Bij de NVO is hiertoe bereidheid getoond. De secretaris zal deze vergelijking maken.

4. Doorlichten Divisies

Het proces van het doorlichten van Divisies op hun activiteiten is afgerond. Dit heeft geresulteerd in het reactiveren van de Divisies Beleid en Organisatie, Curriculum en Methodologie en Evaluatie. De Divisie Leren en Instructie is momenteel het minst actief.

5. Divisietijdschriften

Voorgesteld was na te gaan of er bij Divisies belangstelling bestaat voor reductie-abonnementen op Divisie-gebonden buitenlandse tijdschriften. Dat was inderdaad voor een aantal Divisies het geval. Tijdschriften die genoemd waren, zijn onder andere: Teaching and Teacher Education (TATE) en Human Resource Development Quarterly (HRDQ). Navraag bij de betreffende uitgevers heeft echter duidelijk gemaakt dat de kwantumkortingen zo minimaal zijn, dat er voor de leden van Divisies nauwelijks voordeel te behalen valt. Besloten is voorlopig verder af te zien van reductie-abonnementen op dit soort tijdschriften. Voor de overige tijdschriften die voor VOR-leden met korting verkrijgbaar zijn (Learning & Instruction, International Journal of Educational Research en Educational Research and Evaluation), heeft dit verder geen consequenties.

6. TOR

Gedurende dit jaar heeft de eerste jaargang die bij Ten Brink is uitgegeven zijn beslag gekregen. Het tijdschrift heeft een nieuw uiterlijk gekregen waar positieve reacties op zijn gekomen. Er zijn een aantal kinderziekten geconstateerd die inmiddels zijn opgelost. Er is een probleem met de extra kosten die in rekening worden gebracht voor aanvullende correcties en opmaak. Dit probleem zou kunnen worden opgelost door het manuscript te laten screenen door een corrector. Dat gebeurt momenteel niet, hetgeen als een gemis wordt ervaren. Er wordt momenteel gewerkt aan een vervolgcontract met Ten Brink voor het drukken van de komende jaargangen van het TOR.

Gedurende dit jaar is opnieuw subsidie aangevraagd voor de uitgave van het TOR. Daar is echter een negatief resultaat mee geboekt. Het Ministerie van Onderwijs, Cultuur en Wetenschappen stelt zich op het standpunt dat de vroegere SVO-middelen voor onderwijsonderzoek grotendeels zijn overgedragen aan NWO en dat zij geen directe verantwoordelijkheid meer heeft voor de disseminatie van onderzoeksresultaten via de wetenschappelijke tijdschriften. Besloten is gezien de onaantrekkelijke verhouding tussen kosten en baten hier verder geen vervolg aan te geven.

Het redactiebeleid is een voortdurend punt van aandacht. De beleidswijzigingen die enkele jaren geleden zijn ingezet, worden verder doorgevoerd, hetgeen inmiddels resultaten begint af te werpen. Er is de afgelopen jaren een grotere spreiding van artikelen over Divisies.

Met ingang van januari 1998 is het lidmaatschap van de VOR en het abonnement op het TOR losgekoppeld. Leden kunnen nu voor slechts f 25,- het TOR ontvangen, maar kunnen dat abonnement ook beëindigen. Het is nog niet bekend hoeveel leden van deze laatste optie gebruik zullen maken.

7. Ledenadministratie

Er zijn het afgelopen jaar veel klachten geweest rondom de ledenadministratie. Besloten is de ledenadministratie te koppelen aan het secretariaat. Verwacht wordt dat hiermee de klachten verminderen. Afgesproken is Ten Brink en de uitgeverijen van de andere reductie-abonnementen bijtijds mutaties door te sturen. Hetzelfde geldt voor de factuur-administratie.

8. Financiën

De tariefstructuur voor het jaar 1998 is als volgt:

Basislidmaatschap gewone leden	f 110,-
Basislidmaatschap studenten	30,-
Basislidmaatschap promovendi (AIO-OIO- bursaal)	60,-

Basislidmaatschap combinatieleden NVvO/VELON	75,-
Lidmaatschap van een divisie is gratis	
Voor elk lidmaatschap van een <i>extra</i> divisie betalen de leden	f 10,-
Prijzen van tijdschriftabbonementen:	
Tijdschrift voor Onderwijs Research	f 25,-
International Journal of Educational Research	55,-
Learning and Instruction	45,-
Educational Research and Evaluation	55,-

9. Electronische communicatie binnen de VOR

De Divisie Hoger Onderwijs heeft zich bezig gehouden met de elektronische communicatie binnen de VOR. Afsproken was het eventueel elektronisch uitgeven van het TOR voorlopig uit te stellen. De Divisie zou zich concentreren op het opzetten van een Website. De ontwikkelingen terzake worden door de gehele VOR met belangstelling gevolgd. Zodra er bruikbare voorbeelden zijn ontwikkeld, kunnen andere Divisies deze overnemen.

10. VOR Jaarboek 1998

Zoals dit jaar is afgesproken, is dit jaar ter gelegenheid van de 25e ORD een VOR Jaarboek gemaakt. De bedoeling is dit jaarlijks te doen. De eerste uitgave is een proefuitgave. De bedoeling is om het verder uit te ontwikkelen als het state-of-the-art Jaarboek van de onderwijsresearch. De informatie van de Divisies en instellingen die er momenteel is zit is nog wat onevenwichtig. Bij een volgende uitgave zullen betrokkenen voorbeelden worden verschaft van wat opgenomen kan worden in het Jaarboek.

11. ORD

De ORD in 1997 was georganiseerd door de Katholieke Universiteit Leuven. Voorzitter van de organisatiecommissie was Prof. Dr. E. De Corte. De conferentie was geslaagd.

De ORD in 1998 is de 25e ORD en wordt georganiseerd door de Universiteit Twente. Voorzitter van de organisatiecommissie is Prof. Dr. J.M. Pieters.

Momenteel wordt gesproken over een langere termijnplanning wat betreft de organisatie van de ORD's. Er wordt gesproken over de jaren 1999, 2000 en 2001 met vertegenwoordigers van de Universiteiten van Amsterdam, Leiden en Nijmegen. De VFO heeft voorgesteld de ORD in 2002 in Antwerpen te organiseren. Het Bestuur heeft hier instemmend op gereageerd.

Gekoppeld aan de 25e ORD zou een jubileumnummer van het TOR worden uitgegeven. De uitgave daarvan zou echter samenvallen met het jubileumnummer van PS, dat in 1998 haar 75-jarig jubileum viert. Bovendien bevindt het TOR zich in 1998 in de 23e Jaargang, dus is besloten het jubileumnummer uit te geven in het jaar 2000, als het TOR in haar 25e jaargang is.

12. EERA

De VOR is lid van de EERA. Leden hebben dientengevolge aanzienlijke reductie op de conferentieprijs. Bovendien krijgen de Divisie drie keer per jaar vijf exemplaren van het EERA Bulletin. Individuele leden krijgen momenteel nog geen afleveringen van het EERA Bulletin. Prof. Dr. N. Verloop is lid van de Executive Committee van de EERA en bezoekt de vergaderingen.

De ECER-conferentie in 1999 wordt niet samen met EARLI georganiseerd. De belangen liggen te ver uiteen. De ECER-conferentie in 2000 wordt niet in Parijs maar in Lille georganiseerd.

Wat betreft de kwaliteit van de conferentie: deze was in 1997 (Frankfurt) veel beter dan in 1996 (Sevilla). Er wordt aan gewerkt om de selectiecriteria verder aan te scherpen. De netwerken spelen een meer prominente rol dan bij de conferentie in Sevilla, hetgeen de selectie en programmering ten goede komt.

Dr. M. Mulder is editor geworden van het EERA Bulletin voor de jaren 1998-2001.

13. VOR-symposium Studiehuis

Het eerste VOR-Symposium heeft plaats gevonden in het Educatorium te Utrecht. Het thema was het Studiehuis. Na algemene inleidingen over dit thema zijn vanuit de Divisies resultaten van onderzoek gepresenteerd die bij de invoering van het studiehuis in het onderwijs gebruikt kunnen worden.

Trekkers van dit symposium waren Prof. Dr. J.J.H. van de Akker en Prof. Dr. J.M. Pieters. Er was een grote opkomst en de verdeling van de aanwezigen over praktijk wetenschap was ongeveer gelijk. De gastvrijheid van Universiteit Utrecht werd zeer op prijs gesteld.

14. VOR-prijs

De VOR-prijs is in 1997 niet uitgereikt. De commissie heeft in de loop van het jaar een nieuwe samenstelling gekregen. Prof. Dr. L.F.W. de Klerk, Prof. Dr. L.J.Th van der Kamp en Prof. Dr. B.P.M. Creemers hebben de commissie verlaten en zijn bedankt voor hun inzet voor de jury. Nieuwe leden van de jury zijn: Prof. Dr. P.R.J. Simons, Prof. Dr. T.A.B. Sniijders en Prof. Dr. N. Verloop. Prof. Dr. G.W. Meijnen is benoemd als voorzitter. De zittingstermijn van de leden van de jury is vier jaar.

De regeling van de VOR-prijs is verder aangescherpt: ontvankelijk zijn alleen die proefschriften waarvan de promotor is verbonden aan een Nederlandse of Vlaamse universiteit. En de winnende kandidaat wordt voorafgaand aan de prijsuitreiking op de hoogte gesteld van de uitslag.

Verslag VOR-ledenvergadering 13 mei 1998

18.15 - 19.00 uur Universiteit Twente (tijdens ORD 1998)

Aanwezig

Leden: Th. Bergen, S. Dijkstra, A. van der Hoeven-van Doornum, G. Kanselaar, W. Kuiper, D. Longayroux, G. Lutgens, M. Pakkert, M. Peters-Sips, A. Pilot, F. Riemersma, J. van der Sanden, P.R.J. Simons, J. Veenhoven, C. van Vilsteren, H. Wijffels

Bestuursleden: J.J.H. van den Akker, Tj. Imbos (penningmeester), M. Mulder (secretaris), J.M. Pieters, A. Veerman, N. Verloop (voorzitter)

Bericht van verhindering: A.B. Dijkstra, H. van Hout, J. Huisman, P. Maassen, B. Versloot

1. Opening en agendavaststelling

De voorzitter opent de vergadering en heet de aanwezigen hartelijk welkom.

2. Verslag Algemene Ledenvergadering 21 mei 1997

Abusievelijk was het jaarverslag 1996-1997 meegestuurd met de vergaderstukken. Het correctiestuk, het verslag van de VOR-ledenvergadering van 21 mei 1997 is voor de vergadering uitgedeeld. De voorzitter vraagt een korte leespauze. Na de leespauze wordt het verslag aan de orde gesteld. Het verslag wordt ongewijzigd goedgekeurd en naar aanleiding van het verslag worden geen vragen gesteld.

3. Mededelingen

De voorzitter deelt mee dat wordt nagegaan of er combinatielidmaatschappen mogelijk zijn met de NVO. Dat is momenteel al het geval met de NVvO en VELON. Leden van deze vereniging kunnen tegen een gereduceerd tarief lid worden van de VOR. Om na te gaan of er momenteel al niet veel dubbellidmaatschappen voorkomen, worden binnenkort de ledenbestanden van beide verenigingen naast elkaar gelegd.

Verder deelt de voorzitter mee dat het initiatief om Divisietijdschriften met korting aan te bieden niet tot succes heeft geleid. De desbetreffende uitgevers zijn niet bereid voldoende kortingen te geven. Het bestuur heeft besloten niet verder door te gaan met dit initiatief.

Kanselaar merkt op dat Nederlandse NIP-leden hun eventuele APA-lidmaatschap kunnen betalen via het NIP. Wellicht kan dat ook gedaan worden met de AERA. Er wordt echter opgemerkt dat het AERA-lidmaatschap al bijzonder goedkoop is gezien de hoeveelheid tijdschriften die leden daarvoor krijgen. Aanvullende abonnementen op andere AERA-tijdschriften zijn eveneens goedkoop.

Naar aanleiding van dit punt merkt Mulder op dat de Divisie BBV dit jaar de VOR-presentatie tijdens de AERA heeft verzorgd. Sprekers waren: Kessels, Simons en Mulder en er waren ongeveer veertig belangstellenden. De sessie is georganiseerd in samenwerking met SIG Training and Business and Industry en is verder gesponsord door andere AERA-Divisies. Er was ook een Amerikaanse voorzitter en discussiant (L. Resnick). Deze wijze van organiseren van een VOR-presentatie wordt aanbevolen in de toekomst.

4. Jaarverslag 1997-1998

Het Jaarverslag 1997-1998 wordt aan de orde gesteld. Opgemerkt wordt dat het aantal leden weer stijgt.

Verder wordt meegedeeld dat de volgende ORDs worden georganiseerd in 1999 door de

Katholieke Universiteit Nijmegen, in 2000 door de Rijksuniversiteit Leiden, in 2001 door de Universiteit van Amsterdam, en in 2002 door de Universiteit van Antwerpen.

5. Financiën

Mw. Peters-Sips doet verslag namens de kascontrolecommissie. De commissie heeft op 16 april 1998 de stukken ontvangen van de penningmeester. Uit de winst- en verliesrekening blijkt dat een behoorlijk deel van de contributie van 1996 nog is geïnd in 1997. De balans geeft een gunstig beeld. De commissie heeft de dagafschriften gecontroleerd en op basis daarvan is de commissie tot de conclusie gekomen dat de penningmeester consciëntieus te werk is gegaan.

De begroting voor het jaar 1999 is realistisch gezien de winst- en verliesrekening van 1997. De financiële positie van de VOR kan gunstig worden genoemd. De commissie heeft waardering voor het gevoerde financiële beleid en stelt voor de penningmeester decharge te verlenen.

De voorzitter stelt de vergadering voor dat advies over te nemen. De vergadering besluit dat en dechargeert de penningmeester.

De voorzitter vervolgt met de mededeling dat de huidige kascontrolecommissie (mw. Peters-Sips en de heren Stijnen en Van Wieringen) inmiddels drie jaren in functie is geweest en dat er een nieuwe commissie moet worden gekozen. Het bestuur heeft drie kandidaten gepolst en bereid gevonden zitting te nemen in de kascontrolecommissie. Het zijn de heren Van Bussel (Katholieke Universiteit Brabant en oud-penningmeester van de VOR), Thijssen (Rabobank Nederland) en Van Vilsteren (Universiteit Twente). Er zijn geen bezwaren tegen deze kandidaten en de vergadering stemt in met hun benoeming.

Vervolgens krijgt de penningmeester het woord voor een toelichting op de financiële stukken. Hij brengt naar voren dat er nog twee rekeningen van het TOR moeten worden betaald. De drukker heeft deze onlangs in rekening gebracht. Verder zit er nog voor circa f 9.000,- aan contributiegelden onder de leden. Er wordt een poging gedaan dat geld alsnog te innen, maar een deel van dat bedrag moet als verloren worden beschouwd. Verder is de penningmeester van mening dat de stukken er goed uitzien, hij sluit zich aan bij de constatering van de kascontrolecommissie. De balans 1997 ziet er goed uit en de begroting 1999 is realistisch.

Er worden verder geen opmerkingen gemaakt over de stukken en de vergadering keurt ze vervolgens goed.

6. Verslag ontwikkeling Website Hoger Onderwijs

Pilot doet verslag van de activiteiten met betrekking tot de elektronische communicatie binnen de VOR die zijn uitgevoerd door de Divisie Hoger Onderwijs. Hij merkt op dat de constatering van de vorige ledenvergadering juist was, namelijk dat het opzetten en onderhouden van een systeem voor elektronische communicatie binnen de Divisie een arbeidsintensieve aangelegenheid is. De Divisie heeft gekeken naar de mogelijkheid van een elektronische nieuwsbrief om een website op te zetten. De behoefte aan specifieke Divisie-informatie via het WWW wordt echter deels overbodig gemaakt door het bestaan van andere circuits, zoals Sunco. Er is wel een aanvraag ingediend bij Surf. Deze is ook goedgekeurd. Gedacht wordt aan een vervolg van het initiatief en om berichten over onderzoek op het net te plaatsen. Belangrijk daarbij is de leden af te schermen tegen ongewenste post. Wellicht kan Kirschner aan de Open Universiteit starten met de website, maar voorafgaand aan deze vergadering heeft Pilot helaas geen contact met hem gehad over dit punt.

Opgemerkt wordt dat de Divisie LLG ook belangstelling heeft geuit voor het opzetten van een website. Zij is daarvoor doorverwezen naar de Divisie HO. De Divisie LLG wacht nu op de ontwikkelingen binnen de Divisie HO.

De voorzitter stelt voor het besluit over dit punt van vorig jaar te herhalen: de Divisie HO wordt gevraagd de mogelijkheden verder te exploreren en na te gaan op welke wijze het

Divisie-initiatief kan worden overgedragen naar de andere Divisies.

Opgemerkt wordt dat het VOR Promovendi Overleg ook belangstelling heeft voor een website.

Van den Akker is van mening dat herhalen van het besluit van vorig jaar erg passief overkomt. De VOR zou best actiever kunnen zijn op dit punt en meer kunnen investeren in dit initiatief. Wellicht dat een aantal kosten centraal gedragen kunnen worden. Van belang is ook op VOR-niveau na te denken over dit punt en om het initiatief niet geheel te leggen bij de Divisie(s), vooral ook omdat veel leden hun gegevens onvolledig of te laat hebben ingeleverd. De voorzitter brengt daar tegen in dat wanneer dit onderwerp kennelijk op Divisieniveau onvoldoende prioriteit heeft, het weinig zinvol is hiervoor op VOR-niveau faciliteiten in het leven te roepen. Hij vraagt Van den Akker in elk geval vanuit de Divisie Curriculum (waar dit onderwerp kennelijk wel prioriteit heeft) de initiatieven af te stemmen met die van de Divisie Hoger Onderwijs.

Voorgesteld wordt dit punt verder te bespreken in het Algemeen Bestuur. Simons brengt naar voren dat het van belang is vooral te kijken naar het onderhoud.

Volgens Pilot is een belangrijk punt dat er eigenlijk geen vraag is naar de informatie via een Divisie-website. Simons stelt echter dat een aanbod de vraag ook kan genereren.

De vergadering besluit dit punt verder te bespreken in het Algemeen Bestuur.

7. VOR Jaarboek 1998

De voorzitter deelt mee dat het VOR Jaarboek 1998 is verschenen en hij toont een exemplaar. Hij vermeldt dat dit de eerste versie is en dat het opgevat moet worden als een nul-nummer. Gebaseerd op dit exemplaar zal er jaarlijks een update worden gemaakt. Omdat er veel veranderingen zijn opgetreden sinds de MUB, is de naamgeving van verschillende eenheden niet volstrekt duidelijk en zitten er verder verschillende kleine foutjes in het Jaarboek. Omdat er nu een nulnummer is, kan in de volgende jaren echter worden voortgebouwd op deze versie. De voorzitter bedankt de secretaris en zijn assistenten voor het verrichte werk en verwacht dat de leden reeds veel gebruiksgemak zullen hebben van deze versie.

Opgemerkt wordt dat het Jaarboek veel informatie bevat die op een VOR-website zou kunnen worden gezet.

8. Ledenwerving

De voorzitter deelt mee dat er verschillende acties zijn uitgevoerd om het aantal leden te vergroten. Leden van het VOR Promovendi Overleg, en de AIO's die nog geen VOR-lid zijn, zijn benaderd met het vriendelijke verzoek te overwegen lid te worden van de VOR. De desbetreffende hoogleraren is gevraagd dit te bevorderen, onder meer door de kosten voor het lidmaatschap van de VOR voor promovendi declarabel te maken. Er is een brief verspreid van de Faculteit Toegepaste Onderwijskunde waarin wordt meegedeeld dat Twentse promovendi deze kosten kunnen declareren. Verder wordt gewezen op de actie in de richting van de NVO en verwacht wordt dat ook het Jaarboek 1998, dat op ruime schaal zal worden verspreid, een wervend effect zal hebben.

Uit de vergadering wordt opgemerkt dat het interessant kan zijn afgestudeerde kandidaten informatie te verstrekken over de VOR. Dit gebeurt al bij Pedagogische Studiën. De voorzitter zal de lijst met contactpersonen die dit per universiteit verzorgen ter beschikking stellen. Verder is het interessant de alumni vergaderingen te benaderen met het verzoek informatie over de VOR te verspreiden.

9. Bestuursverkiezing

De voorzitter licht de kandidatuur toe van de verschillende personen. Hij merkt op dat Kuiper, Mulder en Veerman worden voorgedragen door het VOR-bestuur (Veerman is ook voorgedragen door het VOR Promovendi Overleg). De kandidaten Van den Akker, Dijkstra, Pieters, Sanders en Van Wieringen worden voorgedragen door de Divisies.

Verder stelt de voorzitter een wijziging voor in de kandidatenlijst, tenzij de vergadering daar bezwaar tegen maakt. De wijziging is dat Van Hout zich niet herkiesbaar stelt en dat de Divisie HO in zijn plaats Jochems van de Open Universiteit voorstelt. De vergadering maakt geen bezwaar tegen dit amendement. Verder deelt de voorzitter mee dat er geen voorstellen voor tegenkandidaten zijn gedaan en stelt hij voor de voordracht over te nemen. De vergadering neemt dit voorstel over en benoemt de kandidaten.

Het gehele bestuur, alsmede affiliatie en de zittingstermijn van de leden, ziet er nu als volgt uit:

Prof. dr. N. Verloop	Voorzitter (Leiden)	1996-1999
Dr. M. Mulder	Secretaris (Twente)	1998-1999
Dr. W. Kuiper	Penningmeester (Twente)	1998-2001
Prof. dr. J. van den Akker	Divisie Curriculum (Twente)	1998-2001
Dr. A.B. Dijkstra	Divisie Onderwijs & Samenleving (Groningen)	1998-2001
Prof. dr. W. Jochems	Divisie Hoger Onderwijs (Heerlen)	1998-2001
Prof. dr. J. Pieters	Divisie Leren & Instructie (Twente)	1998-1999
Prof. dr. P. Sanders	Divisie Methodologie & Evaluatie (CITO)	1998-2001
Prof. dr. F. van Wieringen	Divisie Beleid & Organisatie (Amsterdam)	1998-2001
Dr. B. Versloot	Divisie Beroepsonderwijs, bedrijfsopleidingen en Volwasseneneducatie (Utrecht)	1996-1999
Prof. dr. Th. Wubbels	Divisie Lerarenopleiding en Leraarsgedrag (Utrecht)	1997-2000
Mw. Drs. A. Veerman	Promovendi	1998-2001

Van den Akker vraagt naar de consequenties van het terugtreden van Van Hout als vice-voorzitter. Verloop merkt op dat de vice-voorzitter door het Algemeen Bestuur uit haar midden wordt gekozen.

Simons merkt op dat in de toekomst ook moet worden gelet op de spreiding van de bestuursleden over de universiteiten. Er zijn nu vier bestuursleden uit Twente en niemand uit Nijmegen. Verloop merkt op dat een deel van dit probleem onoplosbaar is omdat Divisies hun kandidaten kunnen voorstellen. Het bestuur kandideert wel de voorzitter, secretaris en penningmeester. Daarin kan het bestuur dus sturen. Mulder merkt op dat hij nog slechts één jaar beschikbaar is voor het voeren van het secretariaat, teneinde de financiële administratie en de ledenadministratie beter op elkaar af te stemmen. De voorzitter stelt dat dit laatste de komende periode hoge prioriteit moet hebben, gezien het feit dat zich in het recente verleden op dit punt nogal wat problemen hebben voorgedaan. Toegezegd wordt dat bij volgende benoemingen van leden van het Algemeen Bestuur bezien zal worden of rekening gehouden kan worden met de spreiding over universiteiten.

10. Rondvraag

Pieters biedt aan om na te gaan of de VOR-website kan worden gekoppeld aan die van de Faculteit Toegepaste Onderwijskunde. Pilot merkt op dat de Divisie HO bezwaren had tegen het onderbrengen van de website bij de OU. Als er een VOR-website komt, moet er een duidelijk afzonderlijk geheel gebouwd worden dat ook onafhankelijk van de instelling bereikbaar is. Dat is volgens Pieters in principe ook mogelijk. De vergadering stemt in met het voorstel en afgesproken wordt dat dit punt terugkomt in de vergadering van het Algemeen Bestuur.

11. Sluiting

De voorzitter sluit de vergadering en bedankt de aanwezigen voor hun inbreng.

Mededeling

Van de organisatiecommissie ORD'99 ontving de redactie de volgende mededeling.

De 26^{ste} Onderwijs Research Dagen worden het komend jaar georganiseerd door de Katholieke Universiteit Nijmegen en wel op donderdag 20 en vrijdag 21 mei 1999. De Afdeling Onderwijskunde, het Instituut voor Toegepaste Sociale Wetenschappen (ITS), de Universitaire Lerarenopleiding (UNILO) en het Instituut voor Onderzoek van het Wetenschappelijk Onderwijs (IOWO) zorgen samen voor de organisatie.

Als voorzitter van de organisatiecommissie nodig ik u van harte uit om aan de ORD'99 een bijdrage te leveren. Zoals gewoonlijk zullen onderwijsonderzoekers uit Nederland en Vlaanderen hun onderzoeksbevindingen presenteren en delen met collegae en andere geïnteresseerden.

De Onderwijs Research Dagen staan dit jaar speciaal in het teken van samenwerkend leren (met en zonder computers), een uitdagend thema dat de laatste jaren weer sterk in de belangstelling staat. Meer dan in de vorige jaren het geval was willen we op deze ORD aan divisieoverstijgende activiteiten ruimte bieden. Deelnemers worden daarom speciaal aangemoedigd om voorstellen in te dienen die betrekking hebben op het terrein van twee of meer divisies.

Er is voorzien in een diversiteit aan presentatievormen. Naast de inmiddels vertrouwde formats van paper- en poster-sessies en symposia, is er dit jaar de mogelijkheid om onderzoek te bespreken in de vorm van een interactieve ronde-tafel sessie of een forum. Uiteraard zullen eveneens nationaal en internationaal vermaarde gastsprekers worden uitgenodigd om de conferentie toe te spreken. Zo zullen o.a. Mark Smylie (University of Illinois), Geoffrey Walford (Oxford University), en Harvey Goldstein (University of London) hun opwachting maken. Als sociale activiteit zal tijdens de eerste conferentiedag een receptie in het eeuwenoude stadhuis van Nijmegen plaatsvinden, waarna de deelnemers kunnen genieten van het traditionele conferentie-diner.

Alle informatie over deelname is te vinden op de ORD'99-website, en wel op het volgende adres: www.socsci.kun.nl/ped/owk/ord99.

Namens de organisatiecommissie spreek ik de welgemeende hoop uit u komend voorjaar op de terreinen van de KU Nijmegen te kunnen begroeten.

Prof. dr. P. Robert-Jan Simons
Voorzitter Organisatiecommissie ORD'99

Vakkenpakketkeuze en de aansluiting tussen het voortgezet en hoger onderwijs

Rolf. K.W. van der Velden en Maarten H.J. Wolbers

Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), Faculteit der Economische Wetenschappen en Bedrijfskunde, Universiteit Maastricht, Postbus 616, 6200 MD Maastricht.

In dit artikel is nagegaan in hoeverre de vakkenpakketkeuze van invloed is op de aansluiting tussen de tweede fase van het voortgezet onderwijs (HAVO/VWO) en het hoger onderwijs (HBO/WO). Hiertoe hebben we gebruik gemaakt van het schoolverlatersonderzoek 'Registratie van Uitstroom en Bestemming van Schoolverlaters' (RUBS) uit 1997. Uit de resultaten blijkt dat gediplomeerde schoolverlaters van het HAVO of het VWO over het algemeen doorstromen naar een vervolgopleiding in het hoger onderwijs die redelijk 'past' bij het gekozen vakkenpakket. Alleen bij de zogenoemde exacte vakkenpakketten is er sprake van een 'weglekeffect' naar niet-aansluitende vervolgopleidingen. Daarnaast komt naar voren dat de vakkenpakketkeuze zelf en de aansluiting van het vakkenpakket op de vervolgopleiding nauwelijks van belang zijn bij de problemen in de aansluiting tussen het voortgezet en hoger onderwijs. De verschillende maatstaven waaraan dergelijke problemen zijn afgemeten (subjectief ervaren oordeel over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding, voortijdige studie-uitval en spijt van de gekozen vervolgopleiding) hangen niet samen met het gekozen vakkenpakket, of met de aansluiting tussen het vakkenpakket en de gekozen vervolgopleiding. We concluderen dat de onlangs ingevoerde studieprofielen niet tot een verbetering in de aansluiting tussen het voortgezet en hoger onderwijs zullen leiden.

INLEIDING

Het kiezen van een vakkenpakket in de tweede fase van het voortgezet onderwijs (HAVO/VWO) is van belang voor een vervolgopleiding in het hoger onderwijs (HBO/WO). Naast de functie van selectie en allocatie dient de vakkenpakketkeuze vooral om de aansluiting tussen het voortgezet en hoger onderwijs te bevorderen. Dat aan deze aansluiting veel waarde wordt gehecht, blijkt wel uit de herstructurering van de tweede fase van het voortgezet onderwijs die momenteel plaatsvindt. Uitgangspunt van deze herstructurering is de in 1991 door het toenmalige Ministerie van Onderwijs en Wetenschappen uitgebrachte nota 'Profiel van de tweede fase voortgezet onderwijs'. In deze onderwijsnota wordt voorgesteld om de kwalificatiestructuur in de tweede fase van het voortgezet onderwijs te vernieuwen en de aansluiting met het hoger onderwijs te verbeteren. Aanleiding voor de nota vormden de problemen die worden waargenomen in de tweede fase. Het gaat met name om de aansluiting tussen het HAVO/VWO en het hoger onderwijs, naast de problemen van het voortijdig schoolverlaten, de aansluiting tussen het beroepsonderwijs en de arbeidsmarkt, inefficiënte leerwegen en de ongelijke deelname aan het onderwijs tussen verschillende maatschappelijke groeperingen.

Een van de hoofdtaken van het onderwijs in de tweede fase van het voortgezet onderwijs is om jongeren goed toe te rusten voor het hoger onderwijs. Hiervoor dient het onderwijsaanbod te zijn afgestemd op de kwalificaties die nodig zijn voor de verschillende opleidingssectoren in het hoger onderwijs. Deze afstemming zou niet optimaal zijn. Aankomende

studenten zouden kennis en vaardigheden ontberen die noodzakelijk zijn voor het volgen van een vervolgopleiding in het hoger onderwijs. Het gaat hier vooral om studievaardigheden, studieplanning en schriftelijke uitdrukkingsvaardigheden. Daarnaast zouden problemen in de aansluiting tussen de tweede fase van het voortgezet onderwijs en het hoger onderwijs te wijten zijn aan een gebrek aan samenhang tussen de examenvakken in het HAVO en het VWO en een onvoldoende afstemming van de vakinhouden op vervolgopleidingen.

De invoering van zogenoemde studieprofielen sinds het schooljaar 1998-1999 - ter vervanging van het vakkenpakket - moet leiden tot een meer inhoudelijk samenhangend onderwijsaanbod, waarin een evenwicht bestaat tussen verbreding en verdieping van vaardigheden benodigd voor een vervolgstudie in het hoger onderwijs. Er zijn vier studieprofielen onderscheiden: Cultuur en Maatschappij, Economie en Maatschappij, Natuur en Techniek en Natuur en Gezondheid. Ieder studieprofiel bestaat uit drie onderdelen. Allereerst is er een algemeen verplicht deel (bijna 50 procent van het studieprogramma). Daarnaast is er een verplicht gedeelte (ongeveer 30 procent) dat specifiek is voor elk profiel. Ten slotte is er een vrij gedeelte (een kleine 20 procent). Dit gedeelte kan door de leerling worden ingevuld ter verbreding en verdieping van het eigen profiel, maar ook in de vorm van een deel van een ander profiel.

Tegelijkertijd met de komst van de studieprofielen is de tweede fase van het voortgezet onderwijs officieel ingericht als een studiehuis. In dit studiehuis worden scholieren geleerd actief en zelfstandig kennis en vaardigheden tot zich te nemen ('leren leren') (Wijnen, 1997). Om dit te bereiken wordt een deel van de verantwoordelijkheid van de leraar voor het leerproces overgedragen op de leerling. De leraar zal geen leider, maar *begeleider* zijn in het leerproces van leerlingen. Zijn rol zal meer gericht zijn op het stimuleren van kennisverwerving bij leerlingen en minder op het (klassikaal) overdragen van kennis. Hiervoor is een leeromgeving nodig waarbij klassikale instructies worden afgewisseld met werkgroepen waarin samen wordt gestudeerd en individuele studie-uren waarin leerlingen zelfstandig opdrachten uitvoeren met moderne informatievoorzieningen (bibliotheek, computer, internet) binnen handbereik.

Een belangrijke vooronderstelling bij de invoering van de studieprofielen is dat de samenstelling van het vakkenpakket in de oude situatie een belangrijke oorzaak is voor de geconstateerde problemen in de aansluiting tussen het voortgezet en hoger onderwijs. Enerzijds zou de vrijheid van pakketkeuze ertoe kunnen leiden dat leerlingen vakken kiezen die onderling onvoldoende samenhang vertonen. Anderzijds zoekt men de oorzaak in een geringe relatie tussen het gekozen vakkenpakket in het voortgezet onderwijs en de gekozen vervolgopleiding. Door nu bij de opleidingen in het hoger onderwijs richtlijnen op te stellen van het studieprofiel verwacht men de aansluiting tussen het voortgezet en hoger onderwijs in belangrijke mate te kunnen verbeteren.

De vraag is of deze vooronderstelling gerechtvaardigd is. Het is moeilijk op dit moment reeds een definitieve evaluatie te maken van de effecten van de invoering van studieprofielen. De eerste scholieren van deze nieuwe profielen zullen immers op zijn vroegst in het studiejaar 2000-2001 (in het geval van het HAVO) of 2001-2002 (in het geval van het VWO) in het hoger onderwijs terechtkomen. Dan zal ook pas blijken wat de effecten zijn geweest.

Niettemin kan ook nu al een analyse worden gemaakt van de te verwachten effecten. In dit artikel gaan we hiervoor een 'simulatie' uitvoeren van de invoering van de studieprofielen. Op basis van gegevens afkomstig uit een schoolverlatersonderzoek van het Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA) worden de vakkenpakketten van schoolverlaters van het HAVO en het VWO zo goed mogelijk ingedeeld in één van de vier onderscheiden profielen. Enerzijds kan daarmee worden nagegaan in hoeverre de oude pakketten zijn in te delen in één van de vier nieuwe profielen. Dit geeft een aanwijzing in hoeverre in de oude situatie sprake was van een onsamenhangende pakketkeuze. Naarmate de pakketten beter in te delen zijn in één van de vier profielen is er minder reden om een dergelijke samenhang dwingend van bovenaf op te leggen. Anderzijds kan worden bestudeerd in hoeverre een 'verkeerde' pakketkeuze leidt tot aansluitingsproblemen in het hoger onderwijs. Onder 'verkeerd' wordt in

dit geval verstaan een studieprofiel dat niet aansluit bij de gekozen sector in het hoger onderwijs. Bijvoorbeeld: de keuze voor een technische studie met een profiel Economie en Maatschappij of de keuze voor een letterenstudie met een profiel Natuur en Gezondheid. De vraag is of leerlingen die een 'verkeerd' profiel hebben gekozen meer kans hebben om uit te vallen in de vervolgopleiding en of zij vaker een negatief oordeel hebben over de aansluiting tussen het voortgezet en hoger onderwijs. Inzicht hierin kan een indicatie geven van de effectiviteit van de genomen maatregel. Wanneer immers degenen met een 'verkeerd' profiel vaker problemen ervaren in de aansluiting, dan is het gewenst om duidelijke(r) eisen te stellen aan het gekozen vakkenpakket en zal de invoering van studieprofielen leiden tot een verbetering van de aansluiting tussen het voortgezet en hoger onderwijs.

In het artikel beantwoorden we daartoe de volgende onderzoeksvragen. In de eerste plaats wordt bekeken in hoeverre de oude vakkenpakketten zijn in te delen in één van de vier nieuwe studieprofielen. In de tweede plaats wordt bestudeerd in hoeverre de keuze voor een bepaald type vakkenpakket samenhangt met bepaalde kenmerken van leerlingen. In de derde plaats wordt bestudeerd of een vakkenpakket richtinggevend is voor de keuze van een bepaalde vervolgopleiding in het hoger onderwijs. De vraag daarbij is in hoeverre studenten die in het voortgezet onderwijs een bepaald vakkenpakket hebben gevolgd een aansluitende studierichting in het hoger onderwijs kiezen. In de vierde plaats wordt onderzocht of studenten, die een studierichting hebben gekozen die overeenkomt met een bepaald vakkenpakket, vaker tevreden zijn over de aansluiting tussen de afgesloten HAVO- of VWO-opleiding en de vervolgopleiding in het HBO of WO dan degenen die een niet-aansluitende vervolgopleiding hebben gekozen. In de vijfde plaats wordt nagegaan of studenten, die op basis van hun vakkenpakket een aansluitende studierichting in het hoger onderwijs zijn gaan volgen, deze opleiding minder snel voortijdig verlaten dan studenten die een minder goed passende vervolgopleiding hebben gekozen. Bovendien wordt voor de studenten die de opleiding op het moment van ondervraging (nog) niet hebben verlaten in kaart gebracht of zij de vervolgopleiding opnieuw zouden kiezen en in hoeverre dit samenhangt met het vakkenpakket van de vooropleiding.

ONDERZOEKSOPZET

Gegevens en operationalisaties

De gegevens die in dit artikel worden gebruikt, zijn afkomstig uit het schoolverlatersonderzoek *Registratie van Uitstroom en Bestemming van Schoolverlaters* (RUBS) dat in het najaar van 1997 is uitgevoerd door het ROA. Deze schriftelijke enquête, die sinds het begin van de jaren negentig jaarlijks wordt gehouden, heeft betrekking op schoolverlaters van het voorbereidend beroepsonderwijs (VBO), het algemeen voortgezet en voorbereidend wetenschappelijk onderwijs (MAVO/HAVO/VWO) en het secundair beroepsonderwijs ((K)MBO/Leerling-*wezen*). In 1997 zijn zo'n 13.000 schoolverlaters ondervraagd zodat een landelijk representatief beeld kan worden geschetst van de uitstroom en bestemming van deze groepen schoolverlaters. Het moment van ondervraging vindt anderhalf jaar na het voltooien van de opleiding plaats, hetgeen betekent dat de hier gepresenteerde cijfers zijn gebaseerd op schoolverlaters van het schooljaar 1995/1996. De verzamelde informatie heeft met name betrekking op de bestemming van schoolverlaters, waarbij zowel de doorstroom naar het vervolgonderwijs als de intrede op de arbeidsmarkt centraal staan. Daarnaast vindt aan de hand van een aantal vragen een evaluatie plaats van de verlaten opleiding.

Voor de huidige analyse hebben we de schoolverlaters geselecteerd die een HAVO of VWO opleiding hebben voltooid en vervolgens zijn doorgestroomd naar een studie in het hoger onderwijs. Van hen is bekend in welke vakken ze eindexamen hebben gedaan. De betreffende pakketkeuzes zijn ingedeeld in vier typen die zoveel mogelijk overeenkomst vertonen met de onlangs ingevoerde studieprofielen. Respondenten van wie het vakkenpakket niet aan de criteria van één van deze vier typen voldoet, zijn in een restcategorie ondergebracht. Er wordt

nadrukkelijk op gewezen dat de hier gehanteerde indeling niet exact dezelfde is als de indeling in studieprofielen. Dit heeft twee oorzaken. In de eerste plaats komen in de nieuwe profielen vakken voor die nog niet bestaan. Dit betreft met name de vakken Algemene Natuurwetenschappen en Culturele en Kunstzinnige Vorming. In de tweede plaats is een eenvoudige vertaling van de vakkenpakketten naar de profielen niet mogelijk, omdat het aantal vakken in de nieuwe profielen sterk toeneemt. Een zeer strikte vertaalslag zou ertoe leiden dat een groot aantal respondenten in de restcategorie terecht komt. Daarom is geprobeerd om zo veel mogelijk in de geest van de nieuwe profielen criteria te formuleren voor het toekennen van de bestaande vakkenpakketten aan één van de profielen. In bijlage 1 is weergegeven hoe onze indeling van vakkenpakketten samenhangt met de officiële indeling van profielen.

Een kenmerk dat gerelateerd is aan de profielen betreft de keuze voor een aansluitende vervolgopleiding in het hoger onderwijs. Kort gezegd geeft deze variabele aan of een bepaald type vakkenpakket aansluit bij een bepaalde opleidingssector in het hoger onderwijs. Bij de bespreking van tabel 3 zal meer gedetailleerd worden ingegaan op de operationalisatie van deze variabele.

Andere onafhankelijke variabelen die van belang zijn in dit artikel betreffen de aandacht voor studiekeuzebegeleiding in het voortgezet onderwijs (onvoldoende versus (ruim) voldoende), de vraag of de vervolgopleiding op school (bijvoorbeeld met dekaan of mentor) is besproken (ja versus nee), de examencijfers voor de vakken Nederlands en wiskunde¹ (lopend van 1-10) en studievertraging in het voortgezet onderwijs (ja versus nee).

Als achtergrondkenmerken worden in de analyse meegenomen: sekse (man versus vrouw), leeftijd² (17-18 jaar, 19-20 jaar, 21+ jaar) en etniciteit (autochtoon versus allochtoon).

Ten slotte wordt voor schoolverlaters van het VWO aangegeven of de vervolgopleiding betrekking heeft op het hoger beroeps onderwijs of op het wetenschappelijk onderwijs (HBO versus WO).

De afhankelijke variabele waarmee we onze analyse starten is de keuze voor een aansluitende vervolgopleiding in het hoger onderwijs. Daarna worden determinanten van het subjectieve oordeel over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding, de kans op voortijdige studie-uitval in het hoger onderwijs en de kans op het opnieuw kiezen van de vervolgopleiding onderzocht. De oorspronkelijke afhankelijke variabele ter meting van de keuze voor een aansluitende vervolgopleiding in het hoger onderwijs treedt dan op als onafhankelijke variabele. De subjectieve beoordeling van de aansluiting tussen de afgesloten opleiding en de vervolgopleiding is gebaseerd op de vraag aan schoolverlaters of zij de aansluiting met de gevolgd opleiding als goed, redelijk, matig of slecht beschouwen. Studie-uitval is bepaald aan de hand van de vraag of de vervolgopleiding in het hoger onderwijs voortijdig is verlaten. De kans op het opnieuw kiezen van de vervolgopleiding is afgemeten aan het antwoord op de vraag of - indien de opleiding in het hoger onderwijs nog steeds wordt gevolgd - men de vervolgopleiding, achteraf gezien, opnieuw zou kiezen.

Statistische analyse

Voor de statistische analyse van deze afhankelijke variabelen maken we gebruik van multiële regressie-analyse. In het geval van het subjectieve oordeel over de aansluiting tussen de verlaten opleiding en de vervolgopleiding gaat het om lineaire regressie-analyse. In de andere gevallen wordt vanwege het binaire karakter van de afhankelijke variabelen logistische regressie-analyse toegepast. In een logistisch regressiemodel wordt de logit of logodds (= natuurlijke logaritme uit de 'odds' of kansverhouding) over een variabele lineair afhankelijk gesteld van onafhankelijke variabelen (DeMaris, 1992). De interpretatie van de parameters komt overeen met die in een lineair regressiemodel: het intercept geeft de waarde van de logit aan bij $X=0$ en de effectparameter b toont de toename van de logit per eenheid van X . In de praktijk zijn we echter vooral geïnteresseerd in relatieve kansen en niet in logits. Bij het bespreken van de analyseresultaten berekenen we daarom de antilog uit de geschatte parameters (e^b), zodat de effecten in termen van kansverhoudingen kunnen worden gekwantificeerd.

RESULTATEN

Verschillen in vakkenpakketkeuze

De keuze voor een bepaald vakkenpakket hangt samen met verschillende individuele kenmerken (De Jong & Jungbluth, 1989). Allereerst is sekse natuurlijk een belangrijke factor bij de keuze voor een bepaald pakket. Meisjes laten andere motieven meespelen bij de keuze van een vakkenpakket dan jongens (Ten Dam, Van Eck & Volman, 1992; Willems & De Grip, 1993). Deze motieven hangen voornamelijk samen met seksespecifieke rolopvattingen, die zijn aangeleerd tijdens de opvoeding (door de ouders, de school en de inhoud van het onderwijs). Van meisjes wordt minder snel verwacht dat zij zich richten op technische en natuurwetenschappelijke vakken dan jongens. Bovendien waarderen meisjes het succes in deze vakken minder sterk dan jongens. In plaats daarvan richten zij zich op vakken waarin verzorgende en opvoedkundige aspecten een belangrijke plaats innemen.

Tabel 1: De keuze voor een vakkenpakket binnen het HAVO en het VWO naar sekse

Vakkenpakket	HAVO			VWO		
	Man %	Vrouw %	Totaal %	Man %	Vrouw %	Totaal %
Cultuur en Maatschappij	3	14	9	6	19	12
Economie en Maatschappij	23	33	29	24	36	30
Natuur en Techniek	46	9	26	57	25	41
Natuur en Gezondheid	5	12	9	8	14	11
Rest	22	32	28	6	6	6

segregatie-index: HAVO (38%); VWO (32%)

Bron: ROA, schoolverlatersonderzoek RUBS 1997

In tabel 1 is een overzicht weergegeven van de indeling van schoolverlaters naar het type vakkenpakket. Daarbij is een onderscheid gemaakt naar de gevolgde opleiding (HAVO of VWO) en sekse. De tabel laat zien dat de pakketten Economie en Maatschappij en Natuur en Techniek het meest voorkomen. Binnen het HAVO komt het pakket Economie en Maatschappij iets vaker voor dan het pakket Natuur en Techniek (respectievelijk 29 procent en 26 procent van de havisten kiest voor deze pakketten); in het VWO daarentegen wordt het pakket Natuur en Techniek meer gevolgd dan het pakket Economie en Maatschappij (respectievelijk 41 procent en 30 procent). Het valt op dat het percentage leerlingen dat niet kan worden ingedeeld in één van de vier vakkenpakketten bij het HAVO veel groter is dan bij het VWO. Gemiddeld genomen is voor ruim een kwart van de havisten het vakkenpakket niet te herleiden tot een bepaald type. Dit geldt overigens nog sterker voor vrouwen dan voor mannen. Studenten met een VWO-diploma zijn daarentegen veel beter in te delen in de verschillende pakketten. Slechts zes procent van hen - dit geldt zowel voor mannen als voor vrouwen - heeft zo'n gevarieerd vakkenpakket dat de combinatie ervan niet is om te zetten in één van de vakkenpakkettypen.

Uit tabel 1 komt ook duidelijk naar voren hoe groot de verschillen zijn tussen mannen en vrouwen in de keuze van het vakkenpakket. De segregatie-index geeft de som van de absolute verschillen in percentages jongens en meisjes die een bepaald pakket kiezen, gedeeld door twee. Deze index kan eenvoudig worden geïnterpreteerd als het percentage vrouwen dat een ander pakket zou moeten kiezen om tot eenzelfde verdeling te komen als bij mannen. Bij het VWO zou eenderde van de vrouwen een ander pakket moeten kiezen om eenzelfde verdeling te krijgen als bij mannen; bij het HAVO bedraagt dit percentage zelfs 38 procent. Mannen kiezen daarbij veel vaker voor het pakket Natuur en Techniek; vrouwen kiezen vaker voor Natuur en Gezondheid of voor de pakketten Cultuur en Maatschappij en Economie en Maat-

schappij. Dit geldt zowel voor het HAVO als het VWO en bevestigt het bekende sekspecifieke keuze patroon bij de samenstelling van het vakkenpakket.

Daarnaast hangt de keuze voor een bepaald vakkenpakket af van de behaalde schoolprestaties. De algemene veronderstelling is dat de samenstelling van een vakkenpakket in belangrijke mate verband houdt met de leercapaciteiten van leerlingen. Leerlingen die goed presteren in wiskunde hebben meer aanleg voor een exact vakkenpakket en kiezen daarom eerder voor een dergelijk pakket. Daar staat tegenover dat leerlingen met goede prestaties voor het vak Nederlands waarschijnlijk eerder een pakket Cultuur en Maatschappij volgen.

Tabel 2: Het gemiddelde examencijfer voor Nederlands en wiskunde, de gemiddelde leeftijd en het percentage studenten met studievertraging binnen het HAVO en het VWO naar vakkenpakket

Vakkenpakket	HAVO				VWO			
	Nederlands	wiskunde	leeftijd	% studievertraging	Nederlands	wiskunde	leeftijd	% studievertraging
Cultuur en Maatschappij	6,8	6,3	18,9	38	7,0	6,2	19,5	34
Economie en Maatschappij	6,6	6,8	18,8	39	6,9	6,5	19,4	24
Natuur en Techniek	6,3	6,6	18,8	34	6,8	7,1	19,4	24
Natuur en Gezondheid	6,7	6,8	18,8	40	6,8	6,4	19,3	33
Rest	6,7	6,4	19,6	18	7,3	6,5	19,6	20

Bron: ROA, schoolverlatersonderzoek RUBS 1997

In tabel 2 is de samenhang weergegeven tussen de pakketkeuze enerzijds en de prestatiekenmerken examencijfer Nederlands, examencijfer wiskunde, studievertraging in het HAVO/VWO en leeftijd anderzijds. Uit de tabel blijkt dat binnen het HAVO het gemiddelde examencijfer voor het vak Nederlands inderdaad het hoogst is voor leerlingen met het pakket Cultuur en Maatschappij. De verschillen met de andere vakkenpakketten zijn echter niet groot. Een uitzondering hierop vormt Natuur en Techniek. Leerlingen met dit vakkenpakket scoren gemiddeld een half punt lager voor hun examen Nederlands. Binnen het VWO zijn de verschillen tussen de pakketten in het gemiddelde examencijfer voor het vak Nederlands gering.

Voor wat het vak wiskunde betreft geldt een ander beeld. Binnen het HAVO hebben leerlingen met het vakkenpakket Natuur en Gezondheid of Economie en Maatschappij het hoogste cijfer voor wiskunde en zij met het pakket Cultuur en Maatschappij het laagste. Binnen het VWO behalen leerlingen met het pakket Natuur en Techniek gemiddeld genomen de hoogste cijfers voor wiskunde. Leerlingen met het pakket Cultuur en Maatschappij scoren daarentegen het laagst.

Verschillen in leeftijd bestaan nauwelijks tussen de onderscheiden vakkenpakketten. Wel zijn leerlingen van wie het vakkenpakket niet te herleiden valt tot één van de vier pakkettypen enkele maanden ouder (dit geldt zowel voor het HAVO als voor het VWO). Deze vertraging heeft overigens niet tijdens de afgesloten opleiding plaats gehad. Het betreft hier doormerkmers uit een lager schooltype. Deze zitten vaak met een pakket dat voortbouwt op het oude pakket aangevuld met enkele nieuwe vakken. Het percentage leerlingen dat studievertraging heeft opgelopen gedurende de HAVO- of VWO-opleiding is immers het laagst onder leerlingen uit de restcategorie. Ook hebben relatief weinig leerlingen met het vakkenpakket Natuur en Techniek studievertraging opgelopen. Leerlingen met een pakket Cultuur en Maatschappij of Natuur en Gezondheid doen daarentegen het langst over hun opleiding.

Vakkenpakket en de keuze voor een aansluitende vervolgopleiding

Nu een beschrijving is gegeven van de vakkenpakketkeuze en de samenhang met enkele individuele kenmerken kan worden nagegaan of de keuze voor een bepaald vakkenpakket van

invloed is op de keuze voor de onderwijsrichting van de vervolgopleiding in het hoger onderwijs. Daarvoor moet echter worden bepaald welke vervolgopleidingen als aansluitend kunnen worden beschouwd. In tabel 3 is aangegeven welke vakkenpakketten bij welke opleidingssectoren aansluiten³. De gearceerde vlakken geven de aansluitende combinaties weer. Overigens laten we studenten die een vakkenpakket hebben gevolgd dat niet is in te delen in één van de vier typen verder achterwege, omdat voor hen niet kan worden bepaald welke vervolgopleiding als aansluitend kan worden beschouwd. Daarnaast wordt in tabel 3 de verdeling van vakkenpakketten naar vervolgopleiding per opleidingssector getoond, waarbij in iedere rij de bovenste cijfers horizontaal zijn gepercentageerd en de onderste verticaal.

Tabel 3: De sectorkeuze van de vervolgopleiding naar vakkenpakket (een arcering duidt op (voldoende) aansluiting tussen het vakkenpakket en de sector van de vervolgopleiding)

Sector vervolgo- opleiding	Land- bouw	Onder- wijs	Tech- niek	Eco- nomie	Ge- zond- heids- zorg	Gedrag/ maat- schappij	Kunst/ taal/ cultuur	Recht/ open- bare orde	Natuur- weten- schap- pen	Totaal aan- sluiting
Vakkenpakket	%	%	%	%	%	%	%	%	%	%
Cultuur en Maatschappij	0 1	19 26	1 1	25 12	8 9	19 22	18 41	9 26	- -	91
Economie en Maatschappij	1 8	15 55	1 2	49 66	5 16	16 50	6 38	6 52	1 4	92
Natuur en Techniek	6 60	2 10	52 89	12 19	12 40	4 13	2 12	2 17	10 90	79
Natuur en Gezondheid	10 32	7 9	17 8	8 4	35 35	14 15	5 9	2 6	2 6	64
Totaal aansluiting	% 92	% 81	% 97	% 78	% 75	% 72	% 79	% 78	% 96	

Bron: ROA, schoolverlatersonderzoek RUBS 1997

Uit tabel 3 komt naar voren dat de schoolverlaters over het algemeen een vervolgopleiding in het hoger onderwijs kiezen die aansluit bij hun vakkenpakket. Dit geldt met name voor degenen met een pakket Cultuur en Maatschappij of Economie en Maatschappij. Van deze twee groepen kiest meer dan 90 procent een vervolgopleiding die aansluit bij het vakkenpakket. Ook voor de andere twee pakketten geldt dat een meerderheid van de schoolverlaters een aansluitende vervolgopleiding kiest, maar de 'weglek' naar niet-aansluitende sectoren is hier veel groter. Zo kiest van de afgestudeerden met een pakket Natuur en Gezondheid 36 procent voor een niet-aansluitende vervolgopleiding (met name Gedrag en Maatschappij), terwijl dit percentage voor Natuur en Techniek 21 procent bedraagt. Dit hangt uiteraard nauw samen met de veel grotere keuzemogelijkheid die schoolverlaters met een pakket Natuur en Techniek of Natuur en Gezondheid hebben. Zij hebben in de regel toegang tot vrijwel alle opleidingen in het hoger onderwijs, terwijl voor de andere schoolverlaters de keuze voor een belangrijk deel is beperkt tot de opleidingssectoren die hier als aansluitend worden beschouwd.

Vanuit de 'ontvangende' vervolgopleidingen bekeken is de aansluiting tussen het pakket en de opleidingssector over het algemeen hoog tot zeer hoog. De samenhang tussen het gevolgde vakkenpakket en de gekozen vervolgstudie is het sterkst bij de exacte opleidingssectoren.

Zo heeft maar liefst 97 procent van alle studenten die een technische vervolgopleiding in het hoger onderwijs zijn gaan doen, het vakkenpakket Natuur en Techniek (89 procent) of Natuur en Gezondheid (8 procent) gevolgd. Voor vervolgopleidingen in de opleidingssectoren Natuurwetenschappen en Landbouw gelden vergelijkbare percentages (respectievelijk 96 procent en 92 procent). Het percentage voor de sector Gezondheidszorg is iets minder hoog (75 procent). Ook voor niet-exacte opleidingssectoren geldt dat de instromende studenten over een aansluitend vakkenpakket beschikken, al liggen de percentages gemiddeld genomen iets lager. Van de studenten die een opleiding in de sector Onderwijs hebben gekozen, heeft 81 procent tijdens de vooropleiding in het voortgezet onderwijs het pakket Cultuur en Maatschappij (26 procent) of Economie en Maatschappij (55 procent) gevolgd. Voor de opleidingssector Kunst/taal/cultuur is dit percentage 79 procent, voor Recht/openbare orde en Economie 78 procent en voor Gedrag/maatschappij 72 procent.

Tabel 4: Resultaten logistische regressie-analyse van de keuze voor een aansluitende vervolgopleiding

	HAVO B	SE (B)	VWO B	SE (B)
Vakkenpakket				
Cultuur en Maatschappij (referentiegroep)	0		0	
Economie en Maatschappij	-0,47	0,46	0,44	0,40
Natuur en Techniek	-1,40**	0,46	-1,44**	0,35
Natuur en Gezondheid	-1,99**	0,47	-1,68**	0,37
Aandacht studiekeuzebegeleiding				
Onvoldoende (referentiegroep)	0		0	
(Ruim) voldoende	0,04	0,33	0,41*	0,24
Vulgoopleiding besproken				
Nee (referentiegroep)	0		0	
Ja	0,13	0,25	0,11	0,19
Examencijfer Nederlands	-0,28**	0,14	-0,13	0,12
Examencijfer wiskunde	0,49**	0,12	0,28**	0,11
Studievertraging in voortgezet onderwijs				
Nee (referentiegroep)	0		0	
Ja	0,67**	0,26	-0,37*	0,20
Leeftijd				
17-18 jaar (referentiegroep)	0		0	
19-20 jaar	-0,08	0,25	-0,37	0,80
21+ jaar	0,57	0,81	0,12	0,91
Sekse				
Man (referentiegroep)	0		0	
Vrouw	-0,04	0,26	-0,17	0,20
Etniciteit				
Autochtoon (referentiegroep)	0		0	
Allochtoon	0,68	0,80	0,73	0,78
Vulgoopleiding				
HBO (referentiegroep)			0	
WO			0,34*	0,19
Intercept	0,97	1,07	1,35	1,27
Model Chi²	57,47**		93,50**	
df	12		13	
N	625		927	

* = significant op 10%-niveau; ** = significant op 5%-niveau

Bron: ROA, schoolverlatersonderzoek RUBS 1997

In een nadere analyse wordt vervolgens onderzocht in hoeverre het vakkenpakket en andere kenmerken van het individu tegelijkertijd bepalend zijn voor de keuze van een aansluitende opleiding. Tabel 4 toont de uitkomsten van deze analyse. In de tabel is te zien dat schoolverlaters met het pakket Natuur en Techniek of Natuur en Gezondheid significant minder vaak een aansluitende vervolgopleiding kiezen dan schoolverlaters met het pakket Cultuur en Maatschappij. Dit geldt zowel voor het HAVO als voor het VWO. De kans voor studenten met het pakket Economie en Maatschappij om een aansluitende HBO- of WO-opleiding te volgen wijkt daarentegen niet significant af van die voor studenten met het pakket Cultuur en Maatschappij. De meest voor de hand liggende reden waarom studenten met een exact vakkenpakket zich minder sterk laten leiden door hun vakkenpakketkeuze is, zoals gezegd, de grotere keuzevrijheid met betrekking tot de vervolgopleiding. Doorstromers met een exact vakkenpakket hebben wel toegang tot een niet-exacte vervolgstudie in het hoger onderwijs, terwijl een eindexamenpakket zonder exacte vakken geen toegang geeft tot bijvoorbeeld een technische vervolgopleiding.

Bij het VWO bepaalt ook de aandacht voor studiekeuzebegeleiding de keuze van een aansluitende vervolgopleiding. Voor doorstromers die aangeven dat er tijdens de gevolgde VWO-opleiding voldoende of ruim voldoende aandacht werd besteed aan studiekeuzebegeleiding is de kansverhouding van het volgen van een aansluitende studierichting in het hoger onderwijs anderhalf keer groter dan voor hen die vinden dat de mate van studiekeuzebegeleiding onvoldoende is geweest⁴.

Daarnaast zijn schoolprestaties van invloed op de keuze van een aansluitende opleiding. Voor het HAVO geldt dat - gegeven het verschil in de kans op een aansluitende vervolgopleiding tussen de vakkenpakketten - een hoger examencijfer voor Nederlands samengaat met een geringere kans op een aansluitende studierichting in het hoger onderwijs. Een hoog wiskundecijfer tijdens het eindexamen leidt echter tot een grotere kans om de schoolloopbaan voort te zetten in een aansluitende studierichting (dit geldt ook voor het VWO). Blijkbaar kiezen veel scholieren in het voortgezet onderwijs een exact vakkenpakket om alle mogelijkheden voor een vervolgopleiding open te houden, maar besluiten alleen degenen die echt goed zijn in de exacte vakken (hier afgemeten aan het eindexamencijfer voor wiskunde) een aansluitende studierichting in het hoger onderwijs te volgen.

Studievertraging opgelopen tijdens de vooropleiding is ook van invloed op de keuze van een aansluitende vervolgopleiding. Voor het HAVO geldt dat voor gediplomeerden die langer dan de nominale duur over hun opleiding hebben gedaan, de kansverhouding van een aansluitende vervolgopleiding bijna twee keer zo groot is als voor schoolverlaters zonder studievertraging. Binnen het VWO geldt echter het omgekeerde: VWO'ers met studievertraging volgen minder vaak een aansluitende studie in het hoger onderwijs dan VWO'ers zonder vertraging. Tot slot blijkt uit tabel 4 dat schoolverlaters van het VWO die vervolgens naar de universiteit gaan, vaker een studie volgen die past bij hun vakkenpakket dan VWO'ers die na het behalen van hun VWO-diploma doorstromen naar het HBO.

De aansluiting tussen de afgesloten opleiding en de vervolgopleiding

Studenten die op basis van hun vakkenpakket in het HAVO of het VWO een passende studierichting in het hoger onderwijs zijn gaan volgen, blijken de aansluiting tussen de afgesloten opleiding en de vervolgstudie niet positiever te beoordelen dan studenten die een minder goed passende vervolgopleiding hebben gekozen. Uit tabel 5 blijkt dat de invloed van deze factor op het subjectieve oordeel over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding niet significant is. Bovendien bestaat er - op één uitzondering na - geen verschil in de subjectief ervaren aansluiting tussen de onderscheiden vakkenpakketten. Alleen studenten die tijdens hun HAVO-opleiding het pakket Natuur en Gezondheid hebben gevolgd, zijn minder te spreken over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding. Zij zijn ongeveer een half punt minder tevreden over de aansluiting (op een schaal lopend van 1-4) dan de groep die het pakket Cultuur en Maatschappij heeft gevolgd.

Tabel 5: Resultaten lineaire regressie-analyse van het subjectieve oordeel over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding

	HAVO B	SE (B)	VWO B	SE (B)
Vergoopleiding sluit aan bij vakkenpakket				
Nee (referentiegroep)	0		0	
Ja	-0,15	0,10	-0,08	0,07
Vakkenpakket				
Cultuur en Maatschappij (referentiegroep)	0		0	
Economie en Maatschappij	0,00	0,11	-0,08	0,09
Natuur en Techniek	-0,03	0,13	-0,12	0,09
Natuur en Gezondheid	-0,35**	0,14	-0,10	0,11
Aandacht studiekeuzebegeleiding				
Onvoldoende (referentiegroep)	0		0	
(Ruim) voldoende	0,31**	0,10	0,27**	0,08
Vergoopleiding besproken				
Nee (referentiegroep)	0		0	
Ja	0,10	0,08	-0,01	0,06
Examencijfer Nederlands	0,07	0,05	0,04	0,04
Examencijfer wiskunde	0,03	0,04	0,08**	0,03
Studievertraging in voortgezet onderwijs				
Nee (referentiegroep)	0		0	
Ja	-0,21**	0,08	-0,12*	0,06
Leeftijd				
17-18 jaar (referentiegroep)	0		0	
19-20 jaar	0,11	0,08	-0,06	0,23
21+ jaar	-0,02	0,23	-0,10	0,25
Sekse				
Man (referentiegroep)	0		0	
Vrouw	0,24**	0,08	0,05	0,06
Etniciteit				
Autochtoon (referentiegroep)	0		0	
Allochtoon	-0,00	0,19	-0,15	0,18
Vergoopleiding				
HBO (referentiegroep)			0	
WO			-0,34**	0,06
Intercept	2,04**	0,36	2,68**	0,37
Aangepaste R ²	0,05		0,06	
N	625		927	

* = significant op 10%-niveau; ** = significant op 5%-niveau

Bron: ROA, schoolverlatersonderzoek RUBS 1997

De begeleiding bij de studiekeuze is een belangrijke factor bij het subjectieve oordeel over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding. Zowel voor het HAVO als voor het VWO geldt dat de aandacht voor studiekeuzebegeleiding positief samenhangt met de ervaren aansluiting. De vraag is echter in welke causale richting het gevonden verband wijst. In tabel 5 is er van uitgegaan dat een (ruim) voldoende aandacht voor studiekeuzebegeleiding leidt tot een goede aansluiting tussen de afgesloten opleiding en de vervolgoopleiding. Het is echter ook mogelijk dat studenten die de aansluiting tussen de afgesloten opleiding en de vervolgoopleiding als negatief beoordelen, (achteraf) zeggen dat de studiekeuzebegeleiding onvoldoende is geweest.

Voor studenten die een VWO-opleiding achter de rug hebben, speelt het examencijfer voor wiskunde een rol bij het subjectieve oordeel over de aansluiting tussen de afgesloten VWO-opleiding en de vervolgstudie in het hoger onderwijs. Bij ieder punt stijging in het examencijfer neemt het oordeel over de aansluiting met 0,11 punten toe. Daarnaast zien we dat studievertraging een significant negatieve invloed heeft op de ervaren aansluiting. Zowel havisten als VWO'ers die studievertraging hebben opgelopen tijdens hun opleiding beoordelen de aansluiting tussen de afgesloten opleiding en de vervolgopleiding in het hoger onderwijs negatiever dan degenen die geen vertraging hebben opgelopen.

Daarnaast blijkt dat - ondanks correctie voor verschillen in de vakkenpakketkeuze en examencijfer - er bij HAVO-schoolverlaters sprake is van een substantieel sekse-effect: vrouwen geven vaker dan mannen een positief oordeel over de aansluiting tussen het HAVO en de vervolgopleiding in het HBO.

Een laatste conclusie uit tabel 5 heeft betrekking op de keuze van de vervolgopleiding van VWO'ers. VWO-schoolverlaters die zijn doorgestroomd naar het WO zijn minder positief over de aansluiting tussen de afgesloten opleiding en de vervolgopleiding dan zij die een opleiding in het HBO zijn gaan volgen. Dit houdt waarschijnlijk verband met de moeilijkheidsgraad van het WO in vergelijking met die van het HBO. Een universitaire opleiding is minder schools dan een HBO-opleiding en bovendien tilt men in het WO zwaarder aan individuele studievaardigheden dan in het HBO. Zowel het ontbreken van het schoolse karakter als de noodzaak tot zelfstudie op de universiteit kan leiden tot dit negatiever oordeel.

Voortijdige studie-uitval in het hoger onderwijs

Het hoger onderwijs wordt gekenmerkt door een hoge studie-uitval. Uit cijfers blijkt dat het numerieke rendement van het hoger onderwijs de afgelopen decennia laag, maar redelijk stabiel is (Bijleveld, 1990; OC&W, 1993). Gemiddeld genomen haalt zo'n 55 procent van de studenten die aan een bepaalde studie in het hoger onderwijs beginnen na verloop van tijd de eindstreep. Daarbij is de uitval in het WO hoger dan in het HBO. Dit betekent dat de resterende 45 procent ergens tijdens de opleiding afhaakt. Diverse factoren spelen daarbij een rol van betekenis. Zo wijst Prins (1997) onder meer op het belang van studieprestaties, motivatie en aspiraties, studiebegeleiding en studentgerichtheid. De vraag die in dit artikel natuurlijk van belang is, luidt of de kans op studie-uitval samenhangt met het vakkenpakket en de aansluiting van dit pakket op de gekozen vervolgopleiding. Het antwoord op deze vraag is ontkennend. De resultaten in tabel 6 tonen dat studenten die een vervolgopleiding kiezen die inhoudelijk aansluit bij het gevolgde vakkenpakket geen significant kleinere kans hebben om deze voortijdig te verlaten. Daarnaast moet worden geconcludeerd dat er tussen de verschillende vakkenpakketten geen verschil bestaat in de kans op voortijdige studie-uitval.

De examencijfers voor Nederlands en wiskunde hangen wel samen met voortijdige studie-uitval in het hoger onderwijs. Havisten die een hoog indexamencijfer hebben voor het vak Nederlands verlaten minder vaak de vervolgopleiding voortijdig dan zij die een laag cijfer voor het vak Nederlands hebben behaald. Voor VWO'ers geldt juist dat een hoog wiskunde-cijfer een goede voorspelling geeft van het latere studiesucces.

Studievertraging opgelopen tijdens de HAVO- of VWO-opleiding is ook van invloed op voortijdige uitval in het hoger onderwijs. Voor havisten met studievertraging is de kansverhouding van studie-uitval ruim twee zo groot als voor HAVO-schoolverlaters die binnen de nominale duur van de opleiding hun diploma hebben behaald. Binnen het VWO bedraagt het verschil in de overeenkomstige kansverhouding anderhalf. Ook wanneer rekening wordt gehouden met het effect van studievertraging is er een invloed van leeftijd op voortijdige uitval in het hoger onderwijs. Bij het HAVO is de kans op voortijdige studie-uitval groter naarmate men ouder is, een bevinding die ook in eerder onderzoek (Baneke, 1987) is gevonden. Deze bevinding moet vooral worden geïnterpreteerd als een aanwijzing voor (gebrek aan) schoolsucces tijdens de vroege schoolloopbaan. Personen die vóór hun HAVO-opleiding vertraging hebben opgelopen in de loopbaan (door zittenblijven in het basisonderwijs en/of een

omweg via het MAVO) staken hun vervolgstudie in het hoger onderwijs vaker dan studenten zonder dergelijke studievertraging. Ten slotte hebben VWO'ers die naar het WO zijn doorgestroomd een grotere kans om de vervolgstudie voortijdig te verlaten dan VWO'ers die naar het HBO zijn gegaan. De verklaring voor deze bevinding ligt vast en zeker in het feit dat een universitaire opleiding gemiddeld genomen moeilijker is dan een HBO-opleiding.

Tabel 6: Resultaten logistische regressie-analyse van voortijdige studie-uitval in het hoger onderwijs

	HAVO B	SE (B)	VWO B	SE (B)
Vervolgopleiding sluit aan bij vakkenpakket				
Nee (referentiegroep)	0		0	
Ja	0,30	0,35	0,12	0,29
Vakkenpakket				
Cultuur en Maatschappij (referentiegroep)	0		0	
Economie en Maatschappij	-0,14	0,37	-0,31	0,37
Natuur en Techniek	-0,32	0,43	-0,05	0,37
Natuur en Gezondheid	0,22	0,44	0,09	0,42
Aandacht studiekeuzebegeleiding				
Onvoldoende (referentiegroep)	0		0	
(Ruim) voldoende	-0,50	0,31	-0,09	0,31
Vervolgopleiding besproken				
Nee (referentiegroep)	0		0	
Ja	0,25	0,26	0,05	0,23
Examencijfer Nederlands	-0,34**	0,15	-0,10	0,15
Examencijfer wiskunde	0,07	0,12	-0,27**	0,13
Studievertraging in voortgezet onderwijs				
Nee (referentiegroep)	0		0	
Ja	0,81**	0,26	0,43*	0,25
Leeftijd				
17-18 jaar (referentiegroep)	0		0	
19-20 jaar	0,26	0,28	0,26	1,06
21+ jaar	1,36**	0,59	-0,02	1,15
Sekse				
Man (referentiegroep)	0		0	
Vrouw	0,10	0,28	-0,25	0,24
Etniciteit				
Autochtoon (referentiegroep)	0		0	
Allochtoon	-0,54	0,66	-1,02	1,04
Vervolgopleiding				
HBO (referentiegroep)			0	
WO			0,84**	0,26
Intercept	-0,54	1,13	-0,43	1,58
Model Chi ²	33,24**		22,80*	
df	13		14	
N	625		927	

* = significant op 10%-niveau; ** = significant op 5%-niveau

Bron: ROA, schoolverlatersonderzoek RUBS 1997

Achteraf spijt?

Tot slot van dit artikel wordt onderzocht of de studenten die op het moment van ondervraging de vervolgopleiding (nog) niet hebben verlaten, achteraf gezien, dezelfde opleiding opnieuw zouden kiezen. We zijn daarbij vooral geïnteresseerd in de vraag of studenten die een vervolgopleiding hebben gekozen die niet goed past bij het gevolgde vakkenpakket in het voortgezet onderwijs een kleinere kans hebben om deze vervolgopleiding opnieuw te kiezen dan studenten die een aansluitende studie in het hoger onderwijs zijn gaan volgen.

Tabel 7: Resultaten logistische regressie-analyse van het opnieuw kiezen van de vervolgopleiding

	HAVO B	SE (B)	VWO B	SE (B)
Vulgoopleiding sluit aan bij vakkenpakket				
Nee (referentiegroep)	0		0	
Ja	-0,52	0,59	-0,21	0,40
Vakkenpakket				
Cultuur en Maatschappij (referentiegroep)	0		0	
Economie en Maatschappij	-0,46	0,70	-0,42	0,66
Natuur en Techniek	-1,04	0,74	-0,74	0,66
Natuur en Gezondheid	-0,09	0,86	-0,46	0,76
Aandacht studiekeuzebegeleiding				
Onvoldoende (referentiegroep)	0		0	
(Ruim) voldoende	0,77*	0,46	0,55	0,41
Vulgoopleiding besproken				
Nee (referentiegroep)	0		0	
Ja	0,53	0,38	0,06	0,32
Examencijfer Nederlands	0,15	0,24	0,06	0,21
Examencijfer wiskunde	-0,19	0,20	-0,07	0,18
Studievertraging in voortgezet onderwijs				
Nee (referentiegroep)	0		0	
Ja	-0,21	0,40	0,00	0,36
Leeftijd				
17-18 jaar (referentiegroep)	0		0	
19-20 jaar	-1,00**	0,47	-4,33	10,25
21+ jaar	5,66	17,18	-3,85	10,27
Sekse				
Man (referentiegroep)	0		0	
Vrouw	-0,21	0,43	0,45	0,35
Etniciteit				
Autochtoon (referentiegroep)	0		0	
Allochtoon	-2,25**	0,57	-0,11	1,06
Vulgoopleiding				
HBO (referentiegroep)			0	
WO			0,75**	0,33
Intercept	4,07**	2,02	6,75	10,40
Model Chi ²	35,21**		13,84	
df	13		14	
N	523		816	

* = significant op 10%-niveau; ** = significant op 5%-niveau

Bron: ROA, schoolverlatersonderzoek RUBS 1997

Uit tabel 7 blijkt dat het voor het opnieuw kiezen van de vervolgopleiding niet uitmaakt of het vakkenpakket van de vooropleiding aansluit bij de vervolgstudie. Schoolverlaters van wie het vakkenpakket niet aansluit bij de vervolgopleiding zouden niet minder vaak opnieuw dezelfde vervolgopleiding volgen dan schoolverlaters die een aansluitende vervolgstudie hebben gekozen. Daarnaast is het vakkenpakket niet van invloed op het opnieuw kiezen van de vervolgopleiding in het hoger onderwijs: tussen de vier onderscheiden pakketten zijn geen significante verschillen gevonden.

HAVO-schoolverlaters bij wie de aandacht voor studiekeuzebegeleiding (ruim) voldoende is geweest, zouden vaker dezelfde vervolgopleiding opnieuw kiezen dan schoolverlaters van het HAVO die de aandacht voor studiekeuzebegeleiding tijdens de vooropleiding als onvoldoende beschouwen. Daarnaast geldt voor studenten met een HAVO-vooropleiding die op het moment van ondervraging 19 of 20 jaar oud zijn, dat zij - in vergelijking met individuen in andere leeftijdscategorieën - minder vaak opnieuw dezelfde opleiding in het hoger onderwijs zouden kiezen. Bovendien hebben allochtonen afkomstig van het HAVO vaker spijt van de keuze van de vervolgopleiding dan autochtonen.

Ten slotte toont tabel 7 dat de kans om opnieuw dezelfde opleiding te kiezen voor VWO'ers die na hun opleiding naar het WO zijn doorgestroomd groter is dan voor VWO'ers die hebben gekozen voor het HBO. Hieruit kan worden geconcludeerd dat studenten die de overgang van het VWO naar het HBO hebben gemaakt over het algemeen meer spijt hebben van hun keuze dan VWO'ers die voor het WO hebben gekozen, ondanks het feit dat eerstgenoemden - wat betreft hun oordeel over de aansluiting en hun kans op voortijdige studie-uitval - relatief meer succes hebben in de vervolgopleiding.

CONCLUSIES EN DISCUSSIE

In de discussie over de aansluiting tussen het voortgezet onderwijs (HAVO/VWO) en het hoger onderwijs (HBO/WO) wordt onder andere gewezen op het belang van een goede inhoudelijke voorbereiding, door de samenstelling van het vakkenpakket beter af te stemmen op de keuze van de vervolgopleiding. In dit artikel is nagegaan wat de effecten zijn van de vakkenpakketkeuze op een aantal indicatoren die een beeld geven van de aansluiting tussen het voortgezet en hoger onderwijs. Deze vraag is met name relevant in het licht van de onlangs ingevoerde studieprofielen. Hoewel een definitieve evaluatie van deze profielen nog niet goed mogelijk is, kan wel worden geprobeerd om een belangrijke vooronderstelling die ten grondslag ligt aan deze herstructurering te toetsen. Deze vooronderstelling heeft met name betrekking op de mate van samenhang in de vakkenpakketkeuze (of liever: het gebrek daaraan) en de gevolgen van een keuze voor een vervolgopleiding die niet aansluit bij het gekozen vakkenpakket. Hiertoe is een 'simulatie' uitgevoerd op basis van de pakketkeuzes van HAVO- en VWO-schoolverlaters uit de 'oude' situatie.

De analyse levert een aantal opmerkelijke uitkomsten op. In de eerste plaats komt naar voren dat de meeste pakketten redelijk eenduidig zijn in te delen. Bij het VWO blijkt 41 procent te worden ingedeeld bij Natuur en Techniek, 11 procent bij Natuur en Gezondheid, 30 procent bij Economie en Maatschappij en 12 procent bij Cultuur en Maatschappij. Slechts zes procent van de VWO-leerlingen heeft een pakket gevolgd dat niet eenduidig in één van de vier profielen valt onder te brengen. Bij het HAVO ligt deze situatie iets anders. Daar kan 26 procent van de leerlingen worden ingedeeld bij Natuur en Techniek, negen procent bij Natuur en Gezondheid, 29 procent bij Economie en Maatschappij en negen procent bij Cultuur en Maatschappij. Ten slotte heeft 28 procent een pakket gekozen dat als 'overig' kan worden geclassificeerd. Dit betekent dat er eigenlijk weinig reden is om meer stroomlijning in de pakketkeuze tot stand te brengen. Immers, het overgrote deel van de jongeren kiest reeds een pakket dat in de ogen van de overheid al voldoende samenhang heeft. Men moet zich afvragen of in de overige gevallen sturing wenselijk en zinvol is, want er kunnen goede redenen

zijn waarom leerlingen belangstelling hebben voor vakken die ogenschijnlijk ver uiteenlopen. Wanneer dit is gebaseerd op een sterke interesse voor de betreffende vakken, dan is het ongewenst om scholieren daar van af te houden.

Dit neemt niet weg dat het nog steeds verstandig kan zijn om bepaalde ingangseisen te stellen ten aanzien van het studieprofiel. Op die manier zou de aansluiting tussen het voortgezet en hoger onderwijs kunnen worden verbeterd. Wat zijn daar de mogelijke effecten van? Opnieuw zijn de resultaten opmerkelijk. In het algemeen blijken de schoolverlaters van het HAVO en het VWO vervolgoledingen te kiezen die goed aansluiten bij het door hen gekozen vakkenpakkettype. Van de studenten die instromen in een technische opleiding heeft vrijwel zonder uitzondering iedereen een vakkenpakket gekozen dat hier bij aansluit. Alleen bij niet-technische opleidingssectoren in het hoger onderwijs blijkt dat ongeveer een kwart van de studenten een niet-aansluitend vakkenpakket had in het voortgezet onderwijs. Dit laatste heeft echter geen negatieve invloed op de aansluiting met de vervolgoleding. Uit de analyse komt naar voren dat er geen negatief effect is van een niet-aansluitend vakkenpakket op de kans om de opleiding voortijdig te verlaten, noch blijkt er een negatief effect te zijn op het oordeel van de schoolverlaters over de aansluiting tussen het voortgezet onderwijs en de gekozen vervolgoleding in het hoger onderwijs. Ook blijkt er geen enkele samenhang te bestaan met eventuele spijt over de gekozen vervolgoleding. Deze bevindingen worden grotendeels bevestigd in ander recent onderzoek (Den Boer & Guldmond, 1996; Janssen, 1997; Prins, 1997).

We concluderen daarom dat de invoering van studieprofielen niet tot een wezenlijke verbetering van de aansluiting tussen het voortgezet en hoger onderwijs zal leiden. Er kan zelfs worden beargumenteerd dat de invoering van studieprofielen een aantal ongewenste neveneffecten zal hebben. Interessant in dit verband is de vraag wat er gaat gebeuren met de kwantitatieve instroom in de technische opleidingen. In de oude situatie koos ongeveer 40 procent van de VWO-leerlingen voor een vakkenpakket dat vergelijkbaar is met het profiel Natuur en Techniek. Bij het HAVO bedroeg dit ruim een kwart. Niet iedereen die een dergelijk profiel heeft gekozen, gaat vervolgens ook een technische opleiding volgen. Van degenen met een profiel Natuur en Techniek kiest 62 procent voor een echte technische of natuurwetenschappelijke opleiding, 18 procent voor een opleiding in de sector landbouw of gezondheidszorg, terwijl de overige 20 procent een niet-technische opleiding volgt (bijvoorbeeld economie). De feitelijke keuze voor een technische vervolgoleding wordt in belangrijke mate bepaald door de cijfers die men heeft behaald voor de exacte vakken. Degenen met hoge cijfers voor deze vakken gaan vaak een technische of exacte opleiding volgen, terwijl degenen met lage cijfers meestal voor een niet-technische vervolgoleding kiezen.

Men zou dit als volgt kunnen interpreteren. In de oude situatie kozen studenten, indien dat mogelijk was, voor een meer exact georiënteerd pakket om daarmee zo veel mogelijk opties voor vervolgoledingen open te houden. Pas wanneer men goede cijfers had voor de exacte vakken, dan volgde ook daadwerkelijk een keuze voor een exacte vervolgoleding.

Met de invoering van de studieprofielen zal dit waarschijnlijk veranderen. De keuze voor een profiel wordt dan immers meer bepaald door de keuze voor een bepaalde opleidingssector. Degenen die bijvoorbeeld een economische studie willen gaan studeren, zullen dan worden geadviseerd om het pakket Economie en Maatschappij te kiezen. In de oude situatie zou zo'n leerling wellicht toch een meer exact georiënteerd pakket hebben gekozen, waardoor de kans aanwezig was dat hij of zij alsnog voor een technische vervolgoleding zou kiezen. Met de invoering van de studieprofielen wordt het moment van studiekeuze feitelijk vervroegd, hetgeen impliceert dat leerlingen onnodig worden afgehouden van ervaringen met meer exacte vakken. Onnodig, omdat voor het succesvol afronden van een economische studie het profiel Economie en Maatschappij nu eenmaal niet echt noodzakelijk is, terwijl voor het behalen van een technische opleiding het profiel Natuur en Techniek wel een vereiste is. Op grond hiervan kan worden verwacht dat de keuze voor techniek zal afnemen ten gunste van andere profielen. Daarmee staat de invoering van de studieprofielen op gespannen voet met het streven van de overheid om de instroom in technische opleidingen te bevorderen.

NOTEN

1. Indien een leerling geen wiskunde (A of B) in zijn vakkenpakket heeft gehad, dan is het examencijfer voor een ander exact vak (natuurkunde, scheikunde of biologie) genomen.
2. Leeftijd kan ook worden beschouwd als een indicator voor opgelopen studievertraging (door zittenblijven of door omwegen in het onderwijssysteem) vóór de afgesloten HAVO- of VWO-opleiding. Het gaat overigens om de leeftijd op het moment van ondervraging, dat wil zeggen anderhalf jaar na het eindexamen.
3. Daarbij moet worden opgemerkt dat het niet voor alle vervolgopleidingen even gemakkelijk is om te bepalen welke vakkenpakketten al dan niet (voldoende) aansluiten. Met name bij niet-exacte vervolgopleidingen is discussie mogelijk. Zo is het goed te verdedigen dat het pakket Natuur en Techniek aansluit voor degenen die een HBO-lerarenopleiding in een exact vak volgen.
4. De kansverhouding kan worden berekend als $e^{0,41} = 1,51$.

LITERATUUR

- Baneke, J. (1987). *Studiesucces, persoonlijkheid en stress*. Amsterdam: dissertatie UvA.
- Boer, P. den, & Guldemond, H. (1996). *De effecten van vakkenpakketkeuze en omwegen in het voortgezet onderwijs op het studiesucces in het hoger onderwijs*. Groningen: GION.
- Bijleveld, R. (1990). Numeriek rendement en studieduur voor en na de invoering van de tweefasestructuur. *Tijdschrift voor Hoger Onderwijs*, 8, 67-79.
- Dam, G. ten, Eck, E. van, & Volman, M. (1992). *Onderwijs en sekse. Een verkenning van researchprogramma's*. Den Haag: DCE/STEO.
- DeMaris, A. (1992). *Logit modeling. Practical applications*. Sage University Papers series on Quantitative Applications in the Social Sciences, 07-086. Newbury Park, CA: Sage.
- Jansen, E. (1997). Invloed van het vakkenpakket VWO op het studieresultaat in de propedeuse WO. *Tijdschrift voor Onderwijsresearch*, 23, 238-248.
- Jong, U. de, & Jungbluth, M. (1989). De invloed van initiële schoolprestaties, sekse en sociaal milieu op de keuze van het vakkenpakket in het VO. In J. van Damme & J. Dronkers (red.), *Jongeren in school en beroep*. Amsterdam/Lisse: Swets & Zeitlinger.
- 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*. Den Haag: SDU.
- Ministerie van Onderwijs en Wetenschappen (1993). *Feiten en cijfers*. Den Haag: SDU.
- Prins, J. (1997). *Studie-uitval in het wetenschappelijk onderwijs. Studentkenmerken en opleidingskenmerken als verklaring voor studie-uitval*. Nijmegen: Nijmegen University Press.
- 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: SDU.
- Wijnen, W. (1997). Studiehuis: een bruikbaar concept voor het onderwijs. *Tijdschrift voor Hoger Onderwijs*, 11, 47-54.
- Willems, E., & Grip, A. de (1993). *Toekomst in techniek? Waarde-oriëntatie, beeldvorming en keuze-intenties van jongeren*. ROA-R-1993/7. Maastricht: ROA.

Bijlage 1: De relatie tussen de in dit artikel gebruikte indeling van vakkenpakketten en de officiële indeling van profielen

Type vakkenpakket volgens onze definitie:	Verplicht deel profiel volgens officiële definitie:
<p>Cultuur en Maatschappij</p> <ul style="list-style-type: none"> - minimaal drie vreemde talen, géén wiskunde B, géén natuurkunde; of: - geschiedenis, aardrijkskunde, géén economie 1, géén economie 2, géén wiskunde B, géén natuurkunde. 	<ul style="list-style-type: none"> - Frans of Engels (uitgebreid), geschiedenis, wiskunde A, economie 1, culturele en kunstzinnige vorming (uitgebreid) (voor HAVO); - Frans of Duits (uitgebreid) of klassieke taal, geschiedenis, culturele en kunstzinnige vorming (uitgebreid), een van de volgende drie (derde moderne taal, wiskunde A, klassieke taal) (voor VWO).
<p>Economie en Maatschappij</p> <ul style="list-style-type: none"> - economie 1 en/of economie 2, geschiedenis, aardrijkskunde, wiskunde A; of: - economie 1 en/of economie 2, wiskunde A of wiskunde B, géén natuurkunde, géén scheikunde; of: - economie 1 en/of economie 2, Frans óf Duits (niet beide), géén wiskunde B, géén natuurkunde. 	<ul style="list-style-type: none"> - economie 1, economie 2, geschiedenis, aardrijkskunde, wiskunde A.
<p>Natuur en Techniek</p> <ul style="list-style-type: none"> - wiskunde B, natuurkunde, scheikunde; of: - twee van de volgende drie (wiskunde B, natuurkunde, scheikunde), géén biologie; of: - wiskunde A, natuurkunde, scheikunde, géén biologie. 	<ul style="list-style-type: none"> - wiskunde B, natuurkunde, scheikunde.
<p>Natuur en Gezondheid</p> <ul style="list-style-type: none"> - twee van de volgende drie (wiskunde B, natuurkunde, scheikunde), biologie; of: - wiskunde A, natuurkunde, scheikunde, biologie, géén wiskunde B. 	<ul style="list-style-type: none"> - wiskunde B, natuurkunde, scheikunde, biologie.

Verlenging van de expositietijd van ondertitels op televisie: Effecten op leestijd en de herinneringen van kinderen*

C.M. Koolstra, T.H.A. van der Voort (RU Leiden, *Sectie Kind en Media*) en G. d'Ydewalle, (KU Leuven, *Laboratorium voor Experimentele Psychologie*)

(* Met dank aan Fré Meijster (Subtitling International Amsterdam) voor het ter beschikking stellen van de ondertitelapparatuur, Johan van Rensbergen (KU Leuven) voor de datareductie van de oogbewegingsgegevens, Marleen deVijver en stagestudenten (KU Leuven) voor hun hulp bij het verzamelen van de gegevens, en de leerlingen van de basisschool Sancta Maria te Leuven voor hun medewerking als proefpersonen aan het experiment.)

SAMENVATTING

In een tussen-proefpersonen onderzoek opzet werd aan kinderen uit groep 4, 6 en 8 van de basisschool ($N = 54$) een 9 minuten durend fragment uit een Nederlands ondertitelde aflevering van zowel een Amerikaanse actieserie (*Knightrider*) als een Amerikaanse comedyserie (*Full House*) onder één van de volgende drie condities vertoond: (a) de gebruikelijke expositietijd van maximaal 6 seconden per ondertitel; (b) een maximale expositietijd van 8 seconden per ondertitel; en (c) een maximale expositietijd van 10 seconden. Tijdens het bekijken van de fragmenten werd met behulp van oogbewegingsregistratieapparatuur bepaald hoeveel tijd kinderen besteedden aan het lezen van de ondertitels. Onmiddellijk na afloop werd via een herkenningstest bestaande uit meerkeuzevragen de herinnering van de ondertitels vastgesteld. De resultaten wezen uit dat kinderen in de conditie met de hoogste expositietijd (maximaal 10 seconden) significant meer tijd besteedden aan het lezen van de ondertitels dan kinderen waarbij de expositietijd maximaal 8 of 6 seconden was. In de conditie met de hoogste expositietijd werden de ondertitels ook beter herinnerd dan in de twee andere condities, maar dit verschil was niet significant.

1. INTRODUCTIE

Evenals in andere kleinere Europese landen zoals de Scandinavische landen, bestaat in Nederland en België het gebruik om buitenlandse films en televisieprogramma's niet na te synchroniseren maar te ondertitelen. Alleen buitenlandse televisieprogramma's die specifiek bedoeld zijn voor kleuters worden nagesynchroniseerd. Van alle buitenlandse programma's wordt in Nederland 94% ondertiteld (Luyken, Herbst, Langham-Brown, Reid, & Spinhof, 1991).

Bij het ondertitelen van televisieprogramma's worden twee conventies in acht genomen: (a) er worden maximaal twee regels gebruikt die elk bestaan uit maximaal 32 karakters (letters, leestekens en spaties); en (b) bij de vaststelling van de expositietijd van de ondertitels wordt de zogenaamde 6-seconden-regel gehanteerd (o.a., d'Ydewalle & Gielen, 1992; d'Ydewalle, Van Rensbergen, & Pollet, 1987). De 6-seconden-regel houdt in dat de langst mogelijke ondertitel - een tweeregelige ondertitel met 64 karakters - gedurende 6 seconden wordt vertoond, en dat bij kortere ondertitels de expositietijd procentueel afneemt met het aantal karakters. Hoewel de 6-seconden-regel algemeen wordt toegepast in landen waar men gewoon is buitenlandse televisieprogramma's te ondertitelen, is er weinig bekend over de vraag of

deze regel voor de expositietijd van ondertitels optimaal aansluit bij het leestempo van de kijkers (d'Ydewalle & Gielen, 1992).

Met name bij beginnende lezers is het de vraag of de expositietijd van ondertitels voldoende aansluit bij het leestempo. Een expositietijd die optimaal is toegesneden op het nog trage leestempo van jonge kinderen is onderwijskundig van belang, omdat er aanwijzingen zijn dat het lezen van ondertitels buiten schooltijd de leerkracht helpt om kinderen technisch lezen bij te brengen (Koolstra, Van der Voort, & Van der Kamp, 1997). Daarnaast is het met begrip lezen van ondertitels onderwijskundig van belang, omdat kinderen het ondertitelde programma dan beter begrijpen en er ook meer van kunnen leren. Vooral het met begrip kijken naar informatieve programma's kan ertoe leiden dat de televisie een bijdrage levert aan de algemene ontwikkeling van kinderen, hetgeen eveneens het leerproces op school kan ondersteunen (Van der Voort, 1997).

Een tweetal Nederlandse studies laat zien dat een deel van de jonge kijkers inderdaad moeite heeft het hoge tempo van ondertitels te volgen. Van Lil (1988) vroeg aan een representatieve steekproef van kinderen of zij meenden dat zij voldoende tijd hadden om de ondertitels van televisieprogramma's te lezen. Op het niveau van groep 3 zei niet minder dan 90% dat de ondertitels "vaak te vlug gaan", een percentage dat gedurende de basisschoolperiode geleidelijk afnam tot 10% in groep 8.

Peeters, Scherpenzeel, en Zantinge (1988) gingen langs experimentele weg na of het gebruik van ondertitels (in plaats van nasynchronisatie) het begrip van de film hindert. Aan kinderen uit groep 6-8 van de basisschool werden drie versies van een Engelstalige dramaserie voor de jeugd vertoond: (a) de originele Engelstalige versie zonder ondertitels; (b) een Nederlands ondertitelde versie; en (c) een in het Nederlands nagesynchroniseerde versie. De ondertitelde versie werd minder goed begrepen dan de nagesynchroniseerde versie, met name door de jongste kinderen uit het onderzoek (groep 6) en de zwakkere lezers.

In de hierboven genoemde studies werd getracht een beeld te krijgen van mogelijke problemen bij het lezen van ondertitels met behulp van zelfrapportage (Van Lil, 1988) en begripsmeting (Peeters et al., 1988). Een derde manier om eventuele problemen bij het lezen van ondertitels op het spoor te komen bestaat in het registreren van oogbewegingen tijdens het kijken naar ondertitelde televisieprogramma's. Via oogbewegingsregistratie kan exact worden nagegaan of en hoe lang de aandacht van de kijker op de ondertitels is gericht. Tot op heden zijn oogbewegingsregistratiestudies van het kijken naar ondertitelde televisieprogramma's vooral bij volwassenen uitgevoerd (d'Ydewalle & Gielen, 1992; d'Ydewalle et al., 1987; Gielen, 1988). Deze onderzoeken wezen uit dat het lezen van ondertitels bij de meeste volwassenen probleemloos en moeiteloos verloopt. Het lezen van ondertitels leek een grotendeels automatisch uitgelokt gedrag: of men nu wil of niet, men moet zijn aandacht op de ondertitels richten. Dit bleek uit 3 typen bevindingen: (a) direct na het in het beeld verschijnen van een ondertitel richt de blik van de kijker zich op de ondertitel (o.a., d'Ydewalle et al., 1987); (b) kijkers richten hun aandacht ook op ondertitels gesteld in een vreemde, onbekende taal (d'Ydewalle, Praet, Verfaillie, & Van Rensbergen, 1991); en (c) kijkers lezen de ondertitels zelfs bij programma's waarbij zowel geluid als ondertitels in de eigen taal beschikbaar zijn (d'Ydewalle et al., 1987).

De automatisch gelezen ondertitels blijven de volwassen kijker goed bij. In meerkeuzetests waarin de herkenning van ondertitels werd vastgesteld door iedere ondertitel vergezeld te laten gaan van drie sterk overeenkomstige niet-vertoonde formuleringen, werd maar liefst 97% van de vertoonde ondertitels herkend (Gielen, 1988). Uiteraard nam de tijd die volwassenen besteedden aan het lezen van ondertitels toe naarmate het aantal karakters in de ondertitel groter was (d'Ydewalle & Gielen, 1992; d'Ydewalle et al., 1987). Bij 2-regelige ondertitels was de leestijd dus hoger dan bij 1-regelige ondertitels. Merkwaardig genoeg was bij 2-regelige ondertitels behalve de absolute leestijd ook de procentuele leestijd, dit is het percentage van de tijd dat de ondertitel wordt gelezen terwijl deze in beeld is, hoger dan bij 1-regelige ondertitels. Dit betekent dat kijkers voor het lezen van 2-regelige ondertitels behal-

ve absoluut ook *relatief* meer tijd nodig hebben dan voor het lezen van 1-regelige ondertitels.

Tot op heden werd slechts in één studie (d'Ydewalle & Van Rensbergen, 1989) het leesgedrag van ondertitels bij kinderen met behulp van oogbewegingsregistratie onderzocht. In deze studie werd bij basisschoolkinderen uit de groepen 4, 6 en 8 nagegaan of hun leesgedrag vergelijkbaar was met dat van volwassenen. De auteurs concludeerden dat het leesgedrag van de kinderen uit groep 6 en 8 in sterke mate overeenkwam met dat van volwassenen. Evenals bij volwassenen nam bij deze kinderen de leestijd toe met het aantal karakters in de ondertitel; en ook werd er procentueel meer tijd besteed aan 2-regelige dan aan 1-regelige ondertitels. Het leesgedrag van de kinderen uit groep 4 week echter duidelijk af van dat van de andere kinderen. Onder andere werd bij kinderen uit groep 4 nauwelijks verschil gevonden tussen de procentuele leestijd besteed aan 1- en 2-regelige ondertitels. De auteurs concludeerden dan ook dat het ondertitel-leesgedrag bij kinderen uit groep 4 nog weinig systematisch is, waarschijnlijk omdat de gewoonte van het lezen van ondertitels nog onvoldoende is ontwikkeld. In hoeverre de ondertitels van de televisieprogramma's achteraf werden herkend werd in deze studie niet nagegaan.

Omdat tot op heden slechts in één studie het lezen van ondertitels door kinderen is onderzocht, wordt in de huidige studie opnieuw bij kinderen uit groep 4, 6 en 8 nagegaan hoe hun leesgedrag van ondertitels er uitziet. Gelet op het feit dat met name jongere basisschoolkinderen problemen kunnen hebben om de ondertitels bij te houden en in zich op te nemen (Peeters et al., 1988; Van Lil, 1988), is in de huidige studie tevens nagegaan of kinderen baat hebben bij een verlenging van de expositietijd van ondertitels. Hiertoe werd het leesgedrag van kinderen bij het kijken naar ondertitelde programma's geobserveerd onder drie condities: de gangbare expositieduur volgens de 6-seconden-regel en verlengde expositietijden volgens een 8- en 10-seconden-regel. Bij de verlengde expositietijden waren de langst mogelijke ondertitels - twee volle tekstregels van in totaal 64 karakters - gedurende respectievelijk 8 en 10 seconden in beeld, en nam voor kortere tekstregels de expositietijd weer procentueel af met het aantal karakters. Bij de versies waarbij de 8- en 10-seconden-regel is toegepast hadden kinderen dus respectievelijk 33% en 66% meer tijd voor het lezen van de ondertitels dan bij de gangbare 6-seconden-regel.

Met uitzondering van de studie Van Gielen (1988), waarin gebruik werd gemaakt van een detectiveserie, zijn in alle boven besproken oogbewegingsstudies tekenfilms gebruikt. In het huidige onderzoek is het leesgedrag van kinderen tijdens het kijken naar ondertitelde televisieprogramma's bestudeerd onder gebruikmaking van twee andersoortige ondertitelde televisieprogramma's, namelijk een aflevering uit een actieserie en een aflevering uit een comedyserie.

2. HYPOTHESEN

2.1 Absolute leestijd

Omdat een deel van de kinderen naar eigen zeggen onvoldoende tijd heeft om ondertitels te volgen die volgens de gangbare 6-seconden-regel worden gepresenteerd (Van Lil, 1988), verwachtten we dat naarmate de expositietijd hoger is, kinderen in toenemende mate de gelegenheid te baat zullen nemen om tijd te investeren in het lezen van de ondertitels:

H₁: Hoe hoger de expositietijd van de ondertitels, des te meer tijd zullen kinderen aan het lezen van ondertitels besteden.

Omdat in vergelijking met jongere kinderen, oudere kinderen naar eigen zeggen minder moeite hebben met het bijhouden van ondertitels van televisieprogramma's - in de groepen 4, 6 en 8 vindt respectievelijk 63, 33 en 10 procent dat ondertitels te snel gaan (Van Lil, 1988) - verwachtten we te vinden dat met klimmende leeftijd de leestijd zal dalen:

H₂: Hoe hoger het leerjaar waarin kinderen zich bevinden, des te minder tijd zullen zij aan het lezen van ondertitels besteden.

Tenslotte ligt het in de lijn der verwachting dat goede lezers gemiddeld minder leestijd nodig zullen hebben dan zwakke lezers:

H₃: Hoe groter de leesvaardigheid, des te minder tijd zullen kinderen aan het lezen van de ondertitels besteden.

2.2 Procentuele leestijd

De absolute leestijd geeft aan hoeveel seconden een kind gemiddeld per ondertitel besteedt aan lezen, ongeacht het aantal seconden dat de ondertitel in beeld is. De duur van de expositietijd is wel verdisconteerd in de procentuele leestijd die gemiddeld per ondertitel aangeeft welk percentage van de expositietijd besteed wordt aan het lezen van de ondertitels. In een onderzoek uitgevoerd bij universitaire studenten, vonden d'Ydewalle et al. (1987) dat een verhoging van de expositietijd van ondertitels gepaard ging met een *daling* van de procentuele leestijd. Vermoedelijk hadden de betrokken studenten reeds voldoende aan een expositietijd volgens de 6-seconden-regel, met het gevolg dat zij bij hogere expositietijden relatief minder tijd nodig hadden om de ondertitels te lezen. In het huidige onderzoek mag een meerderheid van de kinderen (uit groep 6 en 8) geacht worden ook zeer wel in staat te zijn om ondertitels vertoond volgens de gangbare 6-seconden-regel bij te houden (Van Lil, 1988). In de twee condities waarin de ondertitels langer in beeld staan, zullen deze kinderen, net als volwassenen, relatief minder leestijd gebruiken. Daarom werd verwacht dat:

H₄: Hoe hoger de expositietijd van de ondertitels, des te geringer is de procentuele leestijd. Eerder onderzoek (o.a., d'Ydewalle & Van Rensbergen, 1988; Gielen, 1988) liet zien dat er procentueel meer tijd wordt besteed aan het lezen van 2-regelige ondertitels dan aan het lezen van 1-regelige ondertitels. Een van de mogelijke verklaringen is dat 2-regelige ondertitels meer leestijd vergen omdat zij semantisch en syntactisch complexer zijn dan 1-regelige ondertitels (d'Ydewalle & Van Rensbergen, 1988). Ook in het huidige onderzoek luidde daarom de verwachting:

H₅: Kinderen zullen procentueel meer leestijd besteden aan 2-regelige dan aan 1-regelige ondertitels.

De twee televisieprogramma's die in het onderzoek werden gebruikt, bevatten een verschillend aantal dialogen. In het gebruikte actieprogramma werd relatief weinig gesproken, terwijl in het gebruikte comedy-programma relatief veel werd gesproken. Voor een goed begrip van het geziene lijkt het goed lezen van de ondertitels bij een actieprogramma minder noodzakelijk dan bij een comedy, omdat bij actieprogramma's in vergelijking met comedies het verhaal meer verteld wordt met beelden in plaats van woorden. Daarom verwachtten wij dat kinderen meer aandacht zullen besteden aan het lezen van ondertitels van het comedy-programma dan van het actieprogramma, hetgeen tot uitdrukking kan komen in een hogere procentuele leestijd voor het comedy-programma:

H₆: Kinderen zullen bij de comedy procentueel meer tijd besteden aan het lezen van ondertitels dan bij het actieprogramma.

Omdat oudere kinderen over een betere leesvaardigheid beschikken dan jongere kinderen, verwachtten we daarnaast te vinden dat de procentuele leestijd afneemt naarmate kinderen ouder zijn:

H₇: Hoe hoger het leerjaar waarin kinderen zich bevinden, des te lager zal de procentuele leestijd van ondertitels zijn.

Tenslotte ligt het in de lijn der verwachting dat betere lezers relatief minder tijd besteden aan het lezen van ondertitels dan zwakkere lezers:

H₈: Hoe groter de leesvaardigheid van kinderen, des te lager zal de procentuele leestijd van ondertitels zijn.

2.3 Herkenning

Indien de leestijd onder invloed van een verlenging van de expositietijd van de ondertitels inderdaad omhoog gaat (H₁), hebben kinderen meer tijd om de ondertitels in zich op te nemen,

met als mogelijk gevolg dat de ondertitels ook beter beklijven. We verwachtten daarom te vinden dat:

H₉: Hoe hoger de expositietijd, des te beter zullen kinderen in staat zijn de ondertitels te herkennen.

Omdat de cognitieve ontwikkeling bij oudere kinderen verder is gevorderd dan bij jonge kinderen, verwachtten we:

H₁₀: Hoe hoger het leerjaar waarin kinderen zich bevinden, des te beter zullen kinderen in staat zijn de ondertitels te herkennen.

Tenslotte verwachtten we dat er een positief verband zal zijn tussen leesvaardigheid en de herkenning van de ondertitels:

H₁₁: Hoe groter de leesvaardigheid, des te beter zullen kinderen in staat zijn de ondertitels te herkennen.

3. METHODE

3.1 Deelnemers en onderzoekopzet

Deelnemers waren leerlingen uit de groepen 4, 6 en 8 ($N = 54$; 23 jongens en 31 meisjes) van een Leuvense basisschool die gevestigd was in de buurt van het laboratorium waarin de oogbewegingsregistraties werden uitgevoerd (Laboratorium voor Experimentele Psychologie, KU Leuven). Op grond van een indeling naar lage, gemiddelde en hoge leesvaardigheid verschaft door de leerkracht werden de leerlingen per leerjaar gematched in drietallen van kinderen met een overeenkomstig leesniveau en vervolgens volgens het toeval over drie condities verdeeld waarin werd gekeken naar televisieprogramma's met: (a) een expositietijd van ondertitels volgens de 6-seconden-regel; (b) een expositietijd volgens de 8-seconden-regel; en (c) een expositietijd volgens de 10-seconden-regel. In elke conditie zaten dus 18 leerlingen, 6 kinderen per leerjaar. Om vast te stellen of de matching op leesvaardigheid voor de toewijzing van proefpersonen over condities geleid had tot groepen met een vergelijkbaar niveau van leesvaardigheid, werd achteraf met behulp van de Eén-Minut-Test (Brus & Voeten, 1973) nagegaan of er significante verschillen in technisch lezen tussen de kinderen toegewezen aan de drie condities bestonden. Dit bleek niet het geval te zijn, $F(2, 51) = 0.17$; $p = .84$.

3.2 Stimulusmateriaal

Het stimulusmateriaal bestond uit een fragment van een aflevering van de Amerikaanse actieserie *Knightrider* en een fragment uit een aflevering van de Amerikaanse comedyserie *Full House*. Beide fragmenten waren zodanig gekozen dat zij een op zichzelf staand verhaal vormden en even lang waren. De duur van elk van beide fragmenten was 9 minuten. Het fragment uit *Knightrider* telde 51 ondertitels, terwijl het fragment uit *Full House* er 79 telde. Speciaal ten behoeve van dit experiment werden bij een professioneel ondertitelingsbedrijf (Subtitling International te Amsterdam) van elk fragment 3 versies vervaardigd waarin de expositietijd van de ondertitels werd vastgesteld volgens de 6-, 8- en 10-seconden-regel. De teksten van de ondertitels waren in de drie verschillende versies identiek en in de meeste gevallen ook identiek aan de originele ondertitelteksten gebruikt in de afleveringen zoals die zijn uitgezonden. Bij een klein aantal ondertitels werd de originele tekst in elk van de drie versies enigszins ingekort om te voorkomen dat in de versie met de langste expositietijd (10-seconden-regel) de ondertitels elkaar gingen overlappen.

3.3 Oogbewegingsregistratie

De oogbewegingsregistraties werden uitgevoerd volgens de *pupilmidden-cornea-reflectie methode* (Van Rensbergen, Boutsen, & d'Ydewalle, 1993). Het kind zat op een afstand van 180 cm recht voor het televisiescherm op een stoel met nek- en hoofdsteun zodat het hoofd

zonder moeite stil kon worden gehouden. De oogbewegingsapparatuur (Debic 90) registreerde om de 20 milliseconden de positie op het televisiescherm waarnaar het kind keek. Met behulp van een datareductieprogramma werd voor iedere proefpersoon per vertoonde ondertitel het aantal metingen binnen het ondertitelgebied en het aantal metingen buiten het ondertitelgebied berekend. Het aantal metingen in het ondertitelgebied tijdens het vertonen van de ondertitel werd, na vermenigvuldiging met 20 (milliseconden), opgevat als de leestijd besteed aan de ondertitel. De betrouwbaarheid van de oogmetingen was hoog; Cronbachs alfa was voor *Knightrider* .97 en voor *Full House* .98. De procentuele leestijd werd vastgesteld door de tijd besteed aan het lezen van een ondertitel te delen door de expositietijd van de ondertitel. De betrouwbaarheid van de metingen van de procentuele leestijd was eveneens hoog; Cronbachs alfa was voor elk van beide televisieprogramma's, *Knightrider* en *Full House*, .98.

3.4 Herkenning ondertitels

Om de herkenning van ondertitels te meten werden per programma 15 ondertitels geselecteerd. Bij elke ondertitel werden, naar analogie van de herkenningstest gebruikt in een eerdere studie uitgevoerd bij volwassenen (Gielen, 1988), drie alternatieve formuleringen opgesteld die in hoge mate overeenkwamen met de originele (juiste) ondertitel. Bijvoorbeeld: Bij de ondertitel "Ik heb gescoord" werden als alternatieve formuleringen gekozen: "Ik heb een doelpunt gemaakt", "Ik heb de winnende goal gemaakt" en "Ik heb één keer gescoord." Proefpersonen werd gevraagd van de vier alternatieven de ondertitel aan te kruisen die in het televisieprogramma was gebruikt. De betrouwbaarheid (Cronbachs alfa) was .81 voor *Knightrider* en .85 voor *Full House*.

3.5 Leesvaardigheid

Om naast de grove indicatie van de leerkracht van de leesvaardigheid een meer precieze leesvaardigheidsscore van iedere deelnemer te krijgen werd in de periode van de laboratoriumafnames bij elk kind een individuele Eén-Minuut-Test (Brus & Voeten, 1973) op school afgenomen. Deze test bestaat uit een lijst van 116 ongerelateerde woorden die toenemen in moeilijkheidsgraad. De opdracht is om in één minuut tijd zoveel mogelijk woorden hardop voor te lezen. De score op de test wordt bepaald door per kind het totaal aantal juist opgelezen woorden op te tellen. De score geeft derhalve een maat van zowel de accuratesse als de snelheid waarmee woorden worden ontcijferd. Brus en Voeten (1973) rapporteerden parallel-test betrouwbaarheden van .96 (groep 4), .92 (groep 6), en .89 (groep 8).

3.6 Procedure

Kinderen namen individueel aan het experiment deel. Elk deelnemend kind werd onder schooltijd naar het laboratorium gebracht door een proefleider. Voorafgaand aan de eigenlijke oogbewegingsregistratie werd bij iedere deelnemer een calibratie van de apparatuur uitgevoerd (Van Rensbergen et al., 1993). Vervolgens kreeg het kind een videoband met de twee programmafragmenten te zien. Tijdens het kijken werden zijn of haar oogbewegingen geregistreerd. De volgorde van de twee programmafragmenten (*Full House* en *Knightrider*) werd systematisch gevarieerd. Na afloop van de registratie kreeg het kind de vragenlijst voorgelegd waarmee de herkenning van de ondertitels werd gemeten. Het experiment nam per deelnemer ongeveer 45 minuten in beslag.

4. RESULTATEN

4.1 Absolute leestijd

Om na te gaan of de absolute leestijd besteed aan de ondertitels (a) toenam met de expositietijd (H_1), en (b) afnam met de leeftijd van de kinderen (H_2) werd een variantieanalyse op de absolute leestijd uitgevoerd met als binnen-proefpersonen-factoren regellengte (1- versus

2-regelige ondertitels) en programmatype (actie- versus comedy-serie) en als tussen-proefpersonen-factoren expositietijd (6-, 8-, en 10-seconden-regel) en leerjaar (groepen 4, 6 en 8). Er was sprake van een significant hoofdeffect van expositietijd, $F(2, 45) = 3.23, p = .049$. Post-hoc-analyse met de Scheffé-test ($p < .05$) wees uit dat de tijd besteed aan het lezen van ondertitels in de 10-seconden-regel conditie ($M = 1.52$ seconden per ondertitel) significant hoger was dan in de 6- ($M = 1.13$ seconden) en de 8-seconden-regel-conditie ($M = 1.23$ seconden); de leestijd in de 8-seconden-regel conditie was wel hoger dan in de 6-seconden-regel conditie, maar dit verschil was niet significant (zie Tabel 1). Een tweede hoofdeffect betrof de regellengte, $F(1, 45) = 525.50, p < .001$. Dit effect hield simpelweg in dat er meer tijd werd besteed aan het lezen van 2-regelige ondertitels ($M = 1.84$ seconden per ondertitel) dan van 1-regelige ondertitels ($M = 0.77$ seconden). Er was geen significant hoofdeffect van programmatype, en, anders dan verwacht, was er ook geen hoofdeffect van leerjaar (H_2). Van alle mogelijke interactie-effecten was er één statistisch significant: regellengte met expositietijd, $F(2, 45) = 7.85, p = .001$. Deze interactie hield in dat het verschil tussen de tijd besteed aan het lezen van 1- en 2-regelige ondertitels het grootst was in de conditie met de langste expositietijd (de 10-seconden-regel).

Tabel 1: Gemiddeld aantal seconden besteed aan het lezen van 1- en 2-regelige ondertitels in drie condities

Regellengte	Expositietijd			M
	6-sec-regel	8-sec-regel	10-sec-regel	
1-regelig	0.69	0.75	0.88	0.77
2-regelig	1.60	1.74	2.20	1.84
M	1.13	1.23	1.52	1.29

Om na te gaan of ook het leesvaardigheidsniveau van de kinderen van invloed was op de absolute leestijd (H_3) werd een extra variantieanalyse uitgevoerd. Deze analyse was vergelijkbaar met de zojuist besproken variantieanalyse, met dit verschil dat in plaats van leerjaar het leesvaardigheidsniveau als tussenproefpersonen-factor werd opgenomen. Besloten werd om leesvaardigheid niet samen met leerjaar in één variantieanalyse als factor op te nemen, omdat deze twee variabelen sterk gecorreleerd zijn ($r = .73$). Ten behoeve van deze analyse werden de kinderen aan de hand van hun scores op de Eén-Minut-Test ingedeeld in een drietal groepen met respectievelijk een hoge, gemiddelde en lage technische leesvaardigheid. In deze analyse werden - uiteraard - opnieuw significante hoofdeffecten voor expositietijd en regellengte gevonden.

Tabel 2 vermeldt voor kinderen van verschillende leerjaren en een verschillend leesvaardigheidsniveau hoeveel leestijd zij gemiddeld per ondertitel besteedden bij de drie onderscheiden expositietijden. De laatste kolom van de tabel demonstreert dat kinderen, anders dan verwacht, niet met klimmend leerjaar of leesvaardigheidsniveau minder tijd uittrokken voor het lezen van ondertitels. Gemiddeld genomen besteedden kinderen uit de groepen 4, 6 en 8, evenals kinderen met een laag, midden- of hoog leesvaardigheidsniveau, ongeveer even veel tijd aan het lezen van de ondertitels, een resultaat dat in de twee hierboven besproken variantieanalyses tot uitdrukking kwam in het ontbreken van significante hoofdeffecten voor leerjaar en leesvaardigheid. Tabel 2 wijst echter uit dat de verwachte afname van de leestijd met klimmend leerjaar en leesvaardigheidsniveau zich wel voordeed bij de kinderen die geconfronteerd werden met ondertitels waarbij de 10-seconden-regel was toegepast. Daarentegen nam de leestijd met klimmend leerjaar en leesvaardigheidsniveau toe bij de kinderen die werden blootgesteld aan ondertitels waarbij de gebruikelijke 6-seconden-regel was toegepast.

Tabel 2: Gemiddeld aantal seconden besteed aan het lezen van ondertitels, uitgesplitst naar leerjaar en leesvaardigheidsniveau

	Expositietijd			M
	6-sec-regel	8-sec-regel	10-sec-regel	
Groep				
4	0.83	1.08	1.79	1.23
6	1.19	1.38	1.57	1.38
8	1.37	1.22	1.21	1.27
Leesvaardigheid				
Laag	0.84	1.19	1.89	1.29
Midden	1.19	1.28	1.45	1.31
Hoog	1.31	1.22	1.30	1.28

4.2 Procentuele leestijd

De verwachting was dat kinderen procentueel meer tijd aan het lezen van ondertitels zouden besteden indien: (a) de ondertitels langer in beeld bleven (H_4), (b) de ondertitel uit twee regels in plaats van één regel bestond (H_5), (c) er naar een comedy in plaats van een actiefilm werd gekeken (H_6), en (d) kinderen uit een lager leerjaar afkomstig waren (H_7). Om deze hypothesen te toetsen werd een variantieanalyse op de procentuele leestijd uitgevoerd met als tussen-proefpersonen-factoren expositietijd (6-, 8-, en 10-seconden-regel) en leerjaar (groep 4, 6 en 8), en als binnen-proefpersonen-factoren regellengte (1- versus 2-regelige ondertitels) en programmatype (comedy vs. actieprogramma). Tabel 3 vermeldt de gemiddelde procentuele leestijd voor 1- en 2-regelige ondertitels in elk van de drie experimentele condities. Het hoofdeffect van expositietijd was niet statistisch significant. Conform de verwachting was er echter wel een significant hoofdeffect van regellengte $F(1, 45) = 95.33, p < .001$, inhoudend dat er procentueel meer tijd werd besteed aan het lezen van 2-regelige ($M = 41\%$) dan aan 1-regelige ondertitels ($M = 31\%$). Daarnaast was er sprake van een statistisch significant hoofdeffect van programmatype $F(1, 45) = 9.91, p = .003$. Conform de verwachting werd er procentueel meer tijd besteed aan het lezen van de ondertitels van de comedy ($M = 37\%$) dan aan het lezen van de ondertitels van het actieprogramma ($M = 33\%$). Het hoofdeffect van leerjaar was niet significant. Er werden ook geen significante interactie-effecten aangetroffen.

Tabel 3: Percentage van de expositietijd besteed aan het lezen van 1- en 2-regelige ondertitels in drie condities

Regellengte	Expositietijd			M
	6-sec-regel	8-sec-regel	10-sec-regel	
1-regelig	36	30	28	31
2-regelig	46	38	38	41
M	41	33	33	36

Om na te gaan of de procentuele leestijd samenhang met leesvaardigheid (H_8) werd een variantieanalyse uitgevoerd die vergelijkbaar is met de laatst beschreven analyse, waarbij echter in

plaats van leerjaar het leesvaardigheidsniveau als tussen-proefpersonen-factor werd opgenomen. De uitkomsten van deze analyse kwamen sterk overeen met die van de laatstgenoemde analyse. De hoofdeffecten van regellengte en programmatype waren wederom statistisch significant. Er was geen sprake van een hoofdeffect van leesvaardigheid.

Hoewel anders dan verwacht de procentuele leestijd met klimmend leerjaar en leesvaardigheidsniveau niet afnam, toont Tabel 4 dat een dergelijke afname zich wel voordeed bij kinderen die geconfronteerd werden met de 10-seconden-regel. Bij kinderen die geconfronteerd werden met filmfragmenten waarin de gebruikelijke 6-seconden-regel was toegepast vertoonde de procentuele leestijd met klimmend leerjaar en leesvaardigheidsniveau juist een toename. Dit patroon komt sterk overeen met het eerder besproken patroon aangetroffen in de data met betrekking tot de absolute leestijd (zie Tabel 2).

Tabel 4: Percentage van de expositietijd besteed aan het lezen van ondertitels, uitgesplitst naar leerjaar en leesvaardigheidsniveau

	Expositietijd			M
	6-sec-regel	8-sec-regel	10-sec-regel	
Groep				
4	30	29	39	33
6	44	38	34	38
8	49	33	25	36
Leesvaardigheid				
Laag	30	33	41	35
Midden	44	35	31	37
Hoog	47	33	28	36

4.3 Herkenning

Om na te gaan of de herkenning van ondertitels beter was naarmate zij langer werden vertoond (H_9), en naarmate kinderen in een hoger leerjaar zaten (H_{10}), werd een variantieanalyse uitgevoerd op het percentage herkende ondertitels met als tussen-proefpersonen-factoren expositietijd (6-, 8- en 10-seconden-regel) en leerjaar (groepen 4, 6 en 8). Het percentage herkende ondertitels was het hoogst in de conditie waarin de ondertitels het langst werden vertoond (zie Tabel 5), maar het hoofdeffect van expositietijd was niet statistisch significant, $F(2, 45) = 1.30, p = .283$. Wel was er sprake van een significant hoofdeffect van leerjaar, $F(2, 45) = 14.60, p < .001$. Post-hoc-analyse met de Scheffé-test ($p < .05$) wees uit dat kinderen uit groep 8 meer ondertitels herkenden (85%) dan kinderen uit groep 6 (72%), die op hun beurt meer ondertitels herkenden dan kinderen uit groep 4 (51%). Er werden geen statistisch significante interactie-effecten aangetroffen.

In een aanvullende variantieanalyse werd nagegaan of het percentage herkende ondertitels toenam met leesvaardigheid (H_{11}). In deze analyse werd naast de expositietijd het leesvaardigheidsniveau als tussen-proefpersonen-factor betrokken. De herkenning van ondertitels bleek inderdaad afhankelijk te zijn van leesvaardigheid, $F(2, 45) = 28.46, p < .001$. Kinderen met een lage, gemiddelde en hoge leesvaardigheid herkenden respectievelijk 45%, 73% en 87% van de vertoonde ondertitels. Post-hoc-analyse met de Scheffé-test ($p < .05$) wees uit dat alle onderlinge verschillen tussen de drie leesvaardigheidsniveau-groepen statistisch significant waren.

Tabel 5: Percentage herkende ondertitels, uitgesplitst naar leerjaar en leesvaardigheidsniveau in drie condities

	Expositietijd			M
	6-sec-regel	8-sec-regel	10-sec-regel	
Groep				
4	51	48	53	51
6	71	62	82	72
8	83	84	89	85
Leesvaardigheid				
Laag	51	38	51	45
Midden	66	79	77	73
Hoog	86	84	92	87
M	68	65	75	69

5. DISCUSSIE

De studie liet zien dat een verhoging van de expositietijd van ondertitels door basisschoolkinderen te baat werd genomen om de tijd besteed aan het lezen te verhogen. Er was echter geen significant verschil in leestijd tussen de conditie waarin de gangbare 6-seconden-regel was toegepast en de conditie waarin de 8-seconden-regel werd gehanteerd. Pas in de conditie waarin de 10-seconden-regel werd gevolgd, was er bij een deel van de kinderen sprake van een significante verhoging van de leestijd, namelijk bij kinderen uit groep 4 en 6, en bij kinderen met een lage en gemiddelde leesvaardigheid. Dat juist deze kinderen bij een hogere expositietijd meer tijd besteedden aan het lezen van ondertitels is begrijpelijk, omdat dat deze kinderen de meeste moeite hebben met het lezen van ondertitels (Van Lil, 1988). Kinderen uit groep 8 en goede lezers gebruikten bij elk van de drie expositietijden ongeveer even veel leestijd. Deze groep kinderen had waarschijnlijk geen behoefte aan een verlenging van de expositietijd, omdat zij reeds met een expositietijd volgens de gangbare 6-seconden-regel goed uit de voeten konden.

Merkwaardig genoeg staken de jongste kinderen en de zwakste lezers *minder* tijd in het lezen van ondertitels vertoond volgens de 6-seconden-regel dan oudere kinderen en de betere lezers. Kinderen uit groep 4 en de zwakke lezers besteedden gemiddeld slechts 30% van de tijd dat de ondertitels in beeld stonden aan lezen, terwijl kinderen uit groep 8 en gevorderde lezers respectievelijk 49% en 47% van de expositietijd aan het lezen van de ondertitels besteedden. Dit is opmerkelijk omdat de jonge en de zwakke lezers gelet op hun nog beperkt ontwikkelde leesvaardigheid er verstandig aan zouden doen om bij een korte expositietijd juist *meer* tijd voor het lezen van de ondertitels uit te trekken. Een mogelijke verklaring is dat deze kinderen door de korte expositietijd ontmoedigd raakten, en daarom nauwelijks moeite deden om de vertoonde ondertitels te volgen.

Pas bij een expositietijd volgens de 10-seconden-regel bleken de jongste en de zwakste lezers meer tijd uit te trekken voor het lezen dan oudere kinderen en de beste lezers. Omdat de jongste en de zwakste lezers eerst bij de 10-seconden-regel hun leestijd adequaat aanpassen aan hun beperkte leesvaardigheid, lijken deze kinderen het meest gebaat te zijn bij een verlenging van de expositietijd van de ondertitels. De extra tijd die de jongste en de zwakste lezers bij de hoogste expositietijd aan het lezen van ondertitels besteedden, kwam echter niet

tot uitdrukking in een verbeterde herkenning van de ondertitels. Een mogelijke verklaring is dat de jongste en de zwakste lezers nog in onvoldoende mate beschikken over de minimale vaardigheden die überhaupt nodig zijn om ondertitels met begrip tot zich te kunnen nemen. Mogelijk was de opgave van het decoderen, begrijpen en onthouden van ondertitels die op het televisiescherm worden vertoond voor deze groep kinderen zodanig hoog gegrepen dat zelfs een verlenging van de gangbare expositieduur van ondertitels met maar liefst 66% onvoldoende soelaas bood.

Hoewel er in het onderzoek geen duidelijke samenhang tussen expositietijd en de herkenning van ondertitels werd aangetroffen, bleek de herkenning van ondertitels wel sterk afhankelijk te zijn van de leeftijd en de leesvaardigheid van de kinderen. Conform onze verwachting, nam het percentage herkende ondertitels met klimmend leerjaar (H_{10}) en met een hogere leesvaardigheid (H_{11}) toe. Terwijl kinderen uit groep 4 en de zwakke lezers slechts ongeveer de helft van de ondertitels wisten te herkennen (bij een gokkans van 25%), waren de kinderen uit groep 8 en de betere lezers in staat om ongeveer 85% van de ondertitels te herkennen. De ondertitels werden het best herkend door de goede lezers uit groep 8. Wanneer de 10-seconden-regel werd gehanteerd, wisten de meest leesvaardige kinderen uit groep 8 niet minder dan 92% van de ondertitels te herkennen, een percentage dat het bij universitaire studenten (Gielen, 1988) gevonden herkenningpercentage van 97% dicht benadert.

Evenals in voorgaand onderzoek, bleken 2-regelige ondertitels procentueel meer leestijd te vergen dan 1-regelige ondertitels (H_5). Zoals eerder besproken, vergen 2-regelige ondertitels mogelijk relatief meer leestijd omdat zij syntactisch en semantisch complexer zijn dan 1-regelige ondertitels. Volgens Praet, Verfaillie, De Graef, Van Rensbergen, en d'Ydewalle (1988) is een tweede mogelijke verklaring dat de aanwezigheid van twee ondertitels de aandacht verstoort, omdat bij het lezen van de bovenste regel de aandacht wordt afgeleid door de onderste regel, en de bovenste regel de aandacht afleidt bij het lezen van de onderste regel. Een derde mogelijke verklaring zoeken Praet et al. in de relatie tussen de informatie vervat in beeld en ondertitel. De auteurs vermoeden dat 1-regelige ondertitels slechts betrekkelijk weinig informatie aan het beeld toevoegen, omdat korte ondertitels niet zelden uitroepen, commando's en dergelijke weergeven, waarvan de inhoud ook al opgemaakt kan worden uit de informatie die vervat is in beeld en/of geluid. Wanneer 1-regelige ondertitels inderdaad weinig informatie toevoegen aan de beeldinformatie, is het voor een goed begrip van het vertoonde minder nodig dat de kijker zich inspant om deze ondertitels tot zich te nemen. Hoewel in enkele eerdere studies (d'Ydewalle & Van Rensbergen, 1989; d'Ydewalle, Warlop, & Van Rensbergen, 1987; Praet et al., 1988) gepoogd is de validiteit van de drie genoemde verklaringen vast te stellen, heeft dit tot op heden geen duidelijke aanwijzingen opgeleverd welke verklaring het meest valide is.

Zoals verwacht besteedden de kinderen bij het fragment uit het comedy-programma procentueel meer tijd aan het lezen van de ondertitels dan bij het fragment uit het gebruikte actieprogramma (H_7). Onze hypothese beruiste op de gedachte dat in comedy-programma's meer dialoog en dus ook meer ondertitels voorkomen dan in actieprogramma's, met als gevolg dat een goed begrip van het geziene bij een comedy-programma sterker afhangt van het kennis nemen van de ondertitels dan bij een actieprogramma. Een mogelijke aanvullende verklaring is dat de relatie tussen beeld en tekst in comedies verschilt van die in actieprogramma's. Vermoedelijk heeft de dialoog in comedy-programma's relatief vaak betrekking op zaken die niet in beeld zijn gebracht, en bestaat er in actieprogramma's vaker een directe relatie tussen wat te zien is en wat gezegd wordt. Omdat de ondertitels bij actieprogramma's relatief weinig aan de informatie in het beeld toevoegen, lijkt het voor een goed begrip van de film van minder belang dat kinderen de inhoud van de ondertitels tot zich nemen.

Concluderend kan worden vastgesteld dat een verlenging van de expositietijd van ondertitels voorsnog geen bevredigende oplossing biedt voor de problemen die jonge kinderen bij het lezen van de ondertitels van televisieprogramma's ondervinden. Het nasynchroniseren van buitenlandse televisieprogramma's kan wel een oplossing zijn, en deze methode wordt

reeds toegepast bij programma's bestemd voor kleuters. De praktijk om programma's na te synchroniseren zou kunnen worden uitgebreid naar televisieprogramma's die bedoeld zijn voor jongere basisschoolkinderen, met als nuttig effect dat zij minder moeite hebben het gezienne te begrijpen (Peeters et al., 1988) en vermoedelijk ook meer plezier aan de programma's beleven.

Wij pleiten echter geenszins voor het nasynchroniseren van alle buitenlandse kinderprogramma's, omdat het gebruik van ondertitels naast nadelen ook een aantal onmiskenbare voordelen heeft. Ten eerste bieden ondertitels beginnende lezers volop oefenstof om, met vallen en opstaan, zich te bekwamen in het decoderen van woorden die gedurende slechts korte tijd op het televisiescherm worden vertoond. Er zijn aanwijzingen dat het regelmatig kijken naar ondertitelde televisieprogramma's inderdaad de ontwikkeling van het technisch lezen van kinderen kan bevorderen (Koolstra, Van der Voort, & Van der Kamp, 1997). Ten tweede kunnen ondertitelde televisieprogramma's de kennis van de gebezigde vreemde taal vergroten. Zowel volwassenen (Danan, 1992; Pavakanun & d'Ydewalle, 1995) als kinderen (Koolstra & Beentjes, 1999) leren tijdens het kijken naar een ondertiteld buitenlands programma de betekenis van een deel van de woorden uit de vreemde taal, omdat zij al luisterend naar de dialoog de vertalingen ervan in de moedertaal kunnen meelesen. Een derde voordeel van ondertiteling is dat bij buitenlandse films de originele soundtrack te horen is, hetgeen de authenticiteit van de film verhoogt. Tenslotte is een voordeel dat ondertitelde programma's ook goed kunnen worden gevolgd door doven en slechthorenden (o.a., Kilborn, 1993; Verfaillie & d'Ydewalle, 1987).

LITERATUUR

- Brus, B. Th., & Voeten, M. J. M. (1973). *Eén-Minuut-Test: Verantwoording en handleiding*. Nijmegen: Berkhout.
- Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, 42, 497-527.
- d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415-427). New York: Springer-Verlag.
- d'Ydewalle, G., & Pavakanun, U. (1995). Acquisition of a second/foreign language by viewing a television program. In P. Winterhoff-Spurk (Ed.), *Psychology of media in Europe: The state of the art-perspectives for the future* (pp. 51-64). Opladen: Westdeutscher Verlag.
- d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research*, 18, 650-666.
- d'Ydewalle, G., & Van Rensbergen, J. (1989). Developmental studies of text-picture interactions in the perception of animated cartoons with text. In H. Mandl & J.R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 233-248). Amsterdam, The Netherlands: North-Holland.
- d'Ydewalle, G., Van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J.K. O'Regan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 313-321). Amsterdam, The Netherlands: North-Holland.
- d'Ydewalle, G., Warlop, L., & Van Rensbergen, J. (1987). Verschillen tussen jonge en oudere volwassenen in de verdeling van de aandacht over verschillende informatiebronnen. *Communicatie*, 17, 58-73.
- Gielen, M. (1988). *Perceptie en ondertitels: De paravofeale en perifere informatieverwerking van ondertitels*. Ongepubliceerde licentiaatsverhandeling, Katholieke Universiteit Leuven: Laboratorium voor Experimentele Psychologie.
- Kilborn, R. (1993). 'Speak my language': Current attitudes to television subtitling and dubbing. *Media, Culture and Society*, 15, 641-660.
- Koolstra, C.M., & Beentjes, J.W.J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled TV programs at home. *Educational Technology Research and Development*, 47, 51-60.
- Koolstra, C.M., Van der Voort, T.H.A., & Van der Kamp, L.J.Th. (1997). Television's impact on children's reading comprehension and decoding skills: A three-year panel study. *Reading Research Quarterly*, 32, 128-152.

- Luyken, G., Herbst, T., Langham-Brown, J., Reid, H., & Spinhof, H. (1991). *Overcoming language barriers in television: Dubbing and subtitling for the European audience*. Manchester, England: European Institute for the Media.
- Pavakanun, U., & d'Ydewalle, G. (1991). Watching foreign television programs and language learning. In F.L. Engel, D.G. Bouwhuis, T. Bösser & G. d'Ydewalle (Eds.), *Cognitive modelling and interactive environments in language learning* (pp. 193-198). Berlin: Springer Verlag.
- Peeters, A.L., Scherpenzeel, A.C., & Zantinge, J.H. (1988). *Ondertiteling of nasynchronisatie van kinderprogramma's* (Rapport R88-425). Hilversum: NOS/Kijk- en luisteronderzoek.
- Praet, C., Verfaillie, K., De Graef, P., Van Rensbergen, J., & d'Ydewalle, G. (1988). A one line text is not half a two line text. In R. Groner, G. d'Ydewalle, & R. Parham (Eds.), *From eye to mind: Information acquisition in perception, search and reading* (pp. 205-213). Amsterdam, The Netherlands: North-Holland.
- Van der Voort, T.H.A. (1997). The effects of television on children: Parental perceptions and mediation. *Trends in Communication*, 2, 5-29.
- Van Lil, J.E. (1988). *Beeld van een kind* (Rapport R88-421). Hilversum: NOS/Kijk- en Luisteronderzoek.
- Van Rensbergen, J., Boutsen, L., & d'Ydewalle, G. (1993). *Gebruikershandleiding voor het Leuvense oogbewegingsregistratiesysteem Debic 90* (Psych. Rep. No. 150). Leuven, Katholieke Universiteit Leuven: Laboratorium voor Experimentele Psychologie.
- Verfaillie, K., & d'Ydewalle, G. (1987). *Modality preference and message comprehension in deaf youngsters watching TV* (Psych. Rep. No. 70). Leuven, Katholieke Universiteit Leuven: Laboratorium voor Experimentele Psychologie.

Correspondentie adres:

C.M. Koolstra,

Rijksuniversiteit Leiden, Sectie Kind & Media, Postbus 9555, 2300 RB Leiden,
tel: 071-527 4078, fax: 071-527 3945, email: koolstra@rulfsw.leidenuniv.nl

De kwaliteit van leerlingenoordelen over leraarsgedrag¹

R. de Jong en K.J. Westerhof

GION, Gronings Instituut voor Onderzoek van het Onderwijs, Opvoeding en Ontwikkeling, Rijksuniversiteit Groningen

ABSTRACT

In a research project into the effectiveness of maths' teaching in the first year of secondary education teacher behaviour is rated by external observers and students. The reliability and validity of both methods is established. The results show that teacher behaviour is well assessed when students' opinions are aggregated on classroom level. The quality of aggregated students' opinions is as good as the quality of data of external observers. The predictive validity of aggregated students' opinions is higher than the predictive validity of external observations when subject motivation is taken as dependent variable.

INLEIDING

In onderzoek van het onderwijzen is men geïnteresseerd in de kwaliteit van het onderwijzen. De kwaliteit van het onderwijzen hangt af van factoren die zich op verschillende onderwijsniveaus (leerling, klas, school, context) bevinden. In dit artikel zijn we geïnteresseerd in het niveau van de klas, met name de kwaliteit van de instructie zoals die wordt gerealiseerd door de leraar. Bij de kwaliteit van de instructie gaat het om gedragingen van de leerkracht tijdens het onderwijsleerproces. Deze worden in verband gebracht met leerprestaties. Een eerste onderscheid met betrekking tot de kwaliteit van de instructie is of leraren leertaakgericht zijn of niet. Vervolgens is het de vraag hoe leertaakgerichte leraren hun tijd verdelen over gedragingen als doceren, het stellen van vragen en het geven van feedback. Bijvoorbeeld: er zijn leerkrachten die veel doceren, en er zijn leraren die het doceren afwisselen met het stellen van vragen. Uit overzichten van onderzoek naar het onderwijzen blijkt dat leertaakgerichtheid positief correleert met leerprestaties en dat de tweede, meer interactieve, leerstijl effectiever is dan uitsluitend doceren (Walberg, 1986, Brophy & Good, 1986, Rosenshine & Stevens, 1986, Westerhof, 1992, Creemers, 1994, Scheerens & Bosker, 1997), waarschijnlijk omdat het de leerlingen beter bij de les houdt. Echter niet alleen de frequentie van bepaalde gedragingen is van invloed op de leerprestaties. Ook de manier waarop het gedrag wordt uitgevoerd, de aard van het gedrag, is van invloed. Zo kan de ene leraar tijdens het doceren een duidelijk betoog houden, terwijl een ander onduidelijk is door van de hak op de tak te springen. Ook blijkt uit deze overzichten dat de helderheid waarmee de leerkracht de leerstof presenteert, positief correleert met leerwinst. Interactie tussen frequentie (kwantiteit) en aard (soort, kwaliteit) van een gedraging bepaalt derhalve de kwaliteit van de instructie.

De kwaliteit van de instructie, frequentie en aard, wordt in onderzoek vaak gemeten met behulp van onafhankelijke beoordelaars. Per onderzoek maakt men gebruik van meerdere beoordelaars. Deze moeten worden getraind in het hanteren van een observatie-instrument totdat een redelijke overeenstemming in de beoordeling wordt bereikt. Vanwege de repre-

sentativiteit moeten van een leerkracht meerdere lessen worden geobserveerd. Dit maakt het onderzoek met behulp van externe beoordelaars duur, zeker als de scholen landelijk verspreid zijn. De betrouwbaarheid van de observaties met behulp van externe beoordelaars is niet onproblematisch. Meestal wordt verslag gedaan van de interne consistentie van een beoordelingsinstrument. Men is tevreden als na veldwerk hoge alphas worden gerapporteerd. Hierbij gaat men voorbij aan het feit dat niet alleen het te observeren leraarsgedrag een variantiebron is. Ondanks de training blijven ook de beoordelaars een bron van variantie. Jansen en Westerhof (1998) hebben met een door Heuvelmans en Sanders (1993) voorgestelde methode aan de hand van de trainingresultaten in het School- en Instructie Effectiviteitsproject berekend dat er zes observaties door willekeurig gekozen beoordelaars per leraar nodig zijn om tot een redelijke betrouwbaarheid van observaties te komen. Redenen genoeg om na te gaan of de kwaliteit van de instructie ook op een andere manier betrouwbaar en valide is vast te stellen. Een alternatief is gebruik te maken van de oordelen van leerlingen die door dezelfde leraar worden onderwezen. Leerlingenoordelen hebben in ieder geval het voordeel dat de beoordelaars een representatief beeld hebben van het leraarsgedrag. De gedachte leerlingen te betrekken bij het beoordelen van de kwaliteit van het onderwijs is niet nieuw. In Nederland werd hierover eerder gerapporteerd bijvoorbeeld door van de Grift & Mulder (1986), Wubbels, Creton, Brekelmans & Hooymeyer (1987), Van de Grift, Houtveen & Vermeulen (1997) Houtveen, Vermeulen & Van de Grift (1993).

Binnen het universitair onderwijs wordt regelmatig gebruik gemaakt van oordelen van studenten over hun docenten. Deze oordelen worden gebruikt als hulpmiddel om het instructiegedrag van de docenten te verbeteren. In dit artikel zijn we geïnteresseerd in het meten van de kwaliteit van het leraarsgedrag door middel van leerlingenoordelen. Niet primair om leraarsgedrag te verbeteren, maar om de effectiviteit ervan vast te stellen. Leerlingen in het eerste leerjaar van het voortgezet onderwijs, de doelgroep in dit onderzoek, zijn een stuk jonger dan studenten. Het is de vraag of deze jongere leerlingen in staat zijn leraarsgedrag betrouwbaar en valide vast te stellen. In dit artikel gaan we na welke problemen zich daarbij voordoen. Daartoe zullen oordelen van leerlingen worden vergeleken met die van externe beoordelaars.

THEORETISCH KADER EN VRAAGSTELLING

Het beoordelen van leraarsgedrag door leerlingen heeft volgens Fraser en Walberg (1981, in Fraser, 1991) verschillende voordelen. In de eerste plaats zijn schriftelijke perceptiematen van leerlingen veel doelmatiger dan observatietechnieken waar onafhankelijke waarnemers voor moeten worden ingehuurd en getraind. In de tweede plaats zijn percepties van leerlingen gebaseerd op ervaringen opgedaan tijdens vele lessen. Daardoor hebben leerlingen veel meer ervaring omtrent het gedrag van de leraar in verschillende omstandigheden. Observatiegegevens hebben meestal betrekking op een beperkt aantal lessen. Een externe beoordelaar heeft daardoor geen idee van de representativiteit van het geobserveerde gedrag. De aanwezigheid van een externe beoordelaar kan bovendien het gedrag van de leraar beïnvloeden. In de derde plaats vormen percepties de gezamenlijke oordelen van alle leerlingen in een klas, terwijl observatietechnieken slechts het oordeel van één waarnemer betreffen. In de vierde plaats vormen leerlingpercepties de determinanten van leerlingengedrag en kunnen daarom belangrijker zijn dan door externe beoordelaars geobserveerd gedrag. Tenslotte verklaren perceptiemetingen van de klassenomgeving gewoonlijk meer variantie in leerprestaties dan registraties van gedrag.

Oordelen van leerlingen worden lang niet door iedereen positief gewaardeerd. Aleamoni (1981) noemt een reeks bezwaren die met name onderwijsinstellingen in de USA uiten tegen het gebruik van leerlingenoordelen. De instellingen twijfelen aan de betrouwbaarheid, stabi-

liteit en validiteit van de oordelen. Daarnaast is het de vraag of leerlingen leraarsgedrag wel gedifferentieerd kunnen beoordelen. De meeste beoordelingsschema's van leerlingen "zijn niet meer dan een populariteitswedstrijd waarbij een vriendelijke leraar met humor het steeds wint" (o.c., 112). Studenten zouden tevens leraren die hogere cijfers geven, belonen met een gunstiger oordeel. Ook wordt gewezen op externe factoren die de oordelen van leerlingen zouden beïnvloeden zoals de klassengrootte en sekse. Het voorgaande betekent dat er twijfels bestaan over de kwaliteit van oordelen van studenten. Aleamoni (1981) betwijfelt of de klachten van onderwijsinstellingen terecht zijn. Hij vermeldt diverse onderzoeksresultaten die de klachten van de instellingen ongegrond verklaren. Marsh (1987) stelt dat overzichten van 15 jaar onderzoek in de USA geen eenduidig beeld geven. De resultaten laten volgens hem geen eenduidige conclusies toe. Er zijn zowel onderzoeken met positieve als negatieve resultaten met betrekking tot de betrouwbaarheid en validiteit van leerlingenoordelen over leraarsgedrag.

Uit het voorgaande blijkt dat er tegengestelde posities zijn over de vraag of bij de beoordeling van de kwaliteit van instructie gebruik kan worden gemaakt van leerlingen als beoordelaars. In praktisch opzicht valt er veel voor te zeggen. Bij het beantwoorden van de gestelde vraag lijkt de leeftijd van leerlingen niet onbelangrijk. Veel onderzoek op het gebied van beoordelingen van leraren speelt zich af in het hoger onderwijs. In het onderzoek waarover hier wordt gerapporteerd, figureert een relatief jonge groep namelijk leerlingen met een gemiddelde leeftijd van 12 á 13 jaar. De centrale onderzoeksvraag luidt derhalve of ook relatief jonge leerlingen in staat zijn leraarsgedrag betrouwbaar en valide te beoordelen en welke problemen zich daarbij voordoen.

Analyse van de onderzoeksvraag

Het onderzoek is erop gericht na te gaan welke problemen zich voordoen bij het beoordelen van de kwaliteit van leraarsgedrag door leerlingen. In onderzoek waarbij gebruik wordt gemaakt van externe beoordelaars die afwisselend dezelfde leraren observeren vormen de beoordelaars een variantiebron die het mogelijk maakt de interbeoordelaarsovereenstemming vast te stellen (Heuvelmans & Sanders, 1993, Jansen & Westerhof, 1998). In onderzoek waar leerlingen leraren beoordelen ligt dat anders. Eén groep van leerlingen beoordeelt één leerkracht. Hierdoor missen we een variantiebron die het mogelijk maakt de overeenstemming tussen beoordelaars (groepen van leerlingen) na te gaan. In het onderhavige en soortgelijke onderzoeken zijn hiervoor geen oplossingen aanwezig. Bovendien geldt dat we te maken hebben met een meerniveau probleem. Leerlingen zijn genest binnen leraren. De leerlingoordelen zijn daarom niet onafhankelijk van elkaar. Bij de interpretatie van de resultaten moeten we deze aspecten altijd in de beschouwing betrekken.

De betrouwbaarheid van leerlingenoordelen wordt meestal vastgesteld op basis van itemanalyses (*homogeniteit*) en door de *interbeoordelaarsovereenstemming* tussen leerlingen uit dezelfde klas. Alleen item-analyses uitvoeren geeft een vertekende schatting omdat geen rekening wordt gehouden met de foutkans veroorzaakt door een gebrek aan overeenstemming tussen leerlingen die les krijgen van dezelfde leraar. Item-analyses zijn wel geschikt om vast te stellen of de schalen voldoende van elkaar verschillen (Marsh, 1987).

Marsh stelt dat de beoordelaarsovereenstemming tussen individuele paren van studenten laag is. Hij haalt onderzoek aan van Remmers (1949) en Feldman (1977) waaruit zou zijn gebleken dat de interbeoordelaarsbetrouwbaarheid van een a-select gekozen paar studenten in de .20 is, terwijl de betrouwbaarheid van *het gemiddelde klassenoordeel* afhangt van het aantal leerlingen in een klas. De geschatte betrouwbaarheid van een klas met 50 leerlingen is .95, terwijl de betrouwbaarheid van een klas met 10 leerlingen geschat wordt op .74. Bij een voldoende klassengrootte is de betrouwbaarheid van klasgeaggregeerde oordelen vergelijkbaar met de beste objectieve test (Marsh, 1987, 275). Een ander voordeel van gemiddelde

klassenoordelen wordt genoemd door Whitely and Doyle. Indien factor-analyse worden uitgevoerd op gemiddelde oordelen van de klas dan wordt de invloed van individuele impliciete theorieën van leerlingen op hun oordelen over de leraar gecontroleerd (Whitely and Doyle, in Marsh, 1987, p.272).

Analyse van individuele oordelen van leerlingen kan van belang zijn om na te gaan of individuele oordelen een mediërende invloed hebben op de relatie tussen leraarsgedrag en leerprestaties. Dat is echter niet het thema van dit artikel. Hier zijn we geïnteresseerd in de mening die leden van een groep (de klas) hebben over hun leraar. Dit wordt ook wel 'consensual beta press' genoemd als onderscheid met 'private beta press' (individuele leerlingenoordelen) en 'alpha press' (de klas beoordeeld door een externe waarnemer) (Fraser, 1991, 5).

De validiteit is moeilijker vast te stellen dan de betrouwbaarheid. De eenvoudigste en meest doorzichtige variant van het validiteitsbegrip is dat van de *predictieve validiteit*. De correlatie tussen voorspeller (oordeel over leraarsgedrag) en criterium (leerresultaat) is van beslissende betekenis en kan dienen als een operationele definitie van predictieve validiteit (De Groot, 1961). Marsh (1987) pleit ervoor meer criteria te hanteren dan cognitieve leerprestaties. Vooral affectieve criteria acht hij van belang. Deze criteria worden echter zelden in onderzoek meegenomen.

De *begripsvaliditeit* is een lastiger probleem. In de eerste plaats gaat het erom vast te stellen wat als relevant (te observeren of te beoordelen) leraarsgedrag kan worden beschouwd. Daarvoor kan gebruik worden gemaakt van theorieën over effectief leraarsgedrag en instrumenten die op dit gebied al zijn ontwikkeld. De relevantie van de items kan op logische gronden worden beargumenteerd. Factoranalyse kan worden toegepast om na te gaan of de theoretische begrippen in de werkelijkheid worden aangetroffen. Om na te gaan of de gevonden factoren voldoende eigenheid bezitten (voldoende van elkaar verschillen, gegeven de populariteitshypothese) zal ook de *discriminant validiteit* worden vastgesteld. Dit is een maat voor de gemiddelde overeenkomst van elke afzonderlijke schaal met de andere schalen die leraarsgedrag beogen te meten (Fraser, 1991). Als de correlatie .80 of hoger is, dan wordt verondersteld dat de factoren onderling te weinig verschillen om van afzonderlijke dimensies te spreken.

De *congruentievaliditeit* ofwel soortgenoot validiteit (De Groot, 1961) wordt bepaald door leerlingenoordelen te vergelijken met oordelen van externe waarnemers. Critici ten aanzien van het gebruik van leerlingenoordelen brengen naar voren dat oordelen van leerlingen worden beïnvloed door andere dan feitelijke leraarsgedragingen. Voorbeelden hiervan zijn sekse, leervermogen, klassengrootte en rapportcijfers: Aleamoni (1981) analyseerde 50 studies waarin rapportcijfers van leerlingen werden vergeleken met oordelen van leerlingen over de leraar. In 28 studies werden positieve maar gemiddeld zwakke verbanden ($r=.14$) gevonden. In hoeverre hier sprake is van 'bias' is niet duidelijk. Marsh (1987) is van mening dat adequate definities van 'bias' ontbreken. Het is niet voldoende als wordt aangetoond dat een willekeurige variabele samenhangt met leerlingenoordelen. Er moet ook worden aangetoond dat de betreffende variabele niet samenhangt met effectief leraarsgedrag (o.c., 310).

Op basis van de voorgaande overwegingen wordt de hoofdvraag van het onderzoek als volgt gespecificeerd:

1. Welke dimensies van leraarsgedrag worden door leerlingen onderscheiden en zijn er schalen te contrueren die deze dimensies intern consistent meten?
2. Verschillen de schalen voldoende van elkaar en hebben de schalen een samenhang met cognitieve en affectieve criteria?
3. Hoe verhouden de leerlingenoordelen zich tot de oordelen van externe waarnemers?
4. Verschillen de oordelen van de leerlingen per klas?

5. Worden de leerlingenoordelen beïnvloed door kenmerken die onafhankelijk zijn van het leraarsgedrag (bias op grond van sekse, klassengrootte en gemiddeld proefwerkcijfer van de leerlingen)?

De beantwoording van deze vragen moet leiden tot de conclusie of leerlingen in staat zijn 'effectief leraarsgedrag' vast te stellen.

ONDERZOEKSOPZET

Het onderzoek vond plaats in het eerste leerjaar van het voortgezet onderwijs in schooljaar 1994/1995. De leerlingen waren bij de start van het schooljaar gemiddeld 12,3 jaar oud ($n=1084$). Leerlingen en leraren in 49 klassen (27 scholen) zijn een jaar lang gevolgd tijdens de lessen wiskunde. De klassen waren ongeveer gelijk verdeeld over de stromen vbo/mavo ($n=15$), mavo ($n=17$) en mavo/havo ($n=17$). De leerlingen kregen les uit de methode 'Moderne Wiskunde' (61%) of 'Getal en Ruimte' (39%). Bij het begin van het schooljaar werd de voorkennis van leerlingen op het gebied van rekenen/wiskunde en informatieverwerking gepeild. Aan het eind van het schooljaar werden de prestaties van leerlingen gemeten met een curriculumspecifieke wiskundetoets. Tijdens het schooljaar zijn ook nog andere gegevens over de leerlingen verzameld zoals sekse, etniciteit, intelligentie (ruimtelijk inzicht en reken-snelheid), plezier in wiskunde en proefwerkcijfers.

Aan het eind van het schooljaar is de leerlingen een vragenlijst toegestuurd met vragen over het gedrag van hun leraar. Over de in de vragenlijst gestelde vragen, bestaande uit 55 stellingen, moesten de leerlingen hun oordeel geven. Na invulling van de vragenlijst (als huiswerk) konden de leerlingen deze in een bijgevoegde retourenvelop doen en dichtplakken om te voorkomen dat de leraar hun oordelen kon vaststellen.

Tijdens het schooljaar hebben externe waarnemers het leraarsgedrag tijdens de les op 3 verschillende momenten geobserveerd. Na elke les hebben de waarnemers 40 stellingen beoordeeld die de toestand in de klas beschrijven. De beoordelaars waren getraind. Na de training bedroeg de het coëfficiënt alpha .84. De interbeoordelaarscoëfficiënt p^2 was .82. (methode Heuvelmans en Sanders, 1993, p.443). Tijdens het veldwerk was de coëfficiënt alpha .92. De kenmerken van beide instrumenten worden beschreven in de volgende paragraaf.

Het instrument 'leerlingen over leraarsgedrag'

De vragenlijst die aan leerlingen is voorgelegd bevat 55 stellingen. De leerlingen is een oordeel gevraagd over deze stellingen. De stellingen zijn verdeeld in een 'low inference' deel en een 'high inference' gedeelte. Dit sluit aan bij het onderscheid dat gemaakt is in het kwaliteitsbegrip. Een bijkomend voordeel van het vragen naar frequenties is dat mogelijke 'bias' in de oordelen van leerlingen wordt beperkt. Om de vergelijkbaarheid met het observatie-instrument mogelijk te maken zijn de items gerelateerd aan dit instrument. Niet zozeer qua formulering en schaalvorm maar qua inhoud. Bij het vragen naar een oordeel over de leraar zijn taakrelevante aspecten uit de frequentievragen opgenomen en sociaal-emotionele aspecten uit instrumenten met betrekking tot leerlingenoordelen (Marsh, 1987).

Het leerlingbeoordelingsinstrument heeft de volgende specifieke kenmerken:

a. Frequentievragen:

De eerste 33 stellingen hebben betrekking op leraarsgedrag waarover gevraagd is hoe vaak dit gedrag door de leraar wiskunde wordt vertoond. De leerlingen konden hun antwoord geven door één van de cijfers 1 tot en met 5 te omcirkelen.

Voorbeeld item	minder dan 1x per maand	ongeveer 1 x per maand	ongeveer 1 x per 2 weken	ongeveer 1 x per week	bijna elke les
Onze leraar of lerares wiskunde ...					

Begint de les met aan te geven waar deze les over zal gaan	1	2	3	4	5
--	---	---	---	---	---

b. Oordeelvragen:

De volgende 22 stellingen hebben niet betrekking op de frequentie van gedrag maar op een oordeel over bepaald leraarsgedrag. Hier moesten de leerlingen aangeven in welke mate de uitspraak betrekking had op hun leraar. De leerlingen konden hun antwoord weergeven op een 5-puntsschaal.

Voorbeeld item	klopt helemaal niet	klopt grotendeels niet	klopt een beetje wel, een beetje niet	klopt vrij goed	klopt heel goed
In welke mate kloppen deze uitspraken?					

As wij niet opletten, heeft de leraar/lerares dat snel in de gaten	1	2	3	4	5
--	---	---	---	---	---

Hoewel de beoordelingscategorieën van beide instrumenten op hoofdlijnen overeenkomen zijn er naast vorm en formuleringverschillen enkele aspecten die inhoudelijk verschillen. Sommige onderwerpen kun je moeilijk in dezelfde vorm door waarnemers en leerlingen laten beoordelen zoals bijvoorbeeld of de leerkracht in kleine stapjes door de leerstof gaat, of de leraar de grote lijnen aangeeft of zich verliest in details. Andere thema's kun je wel aan leerlingen vragen maar moeilijk aan externe waarnemers. Bijvoorbeeld wat de leraar doet als de leerling een onvoldoende heeft voor een proefwerk (herhalen, remediëren of doorgaan). Dit soort zaken is door externe beoordelaars meestal niet in een of enkele lessen vast te stellen.

Het instrument 'extern oordeel over leraarsgedrag'

De inhoud van het observatie-instrument is ontleend aan het 'directe instructie model', resultaten van proces-productstudies en theorieën over gebruik van (meta)cognitie tijdens het onderwijsleerproces. Daarbij is een eclectisch gebruik gemaakt van bestaande instrumenten (Veenman e.a., 1986; Westerhof, 1989; Teddlie e.a., 1989; Schaffer & Nesselrodt, 1992). De oriëntatie op het 'directe instructie model' leek ons verantwoord omdat dit model het meest effectief is voor relatief zwakkere leerlingen (zie bv. Corno & Snow, 1986; Stallings & Stipek, 1986) en ook ondersteund wordt door theorieën over effectieve instructie (Rosenshine & Stevens, 1986). De gemiddelde leerling in ons onderzoek is een mavo-leerling en dus enigszins zwakker dan het gemiddelde niveau. Metacognitie is een nog relatief nieuw begrip waarvan wordt verondersteld dat het bij de aanpak van problemen een ondersteunende rol kan spelen (Simons, 1995).

Het instrument bestaat uit 40 items die gerubriceerd zijn in de volgende (theoretische) categorieën: activeren en structureren van voorkennis, introductie van nieuwe kennis, controle op begrijpen, toepassen en oefenen, huiswerk, metacognities, klassenklimaat en taakgerichtheid.

De items zijn uitgewerkt in contrasten (tegenpolen) waarbij '1' en '4' de absolute tegenpolen vormen en '2' en '3' minder sterke vormen van tegenstelling zijn. Een voorbeeld:

Voorbeeld item

1 2 3 4

De leerkracht houdt de leerlingen nauwelijks in de gaten tijdens de les					De leerkracht laat blijken dat hij/zij weet waar iedere leerling mee bezig is
---	--	--	--	--	---

Meer details over de schalen van leerlingen en externe beoordelaars zijn vermeld in de bijlage.

RESULTATEN

De resultaten worden in volgorde van de onderzoeksvragen besproken. De eerste onderzoeksvraag heeft betrekking op de dimensies die door de leerlingen aan het leraarsgedrag worden onderscheiden en de vraag of de onderscheiden dimensies voldoende homogene schalen vormen. Het is een procedure om van a priori geconstrueerde vragenlijsten te komen tot a posteriori geconstrueerde schalen. De te onderscheiden schalen moeten inhoudelijk te duiden en voldoende intern consistent zijn. De vraag welke dimensies leerlingen aan het leraarsgedrag onderscheiden is beantwoord met behulp van factoranalyse. Daarna is met behulp van Cronbachs alpha de interne consistentie van de onderscheiden schalen bepaald. Deze procedure is gevolgd met betrekking tot het leerlingeninstrument en terwille van de vergelijking ook met het instrument voor de externe beoordelaars. Na de factoranalyse is voor beide typen schalen de discriminant- en de predictieve validiteit bepaald en is de congruentievaliditeit van de schalen ten opzichte van elkaar nagegaan. Daarna is voor de leerlingenschalen nagegaan of deze voldoende gevoelig zijn om leraren van elkaar te onderscheiden. Tenslotte is bij de leerlingengegevens nagegaan of het oordeel over de leerkracht door andere factoren dan het leerkrachtgedrag wordt bepaald. Dit betreft de vraag naar de bias in de leerlingenoordelen.

Schaalconstructie leerlingeninstrument

Het leerlingen instrument bestond uit 55 items. De antwoorden van de leerlingen op de 55 items zijn geaggregeerd op klassenniveau. Om te bepalen of er schalen te onderscheiden zijn, is over de geaggregeerde items een factoranalyse uitgevoerd. De factoranalyse leverde 13 factoren op met een hogere waarde dan 1.0. De verklaarde variantie van deze factoren was 81,7%. Van 6 factoren met een voldoende aantal items is de interne consistentie bepaald. De interne consistentie van deze schalen was goed (alpha's tussen .76 en .96). De schalen en items zijn in de bijlage weergegeven. De negatief geformuleerde items zijn omgepoold zodat een hoger getal een positiever oordeel en/of meer frequent positief gedrag betekent. In het volgende worden de schalen inhoudelijk benoemd en wordt de interne consistentie van de schalen weergegeven.

De schalen zijn als volgt inhoudelijk geïnterpreteerd:

1. motiveren en goed uitleggen

In schaal 1 zijn 20 items opgenomen waarvan de homogeniteit goed is ($\alpha = .96$). De meeste items hebben betrekking op oordelen van leerlingen. Enkele items die betrekking hebben op positief of negatief gedrag van de leraar zijn als frequentie gemeten. Kennelijk spelen subjectieve oordelen van leerlingen bij dit soort leraarsgedrag een belangrijke rol.

In deze schaal komen de volgende aspecten van leraarsgedrag naar voren:

- positief gedrag van de leraar t.o.v. leerlingen of de klas (boosheid, negatieve opmerkingen, grapjes maken, tevredenheid uiten, hulpbereidheid)
- motiverend gedrag (saai lesgeven, je vervelen, enthousiast, leuke leraar, boeiend uitleggen)
- orde kunnen houden

- helderheid van de instructie (legt duidelijk uit, is moeilijk te verstaan, te lang aan het woord)
 Voor de leerlingen hangen deze vier aspecten kennelijk samen als zij hun leraar beoordelen.

2. Uitlegvaardigheden

Schaal twee bestaat uit 11 items die betrekking hebben op de frequentie waarmee bepaalde uitlegvaardigheden worden toegepast. De interne consistentie van de items is goed ($\alpha=.91$). De uitlegvaardigheden hebben betrekking op:

- organizers (relatie leggen met vorige les; structuur huidige les)
- concretisering (bordgebruik; gebruik van voorbeelden; voordoen)
- vragen stellen en veel (verschillende leerlingen) beurten geven
- huiswerk bespreken

3. Huiswerkcontrole

Schaal drie heeft slechts 2 items maar de interne consistentie is goed ($\alpha=.83$). De items hebben betrekking op het innemen van huiswerk om het te controleren op fouten en of het netjes is gemaakt. Bij de interpretatie van deze schaal moet ermee rekening worden gehouden dat de spreiding erg gering is. Op een 5-puntsschaal variëren de klassengemiddelden van 1.0 tot 1.8, hetgeen betekent dat de leraren deze vorm van controle minder dan 1 keer per maand of slechts 1 keer per maand uitvoeren.

4. Individuele begeleiding

De vierde schaal heeft betrekking op de frequentie van leerlingbegeleiding. De interne consistentie van de 4 items van deze schaal is goed ($\alpha=.76$). De items hebben betrekking op de frequentie waarmee de leraar door de klas loopt en hulp verleent als dat nodig is.

5. Groepscontrole

Dit is een aspect van managementgedrag. Er wordt gekeken naar de snelheid waarmee de leraar niet-taakbetrokken gedrag signaleert en naar het vermogen om niet-taakbetrokken gedrag non-verbaal te corrigeren. De 2 items hebben een goede interne consistentie ($\alpha=.85$).

6. Metacognitie

De laatste schaal telt 3 items met een goede interne consistentie ($\alpha=.76$). De items hebben betrekking op situaties waarin leerlingen moeten uitleggen hoe ze een opgave hebben gemaakt en waardoor ze fouten hebben gemaakt.

Hoewel klassenoordelen worden gehanteerd, waren we ook geïnteresseerd in de betrouwbaarheid op leerlingniveau. Daartoe is de interne consistentie per schaal, per klas berekend. De schalen 1 en 2 blijken in bijna alle klassen intern consistent te zijn op leerlingniveau. De gemiddelde interne consistentie op leerlingniveau voor alle klassen ligt voor schaal 1 op (α) .79 en voor schaal 2 op .68. Van de vier andere schalen varieert de homogeniteit op leerlingniveau sterk per klas. Dit leidt ook tot gemiddeld lage interne consistenties (respectievelijk .31, .36, .38 en .51). Deze lagere betrouwbaarheden op leerlingniveau zijn ten dele toe te schrijven aan het geringe aantal items per schaal. De eerste twee schalen kunnen ook op leerlingniveau als intern consistent worden beschouwd. De overige vier niet. Dit pleit ervoor bij het beoordelen van leraren door leerlingen uit te blijven gaan van klassenoordelen (geaggregeerde leerlingenoordelen) en niet van de scores op leerlingniveau.

Uit de schaalconstructie blijkt dat leerlingen de leraar niet enkel beoordelen naar de mate van diens populariteit. Er zijn 6 factoren naar voren gekomen die inhoudelijk interpreteerbaar waren. De populariteitshypothese, leerlingen beoordelen alleen of leraren 'leuk' zijn, wordt door de factoranalyse niet ondersteund.

Schaalconstructie observatie-oordelen

De externe beoordelaars hebben gedurende het schooljaar drie keer een les bijgewoond. Na afloop van elke les hebben ze 40 stellingen beoordeeld. Per item is het gemiddelde bepaald van de drie observaties. Over de 40 items is eerst een factoranalyse uitgevoerd. Dit leverde 5 factoren op met een voldoende aantal items en een hogere eigenwaarde dan 1.0. Van de items

die deel uitmaken van deze factoren is de interne consistentie bepaald. De factoren hebben betrekking op de volgende dimensies.

1. *Inzichtbevordering*

Deze schaal heeft 12 items die betrekking hebben op de mate waarin de leraar gedrag vertoont dat gericht is op de bevordering van inzichtelijk leren (cognitie en metacognitie). Dit gedrag is geformuleerd als tegenstelling tot gedrag waarin alleen het leerresultaat (antwoorden en fouten) door de leraar wordt beklemtoond. De interne consistentie is .91.

2. *Duidelijk, taakgericht en positief gedrag*

Deze schaal omvat eveneens 12 items waarin drie aspecten naar voren komen: duidelijk uitleggen, taakgerichtheid bevorderen en positief klimaat. De interne consistentie is .92.

3. *Beurten geven*

Deze schaal heeft 4 items die betrekking hebben op de kwaliteit waarmee beurten worden gegeven zoals veel leerlingen een beurt geven, correct moeten antwoorden en rekening houden met voldoende 'wachtijd' (wait-time). De interne consistentie is .73.

4. *Organizers toepassen*

Deze schaal heeft 3 items die alle betrekking hebben op het hanteren van organizers. De interne consistentie is .84.

5. *Oprachten uitleggen*

Deze schaal heeft 2 items waarin het gaat om duidelijke uitleg van (huiswerk)opgaven. De interne consistentie is .78.

Op basis van de interne consistentie van de verschillende schalen kan worden geconcludeerd dat deze voldoende homogeen zijn. Dit geldt zowel voor de leerlingenschalen (klassengemiddelden) als voor de schalen van de externe beoordelaars.

De validiteit van leerlingenoordelen

In het voorgaande hebben we gezien dat leerlingen de leraar niet enkel beoordelen naar de mate van diens populariteit maar dat diens gedrag aan dat oordeel ten grondslag ligt. In die zin zijn de leerlingenoordelen valide. Als we echter letten op de gemiddelden van de schalen zijn echter duidelijke verschillen te constateren (zie de bijlage). Controle van huiswerk door de leraar wordt bijvoorbeeld als weinig frequent gescoord door de klassen, terwijl de frequenties van andere schalen veel hoger wordt gescoord. Het is mogelijk dat de factoren onderling zeer hoog samenhangen. Om dit na te gaan is de discriminantvaliditeit vastgesteld. Daarnaast is de samenhang van de schalen met leerprestaties, de predictieve validiteit, onderzocht. Daarmee zijn we bij de tweede onderzoeksvraag. Om vergelijkingen mogelijk te maken zijn de analyses uitgevoerd voor de leerlingenschalen en de schalen van de externe beoordelaars.

De discriminantvaliditeit

De gemiddelde correlatie van een bepaalde schaal met de overigen geldt als de maat voor discriminantvaliditeit. Hoe lager de schalen onderling correleren, hoe minder ze hetzelfde meten. De discriminantvaliditeit is berekend voor de schalen leerlingenoordelen en de oordelen van de externe beoordelaars. De leerlingenschalen correleren onderling, afgezien van het teken, tussen de .05 en .56. Van de 15 correlaties zijn er 7 significant. De waarden daarvan liggen tussen .30 en .56. (zie tabel 1). De gemiddelde correlatie is niet erg hoog zodat geconcludeerd kan worden dat de schalen voldoende van elkaar verschillen, alhoewel sommige schalen voor een deel toch hetzelfde meten. Dit geldt bijvoorbeeld voor 'motiveren en goed uitleggen' en 'groepscontrole' die .56 correleren. De schaal 'huiswerkcontrole' vormt een apart verhaal. Deze correleert niet met de andere schalen. Dat komt doordat er nauwelijks leraren zijn die regelmatig huiswerk innemen om te controleren of het netjes gemaakt is en welke fouten voorkomen. Leraren bespreken wel fouten in de klas, maar deze bespreking is niet gebaseerd op een analyse van individuele, specifieke leerproblemen (Jong en Westerhof, 1998).

Tabel 1: Correlatie tussen de schalen 'leerlingenoordelen over leraarsgedrag', de gemiddelde correlatie tussen de schalen, berekend over de waarden afgezien van het teken, en de correlatie met de gemiddelde prestatie- en motivatiescore.

	2	3	4	5	6	Gemiddelde correlatie	Correlatie met wiskunde score	Correlatie met motivatie score
1	.45**	-.13	.36*	.56**	.39**	37.8	.30*	.82**
2		.07	.21	.38**	.51**	32.4	-.03	.34*
3			.07	-.05	.22	10.8	-.40**	-.10
4				.30*	.28	24.4	-.03	.29*
5					.27	31.2	.27	.50**
6						33.4	-.04	.39**

1 = Motiveren en goed uitleggen

2 = Frequentie belangrijke uitlegvaardigheden

3 = Huiswerkcontrole

4 = Individuele begeleiding

5 = Groepscontrole

6 = Metacognitie stimuleren

* significant 5%

** significant 1%

De discriminantvaliditeit tussen de schalen van de externe beoordelaars is gemiddeld lager dan die tussen de leerlingsschalen (de correlaties zijn hoger). Van de 10 correlaties zijn er 8 significant. De range loopt van .36 tot .60. Met name de schalen 'inzichtbevordering' en 'duidelijk, taakgericht en positief doceren' hebben veel variantie gemeen met de overige schalen. Onderling correleren deze schalen .60 (zie tabel 2). Gezamenlijk staan deze schalen waarschijnlijk voor de 'goede' leerkracht.

Tabel 2: Correlatie tussen de schalen 'externe beoordelaars over leraarsgedrag', de gemiddelde correlatie tussen de schalen, berekend over de waarden afgezien van het teken, en de correlatie met de gemiddelde prestatie- en motivatiescore.

	2	3	4	5	Gemiddelde correlatie	Correlatie met wiskunde score	Correlatie met motivatie score
1	.60**	.46**	.43**	.36**	46.3	.18	.15
2		.49**	.38**	.40**	46.8	.34*	.24
3			.10	.02	26.8	.20	.09
4				.38**	32.3	.25	.16
5					29.0	.26	.26

1 = Inzichtbevordering

2 = Duidelijk, taakgericht, positief doceren

3 = Beurten geven

4 = Organisers toepassen

5 = Opgaven uitleggen

* significant 5%

** significant 1%

De discriminantvaliditeit van de schalen van de externe beoordelaars zijn voor de eerste twee schalen iets lager maar voor de rest vergelijkbaar met die van de leerlingsschalen. Ook hier zijn de correlaties van de afzonderlijke schalen met het gemiddelde van de andere schalen niet zo hoog dat aan de eigenheid van de schalen moet worden getwijfeld.

De predictieve validiteit

De predictieve validiteit geeft aan in hoeverre een schaal samenhangt met een criterium. In het onderzoek naar de effectiviteit van het onderwijs zijn criteria veelal effectmaten op het cognitieve domein (Creemers, 1994). Marsh (1987) is van mening dat de effectmaten niet beperkt moeten blijven tot het cognitieve domein. Er zijn echter maar weinig studies waar aandacht wordt besteed aan affectieve criteria. Hier doen we dat wel. We gaan de samenhang na met motivatie² en een leerprestatiescore, beide gemeten in de laatste twee maanden van het schooljaar. Motivatie is gemeten met een schaal bestaande uit 11 items waarin wordt gevraagd hoe plezierig de leerling het vak wiskunde vindt. De homogeniteit is goed ($\alpha = .87$). De schaal is onderdeel van het instrument dat in 1981 door Kremers is ontwikkeld en in 1990 is aangepast door Otten en Boekaerts (1990). De leerprestatie is gemeten met behulp van een test bestaande uit 84 items ($\alpha = .92$). De test is geconstrueerd op basis van beide in het onderzoek betrokken wiskundemethodes. Voor ieder goed antwoord werd een punt toegekend. De maximale score is 84.

Wat betreft de leerlingenschalen valt het op dat deze sterk correleren met de gemiddelde motivatie die de leerlingen rapporteren. Van de zes schalen correleren er 5 positief en significant (zie tabel 1). De hoogste correlatie is voor de schaal 'motiveren en goed uitleggen'. Leerlingen die zeggen gemotiveerd te zijn, zijn van oordeel dat hun leerkracht genoemd gedrag vertoont. Ook 'groepscontrole' en 'het stimuleren van de metacognities' draagt volgens de leerlingen bij aan hun motivatie. Twee van de zes schalen correleren significant met de gemiddelde wiskundescore. 'Motiveren en duidelijk uitleggen' correleert positief en 'huiswerkcontrole' correleert negatief. Het eerste hoeft geen toelichting. Het is in overeenstemming met eerder aangehaalde overzichten waaruit blijkt dat duidelijke, heldere uitleg positief correleert met leerprestaties. De negatieve correlatie met 'huiswerkcontrole' is als volgt te interpreteren. Leerkrachten besteden kennelijk veel tijd aan controle van het huiswerk bij groepen die laag scoren en minder bij groepen die hoog scoren. Dit lijkt een acceptabele verdeling van de aandacht over de leerlingen. Het betekent ook gezien de samenstelling van de onderzoeksgroep dat in het VBO, en het VBO/MAVO kennelijk meer aan huiswerkcontrole wordt gedaan dan in het HAVO. Vanuit theorievorming gebaseerd op effectiviteitsonderzoek lijkt ook dit een goede policy (Creemers, 1994).

De schalen van de externe beoordelaars correleren geen van alle significant met de gemiddelde motivatiescore. Slechts een van de schalen correleert positief en significant met de gemiddelde wiskundescore. Dat is het leerkrachtgedrag dat wordt beoordeeld als duidelijk, positief en taakgericht.

In het voorgaande is nagegaan of op klassenniveau de gemiddelde criteriumscore kon worden voorspeld op grond van de leerlingen- en beoordelaarsschalen. We zijn immers geïnteresseerd in leerlingenoordelen geaggregeerd op klassenniveau. De criteriumscores (cognitief en affectief) zijn echter vooral belangrijk als het gaat om het leerlingenniveau. In het navolgende willen we daarom met behulp van tweenniveau-analyse³ nagaan of en hoe de schalen samenhangen met de individuele motivatie- en wiskundescores. Bij de tweenniveau-analyse wordt nagegaan in hoeverre de leerlingenschalen en de schalen van de externe beoordelaars bijdragen aan het voorspellen van het criterium. Ten behoeve van het vergelijken van de effecten van de onafhankelijke variabelen zijn de oorspronkelijke scores getransformeerd naar z-waarden. Op leerlingenniveau is telkens gecontroleerd voor algemene intelligentie, de voorkennis rekenen en taal en motivatie voor wiskunde. Deze variabelen zijn gemeten aan het begin van het schooljaar. De rekenvaardigheid is vastgesteld met een rekentest ($\alpha = .79$). De taalvaardigheid is gemeten met een toets informatieverwerking ($\alpha = .67$). Van de algemene intelligentie zijn de aspecten visueel vermogen gemeten ($\alpha = .79$) en de snelheid waarmee de vier hoofdbewerkingen rekenen konden worden uitgevoerd ($\alpha = .87$). De motivatie voor wis-

kunde is aangegeven op een likertschaal die loopt van 1 tot 7. De analyses zijn voor beide typen schalen (leerling/externe beoordelaars) en beide criteriummaten afzonderlijk uitgevoerd. Hier beperken we ons tot een tweenniveau-analyse omdat de variantie in de motivatiescore op schoolniveau niet significant is en we primair geïnteresseerd zijn in de verklarende variabelen op klassenniveau. Uit de analyses blijkt dat op klassenniveau alleen de schaal 'motiveren en goed uitleggen' significant bijdraagt aan het voorspellen van de motivatiescore. In tabel 3 wordt daarom alleen de resultaten van de tweenniveau-analyse voor deze schaal en de covariaten op leerlingenniveau als onafhankelijke variabelen en de motivatiescore als afhankelijke variabele weergegeven. In de tabel zijn alleen de covariaten opgenomen die significant met de criteriumvariabelen samenhangen.

Tabel 3: Tweenniveau-analyse met de individuele motivatiescore als afhankelijke variabele en de leerlingenschaal 'motiveren en goed uitleggen' als predictor op klassenniveau (49 klassen, 1084 leerlingen). Variantie van het nul-model .7879.

Fixed part	nul-model	model 1	model 2
Intercept	3.2 (.06)	3.2 (.05)	3.2 (.03)
Leerlingenniveau			
- Voorkennis rekenen		.10 (.03)	.08 (.03)
- Aanvankelijke motivatie		.30 (.02)	.29 (.02)
Klassenniveau			
- Motiveren en goed uitleggen			.26 (.03)
Variantedelen			
Onverklaard			
Klassenniveau %	15.2	11.4	02.7
Leerlingenniveau %	84.8	73.5	73.5
Verklaard %	00.0	15.1	23.8
Significantie test			
χ^2	2717.12	2555.65	2513.26
Reductie		161.47	42.39
Aantal vrijheidsgraden/ p-waarden		2/ .00	1/ .00

Uit tabel 3 blijkt dat de gemiddelde score op de motivatie test 3.2 bedroeg. De standaardfout rond het gemiddelde van het lege model is .06 zodat de werkelijke waarde met 95% zekerheid ligt tussen en 3.1 en 3.3. Op leerlingenniveau zijn er twee factoren die de motivatie significant voorspellen: 'voorkennis rekenen' en de 'motivatie aan het begin van het jaar'. De aanvankelijke motivatie is daarbij het belangrijkste. Opmerkelijk is dat veruit de meeste variantie op leerlingenniveau blijft bestaan. Er is slechts een reductie van 11.3% (van 84.8% naar 73.5%). De variabele 'motiveren en goed uitleggen' op klassenniveau laat de onverklaarde variantie op klassenniveau dalen van 11.4% tot 2.7%, een winst in verklaarde variantie van 8.7%. Samenvattend kunnen we dus zeggen dat de drie factoren gezamenlijk 23.8% van de variantie in de motivatiescore verklaren. Van dit percentage is 8.7% toe te schrijven aan de leerkrachtenfactor. In de volgende analyse gaan we na of de individuele wiskundescore kan worden voorspeld op grond van de leerlingenscalen. Uit de analyses blijkt dat alle covariaten significant bijdragen aan het voorspellen van de wiskundescore en drie van de schalen op klassenniveau: 'motiveren en goed uitleggen', 'huiswerkcontrole' en 'groepscontrole'. De resultaten van de analyses met deze schalen zijn weergegeven in tabel 4.

Uit de tabel blijkt dat de gemiddelde wiskunde score in de verschillende modellen schommelt rond de 47.5. De gemiddelde afwijking rond het gemiddelde is gering. De werkelijke waarde van het gemiddelde ligt met 95% zekerheid tussen 46.1 en 48.5. De variantie in het nul-model bedraagt 164.84. Hiervan bevindt zich 41.1% op klassenniveau en 58.9% op leerlingenniveau. De covariaten voorkennis, intelligentie en motivatie op leerlingenniveau in model 1

Tabel 4: Tweeniveau-analyse met de individuele wiskundescore als afhankelijke variabele en de leerlingenschalen 'motiveren en goed uitleggen', 'huiswerkcontrole' en 'groepscontrole' als predictoren op klassenniveau (49 klassen, 1084 leerlingen). Variantie in het nul-model 164.84. Model 2a en 2b zijn beide vergeleken met model 1. Bij model 2a en 2b: correlatie tussen 'motiveren en goed uitleggen' en 'groepscontrole' is .56.

Fixed part	nul-model	Model 1	model 2a	model 2b
Intercept	47.3 (1.22)	47.6 (.66)	47.6 (.58)	47.7 (.57)
Leerlingenniveau				
Voor kennis rekenen		4.59 (.31)	4.59 (.31)	4.61 (.57)
Taalvaardigheid		1.47 (.28)	1.43 (.28)	1.46 (.28)
Ruimtelijk inzicht		2.43 (.26)	2.44 (.26)	2.43 (.26)
Rekenschap		2.55 (.26)	2.55 (.26)	2.54 (.26)
Aanvankelijke motivatie		0.97 (.25)	0.93 (.25)	0.99 (.25)
Klassenniveau			1.26 (.58)	
Motiveren en goed uitleggen			-1.65 (.59)	-1.73 (.58)
Huiswerkcontrole				1.48 (.56)
Groepscontrole				
Variatie delen				
Onverklaard				
Klassenniveau %	41.1	10.9	08.0	07.6
Leerlingenniveau %	58.9	33.2	33.2	33.2
Verklaard %	00.0	56.0	58.8	59.1
Significantie test				
χ^2	8172.21	7575.89	7563.23	7561.28
Reductie		596.32	12.66	14.61
Aantal vrijheidsgraden/ p-waarden		5/1.00	2/1.00	2/1.00

verklaren 56% van de variantie. Op leerlingenniveau blijft dan 33.2 van de variantie onverklaard en op klassenniveau 10.9. Zoals we reeds eerder zagen in tabel 1, correleren 'motiveren en goed uitleggen' en 'groepscontrole' .56 met elkaar. Indien nu de variabelen 'motiveren en goed uitleggen' en 'groepscontrole' samen in één model worden opgenomen veroorzaakt dat collineariteit zodat de significante samenhang (5%-niveau) van beide schalen met de wiskundescore verdwijnt. Omdat beide variabelen vanuit de theorie omtrent effectief onderwijzen echter zeer belangrijk zijn worden twee modellen gepresenteerd waarin ze afzonderlijk voorkomen. Model 2a laat zien dat leerkrachten, waarvan de leerlingen van mening zijn dat ze motiveren en goed uitleggen, per standaarddeviatie op de onafhankelijke variabele een leerwinst boeken van 1.26 punten op de afhankelijke variabele. Leerkrachten die veel aandacht besteden aan huiswerkcontrole hebben per standaarddeviatie op deze variabele een negatieve invloed op de afhankelijke variabele van -1.65 punten. Dit lijkt een pleidooi om de huiswerkcontrole af te schaffen omdat controle leidt tot slechter presteren. Het is echter waarschijnlijker dat de negatieve samenhang uitdrukt dat leraren in klassen die slecht presteren veel aandacht aan huiswerkcontrole besteden. Indien we ervan uitgaan dat huiswerkcontrole leidt tot het maken van huiswerk dan weten we uit ander onderzoek (De Jong en Westerhof, 1998) dat dat bijdraagt aan het vergroten van de leerwinst. De toevoeging van de schalen 'motiveren en goed uitleggen' en 'huiswerkcontrole' verhogen de verklaarde variantie tot 58.8%. Op leerlingenniveau blijft 33.2% onverklaard en op klassenniveau daalt de onverklaarde variantie van 10.9 tot 8.0%. Model 2b laat ten aanzien van de variabele 'huiswerkcontrole' hetzelfde zien als 2a en de invloed van de variabele 'groepscontrole' is te vergelijken met de invloed van de variabele 'motiveren en goed uitleggen'. De niet verklaarde variantie in model 2b is nog iets kleiner dan in model 2a. Het verschil op klassenniveau bedraagt 0.4%. Resumerend kunnen we zeggen dat de schaal 'motiveren en goed uitleggen' predictieve validiteit bezit ten aanzien van de beide criteriumvariabelen. De schalen 'huiswerkcontrole' en 'groepscontrole' bezitten bovendien predictieve validiteit ten aanzien van de wiskundescore.

Tweeniveau-analyse met de individuele motivatiescore als afhankelijke variabele en de externe beoordelaarschalen op klassenniveau laat zien dat alleen de manier waarop opgaven worden opgegeven samenhangt met de motivatie van de leerlingen. Het effect is net zo groot als het effect van 'voorkennis rekenen'. Toevoegen van de variabele 'opdrachten uitleggen' laat de onverklaarde variantie op klassenniveau dalen met 1.2%. Het overgrote deel van de variantie blijft onverklaard. Zie tabel 5.

Tabel 5: Tweeniveau-analyse met de individuele motivatiescore als afhankelijke variabele en de externe beoordelaarschaal 'opdrachten uitleggen' als predictor op klassenniveau (49 klassen, 1084 leerlingen). Variantie in het nul-model .7879.

Fixed part	Nul-model	Model 1	Model 2
Intercept	3.2 (.06)	3.2 (.05)	3.2 (.05)
Leerlingenniveau			
Voorkennis rekenen		.10 (.03)	.10 (.03)
Aanvankelijke motivatie		.30 (.02)	.31 (.02)
Klassenniveau			
Opdrachten uitleggen			.10 (.05)
Variantiedelen			
Onverklaard			
Klassenniveau %	15.5	11.4	10.2
Leerlingenniveau %	84.8	73.5	73.4
Verklaard %	00.0	15.1	16.4
Significantie test			
χ^2	2717.12	2555.65	2551.36
Reductie		161.47	4.29
Aantal vrijheidsgraden/ p-waarden		2/ .00	1/ .05

De tweeniveau-analyse met de individuele wiskundescore als afhankelijke variabele en de externe beoordelaarschalen als onafhankelijke variabelen geeft als resultaat dat de schalen 'duidelijk taakgericht, positief doceren', 'beurten geven' en 'organizers toepassen' significant samenhangen met de wiskundescore als ze afzonderlijk in een model worden opgenomen. Omdat deze variabelen onderling hoog correleren, zie tabel 2, verdwijnt de significantie van het effect als de eerst genoemde variabele met een van de twee andere in een model wordt geplaatst. 'Beurten geven' en 'organizers toepassen' correleren onderling zo laag dat ze in hetzelfde model significant zijn. De resultaten van de analyse zijn samengevat in tabel 6.

Resumerend kunnen we opmerken dat de schalen die vanuit de theorie over effectief onderwijzen belangrijk zijn predictieve validiteit bezitten ten opzichte van de wiskundescore. De motivatie wordt verhoogd indien duidelijk is hoe de opgaven moeten worden gemaakt. Vergelijking van de tabellen 3 en 5 leert dat de variabelen die positief met de motivatie samenhangen in beide gevallen te maken hebben met de duidelijkheid die de leraar verschaft ten aanzien van de leertaak. Vergelijking van de tabellen 4 en 6 geeft aan dat leerwinst samenhangt met duidelijke uitleg, het scherp houden van de leerlingen door toezicht op het leerproces en de kwaliteit van het geven van beurten.

Congruëntievaliditeit

De derde onderzoeksvraag betreft de overeenkomst tussen de leerlingenoordelen en de oordelen van de externe beoordelaars. Beide instrumenten beogen hetzelfde te meten en moeten dus hoog correleren. Het is de vraag naar de congruentievaliditeit. Op basis van de inhoud van de schalen van leerlingen en externe beoordelaars kunnen verwachtingen worden geformuleerd over de te verwachten relaties tussen de schalen. Schaal 1 van de leerlingen en schaal 2 van de externe beoordelaars hebben beide betrekking op duidelijk uitleggen, positief gedrag

van de leraar en orde houden c.q. de taakgerichtheid bevorderen. Deze schalen moeten positief correleren.

Schaal 2 van de leerlingen gaat over de frequentie waarmee belangrijke uitlegvaardigheden worden toegepast, waaronder het toepassen van organizers. De kwaliteit van uitleggedrag komt naar voren in schaal 2 en 5 van de externe beoordelaars. In schaal 4 van de externe beoordelaars gaat het over organizers. Er wordt dus een positieve correlatie verwacht tussen leerlingsschaal 2 enerzijds en de externe beoordelaarschalen 2, 4 en 5 anderzijds.

Schaal 5 van de leerlingen (groepscontrole) komt bij de externe beoordelaars voor in schaal 2 als orde kunnen houden en de kwaliteit van het beurten geven zodanig dat iedere leerling bij het onderwijsleerproces betrokken is. We verwachten derhalve een positieve correlatie tussen enerzijds leerlingsschaal 5 en de schalen 2 en 3 van de externe beoordelaars.

Tabel 6: Tweeniveau-analyse met de individuele wiskunde score als afhankelijke variabele en de externe beoordelaarschalen 'duidelijk, taakgericht, positief doceren', 'beurten geven' en 'organizers toepassen' als predictoren op klassenniveau (49 klassen, 1084 leerlingen). Het nul-model en model 1 van tabel 4 gaan aan elk van de in de tabel 6 weergegeven modellen vooraf. De variantie in het nul-model is 164.84, de deviance van model 1 is 7575.89

Fixed part	model 2a	Model 2b	model 2c	model 2d
Intercept	47.6 (.63)	47.6 (.63)	47.6 (.61)	47.6 (.58)
Leerlingenniveau				
Voorkennis rekenen	4.58 (.31)	4.61 (.31)	4.60 (.31)	4.63 (.31)
Taalvaardigheid	1.46 (.28)	1.49 (.28)	1.48 (.28)	1.50 (.28)
Ruimtelijk inzicht	2.44 (.26)	2.44 (.25)	2.46 (.26)	2.46 (.26)
Rekensnelheid	2.54 (.25)	2.56 (.26)	2.56 (.26)	2.57 (.26)
Aanvankelijke motivatie	0.96 (.25)	0.96 (.25)	0.95 (.25)	0.94 (.25)
Klassenniveau				
Duidelijk, taakgericht, positief doceren	1.48 (.63)			1.29 (.59)
Beurten geven		1.46 (.63)		1.59 (.60)
Organizers toepassen			1.46 (.63)	
Variantiedelen				
Onverklaard				
Klassenniveau %	09.6	09.6	09.0	08.0
Leerlingenniveau %	33.2	33.2	33.2	33.2
Verklaard %	57.2	57.2	57.8	58.8
Significantie test				
χ^2	7570.59	7570.82	7568.62	7564.02
Reductie	5.30	5.02	7.27	11.87
Aantal vrijheidsgraden/ p-waarden	1/ .03	1/ .03	1/ .01	2/ .01

Voor wat betreft de andere schalen is er geen duidelijke verwantschap. Huiswerkcontrole en individuele begeleiding (leerlingsschalen) zijn bij de externe beoordelaars geen schalen omdat er slechts één enkele stelling op betrekking heeft. Het stimuleren van metacognitie (een leerlingsschaal) is bij de externe beoordelaars (gedeeltelijk) terug te vinden in de eerste schaal met betrekking tot inzichtbevordering. De meeste items van deze schaal hebben evenwel meer betrekking op inzichtbevordering (cognitie) dan op metacognitie. In tabel 7 zijn de correlaties tussen de schalen van de leerlingen en externe beoordelaars weergegeven.

De veronderstelling over samenhangen tussen schalen van leerlingen en externe beoordelaars wordt bevestigd, alhoewel de gevonden correlaties laag zijn. Voor een deel moet dit worden verklaard uit het feit dat de bewoordingen in de schalen verschillen. Nadere inspectie van de schalen, zie bijlage, leert ook dat er accentverschillen bestaan tussen schalen die worden ver-

ondersteld hetzelfde te meten. Bijvoorbeeld, schaal 1 van de leerlingen en schaal 2 van de externe beoordelaars hebben beide betrekking op duidelijk uitleggen. Toch ligt de nadruk in de leerlingenschaal meer op de attitude van de leraar ten opzichte van de leerlingen, terwijl de schaal van de externe beoordelaars sterker de nadruk legt op een soepel verloop van de les. De sterkte van de verbanden tussen (inhoudelijk overeenkomstige) schalen van leerlingen en externe beoordelaars zijn vergelijkbaar met de correlaties binnen schalen van leerlingen of van externe beoordelaars. Geen van de veronderstelde correlaties gaat in de tegenovergestelde richting. Vijf van de zes veronderstelde correlaties zijn significant. De veronderstelde positieve correlatie tussen 'frequentie belangrijke uitlegvaardigheden' en 'opdrachten uitleggen' is onvoldoende hoog om significant te zijn.

Tabel 7: Correlatie tussen de leerlingenschalen en de schalen van de externe beoordelaars.

Leerlingenschalen	Externe beoordelaars				
	1	2	3	4	5
1	.24	.38**	.24	.24	.13
2	.21	.29*	.15	.31*	.22
3	-.01	-.12	-.06	-.17	-.05
4	-.13	.19	.09	.02	.17
5	.20	.34*	.37**	.21	.12
6	.07	.03	.12	.01	-.08

Leerlingenschalen:

- 1 = Motiveren en goed uitleggen
- 2 = Frequentie belangrijke uitlegvaardigheden
- 3 = Huiswerkcontrole
- 4 = Individuele begeleiding
- 5 = Groepscontrole
- 6 = Metacognitie stimuleren

Externe beoordelaars:

- 1 = Inzichtbevordering
- 2 = Duidelijk, taakgericht, positief doceren
- 3 = Beurten geven
- 4 = Organizers toepassen
- 5 = Opdrachten uitleggen

* significant 5%

** significant 1%

Gevoeligheid van leerlingenoordelen

De vierde onderzoeksvraag betreft de gevoeligheid van de leerlingenoordelen. Met gevoeligheid wordt bedoeld de vraag naar de mate waarin oordelen van leerlingen verschillen tussen klassen c.q. leraren laten zien. Daartoe is per leerlingenschaal, maar dan met de individuele scores als uitgangspunt, nagegaan of klassen verschillend oordelen over hun leraar. Hierbij is gecontroleerd voor verschillen tussen klassen in voorkennis, intelligentie en motivatie voor wiskunde bij de start van het schooljaar. De verschillen tussen klassen zijn bij alle schalen significant. Verschillen tussen klassen zijn het sterkst bij schaal 1 ($F=13,8$), schaal 2 ($F=10,0$) en het zwakst bij schaal 3 (huiswerkcontrole; $F=1,9$) en schaal 5 (groepscontrole; $F=3,2$). Lezers die geïnteresseerd zijn in meer details, worden naar de bijlage verwezen.

Bias in leerlingenoordelen

De laatste onderzoeksvraag thematiseert de kritiek op het gebruik van leerlingenoordelen over leraren. De kritiek is dat leerlingenoordelen worden beïnvloed door andere kenmerken dan het gedrag van de leraar. De oordelen vertonen dan 'bias'. Marsh (1987, p 305) noemt als mogelijke biasfactoren onder andere: klassengrootte, populariteit van de docent, voorkennis, motivatie voor het begin van de cursus en het gemiddelde proefwerkcijfer in een klas. Ten aanzien van het laatst genoemde leeft de verwachting dat leraren die hogere cijfers geven hiervoor beloofd worden met een hogere beoordeling door de leerlingen (Aleamoni, 1981, Marsh, 1987). Bias zou moeten blijken uit het feit dat de scores van individuele leerlingen op de leerlingenschalen zijn te voorspellen op grond van de variabelen waarvan aangenomen wordt dat

ze bias veroorzaken. Hier is een parallel te trekken met de criteriummaten. Deze worden voor een deel bepaald door het gedrag van de leerkracht en voor een deel door factoren die daarmee niets van doen hebben zoals bijvoorbeeld de voorkennis of de algemene intelligentie van de leerling. Het kan dus verhelderend werken de invloed van mogelijk bias veroorzakende factoren op de criteriummaten en de leerlingenscalen te vergelijken. Het SIE-materiaal maakt het mogelijk na te gaan of de klassengrootte, sekse van de docent en de leerling, voorkennis rekenen en taal, de algemene intelligentie en het gemiddeld proefwerkcijfer samenhangen met de leerlingenscalen en met de criteriummaten op leerlingenniveau. Als ze op leerlingenniveau samenhangen met de leerlingenscalen vormen ze een mogelijke bron van bias. De analyses zijn weer uitgevoerd in de vorm van tweenniveau-analyses, leerling en klas⁴. De onafhankelijke variabelen op intervalniveau zijn getransformeerd naar z-waarden en groeperingsvariabelen geslacht van leerling en docent zijn 0 (=mannelijk) en 1 (=vrouwelijk) gecodeerd. Omdat op leerlingenniveau alleen de schalen 1 en 2 ('motiveren en goed uitleggen' en 'frequentie belangrijke uitlegvaardigheden') voldoende betrouwbaar en intern consistent zijn (zie schaalconstructie leerlinginstrument), zijn alleen deze naast de individuele wiskunde- en motivatiescore in de analyse betrokken. De resultaten worden weergegeven in de tabellen 8, 9 en 10. In tabel 8 worden de nulmodellen gepresenteerd. In tabel 9 zijn de covariaten op leerlingenniveau in het model opgenomen en in tabel 10 zijn de variabelen op klassenniveau toegevoegd. In de overeenkomstige kolommen van de successieve tabellen hebben we dus een model 0, een model 1 en een model 2. Om te zien of een volgend model significant afwijkt van het voorgaande moet vergeleken worden met de overeenkomstige kolom in de voorgaande tabel.

Tabel 8: Nul-modellen met de leerlingenscalen en criteriummaten als afhankelijke variabelen (49 klassen, 1084 leerlingen)

	Afhankelijke variabelen			
	Leerlingenschaal		criteriummaat	
	1	2	1	2
Vast deel				
Intercept	3.5 (.06)	3.7 (.02)	47.3 (1.22)	3.2 (.06)
Variantiedeel				
Absoluut	0.4398	0.5899	164.84	0.7897
% klassenniveau	39.0	28.5	41.1	15.2
% leerlingenniveau	61.0	71.5	58.9	84.8
χ^2	17.81.94	2251.25	8172.21	2717.12

Leerlingenscalen:

1 = Motiveren en goed uitleggen

2 = Frequentie belangrijke uitlegvaardigheden

Criteriummaten

1 = wiskundescore

2 = motivatiescore

Uit tabel 8 blijkt dat de variantie in de leerlingenscalen en de criteriummaten voor het grootste deel moet worden toegeschreven aan het leerlingenniveau. De percentages variantie voor de leerlingenscalen op leerlingenniveau zijn 61.0 en 71.5%. Dit duidt erop dat het oordeel over de leerkracht slechts voor een deel wordt bepaald door de leerkracht. Andere factoren zijn gezamenlijk minstens zo belangrijk. Bij de interpretatie van de resultaten is dat een belangrijk gegeven. De variantie in de criteriummaten, wiskundescore en motivatiescore, zijn voor respectievelijk 58.9 en 84.8% toe te schrijven aan het leerlingenniveau. Kennelijk doet de leerkracht er iets toe als het gaat om de wiskundeprestatie. Motivatie daarentegen lijkt meer individueel bepaald. In het volgende gaan we na of er biasveroorzakende factoren kunnen worden onderkend. Eerst op leerlingenniveau, dan op klassenniveau

Bias veroorzaakt door factoren op leerlingenniveau

Het opnemen van de covariaten leidt in alle gevallen tot een significante verbetering van de databeschrijving. De resultaten zijn weergegeven in tabel 9. De betekenis van een effect in tabel 9 en volgende is dat een winst van 1 standaarddeviatie op de onafhankelijke variabele leidt tot een groei ter grootte van het effect op de afhankelijke variabele. Bijvoorbeeld 1 standaarddeviatie winst op de voorkennis rekenen leidt tot een 4.60 punt hogere wiskundescore.

Tabel 9: modellen met verklarende variabelen op leerlingenniveau en de leerlingenschalen en criteriummaten als afhankelijke variabelen (49 klassen, 1084 leerlingen)

	Afhankelijke variabelen			
	Leerlingenschaal		criteriummaat	
Vast deel	1	2	1	2
Intercept	3.5 (.06)	3.8 (.07)	47.7 (.71)	3.3 (.06)
Leerlingenniveau				
Voorkennis				
- Rekenen	-.03 (.02)	-.02 (.03)	4.60 (.32)	.05 (.03)
- Taal	.02 (.02)	-.03 (.02)	1.48 (.28)	.00 (.02)
Algemene intelligentie				
- Ruimtelijk inzicht	-.00 (.02)	.01 (.02)	2.43 (.26)	.02 (.03)
- Rekensnelheid	-.00 (.02)	-.03 (.02)	2.56 (.26)	.02 (.03)
Aanvankelijke motivatie	.09 (.02)	.04 (.02)	0.96 (.25)	.29 (.02)
Geslacht (0=jongen, 1=meisje)	-.03 (.03)	-.11 (.03)	-0.10 (.50)	-.20 (.05)
Varianteedeel				
% verklaard	04.3	02.2	53.5	16.1
Onverklaard				
% klassenniveau	36.3	27.7	11.4	11.8
% leerlingenniveau	59.4	70.1	35.0	72.1
Significantietest				
χ^2	1752.06	2234.73	7575.85	2537.88
Reductie	29.88	16.52	596.36	197.24
Vrijheidsgraden / p-waarden	6 / .00	6 / .03	6 / 00	6 / .00

Leerlingenschalen:

1 = Motiveren en goed uitleggen

2 = Frequentie belangrijke uitlegvaardigheden

Criteriummaten:

1 = wiskundescore

2 = motivatiescore

Uit tabel 9 blijkt dat de verklaarde variantie het hoogst is voor de criteriummaten. De wiskundescore wordt voor 53.5% verklaard door de covariaten en de motivatiescore voor 16.1%. Wat betreft de wiskundescore zijn er effecten toe te schrijven aan de voorkennis rekenen (4.60), de algemene intelligentie (ruimtelijk inzicht 2.43 en rekensnelheid 2.56) en taalbeheersing (1.48). Ook het effect van de aanvankelijke motivatie is significant en positief (.96). Het effect van het geslacht is niet significant. De motivatiescore kent twee significante effecten: de aanvankelijke motivatie heeft een positief effect (.29) en het geslacht heeft een negatief effect (-.20). Het laatste betekent dat meisjes na een jaar voortgezet onderwijs zeggen minder gemotiveerd te zijn voor wiskunde dan jongens.

Het opnemen van de covariaten levert voor de leerlingenschalen weinig verklaarde variantie op: 4.3% voor de schaal 'motiveren en goed uitleggen' en 2.2% voor de schaal 'frequentie belangrijke uitlegvaardigheden'. Voor beide schalen is er een significant effect van de aanvankelijke motivatie: .09 voor de schaal 'motiveren en goed uitleggen' en .04 voor de schaal 'frequentie belangrijke uitlegvaardigheden'. Voor de schaal 'frequentie belangrijke uitlegvaardigheden' is er een significant effect van het geslacht (-.11).

Als we de invloed van mogelijk biasveroorzakende factoren op de afhankelijke variabelen vergelijken dan valt op dat de wiskundescore aan de grootste 'bias' onderhevig is. Het

grootste deel van de variantie in de wiskundescore wordt verklaard door factoren die (relatief) onafhankelijk zijn van het leerkrachtgedrag, namelijk voorkennis, algemene intelligentie en de oorspronkelijke motivatie. Dat is niets nieuws en de reden waarom altijd voor deze factoren wordt gecorrigeerd om de invloed van de leerkracht, klas en school te bepalen. De criteriumvariabele motivatie wordt beïnvloed door twee leerkrachtonafhankelijke factoren: aanvankelijke motivatie en geslacht. De effecten zijn redelijk groot te noemen respectievelijk .29 en -.20. De effecten van biasveroorzakende variabelen voor de leerlingenschaal zijn klein. Een groei van 1 standaarddeviatie op de variabele aanvankelijke motivatie leidt tot .09 punt winst op de schaal 'motiveren en goed uitleggen' en tot .04 punt winst op de schaal 'frequentie belangrijke uitlegvaardigheden'. Het geslacht veroorzaakt bias in de schaal 'frequentie belangrijke uitlegvaardigheden'. Het effect is bescheiden. Meisjes beoordelen de leerkracht .11 punt lager dan jongens.

Bias veroorzaakt door factoren op klassenniveau

Tabel 10 bevat de modellen waarop op klassenniveau de variabelen klassengroote, sekse van de docent en het gemiddelde proefwerkcijfer zijn opgenomen. In 8 van de 49 klassen werd onderwezen door een vrouwelijke docent. De gemiddelde klassengroote was 25.7 leerlingen met een standaard deviatie van 3.60. Het gemiddelde proefwerkcijfer over 10 proefwerken per klas was 6.7 met een standaard deviatie van .40.

Tabel 10: Modellen met verklarende variabelen op leerlingen- en klassenniveau en de leerlingenschaal en criteriummaten als afhankelijke variabelen (49 klassen, 1084 leerlingen)

	Afhankelijke variabele			
	Leerlingenschaal		Criteriummaat	
	1	2	1	2
Vast deel				
Intercept	3.5 (.06)	3.7 (.07)	47.8 (.70)	3.3 (.05)
Leerlingenniveau				
Voorkennis				
- Rekenen	-.03 (.02)	-.02 (.03)	4.54 (.32)	.04 (.03)
- Taal	.02 (.02)	-.03 (.02)	1.42 (.28)	.00 (.03)
Algemene intelligentie				
- Ruimtelijk inzicht	.00 (.01)	.01 (.02)	2.43 (.26)	.02 (.03)
- Rekensnelheid	.00 (.01)	-.03 (.03)	2.56 (.26)	.02 (.03)
Aanvankelijke motivatie	.09 (.02)	.04 (.02)	0.97 (.25)	.29 (.02)
Geslacht	-.04 (.03)	-.11 (.04)	-0.10 (.50)	-.20 (.05)
Klassenniveau				
Klassengroote	.01 (.05)	-.02 (.06)	1.89 (0.60)	-.07(.04)
Gemiddeld proefwerkcijfer	.16 (.05)	-.01 (.06)	1.06 (0.60)	.11 (.04)
Sekse docent (0=man, 1=vrouw)	.20 (.14)	.09 (.17)	0.59 (1.60)	.28 (.11)
Variantiedeel				
Verklaard	11.9	02.2	56.3	19.8
Onverklaard				
% klasenniveau	28.7	27.4	08.7	08.1
% leerlingenniveau	59.4	70.4	35.0	72.1
Significantietest				
χ^2	1741.31	2234.35	7563.97	2524.55
Reductie	11.75	00.38	11.88	13.33
Vrijheidsgraden / p-waarden	3 / .01	3 / ns	3 / .01	3 / .01

Leerlingenschaal:
1 = Motiveren en goed uitleggen
2 = Frequentie belangrijke uitlegvaardigheden

Criteriummaten:
1 = wiskundescore
2 = motivatiescore

Uit tabel 10 blijkt dat de criteriumvariabelen 'wiskundescore' en 'motivatiescore' een betere modelfit krijgen door de klassenvariabelen in het model op te nemen. Hetzelfde geldt voor de leerlingsschaal 'motiveren en goed uitleggen'. Er is een geringe winst in verklaarde variantie voor deze afhankelijke variabelen (respectievelijk 2,8, 3,7 en 7,6%).

Als we kijken naar de criteriummaten dan zien we dat klassengrootte een positief, significant effect heeft op de wiskundescore (1,89). Grotere klassen scoren beter. Verder zijn er voor de wiskundescore geen significante effecten, ook niet het gemiddelde proefwerkcijfer. De motivatiescore wordt wel significant hoger door het gemiddelde proefwerkcijfer. Het effect is echter klein (.11). Opmerkelijk is de positieve invloed die vrouwelijke docenten hebben op de motivatie. Vergeleken met de mannen realiseren zij een winst van .28 (bij een gemiddelde motivatiescore van 3,3). Als we kijken naar de leerlingenschalen zien we een gunstiger oordeel op de schaal 'motiveren en goed uitleggen' op grond van het gemiddelde proefwerkcijfer. Het is een relatief klein effect (.16 bij een gemiddelde van 3,5). Als we de effecten van de criteriummaten en de leerlingenschalen vergelijken, valt ten eerste op dat de klassengrootte een belangrijke bias is voor de wiskundescore en het geslacht van de docent voor de motivatiescore. Daarnaast is er voor de criteriummaat motivatie en de leerlingsschaal 'motiveren en goed uitleggen' een lichte bias op grond van het gemiddelde proefwerkcijfer. Dat het gemiddelde proefwerkcijfer niet significant samenhangt met de wiskundescore zou kunnen duiden op het verschijnsel vijven en zessen: het verschijnsel dat leraren cijferen rond een bepaald gemiddelde ongeacht de 'objectieve' resultaten van de leerlingen zoals blijkend uit de wiskunde test (De Groot & Wijnen, 1983). Het is ook goed mogelijk dat de variantie rond de gemiddelde klassengemiddelden in dit geval te gering is om een effect te tonen.

Samenvattend kunnen we concluderen dat vergeleken met de criteriummaten slechts een gering deel van de totale variantie in de leerlingenschalen wordt verklaard door de mogelijk biasveroorzakende variabelen. Op leerlingenniveau zijn er kleine effecten die het oordeel over de leerkracht negatief of positief beïnvloeden. De oorspronkelijke motivatie beïnvloedt het oordeel over de leraar in positieve zin en meisjes neigen ernaar de leerkracht slechter te beoordelen. Op klassenniveau leidt een hoger gemiddeld proefwerkcijfer tot een hogere score op 'motiveren en goed uitleggen'. De gevonden biaseffecten zijn echter klein.

CONCLUSIES

In de literatuur worden tegenstrijdige meningen en onderzoeksresultaten aangetroffen over de vraag of leraarsgedrag goed kan worden beoordeeld door leerlingen. In dit onderzoek zijn in het eerste jaar van het voortgezet onderwijs binnen het vak wiskunde 49 leraren met gezamenlijk 1084 leerlingen geobserveerd. Deze leraren werden beoordeeld door externe beoordelaars en door hun leerlingen. Daarbij stelden we vragen met betrekking tot de dimensies die door leerlingen worden onderscheiden en met betrekking tot de validiteit en betrouwbaarheid van de gebruikte leerlingenschalen. Deze vragen kunnen nu worden beantwoord. Eerst moeten we echter een paar beperkingen van het onderzoek noemen die de conclusies hoe dan ook relativeren.

De eerste beperking is dat het design het niet toestaat de overeenstemming tussen beoordelaars (groepen van leerlingen) na te gaan. Zoals we bij de analyse van de onderzoeksvraag opmerkten zijn hiervoor binnen het gegeven design geen oplossingen aanwezig. Pas als we het gebruikte design uitbreiden en meerdere klassen dezelfde leraar laten beoordelen is het bepalen van een maat voor de interbeoordelaarsbetrouwbaarheid mogelijk. De tweede beperking die inherent is aan het gebruikte design, is dat we te maken hebben met een meerniveau probleem. Leerlingen zijn genest binnen leraren. Dit is problematisch omdat het grootste deel van de variantie in de leerlingenschalen onverklaard blijft. Van deze variantie bevindt zich een belangrijk deel op klassenniveau. Daardoor blijft het mogelijk dat niet het gedrag van de leraar,

maar een groeps sentiment de doorslag geeft bij de beoordeling. Een ander probleem is dat het grootste deel van de onverklaarde variantie zich op leerlingenniveau bevindt, zodat we eigenlijk ook nog onvoldoende weten wat op leerlingenniveau de factoren zijn die het oordeel over de leerkracht vormen. Met dit in gedachte formuleren we nu de resultaten van het onderzoek.

1. De eerste vraag had betrekking op de dimensies in het leraarsgedrag die door de leerlingen worden onderscheiden en de mogelijkheid daaruit schalen te construeren die voldoende betrouwbaar en valide zijn. Uit de factoranalyse blijkt dat het mogelijk is op grond van de leerlingenoordelen 6 schalen te onderscheiden. De populariteitshypothese die stelt dat leerlingen leraren op grond van hun populariteit unidimensioneel beoordelen wordt in dit onderzoek niet bevestigd. De belangrijkste leerlingenschaal bevat 4 aspecten die voor de leerling van belang zijn bij het beoordelen van de leerkracht: positieve houding tegenover de leerlingen, motiverend lesgeven, duidelijk uitleggen en orde kunnen houden. Factoranalyse op de data van de externe beoordelaars laat eveneens zien dat dit een belangrijke factor is. Naast motiveren en goed uitleggen beoordelen de leerlingen de leraar op: uitlegvaardigheden, huiswerkcontrole, individuele begeleiding, groepscontrole en meta-cognitie. Het blijkt dat de homogeniteit van de leerlingenschalen voldoende is als ze worden geaggregeerd op klassenniveau, de α 's lopen dan van .76 tot .85. De homogeniteit op leerlingenniveau varieert sterk per klas. Alleen de schalen 1 'motiveren en goed uitleggen' en 2 'frequentie van belangrijke uitlegvaardigheden' zijn op leerlingenniveau voldoende homogeen. Dit pleit ervoor leerlingenoordelen, in ieder geval voorlopig, te blijven aggregeren op klassenniveau als het de bedoeling is de kwaliteit van de instructie te meten.
2. De resultaten met betrekking tot de discriminant- en soortgenootvaliditeit zijn niet geheel eenduidig. De leerlingenschalen 1, 2, 4, 5 en 6 correleren onderling significant. We kunnen dus aannemen dat ze gedeeltelijke hetzelfde meten. Anderzijds zijn de correlaties zo laag dat dus aangenomen mag worden dat ze in belangrijke mate iets anders meten. Hetzelfde geldt voor de schalen van de externe beoordelaars. Hier was de gemiddelde correlatie tussen de schalen nog iets hoger dan bij de leerlingenschalen. Het samenvattende oordeel is dat de discriminantvaliditeit voldoende is, maar nog verbetering behoeft. Om de soortgenoot validiteit te bepalen zijn de leerlingenschalen en de schalen van de externe beoordelaars met elkaar gecorreleerd. De veronderstelde samenhang tussen de schalen werd bevestigd. Van de schalen waarvan werd verondersteld dat ze hetzelfde meten waren vijf van de zes correlaties significant. De correlaties waren echter niet hoog zodat de schalen die hetzelfde meten ook in belangrijke mate nog iets anders meten. De conclusie luidt dat de soortgenootvaliditeit voldoende is, maar ook nog verbetering behoeft. Het totaal beeld dat door de discriminant- en soortgenootvalidering is ontstaan luidt dat leerlingenoordelen kunnen worden gebruikt om de kwaliteit van het leraarsgedrag te meten. Om dat geheel valide te doen moet aan de hier gebruikte instrumenten echter nog het nodige worden verbeterd.
3. De predictieve validiteit van de leerlingenschalen en de schalen van de externe beoordelaars is op twee manieren na gegaan. Een keer met geaggregeerde gegevens op klassenniveau waarbij de schaalscores worden gecorreleerd met de gemiddelde wiskunde- en motivatiescores en een keer met behulp van tweenniveau-analyse waarbij de individuele leerlingenscores als criteriummaat dienden. Uit een eerste analyse met geaggregeerde criteriummaten bleek dat de leerlingenschalen, met uitzondering van de schaal 'huiswerkcontrole' positief en significant correleren met de motivatiescore. Twee van deze schalen correleren positief en significant met de wiskundescore. De schalen van de externe beoordelaars correleren geen van alle met de motivatiescore en slechts één met de wiskundescore. De conclusie luidt dat wanneer de geaggregeerde criteriummaten worden gehanteerd, de leerlingenschalen voldoende predictief valide zijn, de schalen van de externe beoordelaars echter niet. Uit de tweenniveau-analyse blijkt dat leerlingenschalen 'motiveren en goed uitleggen', 'groepscontrole' en 'huiswerkcontrole'

de wiskundescore significant voorspellen. Een leerlingenschaal is een significante voorspeller voor de motivatie, n.l. 'motiveren en goed uitleggen'. Van de externe beoordelaarschalen zijn er eveneens drie die significant samenhangen met de wiskundescore. Dat zijn 'duidelijk, taakgericht, positief doceren', 'beurten geven' en 'organizers toepassen'. Van de externe beoordelaars hangt de schaal 'opdrachten uitleggen' significant samen met de motivatiescore. Als we de resultaten van de tweenniveau-analyses overzien dan valt op dat de schalen die predictief valide zijn ten opzichte van de wiskundescore te maken met de duidelijkheid van de instructie. Duidelijk uitleggen en de helderheid waarmee opdrachten worden verstrekt, dragen significant bij aan het verhogen van de motivatie. Het zijn allen variabelen die een belangrijke plaats innemen in theorieën over effectief onderwijzen.

4. Met behulp van variantie-analyse is nagegaan of de leerlingenschalen verschillen tussen klassen, c.q. leraren, weergeven. Uit analyse blijkt dat de leerlingenschalen gevoelig genoeg zijn om verschil tussen klassen waar te nemen. De verschillen tussen klassen zijn bij alle schalen significant.

5. Bij de laatste vraag ging het erom of er andere factoren zijn dan het leerkrachtgedrag die het oordeel van de leerlingen over de leraar bepalen. Deze vraag is beantwoord met behulp van tweenniveau-analyse waarbij werd getracht de individuele scores op de leerlingenschalen te voorspellen vanuit mogelijk 'bias' veroorzakende variabelen. Hierbij beperkten we ons tot de schalen 'motiveren en goed uitleggen' en 'frequentie van belangrijke uitlegvaardigheden' omdat alleen deze schalen voldoende homogeen zijn op leerlingenniveau. Uit deze analyses blijkt dat slecht een gering deel van de variantie in de leerlingenschalen kan worden verklaard door mogelijk biasveroorzakende variabelen. Er zijn kleine effecten gevonden die doen vermoeden dat meer voorkennis leidt tot een negatiever oordeel over de leraar, dat gemotiveerde leerlingen hun leraren wat gunstiger zullen beoordelen, dat meisjes wat negatiever zullen zijn en dat leerlingen in klassen die een hoger proefwerkcijfer krijgen iets milder zullen zijn. De effecten zijn echter klein zodat de conclusie luidt dat het door ons geanalyseerde materiaal geen sterk versturende bias in de leerlingenoordelen laat zien.

Bovenstaande conclusies leiden tot het eindoordeel dat leerlingenoordelen verzameld met onze vragenlijst en geaggregeerd op klassenniveau zijn te gebruiken om de kwaliteit van het instructiegedrag te meten, maar dat de gebruikte instrumenten nog verbetering behoeven. Omdat het onderzoek is uitgevoerd in de eerste klas van het voortgezet onderwijs kan gezegd worden dat de centrale onderzoeksvraag of relatief jonge leerlingen in staat zijn leraarsgedrag betrouwbaar en valide te beoordelen, voorzichtig positief wordt beantwoord.

NOTEN

1. Dit onderzoek werd uitgevoerd binnen het School- en Instructie Effectiviteitsproject (SIE-project) dat werd gesubsidieerd door de Stichting Gedragswetenschappen (SGW) en NWO (projectnummer 575-90-801).

2. In onderzoek naar de beïnvloeding van het onderwijsklimaat is gebleken dat 'plezier in het vak wiskunde' verreweg de sterkste factor was bij het verklaren van het proefwerkcijfer (Houtveen e.a. 1993, Van der Grift e.a. 1997).

3. Het SIE-materiaal maakt het mogelijk een drienniveau analyse (leerling, klas en school) uit te voeren. De variantie in de motivatiescore is echter niet significant op schoolniveau. Dit geldt ook voor de later te bespreken leerlingenschalen. Alleen de variantie in de wiskundescore kent een significante variantiecomponent op schoolniveau. Scholen die expliciet aandacht besteden aan sturing van de inhoud van het onderwijs, leerlingengedrag, nascholing en functioneringsgesprekken presteren beter.

4. zie noot 3.

REFERENTIES

- Aleamoni, L.M. (1981). Student ratings of instruction. In: J. Millman (ed.), *Handbook of Teacher Evaluation*, 110-145. London: Sage.
- Borich, G.D. & S.K. Madden (1977). *Evaluating Classroom Instruction. A source book of instruments*. Massachusetts: Addison-Wesley.
- Brophy, J.E. & Th.L. Good (1986). Teacher behavior and student achievement. In: M.C. Wittrock (ed.), *Handbook of research on teaching* (third edition). New York: Macmillan Publications, 328-375.
- Corno, L. & R.E. Snow (1986). Adapting teaching to individual differences among learners. In: M.C. Wittrock (ed.), *Handbook of research on teaching* (third edition). New York: Macmillan Publications, 605-629.
- Creemers, B.P.M. (1994). *The effective Classroom*. London: Cassell.
- Fraser, B.J. (1991). Two decades of classroom environment research. In: B.J. Fraser & H.J. Walberg, *Educational Environments*. Oxford: Pergamon Press, 2-27.
- Fraser, B.J. (1995). Student perceptions of classrooms. In: L.W. Anderson, *International*
- Good, Th.L. & C. Mulryan (1990) Teacher ratings: A call for teacher control and self-evaluation. In: J. Millman & L. Darling-Hammond, *The new handbook of teacher evaluation*. London: Sage, 191-215.
- Groot, A.D. de (1961). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*. 's Gravenhage: Mouton & Co.
- Groot, A.D. & W.H.F.W. Wijnen (1983)10. Vijven en zessen. Cijfers en beslissingen: het selectieproces in ons onderwijs. Groningen: Wolters-Noordhoff.
- Grift, W. van de & Mulder, H.P. (1986). Het meten van leerlingpercepties van het onderwijs-klimaat in de wiskunde. In: Voogt, J.C. & Reints, A. (red.) *Naar beter onderwijs*. (p. 144-159). Tilburg: Zwijssen.
- Grift, W. van de, Houtveen, Th. & Vermeulen, C. (1997). Instructional climate in dutch secondary education. *Schooleffectiveness and schoolimprovement* (8) 4, 449-662
- Jansen, M. & K.J. Westerhof (1998). Interbeoordelaarsovereenstemming bij het meten van de kwaliteit van het onderwijzen met behulp van een beoordelingsschaal. *Tijdschrift voor Onderwijsresearch* (23), nr. 3 (ter perse).
- Heuvelmans, A.P.J.M., P.F. Sanders (1993). Beoordelaarsovereenstemming. In: T.J.H.M. Eggen, P.F. Sanders (eds.), *Psychometrie in de praktijk*. Arnhem: CITO Instituut voor toetsontwikkeling.
- Houtveen, Th. Vermeulen, C. & Grift, W. van de (1993). *Bouwstenen naar onderzoek naar de kwaliteit van scholen. Constructie van een instrument voor het meten van het onderwijsklimaat*. Utrecht: ISOR
- Jong, R. de, K.J. Westerhof (1998). Huiswerk en leerprestaties. *Pedagogische Studiën* (75), nr. 4, 262-276.
- Kremers, E.J.J. (1981). De wiskunde-attitudeschaal: een voorbeeld van een instrument voor het evalueren van affectieve doelstellingen. In: P. Weeda, *Aspecten van leerplanevaluatie. Onderwijskundige studies*. Malmberg.
- Marsh, H.W. (1987). Students' evaluations of University Teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational Research* (11) 3. Oxford: Pergamon.
- Otten, R. & M. Boekaerts (1990). Schoolvakbeleving bij geschiedenis, Nederlands en wiskunde in de brugklas. In: M. Boekaerts & E. de Corte, *Onderwijsleerprocessen*. Nijmegen: ITS.
- Rosenshine, B.V. (1979). Content, time and direct instruction. In: P.L. Peterson & H.J. Walberg, *Research on teaching: concepts, findings and implications*. Berkeley: McCutchan.
- Rosenshine, B. & R. Stevens (1986). Teaching functions. In: M.C. Wittrock (ed.), *Handbook of research on teaching* (third edition). New York: Macmillan Publications, 376-391.
- Schaffer, E.C. & P.S. Nessenrodt (1992). *The development and testing of the special strategies observation system*. San Francisco: AERA-paper.
- Scheerens, J. & R.J. Bosker (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Scheurich, V., B. Graham & M. Drolette (1983). Expected grades versus specific evaluations of the teacher as predictors of students' overall evaluation of the teacher. *Research in Higher Education* (19) 2, 159-173.
- Simons, P.R.J. (1995). Leerlingkenmerken. In: J. Lowyck & N. Verloop, *Onderwijskunde*. Groningen: Wolters-Noordhoff, 15-40.
- Stallings, J.A. & D. Stipek (1986). Research on early childhood and elementary school teaching programs. In: M.C. Wittrock (ed.), *Handbook of research on teaching* (third edition). New York: Macmillan Publications, 727-753.
- Teddlie, C., P.C. Kirby & S. Strinfield (1989). *Effective versus ineffective schools: observable differences in the classroom*. University of Chicago Press.
- Veenman, S., P. Lem, M. Voeten, G. Winkelmolen & H. Lassche (1986). *Onderwijs in combinatieklassen*. 's Gravenhage: SVO.
- Walberg, H.J. (1986). Synthesis on research on teaching. In: M.C. Wittrock (ed.), *Handbook of research on teaching* (third edition). New York: Macmillan Publications, 214 - 229.

- Walberg, H.J. (1991). Educational productivity and talent development. In: B.J. Fraser & H.J. Walberg, Educational Environments. Oxford: Pergamon Press, 93-109.
- Westerhof, K.J. (1989) Effectiviteit van leerkrachtgedrag. Een empirische studie naar de relatie tussen leerkrachtgedrag en leerwinst. Proefschrift. Groningen: Rijksuniversiteit.
- Westerhof, K.J. (1992). On the effectiveness of teaching: direct versus indirect teaching: In: Schooleffectiveness and School Improvement. An International Journal of Research, Policy and Practice, Vol 3, 3, 204-215.
- Wittrock, M.C. (1986). Students' thought processes. In: M.C. Wittrock (ed.), Handbook of research on teaching (third edition). New York: Macmillan Publications, 297-314.
- Wubbels, Th., H.A. Creton, J.M.G. Brekelmans, H.P. Hooymayers (1987) De perceptie van de leraar-leerlingrelatie; constructie en kenmerken van een instrument. Tijdschrift voor Onderwijsresearch, 12, 3-16.

BIJLAGE

Schalen en items van het leerlingen- en externe beoordelaarsinstrument.

De frequentie van leraargedrag beoordeeld door leerlingen 1 = minder dan 1 x per maand 2 = ongeveer 1 x per maand 3 = ongeveer 1 x per 2 weken 4 = ongeveer 1 x per week 5 = bijna elke les	Het oordeel van de leerlingen 1 = klopt helemaal niet 2 = klopt grotendeels niet 3 = klopt een beetje wel/niet 4 = klopt vrij goed 5 = klopt heel goed			
Schaal 'motiveren en goed uitleggen' (factor 1); 20 items; a=.96	M	SD	min	max
Hoe vaak:				
26* maakt de leraar zich boos op de klas	2.8	.9	1.3	4.7
29 maakt de leraar grapjes	2.8	.9	1.5	5.0
30 laat de leraar duidelijk blijken dat hij tevreden is als de klas goed zijn best doet				
32* is de leraar boos op jou	3.2	.7	1.8	4.5
27* geeft de leraar saai les	4.3	.4	3.5	4.9
In welke mate kloppen deze uitspraken?				
40 tijdens wiskundelessen verveel ik me nooit				
46 aan onze leraar kun je merken dat deze wiskunde een leuk vak vindt om te geven	3.0	.5	1.9	3.8
50 de leraar heeft mijn interesse in wiskunde vergroot				
56 onze leraar kan boeiend uitleggen	3.8	.5	2.6	4.7
57 wij hebben een leuke wiskunde leraar	2.7	.4	1.7	3.5
38 mijn leraar wil graag dat wij het zeggen als wij iets niet begrijpen	2.6	.6	1.4	3.9
41* ik begrijp meestal weinig van de uitleg van mijn leraar	3.4	.7	1.9	4.9
43 als je bij het maken van wiskunde-opgaven iets niet begrijpt, legt de leraar het duidelijk uit	4.1	.3	3.3	4.5
49* onze leraar legt nieuwe wiskunde onderwerpen te snel uit				
51 ik kan bij de leraar terecht als ik hulp nodig heb bij het maken van wiskunde-opgaven	3.7	.4	2.8	4.4
54* onze leraar wiskunde is moeilijk te verstaan				
39* de leraar maakt vaak negatieve opmerkingen over ons	4.2	.4	3.2	4.8
44* mijn leraar is te lang aan het woord bij het uitleggen van wiskunde	4.2	.6	1.9	4.9
49* onze leraar is streng met het beoordelen van proefwerken wiskunde	3.6	.6	2.1	4.8
55* onze leraar kan niet goed orde houden	3.1	.5	1.4	4.3
Totale schaal (verschil tussen klassen $F_{48,1030}=13,8;p<.01$)	3.8	.7	2.1	4.8
	3.5	.4	2.7	4.3

De kwaliteit van leerlingenoordelen over leraarsgedrag / 353

De frequentie van leraargedrag beoordeeld door leerlingen 1 = minder dan 1 x per maand 2 = ongeveer 1 x per maand 3 = ongeveer 1 x per 2 weken 4 = ongeveer 1 x per week 5 = bijna elke les	Het oordeel van de leerlingen 1 = klopt helemaal niet 2 = klopt grotendeels niet 3 = klopt een beetje wel/niet 4 = klopt vrij goed 5 = klopt heel goed			
Schaal 'uitlegvaardigheden' (factor 2); 11 items; $\alpha=.91$	M	SD	min	max
Hoe vaak:				
1 geeft de leraar een korte herhaling van wat in de vorige les is besproken	3.2	.7	1.6	4.4
2 begint de leraar met aan te geven waar de les over zal gaan	2.9	.8	1.3	4.7
3 gebruikt de leraar het bord bij het uitleggen van wiskunde	4.7	.3	3.5	5.0
4 gebruikt de leraar concrete voorbeelden om moeilijke stof duidelijk te maken	4.1	.5	2.5	4.7
5 laat de leraar op het bord zien hoe een nieuwe opgave moet worden gemaakt	3.8	.6	2.3	4.8
6 geeft de leraar uitleg over de stof aan de hele klas	4.5	.5	2.4	5.0
8 stelt de leraar vragen over de wiskundestof	3.2	.6	1.5	4.3
9 stelt de leraar een vraag aan jou om te kijken of je de opgave begrijpt	3.0	.6	1.4	4.1
12 geeft de leraar verschillende leerlingen een beurt tijdens uitleg	4.2	.6	1.9	4.9
16 geeft de leraar aan hoe het huiswerk gemaakt moet worden	3.0	.7	1.3	4.6
22 bespreekt de leraar fouten die in het huiswerk zijn gemaakt	4.0	.6	1.4	4.8
Totale schaal (verschil tussen klassen $F_{48,1030}=10,0;p<.01$)	3.7	.4	2.0	4.5
Schaal 'huiswerkcontrole' (factor 3); 2 items; $\alpha=.83$				
Hoe vaak:				
18 neemt de leraar het huiswerk mee om het na te kijken op fouten	1.2	.2	1.0	1.9
20 neemt de leraar het huiswerk mee om te kijken of het netjes is gemaakt	1.2	.2	1.0	1.6
Totale schaal (verschil tussen klassen $F_{48,1030}=1,9;p<.01$)	1.2	.2	1.0	1.8
Schaal 'individuele begeleiding' (factor 4); 4 items; $\alpha=.76$ ['llork4']				
Hoe vaak:				
11 helpt de leraar een leerling als die het antwoord op een vraag niet weet	4.4	.4	3.3	5.0
24 loopt de leraar door de klas als wij opgaven maken	3.9	.9	1.6	5.0
25 komt de leraar direct bij ons als wij hulp nodig hebben	4.0	.6	2.6	5.0
33 geeft de leraar jou hulp bij het maken van opgaven	3.0	.5	1.9	4.1
Totale schaal (verschil tussen klassen $F_{48,1030}=9,1;p<.01$)	3.9	.5	2.7	4.6
Schaal 'Groepscontrole' (factor 5); 2 items; $\alpha=.85$; ['llork5']				
Oordeel over de volgende stellingen:				
42 als je niet oplet tijdens de les, kan de leraar je met een enkele blik weer bij de les betrekken	3.2	.4	2.3	3.8
52 als wij niet opletten, heeft de leraar dat snel in de gaten	3.7	.4	2.8	4.6
Totale schaal (verschil tussen klassen $F_{48,1030}=3,2;p<.01$)	3.4	.4	2.5	4.1
Schaal 'metacognitie stimuleren' (factor 7); 3 items; $\alpha=.7$ ['llork7']				
Hoe vaak:				
13 laat de leraar een leerling voor het bord komen om uit te leggen hoe hij een som heeft gemaakt	2.1	.9	1.0	4.4
14 laat de leraar 2 of meer leerlingen vertellen hoe ze een opgave hebben gemaakt	2.9	.7	1.3	4.3
15 vraagt de leraar jou uit te leggen wat je fout hebt gedaan als hij ziet dat je in een opgave een fout hebt gemaakt	2.8	.5	1.7	3.9
Totale schaal (verschil tussen klassen $F_{48,1030}=7,7;p<.01$)	2.6	.6	1.5	4.8

Beoordeling leraren wiskunde door externe beoordelaars

Items geformuleerd als tegenpolen [eerste stelling helemaal van toepassing = 1; tweede contrasterende stelling = 4; indien één van beide stellingen enigszins van toepassing is, wordt een '2' of '3' gescoord]

Schaal * = gehercodeerd	M	SD	min	max
1. Inzichtbevordering (leraar gericht op (meta)cognitie)	2,9	0,5	1,8	3,8
- huiswerk inhoudelijk bespreken - alleen antwoorden controleren*	3,0	0,7	1,3	4,0
- leerlingen wijzen op fouten – fouten bespreken	3,1	0,7	1,9	4,0
- begrip controleren - geen controle begrip*	3,2	0,6	2,0	4,0
- behandeling methode oplossen probleem - methode zelf ontdekken*	2,9	0,7	1,5	4,0
- onderwijsleergesprek – monoloog*	2,7	0,8	1,0	4,0
- leerlingen moeten laten zien hoe ze aan een antwoord komen - leraar gaat zelden na hoe een antwoord tot stand is gekomen*	3,1	0,8	1,0	4,0
- signaleert alleen fouten - bij fouten wordt de leerling geholpen	3,1	0,6	1,5	4,0
- herhaalt uitleg als de stof niet is begrepen - gaat door ook als leerlingen de stof niet begrijpen*	3,3	0,4	2,3	4,0
- gericht op inventariseren goede antwoorden - laat oplossingswijzen vergelijken	2,2	0,8	1,0	4,0
- oplossingen van leerlingen vergelijken met experts - geen aandacht voor andere oplossingswijzen dan van leerlingen zelf*	2,5	0,8	1,0	4,0
- volstaat met controle hoe een probleem is aangepakt - laat verwoorden hoe een probleem is aangepakt	2,7	0,7	1,0	4,0
- leerlingen geregeld laten uitleggen wat fout is gegaan - volstaan met wijzen op fouten*	2,5	0,7	1,0	4,0
2. Duidelijke, taakgericht, positief doceren	3,2	0,4	1,7	4,0
- uitleg duidelijk en gestructureerd – uitleg is slecht te begrijpen*	3,5	0,6	2,0	4,0
- geeft de grote lijnen aan – verliest zich in details*	3,4	0,6	1,7	4,0
- leerlingen onderbreken de leraar alleen bij vingers – leerlingen onderbreken de leraar ten allen tijde*	3,0	0,7	1,0	4,0
- geeft niet inspirerend les - geeft enthousiast les	3,1	0,8	1,0	4,0
- veel leerlingen druk, afgeleid - werken rustig en taakgericht	2,5	1,0	1,0	4,0
- ontspannen sfeer tussen leraar en leerlingen – gespannen sfeer*	3,5	0,7	1,0	4,0
- prettige sfeer leerlingen - geen prettige sfeer tussen leerlingen onderling*	3,3	0,6	1,5	4,0
- weet waar iedere leerling mee bezig is - houdt leerlingen nauwelijks in de gaten*	3,3	0,6	2,0	4,0
- leerlingen weten precies wanneer er een nieuwe lesfase aanbreekt - leerlingen zijn onzeker over het begin van een nieuwe lesfase*	3,6	0,5	1,7	4,0
- lesovergangen verlopen vlot – lesovergangen verlopen traag*	3,3	0,7	1,0	4,0
- roept afgeleide leerlingen onmiddellijk tot de orde - laat leerlingen de vrijheid zelf te bepalen of ze bij de les zijn*	3,2	0,7	1,7	4,0
- les is afwisselend, levendig, vlot - les is traag en langdradig*	3,1	0,7	1,0	4,0
3. Kwaliteit beurten geven	3,4	0,4	2,4	4,0
- eist dat leerlingen volledig en correct antwoorden - neemt genoeg met vage, onvolledige antwoorden*	3,4	0,5	2,3	4,0
- het is leerlingen duidelijk hoe ze de opdrachten moeten maken - de leraar onderbreekt het werken aan opdrachten herhaaldelijk voor nieuwe aanwijzingen*	3,5	0,5	2,5	4,0
- geeft iedereen de beurt – alleen leerlingen die vingers opsteken krijgen de beurt*	3,3	0,6	2,0	4,0
- geeft leerlingen de tijd om na te denken alvorens te antwoorden - als een leerling het antwoord niet direct weet, krijgt een andere leerling de beurt*	3,2	0,6	2,0	4,0

De kwaliteit van leerlingenoordelen over leraarsgedrag / 355

Beoordeling leraren wiskunde door externe beoordelaars				
Items geformuleerd als tegenpolen [eerste stelling helemaal van toepassing = 1; tweede kontrasterende stelling = 4; indien één van beide stellingen enigszins van toepassing is, wordt een '2' of '3' gescoord]				
Schaal * = gehercodeerd	M	SD	min	max
4. Advance organisers toepassen	2,3	0,8	1,0	3,8
- herinnert leerlingen aan stof van de vorige les - leraar gaat door waar deze de vorige keer is gebleven*	2,4	1,0	1,0	4,0
- herhaalt stof van de vorige les – verwijst niet naar voorafgaand leren*	2,5	0,8	1,0	4,0
- maakt duidelijk welke stof tijdens de les geleerd moet zijn – zegt niets over wat er van leerlingen wordt verwacht*	2,1	0,9	1,0	3,7
5. Opdrachten uitleggen	2,1	0,8	1,0	4,0
- duidelijke uitleg hoe een opgave gemaakt moet worden - leerlingen moeten dit zelf uitzoeken*	2,1	0,9	1,0	4,0
- nauwkeurig aangeven hoe het huiswerk moet worden gemaakt - leraar geeft niet aan hoe het huiswerk moet worden gemaakt*	2,0	0,9	1,0	4,0

Mogelijkheden en Beperkingen van Statistische Uitzuivering binnen Schooleffectiviteitsonderzoek

Eric Maris (*Department of Mathematical Psychology*)
Nijmegen Institute for Cognition and Information (NICI) and National Institute for Educational Measurement (CITO), Nijmegen and Arnhem, The Netherlands, 3 mei 1999

The objective of this article is to give a didactic presentation of the contribution of statistics to school effectiveness research. Our advice for the methodology of school effectiveness research is not new. The contribution of this article is twofold: (a) a precise formulation of the research question and (b) an evaluation of the existing methods as an answer to this question. We start from a precise definition of what we want to compute (estimate): the school effect. To make an unbiased estimate of the school effect, one has to take into account differences between the schools in incoming students and differences between the schools in drop-outs. The starting-point for a possible solution for both of these problems is prediction on the basis of covariates. It is not guaranteed that this solution produces an unbiased estimate, but the solution will be better as more covariates are used as a basis for the prediction.

Onderzoek naar schooleffectiviteit staat weer in de aandacht. De rangorde naar effectiviteit van de Nederlandse scholen, in het najaar van 1997 gepubliceerd in het dagblad Trouw, heeft mede de discussie gestimuleerd. Onderzoek naar schooleffectiviteit is moeilijk. Daar zijn twee redenen voor:

1. Scholen hebben een verschillende instroom. Van een school met veel allochtone en sociaal zwakke leerlingen kan men niet verwachten dat zij even hoog scoort op het centrale eindexamen dan een "blanke" school met alleen leerlingen uit de hogere sociale klasse.
2. Leerlingen verdwijnen soms van school (de uitvallers). Van deze uitvallers weten we niet hoe ze zouden scoren op het eindexamen. Indien scholen nu van elkaar verschillen wat betreft de uitvallers (aantal en type), dan zijn de gemiddelde eindexamencijfers niet meer vergelijkbaar. Bijvoorbeeld, in de ene school kunnen het vooral de slechte leerlingen zijn die uitvallen, terwijl in de andere school zowel goede als slechte leerlingen uitvallen.

De statistiek kan ons helpen bij het onderzoek naar de schooleffectiviteit. Zij doet dat op verschillende manieren:

1. Door het probleem helder te formuleren. In het dagelijks taalgebruik is het begrip *schooleffectiviteit* niet altijd helder omschreven. Het gebruik van statistische methoden vereist dat we precies omschrijven wat we met dit begrip bedoelen. Verder in dit artikel zullen we het onderzoek naar schooleffectiviteit statistisch vertalen in het schatten van *het school-effect*.
2. Door ons technieken aan te reiken waarmee we het schooleffect kunnen schatten en toetsen. Deze technieken zijn speciaal ontwikkeld om een oplossing te bieden voor problemen zoals een verschillende instroom en een verschillend aantal/type uitvallers.
3. Door ons een theorie te geven die ons zegt wanneer de gebruikte technieken een zuivere schatting geven van het schooleffect.

Het doel van dit artikel is het geven van een inleiding in hoe men gebruik kan maken van statistiek in het onderzoek naar schooleffectiviteit. De statistiek die we zullen voorstellen is anders dan de klassieke hypothesetoetsing ("Is dit verschil significant?") waar de meeste lezers wel vertrouwd mee zullen zijn. In dit artikel gaat onze aandacht uit naar het *zuiver schatten van het schooleffect*.

1. HET SCHOOLEFFECT

Onze aandacht gaat uit naar het verschil in effectiviteit van twee scholen, A en B. Deze situatie is eenvoudiger dan wanneer we de effectiviteit van een groot aantal scholen met elkaar zouden vergelijken. Onze beperking tot slechts twee scholen is ingegeven uit didactische overwegingen; de technieken die we zullen voorstellen zijn eenvoudig uit te breiden naar de situatie met een groot aantal scholen.

Wanneer we spreken over een schooleffect, dan gaat het impliciet altijd over een *vergelijking* van scholen: Als er maar één school is, dan is er ook geen schooleffect, want die ene school kunnen we nergens mee vergelijken. We gaan er nu vanuit dat elke leerling gekenmerkt wordt door *twee* eindexamencijfers: het cijfer dat hij zou halen als hij naar school A zou gaan en het cijfer dat hij zou halen als hij naar school B zou gaan. Het eerste cijfer duiden we aan met Y_A , en de tweede met Y_B . Op het eerste zicht is het een beetje vreemd dat we voor iedere leerling een score hebben op *beide* variabelen, Y_A en Y_B , ook al is er maar één school waar ze effectief naar toe gaan. Het bestaan van deze twee scores per leerling is echter de basis van het schooleffect. Het schooleffect is immers het verschil ($Y_A - Y_B$): Als de score in school A groter is dan die in school B, dan is dit verschil positief, en in het andere geval is het negatief. Het feit dat voor elke leerling slechts ofwel Y_A , ofwel Y_B geobserveerd wordt, maakt dat het schooleffect niet rechtstreeks observeerbaar is. Als het schooleffect wél rechtstreeks observeerbaar zou zijn (door leerlingen te "brainwashen", ze zes jaar jonger te maken en naar de andere school te sturen), dan zou het bepalen van het schooleffect een simpele zaak zijn. Omdat dit niet kan, moeten we het schooleffect *schatten*. De technieken die hiervoor bestaan leveren, onder bepaalde voorwaarden, goede schatters (zie verder).

Het hoeft niet zo te zijn dat ($Y_A - Y_B$) hetzelfde is voor alle leerlingen: Voor sommige leerlingen kan school A de betere school zijn, terwijl voor andere leerlingen school B de betere is. Het *gemiddelde* van de verschilvariabele ($Y_A - Y_B$) geeft een aanduiding van het schooleffect voor alle leerlingen samen. Dit gemiddelde duiden we aan met ($\bar{Y}_A - \bar{Y}_B$). Wanneer we verder over *het schooleffect* spreken, dan bedoelen we ($\bar{Y}_A - \bar{Y}_B$).

Deze manier van definiëren van een effect, vertrekkende van twee *mogelijke* observaties, werd in de gedragswetenschappen geïntroduceerd door Rubin (1974). Vanuit dezelfde gezichtshoek, verschenen er later nog artikelen van Holland (1986) en Rubin (1977, 1978).

Het schooleffect is niet rechtstreeks observeerbaar, gewoon omdat iedere leerling ofwel naar school A, ofwel naar school B gaat. In plaats van Y_A , observeren we het gemiddelde van Y_A voor de leerlingen die naar school A gaan. Dit gemiddelde duiden we aan met $\bar{Y}_A | A$. En in plaats van Y_B , observeren we het gemiddelde van Y_B voor de leerlingen die naar school B gaan. Dit gemiddelde duiden we aan met $\bar{Y}_B | B$.

We kunnen er niet zomaar vanuit gaan dat ($\bar{Y}_A | A - \bar{Y}_B | B$) gelijk is aan ($\bar{Y}_A - \bar{Y}_B$). Bijvoorbeeld, als de goede leerlingen vooral in school A terecht komen, dan zal $\bar{Y}_A | A$ groter zijn dan \bar{Y}_A , want \bar{Y}_A is het gemiddelde over *alle* leerlingen, en daar zitten ook de minder goede bij. Als de goede leerlingen vooral in school A terecht komen, dan komen de *minder goede* leerlingen vooral in school B terecht. In dit geval zal $\bar{Y}_B | B$ *kleiner* zijn dan \bar{Y}_B , want \bar{Y}_B is het gemiddelde over *alle* leerlingen, en daar zitten ook de goede bij.

2. DE ZEGENING VAN DE ZUIVER TOEVALLIGE TOEWIJZING

Zuiver toevallige toewijzing betekent dat de leerlingen (en hun ouders) niet zelf mogen kiezen uit school A en school B, maar dat de onderzoeker dit bepaalt, en dit doet o.b.v. *toeval*. Concreet, kan de onderzoeker dit doen door voor elke leerling een muntje op te gooien, en de leerling naar school A te sturen als hij "kruis" gooit, en naar school B als hij "munt" gooit. Hierbij is het niet van belang dat de kans om naar school A gestuurd te worden precies gelijk is aan de kans om naar school B gestuurd te worden (nl., 1/2). Als school A twee keer zo groot is als school B, dan kan men leerlingen met kans 2/3 naar school A sturen en met kans 1/3 naar school B. (Dit kan bijvoorbeeld met een dobbelsteen, waarbij men een leerling naar school A stuurt als er meer dan twee ogen boven liggen, en naar school B als dat niet zó is.) Het belangrijke punt hier, is dat de toewijzing volledig van het toeval afhangt, en van niets anders. De kans waarmee deze toewijzing gebeurt, doet er niet toe.

Als de toewijzing zuiver toevallig is, dan zal $\bar{Y}_A | A$ ongeveer gelijk zijn aan \bar{Y}_A , en zal $\bar{Y}_B | B$ ongeveer gelijk zijn aan \bar{Y}_B . De reden waarom ze niet *precies* gelijk zijn, is dat $\bar{Y}_A | A$ en $\bar{Y}_B | B$ bepaald werden op slechts een *deel* van de personen waarop \bar{Y}_A en \bar{Y}_B gebaseerd zijn. Als het aantal personen waarop $\bar{Y}_A | A$ en $\bar{Y}_B | B$ bepaald werden groot is (bijv., 100), dan zal het verschil met, respectievelijk, \bar{Y}_A en \bar{Y}_B altijd klein zijn.

Zuiver toevallige toewijzing zorgt ervoor dat de eindexamencijfers op basis waarvan men $\bar{Y}_A | A$ berekent een zuiver toevallige *trekking* zijn uit alle eindexamencijfers die men in school A zou kunnen observeren. Dat het hier gaat om een zuiver toevallige trekking van eindexamencijfers, impliceert dat de leerlingen op basis waarvan men $\bar{Y}_A | A$ berekent niet systematisch verschillen van de leerlingen in de volledige groep. Hetzelfde geldt voor de leerlingen op basis waarvan men $\bar{Y}_B | B$ berekent. Hierdoor zullen $\bar{Y}_A | A$ en $\bar{Y}_B | B$ slechts *toevallige* verschillen vertonen met \bar{Y}_A en \bar{Y}_B . Hieruit volgt dat $(\bar{Y}_A | A - \bar{Y}_B | B)$ een zuivere schatter is van $(\bar{Y}_A - \bar{Y}_B)$, het schooleffect.

Deze zegening van de zuiver toevallige toewijzing helpt ons echter niet veel in de praktijk, want het zijn altijd de leerlingen zelf en hun ouders die bepalen naar welke school ze gaan. En niets belet hen om dat niet zuiver toevallig te doen!

3. NIET-EQUIVALENTE GROEPEN EN VOORSPELLING OP BASIS VAN COVARIATEN

We spreken van *niet-equivalente groepen* als het mogelijk is dat er systematische verschillen zijn tussen de leerlingen in de twee scholen. Dit gevaar bestaat altijd als de leerlingen en hun ouders zelf hun school mogen kiezen. Er bestaat een techniek waarmee men probeert om in deze situatie toch een zuivere schatting te maken van het schooleffect. Deze techniek is gebaseerd op het voorspellen van het eindexamencijfer dat men *geobserveerd zou hebben* als de leerling naar de andere school gegaan zou zijn.

Stel bijvoorbeeld dat men van elke leerling de volgende informatie heeft: (a) zijn/haar score op de Eindtoets Basisonderwijs, (b) zijn/haar score op een IQ-test, en (c) het opleidingsniveau van de ouders. Dergelijke variabelen worden *covariaten* genoemd. Men kan een vergelijking opstellen waarmee men, o.b.v. deze covariaten, de eindexamencijfers kan voorspellen, één voor elk van de twee scholen. Op precies dezelfde manier als in een regressie analyse berekent men een formule waarmee men de eindexamencijfers zo goed mogelijk kan voorspellen o.b.v. de covariaten (de regressievergelijking). De covariaten duiden we aan met Z en de regressievergelijkingen met $\bar{Y}_A(Z)$ en $\bar{Y}_B(Z)$, voor, respectievelijk, school A en school B. Deze regressievergelijkingen worden bepaald met afzonderlijke regressie analyses op de gegevens (covariaten en eindexamencijfers) van, respectievelijk, school A en school B¹. Met deze formules beschikken we over een middel om voor elke leerling het verschil $(Y_A - Y_B)$ te *schatten*. Voor de leerlingen die naar school A gaan, *observeren* we Y_A en *voorspellen* we Y_B

met $\hat{Y}_B(Z)$. Het geschatte verschil is dan $(Y_A - \hat{Y}_B(Z))$. En voor de leerlingen die naar school B gaan, *observeren* we Y_B en *voorspellen* we \hat{Y}_A met $\hat{Y}_A(Z)$. Het geschatte verschil is dan $(\hat{Y}_A(Z) - Y_B)$. We kunnen nu een schatting maken van $(\hat{Y}_A - \hat{Y}_B)$, het schooleffect, door het gemiddelde te nemen van al deze geschatte verschillen $((\hat{Y}_A - \hat{Y}_B(Z))$ voor de leerlingen van school A en $(\hat{Y}_A(Z) - Y_B)$ voor leerlingen van school B). Dit gemiddelde noemen we de *voor-covariaten-aangepaste (CA) schatter* van het schooleffect, kortweg aangeduid als *CA-schatter*. Een andere naam die men wel eens tegenkomt, is *voor-covariaten uitgezuiverde schatter*.

De CA-schatter wordt ook berekend in de bekende techniek *covariantie analyse*, die je kunt uitvoeren met vrijwel elk data-analyse pakket (SPSS, SAS, . . .). Naast het berekenen van de CA-schatter, kan men met de covariantie analyse ook statistisch toetsen of de CA-schatter significant verschilt van 0.

De cruciale vraag is of de CA-schatter een zuivere schatter is van het schooleffect. Onder *zuivere schatter* verstaan we dat hij geen systematische onder- of overschatting geeft van het schooleffect. Dit betekent dat de kans dat de CA-schatter groter is dan het schooleffect, even groot is dan de kans dat hij kleiner is. Welnu, men kan aantonen dat de CA-schatter slechts dan een zuivere schatter is van het schooleffect als, gegeven de covariaten op basis waarvan men voorspelt, *geen enkele andere variabele nog een systematisch verband heeft met de schoolkeuze*. Hiermee bedoelen we dat, binnen een groep leerlingen die allemaal dezelfde score hebben op de covariaten, er geen verband bestaat tussen schoolkeuze en enige andere variabele. Bijvoorbeeld, stel dat we kijken naar groepen leerlingen die allemaal dezelfde score hebben op de drie covariaten in ons voorbeeld (bijv., 40 op de Eindtoets Basisonderwijs, 115 op de IQ-test, en 5 voor het opleidingsniveau van de ouders). Als er *binnen deze groepen* een verschil is tussen het aantal allochtonen op school A en school B, dan heeft de variabele *allochtoon-autochtoon* een systematisch verband met de schoolkeuze dat niet verklaard wordt door de covariaten. Hier is dus niet voldaan aan de voorwaarde waaronder de CA-schatter een zuivere schatter is van het schooleffect. Het bewijs van het feit dat de bovenstaande voorwaarde voldoende is voor het zuiver zijn van de CA-schatter, werd gegeven door Rubin (1977), en een meer toegankelijke versie ervan kun je vinden bij Maris (1998).

Het grote probleem met deze voorwaarde voor de zuiverheid van de CA-schatter, is dat je er in de praktijk nooit honderd procent zeker van kunt zijn dat eraan voldaan is. Concreet, je bent nooit helemaal zeker of er buiten de covariaten op basis waarvan je voorspelt, niet nog andere variabelen een rol spelen. Het maximaal haalbare, is alle beschikbare variabelen als covariaat opnemen. Een dergelijke werkwijze is verdedigbaar als (a) alle covariaten verzaamd werden voorafgaand aan de schoolkeuze, en (b) het aantal covariaten veel kleiner is dan het aantal leerlingen (maximaal 1 covariaat per 50 leerlingen). Het blijft echter zó dat de resulterende CA-schatter niet gegarandeerd zuiver is, en hierdoor blijft er ruimte voor kritiek vanwege schooldirecties die ontevreden zijn over het slechte "rapportcijfer" van hun school.

Als zekerheid niet haalbaar is, dan stelt zich de vraag hoe groot de vertekening (onzuiverheid in de CA-schatter) kan zijn als een bepaalde covariaat niet opgenomen wordt. In dit geval spreekt men van een *verborgen covariaat*. Zonder bijkomende veronderstellingen, is het niet mogelijk om de invloed van een verborgen covariaat te kwantificeren. Als men daarentegen bereid is om te vertrekken van een bepaald maximaal verband tussen deze verborgen covariaat en de onafhankelijke variabele (in ons geval, school A of B), kan men deze invloed *wel* kwantificeren. Hoe dit precies in zijn werk gaat, valt buiten het bestek van dit artikel. Men vindt hierover meer in hoofdstuk 6 van het boek "Observational Studies" van Paul Rosenbaum (1995).

Zonder bijkomende veronderstellingen blijft men echter in het ongewisse over de kwaliteit van de CA-schatter. Een manier om om te gaan met deze onzekerheid, is de CA-schatter als *voorlopig* te beschouwen. Zolang er geen onderzoek verschijnt waarin *meer* covariaten opgenomen worden dan in het huidige onderzoek, neemt men de huidige CA-schatter voorlopig voor waar aan. Een onderzoeker of onderzoekster die denkt dat deze CA-schatter onzuiver is omdat er geen rekening gehouden werd met een bepaalde covariaat, moet aantonen dat dit ook werkelijk het geval is. Als hij of zij met deze extra covariaat een andere CA-schatter vindt, dan nemen we deze laatste schatter voorlopig voor waar aan.

4. NIET-EQUIVALENTE GROEPEN EN VERSCHILSCORES

Naast de CA-schatter, kan men ook gebruik maken van *verschilcores* om het schooleffect te schatten. Voor elke leerling is er een voor- en een nameting: De voormeting gebeurt bij aanvang van de opleiding en de nameting op het einde. In ons geval is de nameting het centrale eindexamen. Als voormeting kiest men een instrument waarmee men dezelfde vaardigheid meet als het eindexamen en dat bovendien op dezelfde schaal doet. De score op de voormeting duiden we aan met X , en het gemiddelde van deze variabele in school A en B, duiden we aan met, respectievelijk, $\bar{X}_A | A$ en $\bar{X}_B | B$. De idee is nu om gebruik te maken van deze voormeting om te corrigeren voor de verschillen in initiële vaardigheid tussen de twee scholen: We kijken niet gewoon of de leerlingen in school A een hoger eindexamencijfer hebben dan de leerlingen in school B, maar wel of ze meer *vooruit gegaan* zijn in vergelijking met de voormeting. Dus, in plaats van $(\bar{Y}_A | A - \bar{Y}_B | B)$, gebruiken we nu $[\bar{Y}_A | A - \bar{X}_A | A] - [\bar{Y}_B | B - \bar{X}_B | B]$ als schatter van het schooleffect. Deze laatste schatter noemen we de *verschilcore-schatter*.

Het is een punt van discussie hoe we kunnen bepalen of een voormeting dezelfde vaardigheid meet als de nameting (het eindexamen). Wat men dient te verstaan onder *dezelfde vaardigheid meten*, verschilt naargelang men de *Klassieke Test Theorie* (KTT) of de *Item Respons Theorie* (IRT) aanhangt. In de KTT, meten twee tests dezelfde vaardigheid als ze dezelfde ware scores hebben. Dit kan onderzocht worden als men de betrouwbaarheden van deze tests kent alsook hun onderlinge correlatie. In de IRT, meten twee tests dezelfde vaardigheid als de items van beide tests op één en dezelfde dimensie liggen. Het beantwoorden van deze vraag kan m.b.v. een zogenaamd *unidimensionaliteitsonderzoek*, wat een standaard onderdeel is van een IRT-analyse.

Het al-dan-niet zuiver zijn van de verschilcore-schatter blijkt af te hangen van een, op het eerste zicht, vrij abstracte voorwaarde: Blijft het verschil tussen voor- en nameting in de twee groepen $(\bar{Y}_A | A - \bar{X}_A | A)$ en $(\bar{Y}_B | B - \bar{X}_B | B)$ hetzelfde als de leerlingen naar de andere school zouden gaan? Indien dit verschil inderdaad hetzelfde blijft, dan zou men in plaats van $(\bar{Y}_A | A - \bar{X}_A | A)$ en $(\bar{Y}_B | B - \bar{X}_B | B)$, de volgende gemiddelde verschilcores observeren: $(\bar{Y}_B | A - \bar{X}_B | A)$ en $(\bar{Y}_A | B - \bar{X}_A | B)$. Waarom deze voorwaarde voldoende is voor het zuiver zijn van de verschilcore-schatter wordt aangetoond door Maris (1998).

Net zoals bij de CA-schatter, hebben we ook hier te maken met een voorwaarde voor zuiverheid die niet toetsbaar is, en waarvoor we dus nooit echt hard kunnen maken of hij opgaat of niet. Een ontevreden schooldirectie kan dus ook de verschilcore-schatter op de korrel nemen.

Een opvallend punt is dat de vraag of we dezelfde vaardigheid meten met de twee tests strikt genomen niet van belang is voor de vraag of de verschilcore-schatter een zuivere schatter is van het schooleffect. Anders uitgedrukt, voor de bovenstaande voorwaarde voor de zuiverheid van de verschilcore-schatter doet het er niet toe *wat* er precies gemeten wordt bij de voor- en de nameting. Dit impliceert dat we onze kansen op een zuivere schatting niet kunnen verhogen door eerst een KTT- of IRT-analyse uit te voeren waarmee we eventueel kunnen aantonen dat de twee tests hetzelfde meten. Dit is anders dan bij de CA-schatter: Daar kunnen we, door eigen toedoen, ten minste de *kans* op onzuiverheid verlagen, namelijk door meer covariaten op te nemen.

5. UITVALLERS EN NOG EENS VOORSPELLING OP BASIS VAN COVARIATEN

Het tweede probleem in het onderzoek naar schooleffectiviteit, naast de verschillende instroom (de niet-equivalente groepen), is dat leerlingen soms van school verdwijnen (de uitvallers). Met het oog op een zuivere schatting van het schooleffect, mag men die uitvallers zeker niet buiten beschouwing laten. Het bekende voorbeeld is dat van een school die alle slechte leer-

lingen aan de deur zet en daardoor zijn gemiddelde eindexamencijfer kunstmatig opschroeft. Het omgekeerde fenomeen is natuurlijk ook mogelijk: Een zeer effectieve school met een slecht imago ziet zijn beste leerlingen verdwijnen naar scholen met een beter imago.

Een mogelijke oplossing voor dit probleem is het voorspellen van het eindexamencijfer op basis van covariaten. Voor de uitvallers uit school A, voorspellen we Y_A op basis van de covariaten Z met de formule $\hat{Y}_A(Z)$, en voor de uitvallers uit school B, voorspellen we Y_B op basis van de covariaten Z met de formule $\hat{Y}_B(Z)$. Het bepalen van de voorspellingsformules gebeurt met regressie analyse, net zoals bij de CA-schatting. Merk op dat we hier een voorspelling maken van het eindexamencijfer dat men gehaald zou hebben in de school die men verlaten heeft. Dit is anders dan bij de CA-schatting, waar we een voorspelling maakten van het eindexamencijfer dat men gehaald zou hebben als men voor de andere school zou hebben gekozen. Het ligt echter voor de hand om beide voorspellingen te combineren in één schatting van het schooleffect. Deze schatting is gebaseerd op de volgende berekeningen:

- Voor de niet-uitvallers in school A observeert men Y_A , voorspelt men Y_B met de formule $\hat{Y}_B(Z)$, en berekent men het verschil $(Y_A - \hat{Y}_B(Z))$.
- Voor de uitvallers in school A voorspelt men Y_A met de formule $\hat{Y}_A(Z)$, voorspelt men Y_B met de formule $\hat{Y}_B(Z)$, en berekent men het verschil $(\hat{Y}_A(Z) - \hat{Y}_B(Z))$.
- Voor de niet-uitvallers in school B voorspelt men Y_A met de formule $\hat{Y}_A(Z)$, observeert men Y_B , en berekent men het verschil $(\hat{Y}_A(Z) - Y_B)$.
- Voor de uitvallers in school B voorspelt men Y_A met de formule $\hat{Y}_A(Z)$, voorspelt men Y_B met de formule $\hat{Y}_B(Z)$, en berekent men het verschil $(\hat{Y}_A(Z) - \hat{Y}_B(Z))$.

De schatting van het schooleffect is het gemiddelde (over alle leerlingen) van deze verschillen. We noemen deze schatting de *voor-covariaten- en-uitvallers-aangepaste (CUA) schatting* van het schooleffect, kortweg aangeduid als *CUA-schatting*.

De belangrijke vraag is opnieuw of de CUA-schatting een zuivere schatting is van het schooleffect (geen systematische onder- of overschatting van het schooleffect). Welnu, men kan aantonen dat, onder twee voorwaarden, de CUA-schatting een zuivere schatting is van het schooleffect. De eerste voorwaarde is dezelfde als voor de CA-schatting: Buiten de covariaten op basis waarvan men een voorspelling maakt van de eindexamenuitslag die men in de andere school gehaald zou hebben, heeft geen enkele andere variabele nog een systematisch verband met de schoolkeuze. De tweede voorwaarde is specifiek voor de CUA-schatting, en hij heeft te maken met het verband tussen het al-dan-niet uitvallen, de covariaten, en het eindexamencijfer. De voorwaarde bestaat erin dat, gegeven de scores op de covariaten, er geen verband bestaat tussen het eindexamencijfer en het al-dan-niet uitvallen. Dit betekent dat, binnen een groep leerlingen die allemaal dezelfde score hebben op de covariaten, er geen verschil is tussen het gemiddelde eindexamencijfer van de uitvallers en dat van de niet-uitvallers. Deze laatste voorwaarde wordt in de statistische literatuur aangeduid als *missing-at-random*. Deze term werd geïntroduceerd door Rubin (1976). In het boek van Little en Rubin (1987) wordt een uitgebreide beschrijving gegeven van verschillende methoden om met ontbrekende gegevens om te gaan (waaronder de methode die hier beschreven werd) en van de voorwaarden waaronder deze methoden tot zuivere schattingen leiden.

In deze tweede voorwaarde gaat het om de eindexamencijfers die de uitvallers gehaald zouden hebben als ze niet waren uitgevallen. Het gaat hier dus over niet-observeerbare scores. Daarom is het in de praktijk onmogelijk om echt hard te maken dat aan deze voorwaarde voldaan is. Concreet, je bent nooit helemaal zeker of, gegeven de covariaten op basis waarvan je voorspelt, de uitvallers en de niet-uitvallers gemiddeld dezelfde eindexamencijfers behalen. Het maximaal haalbare, is de eindexamencijfers zo goed mogelijk proberen te voorspel-

len door alle beschikbare variabelen als covariaat op te nemen. Immers, naarmate je de eind-examencijfers beter kunt voorspellen, is er minder ruimte voor verschillen tussen uitvallers en niet-uitvallers *binnen een groep die allemaal dezelfde score hebben op de covariaten*. Een dergelijke werkwijze is verdedigbaar als het aantal covariaten veel kleiner is dan het aantal leerlingen. Het blijft echter zó dat de resulterende CUA-schatter niet gegarandeerd zuiver is, en hierdoor blijft er ruimte voor kritiek.

6. EN WAT MOETEN WE NU DOEN?

Het doel van dit artikel was te laten zien wat de statistiek ons te bieden heeft voor het onderzoek naar schooleffectiviteit. We kunnen het volgende besluiten:

1. Omdat we de leerlingen niet op zuiver toevallige wijze kunnen toewijzen aan de scholen, moeten we van een andere schatter gebruik maken dan het gewone verschil tussen de gemiddelde eindexamencijfers.
2. Voor de CA-schatter en de verschilscore-schatter kunnen we nooit echt hard maken dat ze zuiver zijn, omdat dit afhangt van niet-toetsbare voorwaarden.
3. Het wordt meer aannemelijk dat voldaan is aan de voorwaarde voor zuiverheid van de CA-schatter naarmate we voorspellen o.b.v. meer covariaten.
4. Men kan rekening houden met de storende invloed van uitvallers door de ontbrekende eind-examencijfers van de uitvallers te voorspellen. Dit resulteert in de CUA-schatter.
5. We kunnen nooit echt hard maken dat de CUA-schatter zuiver is, omdat dit afhangt van niet-toetsbare voorwaarden.
6. Het wordt meer aannemelijk dat voldaan is aan de voorwaarde voor zuiverheid van de CUA-schatter naarmate we voorspellen o.b.v. meer covariaten.

De statistiek leert ons dat we bescheiden moeten zijn: We kunnen niet echt hard maken dat onze schatters zuiver zijn. Ze leert ons echter ook dat we het beter zullen doen naarmate we ijveriger zijn: Naarmate we meer covariaten betrekken bij onze voorspelling van ontbrekende scores, is het minder aannemelijk dat de CA- en de CUA-schatter erg onzuiver is. Een belangrijke groep covariaten zijn in principe beschikbaar (score op de Eindtoets Basisonderwijs, opleidingsniveau van de ouders, allochtoon-autochtoon, . . .). Misschien krijgen we de minister ooit nog zover dat we niet alleen de eindexamencijfers krijgen maar ook deze achtergrondgegevens.

REFERENTIES

- Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, 78, 945-970.
- Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- Maris, E. (1998). Covariance adjustment versus gain scores - Revisited. *Psychological Methods*, 3, 309-327.
- Rosenbaum, P. R. (1995). *Observational Studies*. New York: Springer.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 3, 581-592.
- Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-26.
- Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, 6, 34-58.

NOTEN

1. Er zijn hier twee varianten mogelijk: (a) regressievergelijkingen waarin zowel het intercept als de regressiecoëfficiënten kunnen verschillen over de scholen en (b) regressievergelijkingen waarin alleen het intercept kan verschillen over de scholen en de regressiecoëfficiënten gelijk blijven. De tweede variant wordt het meest toegepast. Hij wordt vaak aangeduid als *covariantie analyse met homogene hellingsgraden*. De eerste variant wordt vaak aangeduid als *covariantie analyse met heterogene hellingsgraden*.

Boekbesprekingen

Jeroen Onstenk

Lerend leren werken. Brede vakbekwaamheid en de integratie van leren, werken en innoveren
 Delft: Eburon, 1997, ISBN 90-5166-597-0

Inhoud

"Technologische en marktontwikkelingen alsmede de opkomst van nieuwe organisatie- en productieconcepten leiden in veel beroepen en functies tot het complexer en veelzijdiger worden van de kernproblemen van het werk, en daarmee van de eisen aan vakbekwaamheid. Leren op de werkplek kan een essentiële bijdrage leveren aan het verwerven van die vakbekwaamheid. Het speelt tevens een belangrijke rol in het innoveren van arbeid en organisatie" (p. 1). Op deze wijze geeft de auteur zelf een kernachtige typering van de thematiek die in zijn proefschrift centraal staat. Het proefschrift gaat over het leren op de werkplek van werknemers in arbeidsorganisaties en van leerlingen uit het MBO in het kader van stages. In hoofdstuk 1 ("*Lerend leren werken als innovatie*") worden de achtergronden van de toenemende belangstelling van leren op de werkplek en beroepspraktijkvorming verder belicht. Tevens wordt de probleemstelling uitgewerkt in vijf onderzoeksvragen: Wat zijn de gevolgen van innovatie van arbeid en organisatie voor vakbekwaamheid en wat moet worden verstaan onder brede vakbekwaamheid? Welke bijdrage kan leren op de werkplek leveren aan competentie-ontwikkeling? Hoe kan het "leerpotentieel" van arbeidssituaties worden omschreven? Hoe verloopt het proces van de ontwikkeling van brede vakbekwaamheid, enerzijds bij werknemers in innovatieve arbeidsorganisaties, anderzijds bij leerlingen in het kader van de praktijkcomponent van beroepsopleidingen? Om antwoorden te vinden op de gestelde vragen gaat Onstenk te rade bij zowel de arbeidssociologie, de leerpsychologie, de opleidingskunde als de antropologie. De activiteitstheorie moet daarbij als bindmiddel dienen: "Arbeid en leren worden in deze theorie opgevat als constructieve activiteiten, waarbij de arbeidende/lerende ingrijpt op zijn omgeving en deze mede vorm geeft" (p. 24). Het accent ligt op het leren en opleiden van mensen met het oog op de uitoefening van functies op laag- en middelbaar niveau.

In hoofdstuk 2 ("*Beroepsactiviteiten en kernproblemen*") wordt de basis gelegd voor de typologie van competenties die in het volgende hoofdstuk wordt gepresenteerd. In navolging van Engeström (1987) wordt de arbeids- of beroepspraktijk opgevat als het uitvoeren van activiteiten in een activiteitensysteem. Daaraan worden zes onderling samenhangende componenten onderscheiden: 1) het handelend subject; 2) de gehanteerde gereedschappen en 3) het object (product) van het handelen (deze drie vormen samen het *productiesysteem*); 4) de wijze waarop arbeidstaken over mensen worden verdeeld en 5) de regels die gelden voor de arbeidsactiviteit (*het beheers- of reguleringssysteem*); 6) de gemeenschap waarbinnen de activiteit plaatsvindt (*het communicatief-cultureel systeem*). Elk van deze drie deelsystemen wordt in hoofdstuk 2 verder in kaart gebracht. Daarbij wordt de nodige aandacht geschonken aan de sociotechniek, het belang van volledige (zowel uitvoerende, voorbereidende als ondersteunende taken omvattende) functies en van voldoende regel-, leer- en samenwerkingsmogelijkheden. Bij de bespreking van het productiesysteem wordt een onderscheid gemaakt in het werken met dingen, symbolen en mensen. Bij de analyse van het reguleringssysteem wordt onder andere ingegaan op het verschil tussen problemen waarop geanticipeerd kan worden (planningsproblemen) en onverwachte problemen (incidenten). De cultureel-communicatieve dimensie van het arbeidshandelen wordt essentieel geacht voor zingevingsprocessen en identiteitsontwikkeling. In dat kader wordt bijvoorbeeld aandacht besteed aan formele versus informele communicatie en, onder verwijzing naar Lave & Wenger (1991), aan het proces van inwijding in een bepaalde praktijkgemeenschap. Onstenk beschouwt de drie deelsyste-

men als de belangrijkste determinanten van de kernproblemen die in arbeidssituaties een rol kunnen spelen. Hij definieert kernproblemen als "...kenmerkende en centrale beroepssituaties, waarin a) complexe problemen aangepakt moeten worden en b) rekening gehouden moet worden met de specificiteit van de situatie" (p. 77). In het verlengde van de in hoofdstuk 2 verrichte analyses worden zes soorten kernproblemen onderscheiden: productie-, organisatie-, sociaal-culturele, regel-, samenwerkings- en lidmaatschapsproblemen.

Competentie wordt gezien als een geïntegreerd geheel van kennis, vaardigheden en houdingen met behulp waarvan handelingen kunnen worden verricht en problemen kunnen worden opgelost. Toegespitst op arbeid gaat het uiteraard om arbeidshandelingen en arbeidsproblemen, waarbij de eerder onderscheiden kernproblemen een centrale plaats innemen en het niet alleen gaat om het expliciet en bewust toepassen van regels, maar ook om intuïtieve patroonherkenning ("tacit skills"). Voortbouwend op het classificatieschema van kernproblemen ontwikkelt Onstenk in hoofdstuk 3 ("*Brede vakbekwaamheid*") een classificatieschema voor competenties. Dat leidt tot zes typen competenties, waaraan leer- en vormgevingscompetenties ("...gericht op het kunnen oplossen van leer- en ontwikkelingsproblemen...") (p. 101) worden toegevoegd. Elk van de zeven competenties wordt verder uitgewerkt, waarbij steeds twee of meer deelaspecten worden onderscheiden. Bij de vakmatige competentie (het kunnen oplossen van productieproblemen) wordt bijvoorbeeld een onderscheid gemaakt in technisch-praktische, informatieverwerkende en omgangsvaardigheden. Bij de bespreking van de leer- en vormgevingscompetentie wordt aandacht besteed aan het leren van ervaringen, transfervaardigheden en het vermogen zelf actief bij te dragen aan de vormgeving van de arbeid ("vormgevingscompetentie"). Onstenk benadrukt dat vakbekwaamheid breed moet worden opgevat en niet moet worden beperkt tot de vakmatige competentie. Brede vakbekwaamheid omschrijft hij als: "...een multidimensionale, gestructureerde en samenhangende set vakmatige, methodische, bestuurlijke, strategische, sociaal-communicatieve, normatief-culturele en leercompetenties, gericht op een adequate aanpak van de kernproblemen van het beroep" (p. 128).

In hoofdstuk 4 ("*Leren op de werkplek in de theorie*") gaat Onstenk in op de literatuur over leren op de werkplek. Hij maakt daarbij een onderscheid in vier benaderingen. De eerste, *immanent en expansief leren*, ligt in het verlengde van de activiteitstheorie. De nadruk wordt gelegd op de impliciete, in het werk geïntegreerde competentie-ontwikkeling die het gevolg is van het uitvoeren van in complexiteit verschillende arbeidstaken en het zoeken naar of verkennen van nieuwe werkvormen. De tweede benadering, *gesitueerd leren*, sluit aan bij de sociaal-constructivistische visie op leren als een geleidelijk proces van inwijding in de kernactiviteiten en gebruiken van een praktijkgemeenschap. Het gaat daarbij vooral om het leren (in de directe nabijheid) van qua competentie verschillende collega's, onder andere van de verhalen ("war stories") die zij te vertellen hebben. Typerend voor de derde benadering, *aangeduid als zelfstandig leren*, is de accentuering van het intentionele en zelfgestuurde leren door de werknemer als lerende, waarbij het benutten van leermogelijkheden en reflectie centraal staan. In de vierde benadering wordt leren op de werkplek gezien als *arbeidssocialisatie* met specifieke aandacht voor de ontwikkeling van normatief-culturele competenties.

Hoofdstuk 5 ("*Het leerpotentieel van arbeidssituaties*") gaat over de verschillende determinanten van leren op de werkplek. De kans dat er op de werkplek geleerd wordt, ofwel het leerpotentieel van de arbeidssituatie, wordt volgens Onstenk bepaald door "... 1) de reeds bestaande competentie en leervaardigheden van de werknemer; 2) de leerbereidheid van de werknemer; 3) het leeraanbod en het 4) opleidingsaanbod op de werkplek, alsmede de relaties en wisselwerkingen tussen deze determinanten" (p. 213). Het leeraanbod is afhankelijk van de kenmerken van de functie, de aanwezige informatie en de sociale omgeving. Het opleidingsaanbod wordt bepaald door de mate waarin leermogelijkheden gestructureerd zijn, de

mogelijkheden tot participatie in innovatie en de beschikbaarheid en aard van gestructureerde opleidingen op de werkplek.

In hoofdstuk 6 ("*Leren op de werkplek in nieuwe productieconcepten*") wordt een antwoord gezocht op de vraag in hoeverre door middel van veranderingen in de organisatie van de arbeid het leeraanbod in de betrokken arbeidssituaties vergroot kan worden. Daartoe worden de sociotechniek en het daarop geïnspireerde WEBA-instrument, enkele Zweedse ontwikkelingsprogramma's en de Japanse modellen voor slanke productie en continue verbetering aan een nadere analyse onderworpen. Onstenk concludeert dat het leeraanbod vooral vergroot kan worden door functieverbreiding en -integratie, taakgroepen en een coachende rol van collega's en chefs. Hij beschouwt deze instrumenten als "...bouwstenen voor de kwalificerende organisatie" (p. 253).

Het verrichte empirisch onderzoek wordt gepresenteerd in hoofdstuk 7 ("*Leren in kwalificerende organisaties: casestudies*") en hoofdstuk 8 ("*Leren in de stage*"). De case-studies betreffen twee bedrijven uit de procesindustrie en de automatiseringsafdeling van een bank. De door middel van document-analyses, interviews en observaties op de werkplek verzamelde informatie is gebruikt om een aantal in de vorige hoofdstukken behandelde thema's en trends te illustreren, waaronder de invloed van de sociotechnische benadering, het streven naar kwaliteitsverbetering en verbreding van vakbekwaamheid, de ontwikkeling van meer volledige functies en de integratie van diverse opleidingsondersteunende en leerbevorderende activiteiten en hulpmiddelen in de arbeidssituatie. In hoofdstuk 8 wordt verslag gedaan van een onderzoek naar leerprocessen en leereffecten van stages in het MBO. Informatie is verkregen uit interviews met 35 stagiairs voor, tijdens en na hun stage, alsmede uit bezoeken aan de stageplaats en gesprekken met stagebegeleiders. Het onderzoek heeft geleid tot vergroting van het inzicht in de kernproblemen van de betrokken beroepen en de factoren die bijdragen aan de effecten van stages op de ontwikkeling van brede vakbekwaamheid. Aan het eind van het hoofdstuk presenteert Onstenk een model voor het leren in de stage.

Algemeen oordeel

Het boek is een aanrader voor hen die zich interesseren voor het leren en opleiden van volwassenen in veranderende arbeidsorganisaties en voor de rol die de (toekomstige) werkplek kan spelen bij de ontwikkeling van startcompetenties in het beroeps onderwijs. Onstenk beweegt zich op een terrein waar sprake is een grote verscheidenheid aan (wellicht zelfs wildgroei van) normatieve modellen. Dit verschijnsel maakt duidelijk dat we met weerbarstige materie te maken hebben. Er zijn veel invalshoeken mogelijk, het object van interesse is moeilijk af te grenzen, er zijn veel vragen. Onstenk's poging om een aantal fundamentele vragen met een multidisciplinaire aanpak te benaderen levert een goed gestructureerd en boeiend overzicht op dat naar mijn mening zowel interessant is voor "beginners" als voor de op het terrein reeds ingevoerde lezer die zoekt naar synthese. Onstenk levert met zijn proefschrift een belangrijke bijdrage aan het debat over competentie-ontwikkeling en beroeps pedagogiek. Daarbij geeft hij impliciet of expliciet interessante suggesties voor empirisch onderzoek naar leerprocessen in het kader van het leren en opleiden in arbeidsorganisaties en het beroeps onderwijs.

Specifieke evaluatieve opmerkingen

- De "kernproblemen van het beroep" kunnen een vruchtbaar uitgangspunt zijn voor curriculumontwikkeling, begeleiding van lerenden/werkenden op de werkplek (ook in het kader van de stage) en de ontwikkeling van beoordelingscriteria. Zij zouden bijvoorbeeld de kern kunnen zijn van epitomen in elaboratief gesequentieerde curricula en aldus de lerenden kunnen helpen bij het opbouwen van (voorlopige) mentale modellen van een vakgebied of arbeidssituatie.

- Op een terrein dat zich moeilijk laat afgrenzen en waar nog relatief weinig empirisch onderzoek heeft plaatsgevonden, is een fenomenologisch-antropologische invalshoek erg waardevol. Op verschillende plaatsen in het proefschrift komt de waarde van deze benadering naar voren. Ik denk hierbij aan de passages waarin de rol van verhalen ("war stories"), "eye-openers" en "incidenten" bij de ontwikkeling van competenties wordt belicht.
- Het is vruchtbaar om verschillende invalshoeken te combineren. De fenomenologisch-antropologische en sociaal-constructivistische benaderingen laten zich naar mijn mening echter niet zo gemakkelijk verenigen met de meer objectivistische invalshoek van de activiteitstheorie. Mogelijk is dit gegeven er mede de oorzaak van dat Onstenk relatief weinig aandacht schenkt aan individuele verschillen (bijvoorbeeld) in doeloriëntaties, epistemologische opvattingen en de betekenis die (aanstaande) beroepsbeoefenaren toekennen aan werk. Ook wordt niet ingegaan op de invloed van het psychologisch contract (Freese & Schalk, 1995) tussen werknemer en werkgever op de bereidheid om gebruik te maken van leer- en opleidingsmogelijkheden.
- In het kader van het stage-onderzoek zijn interessante interviewgegevens verzameld over (onder andere) kenmerken van stageplaatsen, begeleidingsmaatregelen, leeractiviteiten en leereffecten. De verkregen gegevens zijn aan de hand van een a priori classificatieschema geanalyseerd. Het zou mijns inziens echter de moeite waard zijn (een deel van) de data te heranalyseren met behulp van fenomenografische technieken.
- Terecht beschouwt Onstenk leren en ontwikkeling als een integraal onderdeel van brede vakbekwaamheid. Ik vind het echter minder gelukkig dat hij in zijn analyseschema van competenties de leer- en vormgevingscompetentie als een aparte categorie opneemt naast de op specifieke leerinhouden betrekking hebbende competenties. Het is mijns inziens gepaster om leercompetentie als een integraal onderdeel te zien van elk van de onderscheiden competenties. Dat sluit bovendien ook beter aan bij de uitgangspunten van de activiteitstheorie, waarin ontwikkeling als een kenmerk van competentie wordt gezien.
- Ik vind het prijzenswaardig dat Onstenk heeft geprobeerd relaties te leggen tussen de uit interviewgegevens naar voren gekomen (impliciete en expliciete) leeractiviteiten van MBO-leerlingen en leereffecten of competentie-ontwikkeling. Het is jammer dat iets dergelijks niet heeft plaatsgevonden bij de case studies in de bedrijven. Het verbaast me ook enigszins dat bij de suggesties voor vervolgonderzoek de relatie tussen leeractiviteiten en leerresultaten niet genoemd wordt. Hiernaar zou in de context van het leren en opleiden in het beroepsopleidings- en in arbeidsorganisaties veel meer onderzoek moeten worden gedaan.

Literatuur

- Engeström, Y. (1987). *Learning by expanding. An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy.
- Freese, C. & Schalk, R. (1995). Het psychologisch contract en leeftijdsbewust personeelsmanagement. In: R. Schalk (Red.), *Oudere werknemers in een veranderende wereld* (pp. 207-223). Utrecht: Lemma.
- Lave, J. & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Johan van der Sanden
 KUB: vakgroep psychologie TUE
 (Technische Universitaire Lerarenopleiding)
 Postbus 513 5600 MB Eindhoven
 e-mail: j.m.m.v.d.sanden@phys.tue.nl

Marjan Vermeulen (1997)

De school als arbeidsorganisatie: Personeelsbeleid, organisatiekenmerken en arbeidsbeleving van leraren in scholen voor voortgezet onderwijs.

Academisch Proefschrift Universiteit van Amsterdam

Uitgegeven in eigen beheer.

De school is óók een arbeidsorganisatie waarin personeelsbeleid een instrument kan zijn om leraren te ontwikkelen en daarmee het onderwijs te verbeteren. Personeelsbeleid kan de verbinding leggen tussen de onderwijskundige doelen en de behoeften en kwalificaties van het personeel. De autonome scholen, vooorzien van lump-sum financiering en professioneler wordend management, pogen hier gestalte aan te geven. Personeelsbeleid is een middel om het leraarsberoep meer elan te geven en om 'psychologische veroudering' tegen te gaan. Marjan Vermeulen, die een bijdrage leverde aan de Commissie Toekomst Leraarschap, schreef een dissertatie met als hoofdvraag: Welke kenmerken van de school als arbeidsorganisatie, in het bijzonder het personeelsbeleid, hangen samen met de arbeidsbeleving van leraren?

Conceptueel model

Vermeulen omschrijft haar studie als exploratief maar laat zich wel sturen door een aantal inhoudelijke noties. In het theoretische hoofdstuk passeert een aantal organisatie-theorieën de revue die fraai blijken te passen in de effectiviteitskwadranten van Quinn & Rohrbaugh (1983) met als hoofd-dimensies: 'intern versus extern gericht' en 'flexibiliteit versus controle/beheersing'. Het human relations model valt in het kwadrant 'intern, flexibel', het open systeem model is 'extern, flexibel', het intern proces model (vergelijkbaar met Mintzberg's professionele bureaucratie) is te kenschetsen als 'intern, controle' en het rationele doel model (machinebureaucratie; schooleffectiviteitsbenadering) is extern en controle-gericht. Eveneens in navolging van Quinn e.a. worden acht managementrollen onderscheiden die te plaatsen zijn in genoemd raamwerk: bestuurder, producent, monitor, coördinator, mentor, stimulator, vernieuwer, onderhandelaar.

Vanzelfsprekend wordt aandacht besteed aan wat nogal groots aangeduid wordt met 'Human Resource Management theorieën'. Deze benadering benadrukt de wisselwerking tussen het strategische beleid van de organisatie en personeelsmanagement. Daarbij wordt een ontwikkelingsgang onderscheiden: van personeelsbeheer via personeelszorg naar personeelsmanagement. In het raamwerk past HRM bij de kwadranten 'human relations' en 'rationele doel', wat illustreert dat HRM concurrerende waarden behelst.

De theoretische verkenning mondt in hoofdstuk 3 uit in een conceptueel model met vier hoofdcomponenten: situationele kenmerken van de school - de interne organisatie van de school (structuur, cultuur, primair proces, management en beleid) - personeelsbeleid (omvat instrumenten als training en ontwikkeling) - arbeidsbeleving (gemeten als demotivatie en arbeidssatisfactie). Arbeidsbeleving - opgevat als de perceptie van de werknemer van zijn arbeidssituatie - vormt de afhankelijke variabele en personeelsbeleid de intermedierende variabele tussen schoolorganisatiekenmerken en arbeidsbeleving.

Methode

Er is gebruik gemaakt van bestaande data: schoolleiders 1991 (n=188), leraren 1994 en leraren 1995 (gekoppeld aan enkele schoolleidersgegevens). Bij beide lerarenbestanden is gekozen voor een gemiddelde op schoolniveau wat leidt tot een aanzienlijke reductie van aantallen, respectievelijk tot 48 en 115. Als argument voor het onbenut laten van het multi-niveau-karakter van de data wordt gegeven dat de interesse niet uitgaat naar analyses op individueel niveau in relatie tot het schoolniveau. Gevolg is dat de resultaten betrekking hebben op de *gemiddelde* arbeidsbeleving van leraren binnen de school. In de analyses is geen rekening gehouden met het gegeven - bijvoorbeeld via weging - dat het aantal respondenten per school nogal uiteenloopt: 1 à 2 tot 6 of meer. De analyses hadden op dit punt verfijnder gekund. Schrijver

meldt wel dat de respondenten van eenzelfde school de situatie soms anders beoordelen. Zij interpreteert deze uitkomst als bevestiging van het vermoeden dat het personeelsbeleid nog weinig consistent in scholen gestalte heeft gekregen.

Bij de operationalisatie moest worden uitgegaan van bestaande gegevens. De arbeidsbeleving is gemeten met twee specifieke schalen voor leraren: de demotivatieschaal (Van Ginckel; De Jong) en de arbeidssatisfactieschaal (Prick; De Jong). Vergelijking van de in deze studie verkregen schaalcores met eerdere gegevens laat zien dat de demotivatie bij leraren nagenoeg gelijk is gebleven en dat de arbeidssatisfactie na een dip in de jaren negentig weer op niveau van 1982 is. De samenhang tussen beide schalen is hoog (resp. -0.78 en -0.85 , twee databestanden), wat principale componentenanalyse bevestigt. Ondanks de inhoudelijke overlap is gekozen om beide arbeidsbelevingsschalen apart te hanteren, omdat zo vergelijking met ander onderzoek mogelijk is.

Bij de samenvatting van de resultaten gaan we uit van Vermeulen's deelvragen.

In hoeverre is er sprake van personeelsbeleid binnen schoolorganisaties?

HRM onderscheidt drie vormen van personeelsbeleid. De data van 1991 laten zien dat 12 procent als 'beheersschool' te typeren is (gebruikt nog geen enkel personeelsinstrument), 84 procent is 'overgangsschool' (enkele instrumenten) en bijna vijf procent is 'personeelszorgschool' (inzet van alle instrumenten). Slechts twee scholen komen in de buurt van 'personeelsmanagement'.

Is er samenhang tussen de mate van ontwikkeling van personeelsbeleid en kenmerken van de schoolorganisatie?

Na vergelijking van de drie typen scholen op een aantal variabelen concludeert Vermeulen dat beheers- en overgangsscholen kenmerken vertonen van de segmentale organisatie, terwijl zorgscholen die beleidsontwikkeling kennen, gelijkenis tonen met niet-segmentale schoolorganisatie zoals omschreven door E. Marx. Theoretisch is dit interessant: er is niet alleen een verband (contingentie) tussen organisatie-model en onderwijskundige uitgangspunten zoals Marx leert, maar ook met de uitwerking van het personeelsbeleid. Naarmate een organisatie complexer wordt, wordt sterker een appel gedaan op het beleidsvormend vermogen wat mede personeelsbeleid kan triggeren dat de organisatie versterkt.

Is er samenhang tussen aspecten van personeelsbeleid, de inzet van personeelsinstrumenten en de arbeidsbeleving van leraren?

Er blijken veel verbanden te zijn tussen de inzet van specifieke personeelsinstrumenten (waaronder functioneringsgesprekken) en beide arbeidsbelevingsindicatoren. De instrumenten onder de noemer van training en (loopbaan)ontwikkeling hebben de duidelijkste invloed. Er is een personeelsbeleidsontwikkelingsschaal geconstrueerd bestaande uit een aantal (krachtige) instrumenten. Deze index correleert -0.41 met demotivatie en $.45$ met satisfactie. Dit betekent dat naarmate meer instrumenten worden ingezet, de positieve arbeidsbeleving toeneemt. Uit de toegenomen correlatie bij gebruik van de index wordt geconcludeerd dat maatregelen elkaar ondersteunen en dat het in samenhang invoeren tot het beste resultaat leidt. Frappant is dat ondanks de gevonden verbanden minder dan een kwart van de respondenten het eens is met de uitspraak dat het gevoerde personeelsbeleid de eigen arbeidsmotivatie bevordert.

Is er samenhang tussen kenmerken van de schoolorganisatie en arbeidsbeleving van leraren?

Van de structuurkenmerken hebben vooral uitwerking op de arbeidsbeleving: de mate van invloed, de inbreng in de besluitvorming en de duidelijkheid van de taakverdeling. Van de cultuurkenmerken heeft een positieve invloed het delen van gezamenlijke uitgangspunten en visie. De schaal 'goede sfeer' leidt eveneens tot hoge correlaties (-0.42 en $.47$). Van de onderwijsproceskenmerken zijn vooral de motivatie van en het contact met leerlingen en ouders van belang voor de arbeidsbeleving. Van de managementrollen gaat een positieve werking uit van de mentorrol (-0.46 en $.50$), de stimulator en de monitorrol. Als de schoolleidersstijl als 'doeltreffender' wordt gepercipieerd, gaat dat samen met positievere arbeidsbeleving. Er is geen lineair verband tussen schoolgrootte en arbeidsbeleving.

Wat is het relatieve belang van personeelsbeleid ten opzichte van andere schoolkenmerken in de verklaring van de arbeidsbeleving van leraren? (hoofdstuk 8)

De personeelsbeleidsontwikkelingsschaal (index is een betere term, want een somscore) is de belangrijkste indicator die aangeeft in welke mate een school personeelsinstrumenten inzet. De index verklaart 15.3% van de variantie van de demotivatie en 20.0% van de variantie van de satisfactie.

Met variantie-analyse worden verbanden onderzocht tussen de mate van ontwikkeling van personeelsbeleid en arbeidsbeleving, waarbij een aantal covariaten is ingevoerd en gecheckt op significantie van hun bijdrage. Variabelen die met de organisatiecultuur te maken hebben, blijken gezamenlijk belangrijke covariaten voor de relatie personeelsbeleidsontwikkeling - arbeidsbeleving. Het betreft: 'goede sfeer' en 'als leraar klaar staan voor de school'.

Variabelen die het personeelsbeleid als zodanig krachtig beïnvloeden (correlaties $> .50$) zijn: de invloed bij de besluitvorming en de mentor- en monitorrol van het schoolmanagement. Een goed onderhouden gebouw blijkt een belangrijke covariaat te zijn bij de beoordeling van het personeelsbeleid. Deze sterke verbanden brengen Vermeulen er toe om in navolging van anderen het scala aan personeelsinstrumenten uit te breiden met: ondersteuning van de schoolleiding als mentor en monitor, en participatie in besluitvorming. De mentor houdt zich bezig met ondersteunen en begeleiden van personeel, terwijl de monitorrol voorwaarde is voor een goedlopende organisatie en ontwikkeling van personeelsbeleid. Geconcludeerd wordt dat de schoolleiding van eminent belang is voor de ontwikkeling van personeelsbeleid. De auteur laat in een fraai figuur zien welke managementrollen een bijdrage blijken te leveren aan het personeelsbeleid en - in minder mate - aan de arbeidsbeleving.

Als toegift wordt de effectiviteitsmaat van deze studie 'arbeidsbeleving' vergeleken met de 'leerlingeffectiviteit' uit de schooleffectiviteitsliteratuur. Vastgesteld wordt dat er overeenkomsten zijn tussen scholen die leerlingeffectief zijn en effectief in de zin van een positieve arbeidsverdeling. Het betreft kenmerken als gezamenlijke uitgangspunten, een krachtige schoolleiding, een ordelijk en plezierig schoolklimaat.

Kanttekeningen

De studie kent een heldere structuur die alleen bij het belangrijke hoofdstuk 8 enigszins te wensen over laat. Er wordt duidelijk gerapporteerd, hoewel de concentratie van de lezer op de proef wordt gesteld door de vele bivariate verbanden die successievelijk de revue passeren. Hoofdstuk 8 stelt wat dit betreft hoge eisen: men moet scherp opletten welk bestand benut is, of er sprake is van de personeelsbeleidsontwikkelingsschaal dan wel van de personeelsbeleidsbeoordelingsschaal als afhankelijke en welke relatie er precies onderzocht wordt. Gelukkig bieden de diverse samenvattingen uitkomst.

Vermeulen blijft de veelheid van haar gegevens de baas. De gevonden verbanden worden regelmatig tegen het licht gehouden van theoretische noties en enkele onderzoeken. Er wordt weinig gerefereerd aan (internationaal) onderzoek op dit gebied omdat dit er nauwelijks zou zijn. Toch zijn er raakvlakken met bijvoorbeeld onderzoek naar 'school as a workplace' (Rosenholtz), naar professionalisering van leraren en niet te vergeten de algemene literatuur over arbeidsatisfactie. In toekomstig onderzoek verdient het aanbeveling het complexe begrip arbeidsbeleving in meer dan twee facetten te operationaliseren.

Er is gekozen voor betrekkelijke eenvoudige analysetechnieken met de correlatie in de hoofdrol. In hoofdstuk 8 is gewerkt met (co)variantie-analyses. Multiple-regressie-analyses of structurele modellen zouden waarschijnlijk inzichtelijker en efficiënter hebben gewerkt. Eerder noemden we al het negeren van het multi-niveau aspect. Hoewel de gevonden verbanden doorgaans heel plausibel lijken, krijgt men soms toch het gevoel van *data snooping*: waarom dit verband en deze covariaat wel significant en een andere niet? Het gemis aan een expliciete theorie wreekt zich dan.

De studie van Vermeulen heeft verdiensten. Het nog prille personeelsbeleid op scholen blijkt gebaat met gerichte sturing door de schoolleiding, het planmatig aanbieden van scho-

ling en training en het versterken van cultuurkenmerken als sfeer en gezamenlijke visie. Dit proefschrift overtuigt dat personeelsbeleid in scholen een stevige plaats op de onderwijsresearchagenda verdient.

Hans van Gennip,
KUN Nijmegen (ITS)

M.Th. Glaudé (1997)

'Werkplekopleiden als innovatie'
Proefschrift Universiteit Utrecht.

Onderzoek: aanleiding en probleemstelling

De vakgroep Onderwijskunde van de Universiteit Utrecht startte in 1990 een onderzoeksprogramma rondom werkplekopleidingen. Het promotie-onderzoek van dr. Marjan Glaudé, 'Werkplekopleiden als innovatie', is uitgevoerd binnen dit kader. Het onderwerp van de studie is de relatie tussen organisatieverandering en werkplekopleiden. Meer specifiek luidt de hoofdvraag: *'Onder welke condities kan een werkplek-opleiding als effectief management-instrument worden ingezet om gestalte te geven aan vernieuwing van de organisatie?'*

De thematiek is relevant, gezien enerzijds de grote (en groeiende) belangstelling van de praktijk voor opleiden op de werkplek en anderzijds de geringe empirische kennisbasis die over dit thema is ontwikkeld. Leren op de werkplek is steeds meer aan de orde van de dag. Door de huidige economische en technologische ontwikkelingen wordt 'leren' in toenemende mate een onderdeel van het 'werken', in plaats van een activiteit die bedoeld is om operationele taken beter te kunnen uitvoeren. Dat roept een spanningsveld op, constateert de onderzoekster (p. 3): *'Enerzijds dienen medewerkers nieuwe kennis en kunde te verwerven, anderzijds moet mede onder druk van de externe omgeving de productie doorgang vinden. De medewerkers kunnen niet continu uit het productieproces worden gehaald om deel te nemen aan externe opleidingen.'* Werkplekopleiding biedt een potentiële uitweg uit dit spanningsveld, omdat sprake is van een koppeling tussen het verrichten van productieve arbeid en leerprocessen.

Het thema van de dissertatie is vooral interessant gezien de insteek die Glaudé kiest: niet de instrumentele of didactische kant van werkplekopleidingen wordt bekeken ('hoe geef je een werkplekopleiding (vorm)?'), maar de manier waarop deze beleidsmatig in organisaties kunnen worden ingezet om bedrijfsdoelstellingen te (helpen) realiseren, in dit geval doelstellingen rondom innovaties.

Historische en theoretische verkenning werkplekopleiden

Het onderzoeksrapport start, na de inleiding, met een historische en theoretische verkenning. Het historisch overzicht is de moeite waard omdat Glaudé zich niet beperkt tot een chronologisch overzicht van stromingen en ontwikkelingen binnen werkplekopleiden, maar tracht een relatie te leggen tussen ontwikkelingen in de organisatiecontext en de manier waarop deze met werkplekopleidingen omgaan. Haar conclusie is dat er een golfbeweging te herkennen is, waarbij organisaties in sommige decennia veel speelruimte hebben, en in andere juist zeer gebonden zijn. Ruimte betekent hier dat de organisatie grotendeels zelf kan bepalen welke producten ze levert en dergelijke, terwijl gebrek aan ruimte inhoudt dat omgevingsontwikkelingen de inrichting van organisaties sterk beïnvloeden. Glaudé constateert dat in de periodes van weinig speelruimte de relatie tussen 'werken, opleiden en leren' hechter wordt, en wijst de huidige trend richting minder vrijheid van organisaties aan als oorzaak van 'een kentering van het tij' voor werkplekopleidingen: na een periode van relatief weinig aandacht is het weer in de belangstelling teruggekeerd.

Het rapport plaatst werkplekopleidingen niet alleen in een historisch, maar ook in een theo-

retisch perspectief. Geen gemakkelijke opgave, werkplekopleiden is niet makkelijk te definiëren. Wat is bijvoorbeeld 'op de werkplek'? Dient dat beperkt te worden tot de fysieke werkplek van de medewerker, achter de machine, de lopende band of het bureau? Of kunnen we ook van een werkplekopleiding spreken als een medewerker werkt aan een actueel en dringend organisatieprobleem, maar dat doet in een opleidingslokaal of conferentie-oord? Een andere kwestie is wanneer we kunnen spreken van een formele opleiding en wanneer van een informeel leerproces. Glaudé raakt deze kwesties aan, maar maakt helaas niet altijd duidelijke keuzes en/of verantwoordt ze niet expliciet. Zo stelt ze bijvoorbeeld wel dat informele vormen van leren buiten haar definitie van werkplekopleiden vallen, maar vermeldt ze niet waar zij de scheidslijn tussen formeel en informeel leren legt. De onderzoekster beperkt zich echter tot het al dan niet gestructureerd en/of gepland zijn van het leerproces als criteria, maar omzeilt daarmee de definitiekwestie feitelijk door niet verder te specificeren wat ze daaronder verstaat. Immers: wat is structuur? Als er een begin- en eindtijd voor het leertraject is gesteld is er enige mate van structuur aangebracht. Maar meestal spreken we pas over structuur als er ook nog leerdoelen zijn geformuleerd, opleidingsactiviteiten zijn bepaald en in de tijd uitgezet, en/of ondersteuning (materieel en in personele zin) wordt geboden voor die activiteiten (zie bijvoorbeeld Kruijd, 1991; Onstenk, 1994).

Interessant onderdeel van de theoretische verkenning is het onderscheid dat Glaudé aanbrengt tussen didactische en arbeidsorganisatorische structurering van het leerproces. Ze merkt terecht op dat via beide wegen op effectieve wijze structuur in het leerproces kan worden aangebracht, en ondersteuning voor de lerende. Veel traditionele benaderingen van werkplekopleidingen beperken zich echter tot de didactische maatregelen. Dat Glaudé voor een integrale benadering van het concept kiest, past bij haar (te waarderen) keuze om het fenomeen werkplekopleiden in een ruimere context te plaatsen en bestuderen.

Conceptueel model

Het conceptueel model bestaat uit twee kolommen (kokers), die half over elkaar heen vallen. De ene koker representeert het arbeidssysteem, de andere het opleidingssysteem. De werkplekopleiding vindt plaats in het midden, waar beide systemen elkaar overlappen. De kokers bevatten een niveau-indeling: macro, meso en micro niveau. Op elk van de drie niveaus formuleert Glaudé op basis van literatuur condities voor het effectief inzetten van werkplekopleidingen bij organisatievernieuwing. Ze incorporeert een grote hoeveelheid condities, die ze in drie clusters onderbrengt: integratiebevorderende maatregelen; de opleidingscyclus; contingenties tussen organisatie-aspecten. De kracht van het model is dat Glaudé werkplekopleidingen in de context plaatst van de drie organisatieniveaus (macro, meso en micro) en niet blijft steken in het micro-niveau, zoals de meeste (onderzoeks-)publicaties die tot nu toe op dit terrein zijn verschenen. Ze plaatst werkplekopleiden in de brede organisatie- en beleidscontext. Nadeel is dat het model daarmee dusdanig breed wordt, dat dit het onderzoek weinig richting geeft.

Onderzoeksonderwerp, dataverzameling en analyse

Als onderzoeksmethodiek is gekozen voor casestudy-onderzoek, vooral vanwege het nog beperkte inzicht in de omstandigheden waaronder een werkplekopleiding het best functioneert, en het belang van de organisatiecontext. De zes case-organisaties zijn ondernemingen die zich in een organisatievernieuwingproces bevinden, en die daarvoor onder meer een werkplekopleiding gebruiken om de vernieuwingsdoelen te realiseren. Met deze brede opvatting van het fenomeen innovatie (namelijk een organisatieverandering) maakt de onderzoekster het zich niet eenvoudig. Met een keuze voor een type innovatie met minder verstrekkende gevolgen, bijvoorbeeld invoering van een nieuwe machine, was wellicht een meer hanteerbaar onderzoeksterrein afgebakend.

Over de casestudy-bedrijven is een indrukwekkende hoeveelheid informatie verzameld. De opzet voor de dataverzameling is zorgvuldig ontworpen. Er is rekening gehouden met de

juiste respondenten voor de juiste gegevens, en er is triangulatie toegepast (Yin, 1984). Een problematisch punt bij de dataverzameling is echter het grote scala aan op zichzelf brede thema's waarover gegevens zijn verzameld (zoals: externe omgeving organisatie, beleid, organisatie-ontwerp, concept werkplekopleiden). Dit is het resultaat van de veelomvattendheid van de condities uit het conceptueel model en de brede opvatting van innovatie (nl. als organisatievernieuwing). Al met al lijkt het erop dat de kracht van de dissertatie (namelijk de brede benadering van het fenomeen werkplekopleiden) tegelijk het voornaamste probleempunt vormt. Soms dwaalt de dissertatie wel erg ver af van het eigenlijke onderzoeksonderwerp af, namelijk werkplekopleiden. Dit doet zich met name voor waar het gaat om condities op het gebied van integratiebevorderende factoren en de contingentietheorie. Daar typeert Glaudé organisatie-aspecten als de algehele strategievorming en de organisatieomgeving in enkele steekwoorden (bijv. 'adaptief' of 'visionair'). Aan dergelijke typering is een onderzoek op zich te wijden, het is de vraag of zij op een betrouwbare manier te maken zijn via een aantal interviewvragen en bestudering van documenten. Hetzelfde geldt soms voor aspecten die dicht bij het onderwerp staan, vooral bij de effectbepaling van werkplekopleidingen zijn kanttekeningen te plaatsen. De problematiek van het vaststellen van opleidingseffecten is methodologisch zeer complex, zeker als het gaat om effecten op afdelings- en organisatieniveau (zie bijvoorbeeld de dissertatie van Van Sandick & Schaap). Onderhavig onderzoek bekijkt de effecten van de werkplekopleidingen op leergedrag-, werkgedrag-, afdelings- en organisatieniveau via het stellen van enkele korte vragen hierover aan diverse respondenten. Op deze manier kunnen echter niet meer dan *indicaties* voor effecten gevonden worden.

De werkwijze bij de analyse van de enorme hoeveelheid gegevens dwingt op zich respect af (de casestudybeschrijvingen besloegen elk zo'n 100 pagina's) Glaudé heeft deze op een systematische manier weten terug te brengen tot een hanteerbaar en overzichtelijk niveau, gebruikmakend van de matrixmethode van Miles & Huberman. De codering van casestudybeschrijvingen gebeurde door vier personen (Glaudé en drie studenten), de relatief hoge inter- en intrabeoordelaarsbetrouwbaarheid (steeds meer dan 80%) laat zien dat de onderzoekster hiervoor een betrouwbare coderingsmethodiek heeft ontwikkeld. Aan de hand van de codes zijn de gegevens in matrices geplaatst, met behulp waarvan zowel verbanden per case (within-case analysis) als tussen de cases zijn opgespoord (cross-case analysis).

Helaas begeeft de analyse zich bij deze codering enigszins op glad ijs. Om te kunnen komen tot eenduidige uitspraken over de mate waarin de genoemde clusters van voorwaarden in een case te herkennen waren, en over de vraag of een case succesvol was of niet, is een grote diversiteit aan gegevens is teruggebracht tot clustercodes als 'er is aan deze voorwaarden voldaan', 'er is aan de helft voldaan', of 'niet voldaan'. Bijvoorbeeld: om te bepalen of een case-organisatie integratiebevorderende maatregelen neemt is het totaal aantal getroffen maatregelen (bijvoorbeeld 7) gedeeld door het aantal 'mogelijk te treffen maatregelen', in dit geval 14. De score (clustercode) is dan 0,5 wat in het indelingscriterium neerkomt op 'relatief weinig maatregelen'. Het voordeel is dat een zeer complexe werkelijkheid wordt gereduceerd tot één enkele uitspraak, wat vergelijking tussen de cases mogelijk maakt. Het is echter de vraag of de veralgemenisering hier niet te ver doorschiet. De cijfers waarvan gebruik wordt gemaakt zijn immers ook al een interpretatie van de case-informatie. Daar is al een vereenvoudiging van de realiteit gemaakt door te stellen dat een maatregel 'wel' of 'niet' genomen is. De werkelijkheid laat zich niet zo makkelijk beschrijven, maatregelen kun je ook in aangepaste vorm nemen bijvoorbeeld, of in bepaalde tijdsperiodes wel en andere niet, etcetera. De beperkte driepuntsschaal uit dit onderzoek laat weinig ruimte voor dergelijke nuanceringen. Meespelelend probleem is wat de waarde van de clustercode uiteindelijk is: er wordt alleen gekeken naar of de condities voorkomen, niet naar de kwaliteit waarmee die condities zijn uitgewerkt.

Om naast de beschrijving van wat er in de case-organisatie gebeurt ook enkele verklarende conclusies te trekken (d.w.z. relevante condities op te sporen) heeft de onderzoekster een uitspraak gedaan over de mate van succes van de diverse cases, op basis van de mate waarin doelen bereikt zijn (op de onderscheiden niveaus: leergedrag, werkgedrag, afdeling en orga-

nisatie), de mate van ambitie van die doelen en de doorlooptijd van de opleiding. Dit bleek lastig, zoals de onderzoekster zelf ook aangeeft. De succesvolle cases zijn vergeleken met de minder succesvolle, op de geaggregeerde oordelen over de clusters van condities (cluster-codes). Hier kwamen geen eenduidige resultaten uit naar voren. Al met al dient dan ook de conclusie te luiden dat dit verklarende deel minder uit de verf komt. Dat valt wel te verwachten, gezien de complexe aard van de problematiek (met name de gelaagdheid) en de stand van de theorievorming. De vraag is of de probleemstelling dan ook niet beter een puur beschrijvend karakter had kunnen krijgen.

Conclusie

Al met al is de hoofdconclusie over deze dissertatie dat het een waardevolle aanvulling vormt voor het nog relatief jonge onderzoeksgebied van bedrijfsopleidingen. Zeker omdat het een relatief onontgonnen, maar wel belangrijker wordend onderdeel betreft, namelijk werkplekopleiden. De studie van Glaudé verdient waardering vanwege het beschrijvend deel. Mede door de goede leesbaarheid van de case-verslagen geeft het proefschrift een uniek doorkijkje in de manier waarop organisaties met werkplekopleidingen omgaan, en de samenhang (of het gebrek eraan) tussen beslissingen op macro, meso en micro niveau. De studie levert weinig (extra) inzicht op in maatregelen die getroffen moeten worden voor het effectief inzetten van werkplekopleidingen om organisatieveranderingen te bespoedigen, maar geeft wel aan hoe het in de praktijk gesteld is met de maatregelen waarvan in de literatuur wordt vermoed dat ze een belangrijke rol spelen. Daarmee legt het onderzoek een goede basis voor verdere studies op dit terrein.

Literatuur

- Glaudé, M. Th. (1997) *Werkplek-opleiden als innovatie*. Proefschrift. Universiteit Utrecht: Vakgroep Onderwijskunde.
- Kruijd, D. (1991) *Opleiden op de werkplek: een terreinverkenning*. In: *Capita Selecta Opleiders in Organisaties*, aflevering 6. Deventer: Kluwer.
- Onstenk, J. (1994) *Opleiden op de werkplek: een verkenning in zes landen*. Amsterdam: Max Goote instituut.
- Yin, R. (1984) *Case study research: design & methods*. Londen: Sage

Ww. drs. Saskia Tjepkema
Universiteit Twente

Faculteit Toegepaste Onderwijskunde
Postbus 217, 7500 AE Enschede
(tjepkema@edte.utwente.nl)