

Tijdschrift voor onderwijsresearch 1990

<https://hdl.handle.net/1874/208868>

VWB 620

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591 X

Jaargang 15
1990

Swets & Zeitlinger B.V.
Lisse

Redactie:

Bert P.M. Creemers (secretaris)
Christiaan Hamaker
Bernadette van Hout Wolters
Leo J.Th. van der Kamp
Gellof Kanselaar
G. Wim Meijnen
P. RobertJan Simons
Lieven Verschaffel
Marinus J.M. Voeten
Wynand H.F.W. Wijnen

BIBLIOTHEEK DER
RIJKSUNIVERSITEIT
UTRECHT



VERENIGING VOOR ONDERWIJSRESEARCH

Artikelen

Effecten van een training in basale gespreksvaardigheden. <i>S.M. Bögels en R. Kreutzkamp</i>	201
Statistische aspecten van multi-niveau onderzoek. <i>R.J. Bosker en T.A.B. Snijders</i>	317
De ontwikkelingen in het schoolloopbaanonderzoek: een terugblik op een decennium. <i>J. Dronkers</i>	8
Doet godsdienst leren? Schoolloopbanen in de gereformeerde sector van het Nederlandse onderwijs. <i>A.B. Dijkstra</i>	56
Onderwijsonderzoek vanuit multi-niveau perspectief. <i>P. van den Eeden en G.W. Meijnen</i>	259
Enige literatuur over multi-niveau onderzoek	330
The threat of scale drift in item banks. <i>Dato N.M. de Gruijter</i>	104
Het oog van de meester: opstelbeoordeling door schrijvers en leerkrachten. <i>J.B. Hoeksma en K. de Glopper</i>	153
Het effect van leerkrachtverwachtingen en streefniveaus op schoolloopbanen in het basisonderwijs. <i>A.A. van der Hoeven-van Doornum, M.J.M. Voeten en P. Jungbluth</i>	23
Het gebruik van kleurendruk bij zoektaken in schriftelijk studiemateriaal. <i>B.H.A.M. van Hout Wolters en W.M.J. Kerstjens</i>	347
Studie-uitval en vertraging bij zes studierichtingen uit het Wetenschappelijk Onderwijs. <i>U. de Jong en J. Meijer</i>	42
Longitudinaal onderzoek van schoolloopbanen. <i>L.J.Th. van der Kamp en M.J.M. Voeten</i>	3
Gebruik van verzorgingsactiviteiten in het basisonderwijs. <i>D. Koster, Th. van Batenburg en H.H. Tillema</i>	215
Interne structuur en constructvaliditeit van belevingsschalen voor Wiskunde en Engels. <i>Hans Kuhlemeier, Huub van den Berg en Franke Teunisse</i>	110
An experimental study of the when and what of children's questions. <i>Hans van der Meij</i>	162
Effecten van computerprogramma's op ontwikkeling in leesprestaties. <i>T. Mooij</i>	285
Multiniveau programmatuur: enige gegevens	332
De stabiliteit van schooleffecten. <i>J. Roeleveld, U. de Jong en P. Koopman</i>	301
De betrokkenheid van schoolleiders in het V.O. bij vernieuwingen in het onderwijs. <i>P.C. van der Sijde en E. Pruis</i>	228
Sekseverschillen in ruimtelijke vaardigheid: andere ervaringen? <i>M. Taal</i>	355
Effecten van gedifferentieerd wiskundeonderwijs: de toepassing van een model voor multilevelanalyse bij curriculumevaluatie. <i>J. Terwel en P. van den Eeden</i>	273
De doelmatigheid van het algemeen voortgezet onderwijs. <i>A.G.J. van der Torre en B. Kuhry</i>	89
De meerwaarde van een goede probleemaanpak. <i>M.V.J. Veenman en J.J. Elshout</i>	337
Het oplossen van vergelijkingsvraagstukken: een empirische toetsing van het model van Lewis en Mayer. <i>L. Verschaffel, E. de Corte en A. Pauwels</i>	172
Analyse van bijzondere gebeurtenissen in schoolloopbanen: kansen op zittenblijven. <i>M.J.M. Voeten en Th. Boland</i>	73
Intuïtieve, empirische en theoretische redeneringen. <i>H. Vos.</i>	123

Notities en commentaren

Variatie in schoolgemiddelden op de CITO-eindtoets basisonderwijs. <i>Th.A. van Batenburg</i>	362
De eerste ervaringen met een geïndividualiseerde toets bestaande uit 600 vragen. <i>H.J.M. van Berkel, E. de Graaff, J. Sprooten en H.G. Schmidt</i>	189
Onderwijsonderzoek te kijk gezet. Commentaar op het rapport 'Schaal en kwaliteit in het basisonderwijs'. <i>Bert P.M. Creemers en Wijnand Th.J.G. Hoebe</i>	234
Onderwijsonderzoekers geven zich bloot. <i>N.A.J. Lagerweij en J.F.M. Claessen</i>	334

Boekbesprekingen

H. van den Bergh: <i>Examens geëxamineerd</i> . (Michel Zwarts)	196
W. van Esch: <i>Schoolbesturen en de vrijheid van onderwijs. Bestuurlijke en organisatorische factoren van invloed op de benutting van beleidsruimte door bijzondere schoolbesturen</i> . (R.H. Hofman)	242
F.M.E. Gerritsen: <i>VTO taalscreening 3- tot 6-jarigen</i> . (W.H.J. van Bon)	370
Anne Kerkhoff: <i>Taalvaardigheid en schoolsucces</i> . (Johan M. Wijnstra)	137
C.M. van Putten: <i>Leerlingen van het individueel beroepsonderwijs nader beschouwd: Een vergelijkend onderzoek naar cognitieve, sociaal-emotionele en achtergrondkenmerken van leerlingen bij de aanvang van het IBO, LBO en VSO/MLK</i> . (J.H.L. Oud en M.M.B. Oud-de Glas)	192
M. Sprangers: <i>Response shift and the retrospective pretest: on the usefulness of retrospective pretest - posttest designs in detecting training related response shifts</i> . (Gerard J.P. van Breukelen)	248
Anne van Streun: <i>Heuristisch Wiskunde Onderwijs; verslag van een onderwijsexperiment</i> . (J.J. Elshout)	139
B. de Vries: <i>Het leven en de leer. Een studie naar de verbinding van leren en werken in de stage</i> . (A.F.M. Nieuwenhuis en J.P.A. Snoek)	143
J.M.H.W. Willems: <i>Studietaken als instructiemiddel: een onderzoek naar de sturende werking van opdrachten bij studieteksten vanuit een informatietheoretisch kader</i> . (Christa C.J. Teurlings)	244

Redactionele medewerkers

In 1990 werd redactionele medewerking verleend door:

E.M.H. Assink	RU Utrecht
M. Beishuizen	RU Leiden
E. Bol	RU Utrecht
R.J. Bosker	UT Enschede
F.B. Brokken	RU Groningen
A.G. Bus	RU Leiden
J. Cohen-Schotanus	RU Groningen
J. Dronkers	SCO Amsterdam
P. Edelenbos	RION Groningen
P. de Graaf	KUB Tilburg
W. van de Grift	RU Utrecht
E.G. Harskamp	RION Groningen
W.K.B. Hofstee	RU Groningen
J.J. Hox	Uv Amsterdam
R. de Jong	RION Groningen
P. Jungbluth	ITS Nijmegen
K.B. Koster	RU Groningen
E.J.J. Kremers	CITO Arnhem
H. Kuyper	RION Groningen
G.J. Mellenbergh	Uv Amsterdam
S. Oosterloo	UT Enschede
M. Oud-de Glas	ITS Nijmegen
J. Peeck	RU Utrecht
P. Reitsma	VU Amsterdam
G.C.W. Rijlaarsdam	Uv Amsterdam
H.G. Schmidt	RU Limburg
J.H. Slavenburg	SAD Rotterdam
T.A.B. Snijders	RU Groningen
K.M. Stokking	RU Utrecht
H.H. Tillema	RION Groningen
H. Verstralen	CITO Arnhem
T.F.W.P. Willemsen	RION Groningen



BULLETIN

**ORGAAN VAN DE VERENIGING
VOOR ONDERWIJSRESEARCH**

Redactie-adres

Greetje P.C. van der Werf, RION,
Postbus 1286, 9701 BG Groningen,
tel. 050-636657

Jaargang 14

1990

VOR:

Convocatie ledenvergadering mei 1990	148
Beknopt verslag verenigingsjaar 1989	149
Verslag VOR-ledenvergadering d.d. 10-5-1990	252
Congressen, Symposia, Cursussen, Trainingen	254, 376
Ontvangen publikaties	151, 200, 255, 382
Oproep Pedon	152

Van de redactie

Zoals uit de omslag blijkt is het Tijdschrift voor Onderwijsresearch de 15e jaargang ingegaan. Hoewel dit natuurlijk niets is vergeleken met collega-tijdschriften die over onderwijsonderzoek rapporteren, zoals Pedagogische Studiën die inmiddels haar 67e jaargang is begonnen, is het in zoverre toch memorabel dat daarmee de levensvatbaarheid van een specifiek tijdschrift op het terrein van het onderwijsonderzoek wel is aangetoond.

Nog een blijk van continuïteit in het onderwijsonderzoek is, dat op dit moment nog enige medewerkers van het eerste uur aan de redactie verbonden zijn, zoals de vorige redactiesecretaris Rinus Voeten en de huidige redactiesecretaris Bert Creemers.

Het tijdschrift bestaat al zo lang dat zelfs iemand voor de tweede keer lid van de redactie is geworden na er een tijd uit geweest te zijn. Dit geldt voor Wynand Wijnen, die in het afgelopen jaar voorzitter van de redactie is geworden.

In een wisselende samenstelling heeft de redactie langzamerhand vorm gegeven aan een tijdschrift dat, na een sterk technisch-methodologisch begin, geworden is tot een tijdschrift waarin alle aspecten van onderwijsonderzoek, zoals inhoudelijke rapportages over onderwijsonderzoek die interessant zijn voor een breder publiek, methodologisch-statistische problemen en reviews, gepubliceerd worden. Daarnaast worden korte bijdragen naar aanleiding van de actuele ontwikkelingen of reacties op verschenen bijdragen geplaatst in de rubriek Notities en Commentaren.

Ook al is het niet direct aan de in de afgelopen jaargang gepubliceerde boekbesprekingen af te leiden, langzamerhand is ook deze rubriek en daarbij vooral de bespreking van dissertaties op het terrein van het onderwijsonderzoek op gang gekomen. Waarschijnlijk is de redactie in staat om in de loop van dit jaar de achterstand die in de afgelopen jaren was ontstaan weg te werken.

Nogmaals doet de redactie hierbij het verzoek aan ieder die promoveert op het terrein van het onderwijsonderzoek een exemplaar van zijn dissertatie te zenden aan het redactiesecretariaat (adres: Postbus 1286, 9701 BG Groningen).

De belangrijkste verschuiving gedurende de afgelopen jaren is wellicht dat meer gerapporteerd wordt over onderwijsonderzoek, dan ten behoeve van onderwijsonderzoek in technische of methodologische zin. Dat is mede te danken aan de betrokkenheid van de lezers met het tijdschrift, die blijkt uit commentaar en kritiek, maar meer, of in elk geval wat de redactie betreft liever, uit het inzenden van manuscripten. De redactie gaat ervan uit dat ook in de komende jaren dit beleid verder kan worden doorgezet. Zij roept dan ook opnieuw iedereen op om manuscripten op het terrein van het onderwijsonderzoek voor publikatie aan te bieden. Graag zag de redactie dat de toenemende stroom van dissertaties op het terrein van het onderwijsonderzoek zich ook zou weerspiegelen in het aanbod van manuscripten aan het tijdschrift. Niet alleen omdat het mogelijk moet zijn om over dissertatie-onderwerpen een tijdschriftartikel te publiceren waarmee op bredere schaal aan het onderzoek bekendheid kan worden gegeven, maar ook omdat wanneer een dergelijk opus magnum in voorbereiding of afgesloten is, van schrijvers verwacht mag worden dat ze ook in de vorm van manuscripten een bijdrage aan het onderwijsonderzoek kunnen leveren of willen blijven leveren.

Verder nog enige mededelingen van huishoudelijke aard. Sinds jaar en dag is het de gewoonte om iets aan het uiterlijk van het tijdschrift te veranderen, beter ruggetje, ander lettertype en dergelijke. In dit kader dit jaar de aanwijzing dat in het vervolg definitieve versies van manuscripten op floppy en op papier moeten worden aangeleverd. De aanwijzingen daarvoor treft u aan in de binnenzijde van de omslag. De redactie is in de

loop van vorig jaar versterkt met Lieven Verschaffel. Hij breekt met de traditie dat een Belgisch lidmaatschap vooral corresponderend van aard is, tenminste als men brieven over het niet kunnen bijwonen van vergaderingen ook als zodanig opvat, en hij heeft in het afgelopen jaar reeds een hogere frequentie van aanwezigheid bij redactievergaderingen gehaald dan alle voorgaande Belgische redactieleden samen. We verwachten dus ook nog veel van hem en we wensen hem een plezierige tijd in de redactie van het Tijdschrift voor Onderwijsresearch toe.

Longitudinaal onderzoek van schoolloopbanen

Leo J.Th. van der Kamp en Marinus J.M. Voeten

Rijksuniversiteit Leiden en Katholieke Universiteit Nijmegen

ABSTRACT

Longitudinal studies of school careers.

This article serves as an introduction to a special issue of this journal on Dutch educational attainment research using longitudinal data. The first article gives a review of the developments of Dutch educational attainment research during the last decade. In the other four articles recent Dutch research has been reported on specific topics in the field of educational attainment. Not only is there a spread on the themes covered in the four articles, there is also a spread in methodology.

In dit themanummer zijn artikelen opgenomen die ontstaan zijn uit presentaties gehouden op de Onderwijs Research Dagen (ORD) 1989 te Leiden. Drie van de vijf bijdragen maakten deel uit van het symposium over longitudinaal onderzoek van schoolloopbanen, de beide andere bijdragen (De Jong & Meijer; Dijkstra) werden ook op de ORD gepresenteerd, resp. in 1988 en 1989.

In het thema longitudinaal onderzoek van schoolloopbanen zijn tenminste twee kernen te onderkennen: schoolloopbanen en longitudinale data-analyse. Deze inleiding heeft tot doel beide kernen afzonderlijk te belichten en tevens het thema als geheel te introduceren aan de hand van de in dit nummer opgenomen bijdragen.

SCHOOLLOOPBAANONDERZOEK IN NEDERLAND

Een van de eerste Nederlandse schoolloopbaanstudies van enige omvang is het onderzoek ondernomen in het kader van het zgn. Leidse talentenproject onder leiding van de socioloog Van Heek, waaraan onderzoekers met uiteenlopende achtergrond participeerden. Dit project probeerde onder meer de samenhang bloot te leggen tussen sociaal milieu van herkomst en geschiktheid tot het volgen van het middelbaar onderwijs dat toegang verschaft tot het hoger onderwijs. Enerzijds gaf dit project aanleiding tot het opzetten van betere studies, anderzijds had het ook 'enkele nuttige onbedoelde gevolgen'. Tot deze gevolgen rekent Dronkers de opzet van de zgn. 'Van Jaar tot Jaar' studie met een cohort bestaande uit een representatieve steekproef van lagere school leerlingen die tot op heden worden gevolgd. Tegelijkertijd werden andere cohorten gestart. Dronkers schetst in zijn bijdrage een aantal ontwikkelingen in het schoolloopbaanonderzoek, waarbij de nadruk wordt gelegd op schoolloopbaancohorten en minder op de resultaten van schoolloopbaanonderzoek. Het gaat om longitudinale cohorten die op onderzoeksvragen van velerlei aard antwoord kunnen geven. Als zodanig vormt dit type schoolloopbaandata geen zelfstandig thema in het onderwijsonderzoek, noch in het sociologisch onderzoek.

Dat schoolloopbaandata antwoord kunnen geven op onderzoeksvragen van velerlei aard blijkt uit de vier overige bijdragen. Elk van deze bijdragen stelt een bepaalde onderzoeksvraag en tracht deze vraag te beantwoorden aan de hand van data van een bepaald cohort. Deze longitudinale data worden telkens op verschillende manieren geanalyseerd. Inmiddels heeft zich een methodologie van longitudinale data-analyse ontwikkeld mede onder invloed van ontwikkelingen in schoolloopbaanonderzoek. Welke data-analyse gekozen moet worden hangt niet alleen af van de aard van de onderzoeksvraag, maar ook van het typische karakter van de data in kwestie (met name van meetniveau en verdelingsvorm).

In de bijdrage van Van der Hoeven-van Doornum, Voeten en Jungbluth wordt het effect van leerkrachtverwachtingen en streefniveaus onderzocht op schoolloopbanen in het basisonderwijs. Met dit onderzoek leveren de auteurs een bijdrage aan de verklaring van ongelijke onderwijskansen die niet kunnen worden toegeschreven aan verschillen in aanleg. Verwachtingen die leerkrachten van hun leerlingen hebben – en hun streefniveau van de leerlingprestaties – leveren een bijdrage aan de verklaring van de invloed van het sociale milieu op latere leerprestaties en vervolgadvis. Deze studie maakt gebruik van de methodologie van de lineaire structurele vergelijkingen. Er wordt een model gepostuleerd dat met de LISREL procedure op zijn geldigheid wordt onderzocht. De data-analyse is zorgvuldig uitgevoerd onder meer door de totale steekproef in twee delen te splitsen, nl. een exploratie-steekproef en een validatie-steekproef. Wordt de exploratie-steekproef gebruikt om (samen met theoretische bevindingen) het uitgangsmodel te formuleren, dit uitgangsmodel wordt toegepast op de validatie-steekproef en de uiteindelijke parameterschatting en hypothesetoetsing vindt plaats op de gegevens van de totale steekproef. Dit is een van de strategieën voor de systematische uitvoering van een LISREL procedure. Anderson en Gerbing (1988) bevelen een modellering aan in termen van structurele vergelijkingen voor het toetsen en ontwikkelen van theorieën in twee stappen. Hierbij wordt gebruik gemaakt van een reeks geteste modellen en sequentiële chi-kwadraat verschiltoetsen. Naast structurele modellen voor variabelen op intervalniveau gemeten zijn er nieuwe ontwikkelingen in LISREL, onder meer voor de analyse van ordinale variabelen (zie bijv. Jöreskog, 1989).

Een bijzondere vorm van de schoolloopbaan is de studieloopbaan in het universitaire onderwijs. De bijdrage van De Jong en Meijer richt zich op verschillen in studieloopbanen in zes studierichtingen, meer specifiek op uitval en vertraging. De vraag is onder meer op welke wijze de eventuele verschillen kunnen worden geïnterpreteerd in het licht van de doelstellingen van de Wet Tweefasenstructuur. De data betreffen een cohort van studenten die zich in 1982/83 voor het eerst lieten inschrijven bij een van de zes studierichtingen. Gegevens over vooropleiding, leeftijd, geslacht en studieresultaten over de eerste vier jaar werden verzameld. Er werd een structureel model voor de data gespecificeerd en de parameters van dit model werden vervolgens geschat. De uitkomsten van de data-analyse konden antwoord geven op de gestelde onderzoeksvragen. Aanzienlijke verschillen tussen studierichtingen werden geconstateerd.

De bijdrage van Dijkstra betreft een onderzoek in het kader van het project 'Schoolloopbanen in gereformeerd en nationaal onderwijs'. De probleemstelling luidt: 'Wijkt het schoolloopbaanverloop binnen een richting in het bijzonder onderwijs, die zich kenmerkt door een sterke godsdienstige geïnvolveerdheid, af van loopbaanpatronen zoals die voor het overige Nederlandse onderwijs kunnen worden vastgesteld?' Een specifiek cohort, het zgn. LVGS-cohort wordt vergeleken met het SLVO-cohort van het CBS. Relevante verschillen tussen beide cohorten worden geanalyseerd met multiële regressie- en padanalyse. In zekere zin zijn deze analysetechnieken te beschouwen als rudimentaire vormen van causale of structurele vergelijkingen modellering. Ook in deze bijdrage gaat

het om een gerichte substantiële onderzoeksvraag die beantwoord wordt met behulp van technieken voor (longitudinale) data-analyse, toegepast op gegevens van relevante cohorten.

De trits onderzoeksvraag - cohort - longitudinale data-analyse is ook duidelijk te onderscheiden in de bijdrage van Voeten en Boland. De globale onderzoeksvraag luidt: Op welke wijze wordt de kans op zittenblijven beïnvloed door achtergrondvariabelen. Het onderzochte cohort betreft leerlingen die in 1979 naar het eerste jaar van de lagere school gingen. Van deze leerlingen werden de gegevens verzameld tot het tweede jaar van het voortgezet onderwijs. Door als analyse-eenheid een tijdperiode (bijv. een schooljaar) per leerling te nemen, kunnen ook de gegevens van leerlingen worden geanalyseerd die op een later tijdstip in- dan wel uitstromen. Waar het in de analyse op neer komt is het onderzoeken van 'event histories'. De aanpak staat, m.n. in de sociologische literatuur, bekend als 'event history analysis' (zie bijv. Allison, 1984; Blossfeld, Hammerle & Mayer, 1989; Flinn & Heckman, 1982); in de medisch-biologische wetenschappen komt men de methode tegen onder namen als 'life-table regression' of 'survival analysis' (Cox & Oakes, 1984). De aanpak komt neer op regressie-analyse van longitudinale data met als observatie- en analyse-eenheid de bijzondere gebeurtenis in kwestie.

METHODEN VOOR LONGITUDINALE ANALYSE VAN SCHOOLLOOPBANEN

In principe kunnen de methoden en technieken voor longitudinale data-analyse (zie bijv. Plewis, 1985) alle toegepast worden in onderzoek naar schoolloopbanen. Zo er een algemeen geaccepteerde taxonomie zou bestaan voor methoden voor longitudinale data-analyse, dan zou deze verbijzonderd kunnen worden voor schoolloopbaanonderzoek. Het is waarschijnlijk nog te vroeg om een dergelijke taxonomie te formuleren: het aantal toepassingen van longitudinale analysemethoden in schoolloopbaanonderzoek is nog te beperkt, dit geldt ook voor vergelijkende studies naar de bruikbaarheid ervan. Uit de bijdragen aan dit themanummer is evenwel duidelijk dat bekende data-analysemethoden niet zonder meer zinvol toegepast kunnen worden op gegevens uit schoolloopbaanonderzoek. De methoden en technieken dienen aangepast te worden aan de specifieke aard van de longitudinale data. Er zal in ieder geval rekening moeten worden gehouden met de observatie- of analyse-eenheid, met het feit of er continue dan wel categorische variabelen in het spel zijn en met het al dan niet uitgaan van een bepaald model (bijv. het algemene lineaire model) ten grondslag aan de analysetechniek. Het behoeft nauwelijks betoog, dat primair inzicht nodig is in de aard van het te verklaren verschijnsel, voordat een verantwoorde keuze van onderzoeksdesigns en analyseprocedures kan worden gemaakt.

In de bijdragen aan deze bundel zijn aan de orde gekomen structurele vergelijkingsmodellen, event-history analyse middels logistische regressie en andere vormen van regressie-analyse. Maar gegeven de vraagstelling en de aard van de data zullen ook andere analysemethoden bruikbaar zijn voor de analyse van longitudinale cohort-data. Zo is correspondentie-analyse voorgesteld voor de analyse van event-history data (Van der Heijden en De Leeuw, 1989), maar correspondentie-analyse leent zich in het algemeen voor de analyse van categorische gegevens uit schoolloopbaanonderzoek (zie bijv. Greenacre, 1984; Van der Heijden, 1987). En zodra (multiple) correspondentie-analyse toegepast kan worden op bepaalde onderzoeksgegevens, vormt loglineaire analyse een alternatieve benaderingswijze voor dezelfde gegevens. Hiermee is het arsenaal van analysetechnieken voor gegevens uit schoolloopbaanonderzoek zeker niet uitgeput. We volstaan echter met deze opsomming die uitsluitend illustratief is.

Naast verschillen zijn er ook overeenkomsten in design en analyse tussen de vier

onderzoeken. Alle vier kunnen worden opgevat als panel-studies waarin metingen verkregen op twee of meer tijdstippen aan elkaar worden gerelateerd (met uitzondering van de bijdrage van Voeten en Boland, waarin de kans op zittenblijven als een (arbitraire) functie van de tijd wordt geanalyseerd). In alle studies wordt de tijd als discreet opgevat. Dit lijkt de dominante aanpak te zijn in longitudinaal onderzoek van schoolloopbanen. De aandacht lijkt meer uit te gaan naar bereikte status op verschillende punten in de tijd dan naar de bestudering van leren als proces van continue ontwikkeling in de tijd. Uit de biologische (en ontwikkelingspsychologische) traditie van bestudering van groeicurven zijn andere methoden van longitudinale data-analyse voortgekomen. De kern van deze methoden bestaat uit een model van *individuele* groei (Bock, 1976), gericht op het karakteriseren van individuele leertrajecten. Interindividuele verschillen in groeikarakteristieken kunnen vervolgens worden bestudeerd om na te gaan of er systematische relaties bestaan met relevante individuele of omgevingskenmerken (Willett, 1988). Een dergelijke aanpak kan de vorm krijgen van multilevel analyse (Bryk & Raudenbusch, 1987; Raudenbusch & Bryk, 1988; Bock, 1989).

CODA

De vraag of schoolloopbaanonderzoek sociologisch dan wel onderwijsonderzoek is, is op zichzelf weinig interessant. Interessanter is het na te gaan uit welk perspectief, of liever welke perspectieven, schoolloopbaanonderzoek is geëntameerd. Enerzijds is dit vanuit sociologisch perspectief, waarvan het Leidse talenproject een goed voorbeeld is. In beginsel was dit het eerste grootschalige onderzoek naar stratificatie in Nederland. Anderzijds is er het onderwijskundig perspectief waarbij het van groot belang is te onderzoeken welke (beïnvloedbare) kenmerken van de onderwijsomgeving systematisch samenhangen met interindividuele verschillen in groei. Dit betekent dat goede methoden nodig zijn voor de beschrijving van leertrajecten van individuele leerlingen, niet alleen op macro-sociologisch niveau maar ook op micro-niveau. Een voorbeeld waar longitudinale data-analyse overgaat in praktisch onderwijskundige toepassingen, is de ontwikkeling van leerlingvolgsystemen (Oud & Mommers, 1988; Oud, Van den Bercken, & Essers, in druk).

Het is moeilijk te voorspellen in welke richting het schoolloopbaanonderzoek zich zal bewegen. Zonder twijfel zullen de ontwikkelingen zoals geschetst door Dronkers in zijn bijdrage aan dit themanummer voortzetten. Het valt op dat ontwikkelingen in het onderzoek van schoolloopbanen sterk gekoppeld zijn aan de gebruikte methodologie voor de analyse van dergelijke longitudinale data: regressie-analyse, padanalyse, structurele vergelijkingsmodellen, log-lineaire analyse, event-history analyse, (multiple) correspondentie-analyse en multi-level analyse volgen elkaar op of wisselen elkaar af. Veel minder duidelijk is er een ontwikkeling in theorie-vorming te onderkennen. Weliswaar vormen schoolloopbanen geen zelfstandig thema van onderwijsonderzoek, zodat we ook geen theorie van schoolloopbanen kunnen verwachten. Maar we kunnen niet aan de indruk ontkomen, dat de voortgang van het onderzoek weinig door theorie-vorming wordt gestuurd.

LITERATUUR

- Allison, P. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA.: Sage.
- Anderson, J.C., & Gerbing, D.W. (1988). *Structural equation modeling in practice: A review and*

- recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Blossfeld, H.-P., Hammerle, A., & Mayer, R. (1989). *Event history analysis: Statistical theory and application in the social sciences*. Hillsdale, N.J.: Erlbaum.
- Bock, R.D. (1976). Basic issues in the measurement of change. In D.N.M. de Gruijter & L.J.T. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 75-96). New York: Wiley.
- Bock, R.D. (Ed.). (1980). *Multilevel analysis of educational data*. New York: Academic Press.
- Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147-158.
- Cox, D.R., & Oakes, D. (1984). *Analysis of survival data*. London: Chapman and Hall.
- Flinn, Ch.J., & Heckmann, J.J. (1982). New methods for analyzing individual event histories. In: S. Leinhardt (Ed.), *Sociological methodology 1982*. San Francisco: Jossey-Bass.
- Greenacre, M.J. (1984). *Theory and applications of correspondence analysis*. New York: Academic Press.
- Jöreskog, K.G. (1989). *New developments in Lisrel analysis of ordinal variables using polychoric correlations and weighted least squares*. Research Report 89-3. University of Uppsala: Department of Statistics.
- Oud, J.H.L., & Mommers, M.J.C. (1988). Longitudinale computerondersteunende onderkenning van lees- en spellingsmoeilijkheden: een toepassing van het Kalman-filter in de onderwijspraktijk. *Tijdschrift voor Onderwijsresearch*, 13, 31-49.
- Oud, J.H.L., Van den Bercken, J.H., & Essers, R.J. (in druk). Longitudinal factor score estimation using the Kalman-filter. *Applied Psychological Measurement*.
- Plewis, I. (1985). *Analysing change: measurement and explanation using longitudinal data*. New York: Wiley.
- Raudenbush, S.W., & Bryk, A.S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E.Z. Rothkopf (Ed.), *Review of Research in Education*, Vol. 15, 1988-89 (pp. 423-475). Washington: American Educational Research Association.
- Van der Heijden, P.G.M. (1987). *Correspondence analysis of longitudinal categorical data*. Leiden: DSWO-Press.
- Van der Heijden, P., & De Leeuw, J. (1989). Correspondence analysis with special attention to the analysis of panel data and event history data. In C.C. Clogg (Ed.), *Sociological methodology*, Vol. 19, 1989 (pp. 43-87). Oxford: Basil Blackwell.
- Willett, J.B. (1988). Questions and answers in the measurement of change. In E.Z. Rothkopf (Ed.), *Review of Research in Education*, Vol. 15, 1988-89 (pp. 345-422). Washington: American Educational Research Association.

De ontwikkelingen in het schoolloopbaanonderzoek: een terugblik op een decennium

J. Dronkers

Vakgroep Sociologie, Katholieke Universiteit Brabant

ABSTRACT

This article gives a review of the developments in Dutch educational attainment research during the last decennium, specifically research using longitudinal data is reviewed. Longitudinal educational attainment research is now one of the most flourishing fields of educational research. Eight developments are discussed: (1) increased connections with similar research in other countries; (2) broadening the focus of educational attainment research to the whole educational career from kindergarten to university; (3) using more and different indicators for measuring social background; (4) paying attention to differentiation within school types (e.g. student's choices of school subjects); (5) inclusion of school characteristics in the individual attainment model, together with the use of multilevel analysis; (6) comparative studies of schoolcareers of different cohorts of pupils, which differ in time period; (7) extensions of the educational attainment model to include occupational careers; (8) putting schoolcareers of women and ethnic minorities on the research agenda.

1. INLEIDING

Longitudinaal schoolloopbaanonderzoek behoort thans tot een van de bloeiendste gebieden van de onderwijsresearch. Regelmatig verschijnen er artikelen in de vaktijdschriften die gebaseerd zijn op longitudinale schoolloopbaandata. Ook dringen conclusies, bereikt met behulp van longitudinale schoolloopbaandata, geregeld door in de algemene pers. Tenslotte verschijnen er in de internationale vakliteratuur met een zekere regelmaat artikelen waarin gebruik wordt gemaakt van Nederlandse longitudinale schoolloopbaandata.

Toch is deze bloei van pas tamelijk recente datum. Tot de eerste twee decennia na de tweede wereldoorlog werd door een aantal onderzoekers met uiteenlopende achtergronden schoolloopbaanonderzoek uitgevoerd (zie Tesser, 1981). Sommige van de door hen gebruikte data hadden een longitudinaal karakter, maar meestal was de lengte van de onderzochte periode relatief kort. Door een aantal incidentele oorzaken kreeg één van deze schoolloopbaanstudies (Van Heek e.a., 1968) de meeste aandacht en werden de andere studies in mindere of meerdere mate vergeten. De studie die zo centraal kwam te staan was echter zeker niet die met de minst omstreden analyses of met de meest aanvaarde beleidsconclusies. In feite werden de analyses van Van Heek e.a. reeds snel vervangen door betere (o.a. Van Kemenade en Kropman, 1972) en het belangrijkste beleidsadvies (de doorstroomschool) werd nooit gerealiseerd. Maar dit 'Verborgene Talent' project had enkele nuttige onbedoelde gevolgen. Het leidde allereerst tot de zgn. 'Van Jaar tot Jaar'

Adres: Postbus 90153, 5000 LE Tilburg

Een eerdere, Engelstalige, versie van dit artikel werd gepubliceerd in Bakker, Dronkers en Meijnen (1989)

studie. Uit hetzelfde CBS-cohort van lagere schoolverlaters (1965) waarmee het 'Verborgene Talent' project werkte, werd door het I.T.S. een representatieve steekproef van lagere school leerlingen getrokken, die tot op de huidige dag, nu zij allang hun onderwijsloopbaan hebben afgesloten, worden gevolgd (Diederer, 1983, 1989). De dataverzameling, gefinancierd door het Ministerie van Sociale Zaken, maar niet door het Ministerie van Onderwijs en Wetenschappen of S.V.O., verliep de eerste tien jaar in het verborgene. Dit cohort-onderzoek leed toen aan de kwaal waaraan wel meer beginnende cohort-studies leiden: een laag onderzoeksrendement. Om vruchtbaar te kunnen zijn moeten meerdere golven data verzameld zijn. Dit verzamelen vraagt echter tijd en het cohort dreigt daardoor uit de belangstelling van de onderzoek(st)ers en subsidiënten te raken. Pas aan het eind van de jaren zeventig werd het 'Van Jaar tot Jaar' cohort 'opnieuw ontdekt' door onderzoekers en uitvoerig gebruikt voor secundaire analyses.

Deze 'herontdekking' van het 'Van Jaar tot Jaar' cohort viel samen met de start van een nieuw cohort, het zgn. SMVO-cohort. De reden voor dit cohort-onderzoek was duidelijk politiek en niet wetenschappelijk. Op aandrang van en gedeeltelijk gefinancierd door het Ministerie van Onderwijs en Wetenschappen werd door het CBS in 1977 een grote steekproef uit de eerste klassers van het voortgezet onderwijs getrokken en verzorgden het CITO en het ITS de aanvullende dataverzameling. Het Ministerie van Onderwijs en Wetenschappen wenste dit cohort-onderzoek, omdat op deze manier een belofte van minister Van Kemenade aan de kamer, gedaan in het debat rond middenschool en contourennota, ingelost kon worden. Door de medefinanciering van de dataverzameling door het Ministerie van Onderwijs en Wetenschappen was dit cohort tamelijk gemakkelijk bereikbaar voor onderzoek(st)ers. Dit leidde tot een groot aantal secundaire analyses met behulp van deze SMVO-data (zie voor het overzicht tot 1987, CBS, 1988). De schoolverlaters van dit SMVO-cohort zijn gedurende een aantal jaren ook nog onderzocht over hun arbeidservaringen. Echter, zo toegankelijk als het schoolloopbaan-deel van het SMVO-cohort was, zo ontoegankelijk is het beroepsloopbaan-deel van het SMVO-cohort voor de meeste onderzoek(st)ers. De angst voor aantasting van de privacy, die bij het CBS door het zo lang ontbreken van een goede wetgeving grote proporties heeft aangenomen, is de belangrijkste oorzaak van deze ontoegankelijkheid. Het succes van dit SMVO-cohort bracht het CBS ertoe op eigen kracht in 1982 een nieuw cohort te starten, het zgn. SLVO-cohort. Helaas is dit cohort tot nu toe nog niet goed toegankelijk voor secundaire analyse door niet-CBSers. Longitudinale data worden echter niet alleen door overheidsinstanties zoals het CBS of in opdracht van ministeries door bijv. het ITS verzameld. Ook lagere overheden (met name gemeenten) verzamelen gegevens over de schoolloopbanen van leerlingen in hun gemeenten. De bekendste en meest gebruikte gemeentecohorten zijn het VOCAL-cohort (gemeente Amsterdam in de jaren zestig en zeventig) en de GALO-data (gemeente Groningen tussen 1958 en 1970).

Ook universitaire instituten verzamelden longitudinale data over school- en beroepsloopbanen. Peschar (1975, 1978) en Bosman, Louwes en Van der Meer (1980) reconstrueerden met de GALO-data als uitgangspunt de school- en beroepsloopbanen van lagere school-leerlingen uit de stad Groningen in 1958/59 en 1970, teneinde de kracht van het secundaire effect van sociaal milieu en de verandering daarvan vast te stellen, of teneinde het effect van geslacht op de onderwijskansen van hun kinderen vast te stellen. Meijnen (1977, 1984) verzamelde geheel op eigen kracht schoolloopbaangegevens over een cohort lagere school-leerlingen uit de stad Groningen. Dit cohort werd gestart om de loopbanen tijdens de lagere school in kaart te brengen, wat gezien de resultaten van Van Heek een logische stap was. Dit zgn. Groningen-cohort wordt nog steeds onderzocht, ook nu de leerlingen het voortgezet onderwijs volgen (Bosker, Van der Velden en Otten, 1989). Een laatste particulier schoolloopbaancohort is dat van de LVGS (Dijkstra, 1989). Dit cohort

bevat de schoolloopbanen van leerlingen op een aantal vrijgemaakt-gereformeerde lagere scholen in de jaren tachtig. Ook zijn er longitudinale schoolloopbaancohorten, die met het oog op één bepaalde onderzoeksvraag verzameld zijn. Met name gaat het hier om cohorten, bestudeerd in het kader van het leesonderzoek (Van Dongen, 1984), de invoering van de middenschool (Coopmans, 1989) en het onderwijsvoorrangsbeleid. Ten slotte werd ook nog een longitudinaal school- en beroepsloopbaancohort geschapen door de respondenten van een oorspronkelijk als survey opgezette schoolloopbaanstudie opnieuw op te zoeken en te enquêteren (de zgn. Hartog uitbreiding van het Matthijssen-Sonnemans-cohort).

De nadruk in deze inleiding op de schoolloopbaancohorten in plaats van bepaalde onderzoeksresultaten die bereikt zijn met het schoolloopbaanonderzoek is met opzet gekozen. Het schoolloopbaanonderzoek is namelijk geen eigenstandig onderwerp binnen de onderwijsresearch. De verschillen in school- en beroepsloopbanen, die met al deze longitudinale cohorten in kaart gebracht kunnen worden, zijn op zich niet bijzonder interessant. Deze verschillen kunnen echter wel het uitgangspunt bieden voor allerlei soorten vragen, gesteld vanuit zeer verschillende perspectieven, die alleen met longitudinale data min of meer betrouwbaar beantwoord kunnen worden. Schoolloopbaandata zijn dus hulpmiddelen om onderzoeksvragen aan te pakken, niet een eigenstandig thema binnen de onderwijsresearch. Dat blijkt duidelijk als men de lijsten van secundaire analyses met de schoolloopbaandata bestudeert. Men kan bij de lijst voor het SMVO-cohort onderwerpen tegenkomen als: éénouder- en tweeoudergezinnen, regionale verschillen, werkende moeders, hoogbegaafden, prestatiemotivatie, verschillen tussen scholen, werkloosheid van ouders, arbeidsmarkt en etnische minderheden (zie CBS, 1988). Nog duidelijker wordt dat als men de nieuwsbrieven raadpleegt over de 'National Longitudinal Surveys of Labor Market Experience' van het Center for Human Resource Research van de Ohio State University. Dit hulpmiddelkarakter van schoolloopbaandata maakt dat dit overzichtsartikel geen volledige samenvatting geeft van alle resultaten van het schoolloopbaanonderzoek. Een dergelijke samenvatting zou zeer uiteenlopende onderzoeksterreinen moeten omvatten en zeer uiteenlopende theoretische kaders moeten schetsen, wat de omvang van een overzichtsartikel verre te boven zou gaan. Dat wil niet zeggen dat deze diversiteit van onderzoeksterreinen reeds vanaf het begin van het longitudinaal schoolloopbaanonderzoek aanwezig was. In het begin van dit onderzoek in Nederland was de ongelijke onderwijskansen van kinderen uit verschillende sociale milieus hét centrale thema. Dit oorspronkelijke accent binnen het schoolloopbaanonderzoek kan ook nu nog het beeld van de mogelijkheden van het schoolloopbaanonderzoek vertekenen, namelijk dat het alleen maar geschikt zou zijn om de relatie tussen sociaal milieu en onderwijskansen te onderzoeken.

Dit overzichtsartikel wil duidelijk maken dat de ontwikkelingen gedurende het laatste decennium rond het schoolloopbaanonderzoek juist tenderen in de richting van een steeds grotere diversiteit van onderzoeksvragen. Bij deze schets van het schoolloopbaanonderzoek baseer ik mij sterk op het werk van de leden van de SISWO-werkgroep 'Longitudinaal school- en beroepsloopbaanonderzoek'. Deze werkgroep omvat nagenoeg alle onderzoek(st)ers in Nederland en Vlaanderen die met longitudinale schoolloopbaandata zijn niet alleen onderwijsonderzoekers, maar ook sociologen, psychologen, historici, economen en methodologen lid van deze werkgroep, die in 1978 werd ingesteld door de stuurgroep Onderwijssociologie (het interuniversitair samenwerkingsverband van instituties en vakgroepen waarin onderwijssociologie bedreven wordt). De bundel uitgegeven ter gelegenheid van het 10-jarig bestaan van de werkgroep geeft een goed beeld van de

diversiteit van onderzoeksthema's (Bakker, Dronkers en Meijnen, 1989). Het is evenwel goed erop te wijzen dat de onderwijssociologie meer is dan alleen schoolloopbaanonderzoek (Rupp en Wesselingh, 1989).

Acht belangrijke ontwikkelingen in het schoolloopbaanonderzoek kunnen sinds 1978 onderscheiden worden:

1. De toegenomen band met het internationale onderzoek, zowel dat op het terrein van het schoolloopbaanonderzoek zelf als dat rond sociale stratificatie en mobiliteit.
2. De uitbreiding van de aandacht rond de overgang lager onderwijs / voortgezet onderwijs naar de gehele schoolloopbaan van kleuterschool tot universiteit.
3. De verdere verfijning van de meting van het ouderlijk milieu.
4. De invoering van horizontale verschillen in de schoolloopbaanmodellen.
5. De invoering van schoolkenmerken in het individuele schoolloopbaanmodel en het gebruik van multi-level modellen.
6. Het toenemend aantal vergelijkingen tussen cohorten, die verschillen in tijd.
7. De toevoeging van beroepsloopbanen aan het schoolloopbaanmodel.
8. Het opnemen van vrouwen en etnische minderheden op de onderzoeksagenda.

In de volgende paragrafen zal ik deze acht ontwikkelingen kort bespreken.

2. DE SAMENHANG MET HET INTERNATIONALE ONDERZOEK

De ontwikkelingen in het Nederlandse schoolloopbaanonderzoek van het laatste decennium zijn niet te begrijpen zonder kennis te nemen van de ontwikkelingen in het internationale onderzoek, zowel die op het terrein van het schoolloopbaanonderzoek zelf als die rond sociale stratificatie en mobiliteit. Van Heek was in de jaren vijftig een van de stichters van het internationale netwerk rond het sociale stratificatie- en mobiliteitsonderzoek en zijn werk weerspiegelde deze internationale contacten. Deze gingen echter verloren in de loop van de jaren zestig en zeventig. Dit verlies leidde er toe dat in Nederland relatief laat werd gereageerd op de vooruitgang in het ongelijkheidsonderzoek, zoals die het duidelijkst tot uitdrukking kwam in de studie van Blau en Duncan (1967). Pas de politieke invloed van het daarop voortbouwende boek van Jencks (1972) in de strijd rond middenschool en contourennota leidde tot een heropleving van het schoolloopbaanonderzoek. De eerste stap daarbij was de toepassing van de nieuwe buitenlandse modellen en technieken op de Nederlandse data: Dronkers en De Jong (1978) pasten het Jencks model toe en Dronkers (1978) het zgn. Wisconsin-model (zie Sewell en Hauser, 1980). Eenzelfde ervaring in het toepassen van internationale onderzoeksmodellen valt te signaleren in het verwante Nederlandse mobiliteits- en stratificatieonderzoek (Peschar en Ultee, 1978). Allereerst beïnvloedden methodische ontwikkelingen in het buitenland het Nederlandse onderzoek. Padanalyse, gebaseerd op normale multi-variate regressieanalyse, werd zo een populaire techniek om later verdrongen te worden door LISREL. Optimale schalingstechnieken in de vorm van HOMALS en PRINCALS deden hun intree. Loglineaire analyse (met name veel gebruikt in mobiliteitsonderzoek) en recentelijk 'event-history analysis' worden thans ook toegepast in het schoolloopbaanonderzoek. Tenslotte vonden multi-level-analyse technieken hun weg in het schoolloopbaanonderzoek. Maar ook buitenlandse theorieën en debatten beïnvloedden het Nederlandse schoolloopbaanonderzoek. In het eind van de jaren zeventig stond vooral de zgn. reproductie-these centraal, die ontwikkeld was door bijv. Bourdieu en Passeron (1977) en Bowles en Gintis (1976), maar later verschoof de aandacht met name door de studie van Rutter c.s. (1979) en Coleman, Hoffer en Kilgore (1982) naar de effecten van schoolkenmer-

ken op individuele onderwijskansen. Deze sterke internationale invloed op zowel de methoden als op de onderzoeksonderwerpen heeft voor- en nadelen. Het voordeel is dat daardoor het Nederlandse onderzoek gemakkelijker een internationaal niveau kan bereiken. Dat dit tegenwoordig ook daadwerkelijk lukt, mag blijken uit de vele publikaties in internationale tijdschriften en boeken, waarin Nederlandse schoolloopbaandata gebruikt worden. Wel moet nog een zekere achterstand, opgelopen in de jaren zestig en zeventig ten opzichte van collega's in overeenkomstige situaties (b.v. Scandinavië), ingehaald worden. Het nadeel is dat het Nederlandse onderzoek vaak blijft steken in een exegese van buitenlandse resultaten en methoden, zonder dat een originele bijdrage wordt geleverd aan de inhoud van het internationale debat. Het Nederlandse onderzoek is daardoor niet meer dan een zoveelste al dan niet interessante toepassing in een reeds lopend internationaal debat. De grootste kracht van het Nederlandse onderzoek is dan nog alleen de hoge kwaliteit van de schoolloopbaandata. Deze afhankelijke positie van het Nederlandse onderzoek is voor een belangrijke mate terug te voeren op de onbelangrijke culturele en politieke plaats van Nederland in Europa, waardoor het moeilijk is vanuit Nederland de toon te zetten in een internationaal debat. Wellicht doen wij er verstandig aan het unieke van het Nederlandse onderwijsbestel niet meer te benadrukken maar vooral Nederland (en dus ook haar onderwijs) te analyseren en te presenteren als een onderdeel van de Europese Gemeenschap. Daardoor zou het culturele en politieke belang van het Nederlandse sociaal-wetenschappelijk onderzoek voor het internationale debat kunnen toenemen, hetgeen de eigen theorie- en methodenontwikkeling zou kunnen stimuleren. Een eerste stap op deze weg is het systematische en gemeenschappelijke streven om via vertalingen, bewerkingen en publikaties de Nederlandse onderzoeksresultaten ook aan het internationale forum bekend te maken (bijv. Bakker, Dronkers en Ganzeboom, 1984; Bakker, Dronkers en Meijnen, 1989). Het nadeel van deze beoogde buitenlandse oriëntatie kan zijn dat daarmee wel de invloed van het schoolloopbaanonderzoek op het binnenlandse onderwijspolitieke debat, die thans wel aanwijsbaar is (bijv. in het WRR-rapport over basisvorming) kan afnemen, omdat dit debat zich geheel binnen een Nederlandstalig circuit afspeelt.

3. DE UITBREIDING VAN DE SCHOOLLOOPBAAN

Van Heek, maar ook de andere schoolloopbaanonderzoekers uit de jaren vijftig en zestig, concentreerden hun aandacht op de overgang van de lagere school naar het voortgezet onderwijs. Deze concentratie was erg begrijpelijk omdat tijdens deze overgang de openlijke keuze van ouders en leerlingen voor één van de hiërarchisch geordende typen voortgezet onderwijs blijkt. De eerste onderzoekers verwachtten daarom dat tijdens deze overgang getalenteerde kinderen uit de lagere sociale klassen niet die schooltypen kozen die overeenkwamen met hun schoolgeschiktheid. Hun eerste onderzoeksresultaten waren teleurstellend: jongens uit de arbeidersklasse kozen niet duidelijk te laag in vergelijking met hun schoolprestaties. Hoewel belangrijke delen van deze vroege conclusies nu wel definitief verworpen zijn, was een van de resultaten van deze teleurstelling een uitbreiding van de aandacht voor de gehele schoolloopbaan van kleuterschool tot universiteit. Men hoopte zo alsnog de selectie- en keuzeprocessen te achterhalen, waardoor de ongelijke onderwijskansen ontstonden. Meijnen met zijn zgn. Groningen-cohort (1977, 1983) is een van de belangrijkste onderzoekers van de schoolloopbaan gedurende de lagere school. Tot op zekere hoogte bevestigen zijn resultaten de conclusie dat ongelijke onderwijskansen niet ontstaan tijdens belangrijke beslispunten in de schoolloopbaan (zoals de eerste onderzoekers meenden) maar dat de ongelijkheid in schoolprestaties tussen kinderen uit verschillende milieus langzamerhand tijdens de lagere schoolperiode groeit om aan het eind van die lagere school min of meer

'gefixeerd' te worden. Deze 'fixatie' van milieuverschillen in schoolprestatie verschillen (ook wel het primaire milieu-effect genoemd, zie Boudon 1974) was de belangrijkste reden dat Van Heek c.s. zo weinig 'foute' keuzes onder arbeidersjongens vonden. Het mogelijke secundaire milieu-effect (de mate van milieu-invloed op keuzen in het middelbaar en hoger onderwijs, na controle voor schoolprestaties, ook een term van Boudon) was de belangrijkste reden om het schoolloopbaanonderzoek ook tot in het voortgezet en hoger onderwijs voort te zetten. Peschar (1976) toonde als een van de eersten overtuigend aan dat een dergelijk secundair milieu effect ook in Nederland bestond (zie de eerste jaargang van het Tijdschrift voor Onderwijsresearch).

Weliswaar is dit secundaire milieu-effect kleiner dan het primaire milieu-effect maar het is niet te verwaarlozen. Wel zijn er een aantal tekenen die wijzen op een afnemend secundair milieu-effect in de eerste fase van het voortgezet onderwijs (Faasse, Bakker, Dronkers en Schijf, 1986; Willemse, 1987), maar het is heel goed mogelijk dat in de tweede fase van het voortgezet onderwijs (bijv. rond de keuze van een vervolgopleiding) dit secundair effect juist sterker is geworden. De kracht van dit secundair milieu-effect is bovendien mede afhankelijk van de wijze waarop het bereikte onderwijsniveau wordt vastgesteld. Het lijkt erop dat een fijnere niveaumeting, waarin naast onderwijstype ook het vakkenpakket of de studierichting is opgenomen meer secundaire milieu-effecten onthult dan een grovere meting met slechts drie of vier niveaus (zie bijv. Bosker en Van der Velden, 1989). Bij de overgang naar het tertiair onderwijs is geen duidelijk secundair milieu-effect meer aantoonbaar (Kodde en Theunissen, z.j.), maar hierbij moet aangetekend worden dat het onderzoek rond deze overgang onderontwikkeld is. Hetzelfde geldt voor het onderzoek naar de loopbaan in het tertiair onderwijs. Hoewel ook onderontwikkeld, zijn er ook hier tekenen die wijzen op een afwezig secundair milieu-effect (De Jong en Meyer, 1988). Met het oog op een aantrekkelijk overheidsbeleid de studierendementen in het hoger onderwijs te verbeteren, is het van het grootste belang dat dit onderontwikkelde gebied van het schoolloopbaanonderzoek zo spoedig mogelijk ontgonnen wordt.

4. DE VERFIJNING VAN DE METING VAN HET OUDERLIJK MILIEU

De eerste schoolloopbaanonderzoekers gebruikten het beroep van de vader als de enige indicator van de sociale klasse of het sociaal milieu van de ouders. Tegenwoordig worden in het algemeen meer indicatoren gebruikt naast die van vader's beroep. De eerste uitbreiding is het gebruik van vader's opleiding en moeder's opleiding als tweede en derde indicator. Vaak worden deze drie indicatoren gecombineerd in een enkele milieu-index, hoewel dit methodologisch gezien niet meer nodig en theoretisch gezien onaantrekkelijk is. Het is immers zinvoller de afzonderlijke componenten van het ouderlijk milieu te onderscheiden en hun eigen betekenis voor de schoolloopbaan vast te stellen, dan te werken met een ondoorzichtige combinatie van milieu-kenmerken. Bovendien belet een combinatie van indicatoren ons het zicht op verschuivingen in hun betekenis en op de opkomst van nieuwe vormen van milieu-ongelijkheid. De eerste reden voor de uitbreiding van het aantal milieu-indicatoren is het streven naar een verbetering van de verklaring van de variantie in de schoolloopbaan. In het algemeen zijn vader's en moeder's opleiding betere voorspellers van schoolloopbaanvariantie dan vader's beroep (zie bijv. De Graaf, 1987). Moeder's beroep wordt bijna nooit gebruikt als een indicator voor sociaal milieu, waarschijnlijk als gevolg van de relatief lage arbeidsmarktparticipatie van gehuwde vrouwen met kinderen in Nederland. Maar Dronkers (1987) toonde aan dat het beroep van de moeder een zelfstandig effect heeft op de schoolloopbaan van haar kinderen, ook als men controleert voor de andere relevante sociaal milieu-indicatoren.

Een tweede reden voor de opkomst van nieuwe indicatoren van sociaal milieu ligt in de maatschappelijke ontwikkelingen rond gezin en arbeid. Er is aangetoond dat kinderen van eenoudergezinnen, die zijn ontstaan door echtscheiding, een lager onderwijsniveau behalen dan kinderen uit vergelijkbare tweeoudergezinnen (Bosman en Louwes, 1988). Hetzelfde geldt voor kinderen met werkloze vaders of met vaders in de WAO: zij behaalden ook een lager opleidingsniveau dan kinderen uit vergelijkbare gezinnen met werkende vaders (te Grotenhuis en Dronkers, 1989). De opkomst van deze nieuwe indicatoren van sociaal milieu kan wijzen op het ontstaan van een nieuwe sociale klasse van 'marginalen' of 'inactieven', naast de reeds bestaande ongelijkheden op grond van opleiding en beroep. De opkomst van deze nieuwe klasse is zeker niet uniek voor Nederland en wellicht zijn in landen met een minder krachtig uitgebouwd stelsel van sociale zekerheid de negatieve effecten zelfs sterker. Echter, het feit dat ook in de welvaart- en verzorgingsstaat van Nederland werkloosheid, echtscheiding en arbeidsongeschiktheid van ouders kan leiden tot slechtere onderwijsresultaten bij hun kinderen, wijst erop dat het wegnemen van financiële barrières niet voldoende is om het ontstaan van ongelijke onderwijskansen te verhinderen.

De derde reden voor de uitbreiding van het aantal sociaal milieu-indicatoren ligt in de poging het effect van het ouderlijk milieu op de schoolloopbanen beter te verklaren. De toegevoegde milieuvariabelen zijn dan ook geen exogene variabelen maar zijn intermediaire variabelen tussen de sociaal milieu-indicatoren en de afhankelijke schoolloopbaanvariabelen. Het gaat er dan om de effecten van het sociaal milieu te verklaren met behulp van deze intermediaire variabelen, zodat er geen direct effect van milieu op de schoolloopbaan meer gevonden wordt maar al de milieu-effecten indirect via de intermediaire variabelen verlopen. Meijnen (1977) probeerde dit als één van de eersten systematisch met behulp van indicatoren van de gezinscultuur (aspiraties, rolstructuur, cultureel-pedagogisch niveau), overigens zonder volledig in zijn doel te slagen. Daarna zijn meerdere pogingen tot het toevoegen van soortgelijke intermediaire variabelen ondernomen, over het algemeen met wisselend succes (vgl. De Graaf, 1987 met Bosker, Van der Velden en Otten, 1989). Op zich is het gebruik van nieuwe indicatoren als intermediaire variabelen tussen sociaal milieu en schoolloopbanen wetenschappelijk bevredigender dan een blote vaststelling van het effect van de sociaal milieu-indicatoren. Maar wellicht kennen wij de vele mogelijke gezinsprocessen die leiden tot ongelijke onderwijskansen te slecht om het gebruik van milieu-indicatoren als intermediaire variabelen tot een volledig succes te kunnen maken. Dronkers (1989) vond alvast dat de drie traditionele milieukenners (beroep vader, opleiding beide ouders) slechts 50% van de variantie verklaarden van de overeenkomst in onderwijsniveaus van broers en/of zusters.

5. HORIZONTALE VERSCHILLEN IN DE SCHOOLLOOPBAAN

Naast de veel onderzochte verticale differentiatie of hiërarchie in ons onderwijsstelsel kennen wij ook horizontale differentiatie. Binnen elk schooltype moet de leerling kiezen tussen verschillende vakkenpakketten, tussen afstudeerrichtingen, tussen trajecten of studierichtingen, die formeel wel gelijk zijn maar die toch tot verschillende onderwijs- en beroepsloopbanen kunnen leiden. Tot voor kort hadden de meeste onderzoekers alleen oog voor de verticale differentiatie en verwaarloosden ze de gevolgen van de horizontale differentiatie. Bosker en Van der Velden (1989) behoorden tot de eersten die de horizontale en de verticale verschillen probeerden te combineren in één schaal. Zij deden dit op basis van de doorstroommogelijkheden die een bepaalde combinatie van vakkenpakket of studierichting en schooltype biedt. Het belang van het onderscheiden van deze combinaties is dat de horizontale differentiatie een verborgen verticale differentiatie van leerlingen kan maskeren. Een bepaalde combinatie van vakken in een lager schooltype kan zo toch een hoger civiel effect hebben dan bepaalde

andere combinaties van vakkenpakketten in een 'hoger' schooltype. De verwaarlozing van deze horizontale differentiatie kan daardoor leiden tot een onderschatting van de mate van ongelijkheid in het onderwijs. Ook is er groeiende aandacht voor de gevolgen van deze horizontale differentiatie voor de verdere onderwijs- en beroepsloopbaan. Renique (1987) drong terecht aan op meer aandacht voor deze effecten. Bakker (1987) en Baeten, Van Cuyk-Remijssen, Dronkers, en 't Hof (1989) lieten zien dat er inderdaad verschillende gevolgen zijn van deze horizontale differentiatie op de beroepsloopbaan die niet terug te voeren zijn op een openlijke of verborgen verticale differentiatie binnen het onderwijs. Indien de huidige tendens van voortschrijdende horizontale differentiatie in het middelbaar en hoger onderwijs, ten dele als gevolg van de gestegen onderwijsdeelname, zich voortzet zal het belang van meer aandacht voor deze horizontale differentiatie groeien. Dit kan een nieuw onderzoeksterrein worden.

6. SCHOOLKENMERKEN EN INDIVIDUELE LOOPBANEN

De eerste schoolloopbaanonderzoekers, met uitzondering van Vervoort (1968), hielden zich nauwelijks bezig met de mogelijke effecten van schoolkenmerken op de individuele schoolloopbaan. Waarschijnlijk meenden zij dat het gecentraliseerde Nederlandse onderwijsbestel geen belangrijke effecten van schoolkenmerken op individuele schoolloopbanen toeliet. De uitkomsten van Coleman et al. (1966), die aantoonde dat de verschillen in materiële toerusting van scholen geen significant effect hadden op de leerprestaties van de leerlingen (na controle voor de leerlingpopulatie), bevorderden ook niet het onderzoek naar effecten van schoolkenmerken. Pas de verschuiving van de belangstelling voor materiële toerusting naar belangstelling voor schoolklimaat en effectiviteit (Rutter et al. 1979) opende nieuwe perspectieven voor de studie van effecten van schoolkenmerken in het Nederlandse onderwijsbestel. De eerste echte doorbraak rond dit onderwerp maakte Meijnen (1984), die aantoonde dat verschillende curriculumtypen in de bovenbouw van de lagere school uiteenlopende effecten hadden op de leerprestaties van de leerlingen en dat deze effecten niet overeenstemden met de populaire opvattingen daaromtrent. Afgezien van een enkele misser (Wolf, 1984) is het schoolkenmerkenonderzoek tot een bloeiende industrie geworden (zie bijv. Scheerens en Stoel, 1987; Scheerens en Verhoeven, 1989). Hierbij wordt druk gebruik gemaakt van longitudinale gegevens, omdat deze voldoende mogelijkheden geven te controleren voor mogelijke verstoringende effecten. Er zijn twee oorzaken voor deze bloei in het schoolkenmerkenonderzoek. Allereerst biedt dit onderzoek een uitgelezen mogelijkheid een belangrijk sociaal-wetenschappelijk probleem empirisch aan te pakken, nl. de relatie tussen individueel gedrag en contextkenmerken. Psychologisch georiënteerde onderzoekers zijn geneigd de oorzaken van gedrag te reduceren tot individuele kenmerken (het zgn. psychologisme) terwijl sociologisch georiënteerde onderzoekers geneigd zijn die oorzaken uitsluitend bij de structuren of de context te leggen (het zgn. sociologisme). Het schoolkenmerkenonderzoek biedt door de combinatie van individuele en contextgegevens en door de ontwikkeling van multi-level-analyse technieken de gelegenheid dit probleem niet op het niveau van overtuigingen en ideologieën maar op het niveau van empirisch onderzoek vast te stellen. Zo kan het schoolkenmerkenonderzoek niet alleen betekenis krijgen binnen de onderwijsresearch, maar kan zij ook een voorbeeld worden voor soortgelijk onderzoek in andere maatschappelijke sectoren (bijv. arbeids- en organisatiepsychologie en -sociologie), waarin het probleem van de verhouding individu-context ook een belangrijk thema is of behoort te zijn. Ook voor methodologen (bijv. Kreft, 1987) is dit schoolkenmerkenonderzoek een belangrijk terrein omdat juist in de sector onderwijs de beste data voor handen zijn over individueel gedrag en de context waarbinnen dat gedrag zich afspeelt. De tweede oorzaak

voor de bloei van het schoolkenmerkenonderzoek is een politieke. De maatschappelijke belangstelling voor de 'effectieve school' is in de jaren '80 sterk toegenomen, mede onder invloed van de effectieve schoolbeweging in de U.S.A. en de groeiende bereidheid om het verschaffen van collectieve goederen (waaronder onderwijs) te privatiseren of alvast te decentraliseren en te dereguleren. Op zichzelf is een dergelijke politieke steun aan een bepaald onderzoeksterrein niet bezwaarlijk. De publieke aandacht voor de ongelijkheid van onderwijskansen ontstond immers niet door de constatering dat die ongelijkheid bestond of door een mogelijke vergroting van die ongelijkheid maar door de aanvaarding van de 'Human Capital' idee als een van de moderne middelen voor wederopbouw en industrialisatie. Het gevaar van de politieke steun aan het schoolkenmerkenonderzoek ligt echter in de verleiding waaraan onderzoek(st)ers bloot staan om de betekenis van hun vondsten op te blazen. Dit gevaar van het overschatten van eigen resultaten is om vier redenen groot: 1) het geringe effect van schoolkenmerken in verhouding tot individuele effecten; 2) de instabiliteit van de schooleffecten; 3) de contextgebondenheid van schooleffecten; 4) de onduidelijkheid omtrent de mechanismen en processen die de schooleffecten tot stand brengen. Alle schoolkenmerken tezamen lijken slechts vijf tot vijftien procent van de individuele variantie in schoolloopbanen te kunnen verklaren. In verhouding tot de betekenis van de leerlingkenmerken en hun sociale achtergrond is dat niet veel. Bovendien verklaren de gemeten schoolkenmerken dan nog eens slechts ongeveer de helft van die variantie, de rest overlatend voor onbekende schoolkenmerken. Heel krachtige variabelen (in de betekenis van veel variantie kunnen verklaren; Ellemers, 1976) zijn de gemeten schoolkenmerken dus niet en hun betekenis kan derhalve snel overdreven worden. Daar komt nog bij dat de effecten van schoolkenmerken niet erg stabiel in de tijd lijken. Schoolkenmerken die in het ene jaar effect lijken te hebben bij een bepaalde groep scholen, hebben dat vaak in een ander jaar bij diezelfde groep scholen niet (Van den Eeden en Koopman, 1989). Dat betekent nogmaals dat de betekenis van een effectief schoolkenmerk gemakkelijk overdreven kan worden. Bovendien is ook gebleken dat het effect van een schoolkenmerk mede afhankelijk is van de context waarbinnen de vergeleken scholen moeten opereren (bijv. de denominatie van de school heeft in de regio Tilburg een tegengestelde betekenis dan die in de regio Amsterdam). Dat verkleint nogmaals de generaliseerbaarheid van de afzonderlijke onderzoeksresultaten. Tenslotte betekent de vaststelling van een significant schooleffect nog niet dat dan de mechanismen en processen bekend zijn die tot dit effect geleid hebben. Juist omdat nogal uiteenlopende processen en mechanismen verantwoordelijk kunnen zijn voor het gevonden effect, is men niet klaar met het vinden van het effect, hoe belangrijk op zichzelf ook. Zolang deze processen en mechanismen nog onbekend zijn, kan men in het schoolkenmerkenonderzoek nog voor onaangename verrassingen komen te staan en zullen onderzoek(st)ers de verleiding van het opblazen van de betekenis van hun vondsten moeten proberen te weerstaan. Dit wil niet zeggen dat de resultaten verzwegen of slechts in voor specialisten toegankelijke artikelen verstopt moeten worden. De wetenschapsgeschiedenis kent voldoende voorbeelden van maatschappelijk succesvol toegepaste onderzoeksresultaten, waarbij de mechanismen en processen nog onbekend waren (bijv. zwaartekracht), maar diezelfde geschiedenis kent nog meer voorbeelden van snelle toepassingen van onderzoeksresultaten die rampzalig bleken, juist omdat de mechanismen en processen nog onbekend waren.

7. VERGELIJKINGEN TUSSEN COHORTEN

Er bestaat zeker in vergelijking met het buitenland reeds een traditie van het maken van vergelijkingen tussen verschillende cohorten in het Nederlandse schoolloopbaanonderzoek. Deze traditie begon in Nederland al in 1982, voordat Campbell (1983) zijn Amerikaanse

collega's opriep vergelijkingen tussen cohorten te maken op grond van het zogenaamde Wisconsin-model van Sewell en Hauser (1980). De eerste Engelstalige review van deze Nederlandse traditie van vergelijkingen tussen cohorten dateert van 1983 (Dronkers, 1983). Het doel van deze vergelijkingen was het verschaffen van empirische bewijzen voor een mogelijke verschuiving in de balans tussen 'ascriptie' en 'achievement' binnen een onderwijsstelsel, tussen de effecten van milieu en prestaties. Deze mogelijke verschuiving is steeds een belangrijk strijdpunt geweest in het debat tussen de verschillende stromingen in met name de onderwijssociologie. Onderwijshervormingen zoals de mammoetwet, midden-school en basisvorming streven naar een onderwijsstelsel waarin de betekenis van het ouderlijk milieu, het geslacht en de etnische groep (ascriptieve kenmerken) in de schoolloopbaan zou worden teruggedrongen ten gunste van de betekenis van de schoolprestaties van de leerlingen (achievement kenmerk). Het terugdringen van de betekenis van 'ascriptieve' kenmerken ten gunste van de 'achievement' kenmerken wordt ook de meritocratisering (Young, 1958) genoemd. De voorlopige algemene conclusie uit een reeks vergelijkingen tussen verschillende cohorten is dat de balans tussen ascriptie en achievement is verschoven maar dat deze verschuiving niet erg groot is en gedeeltelijk geneutraliseerd wordt. Allereerst lijkt het effect van sociaal milieu op het leveren van schoolprestaties onveranderd (het zgn. primaire effect van milieu). Ten tweede lijkt het effect van sociaal milieu op de verdere schoolloopbaan (het zgn. secundaire effect van sociaal milieu), na controle voor schoolprestaties, te zijn afgenomen, en lijkt het effect van schoolprestaties op de verdere schoolloopbaan toegenomen. De combinatie van deze twee veranderingen lijkt verantwoordelijk voor de stabiliteit van de ongelijke onderwijskansen. De 'winst' van het afgenomen secundaire effect wordt gedeeltelijk teniet gedaan door de combinatie van een ongewijzigd primair effect en een versterkt effect van schoolprestaties. Deze conclusie is echter voorlopig omdat alleen nog maar schoolloopbanen rond de overgang van lager naar voortgezet onderwijs en in de eerste vijf jaren voortgezet onderwijs zijn vergeleken. Bovendien zijn tot nu toe slechts steeds twee cohorten paarsgewijs vergeleken, terwijl het noodzakelijk is om drie of meer cohorten gelijktijdig te vergelijken als men een trend zou willen vaststellen. Ook zouden de vergelijkingen tussen cohorten uitgevoerd moeten worden met een schoolloopbaanmodel waarin schoolkenmerken opgenomen zijn in plaats van het thans nog gebruikte individuele model.

Ondanks deze beperkingen hebben vergelijkingen tussen cohorten een grote maatschappelijke betekenis. De Mamoetwet is ingevoerd zonder een deugdelijke evaluatie van haar effecten op de ongelijkheid van kansen. Pas de hierboven aangeduide vergelijkingen maakten een dergelijke evaluatie mogelijk. Daarbij bleek dat de Mamoetwet door de invoering van het advies van het hoofd der school en de HAVO toch wel enig effect op de ongelijke onderwijskansen heeft gehad (Faasse e.a., 1986; Willemse, 1987). Een soortgelijke vergelijking tussen het fungerend onderwijsstelsel en de experimentele middenschole is echter in belangrijke mate mislukt (Coopmans, 1989), omdat de vrije toegang tot de data door de direct belanghebbenden geweigerd werd. Hopelijk gaat het rond de invoering van de basisvorming in het voortgezet onderwijs beter. De eerste tekenen (Peschar, 1988) wijzen daar alvast wel op.

8. BEROEPSLOOPBANEN

Vanaf het begin van het schoolloopbaanonderzoek in Nederland heeft de wens geleefd de beroepsloopbanen te incorporeren in de analyses. Deze wens vloeiende enerzijds voort uit de geschiedenis van het schoolloopbaanonderzoek. In het mobiliteitsonderzoek, waaruit het voortkwam, is het onderwijs slechts een tussenstap tussen het ouderlijk milieu en het latere

beroep van de kinderen. Anderzijds wordt in veel onderwijshervormingen ook verwezen naar de betekenis van het onderwijs voor de arbeidsmarktpositie van de afgestudeerden. Longitudinale school- en beroepsloopbaangegevens kwamen echter pas in het midden van de jaren tachtig beschikbaar: het 'Van Jaar tot Jaar'-cohort (het laatst ondervraagd toen de leerlingen ongeveer 25 jaar oud waren); het Matthijssen-Sonnemans-cohort (het laatst ondervraagd toen de leerlingen ongeveer 42 jaar oud waren); het schoolverlatersdeel van het SMVO-cohort (slechts de schoolverlaters in de eerste helft van de jaren tachtig zijn ondervraagd over hun arbeidsmarktpositie). Deze longitudinale gegevens maakten een definitieve replicatie van de resultaten van Jencks (1972, 1978) voor Nederland mogelijk (Dronkers en Bakker, 1986; Dronkers, 1987). Langzamerhand verschijnen er nu meer analyses met deze longitudinale school- en beroepsloopbanen (bijv. Meesters en Einerhand, 1986; Meesters en Van de Pol, 1989).

De mogelijkheden van de thans beschikbare longitudinale school- en beroepsloopbaan-data zijn echter nog nauwelijks verkend, maar ze lijken veelbelovend. De toekomstige studies van deze data zullen waarschijnlijk niet zozeer handelen over het effect van het ouderlijk milieu op het beroepsniveau van de kinderen, maar veel meer over de relaties tussen onderwijs en arbeidsmarkt en het functioneren van die arbeidsmarkt. Deze relaties lijken tot een belangrijk knelpunt te groeien, voortkomend uit de combinatie van groeiende onderwijsdeelname en diploma-inflatie, toegenomen meritocratie en hardnekkige al dan niet openlijke werkloosheid. Het longitudinale onderzoek is uitermate geschikt om de complexe selectieprocessen bij de entree op de arbeidsmarkt en gedurende de beroepsloopbaan te ontrafelen, omdat alleen longitudinale data voldoende mogelijkheden bieden om min of meer systematisch voor deze selectieprocessen te controleren. Cross-sectionele data, hoe aantrekkelijk ook door hun grote actualiteitswaarde, hebben op dit punt veel minder mogelijkheden en moeten daarom wel een vertekend beeld van de relaties tussen onderwijs en arbeidsmarkt geven. De analyse van beroepsloopbanen vraagt, doordat het hier om tijdsafhankelijke processen gaat, wel om andere onderzoeksmethoden. Met name 'event-history analysis' lijkt hiervoor een aangewezen methode, waarmee zowel in Nederland (Meesters en Van de Pol, 1989) als in Duitsland (Blossfeld en Mayer, 1988) succesvol beroepsloopbanen geanalyseerd zijn. Naast de studie van effecten van schoolkenmerken, is de introductie van beroepsloopbanen in het schoolloopbaanonderzoek een van de meest veelbelovende ontwikkelingen in het longitudinaal loopbaanonderzoek.

9. VROUWEN EN MINDERHEDEN

De eerste onderzoekers van schoolloopbanen verwaarloosden afwijkende schoolloopbanen van meisjes. Van Heek (1968) vermeldde bijvoorbeeld niet dat hij wél een reserve aan getalenteerde meisjes vond die in staat zouden zijn het VWO te volgen maar die toch een lager onderwijstype volgden. Hij was alleen in jongens geïnteresseerd. Gedurende de jaren zeventig en tachtig kregen de schoolloopbanen van vrouwen meer aandacht, zonder twijfel mede onder druk van de tweede feministische golf. De dissertatie van Diederik (1989) gaat uitvoerig in op die verschillen in schoolloopbanen tussen mannen en vrouwen. Het meest opvallende rond deze verschillen is hun veranderbaarheid én hun stabiliteit. Het is uit tal van studies duidelijk geworden dat het verschil tussen mannen en vrouwen in het bereikt onderwijsniveau snel aan het afnemen is. Deze verandering in ongelijke onderwijskansen tussen mannen en vrouwen is spectaculair in vergelijking met de bijna onveranderde ongelijkheid van onderwijskansen tussen sociale klassen. Toch zijn de oorzaken voor deze verandering onduidelijk. De lage deelname van Nederlandse vrouwen aan betaalde beroepsarbeid was in de jaren vijftig en zestig niet sterk veranderd. Deze stijging trad pas op

nadat het inlopen van de onderwijsachterstand op gang gekomen was. Hetzelfde geldt voor de invloed van de tweede feministische golf, ook die kwam pas op nadat dit inlopen was begonnen. Ook kan men niet beweren dat de machtsrelaties tussen mannen en vrouwen in de afgelopen decennia radicaal gewijzigd zijn. Ook de invoering van co-educatie in de jaren vijftig en zestig kan dit inlopen niet goed verklaren (Dronkers, 1987). Het inlopen van de onderwijsachterstand is geen specifiek Nederlands verschijnsel. Hoewel hier de in te lopen achterstand extra groot was, treedt dit verschijnsel in heel Europa op. Het ontrafelen van de oorzaken van dit inlopen van de onderwijsachterstand door vrouwen is erg belangrijk, omdat het ons iets kan leren over de mechanismen en processen in de samenleving die verandering van ongelijke onderwijskansen kunnen bevorderen of neutraliseren.

Echter, in een ander opzicht is het verschil tussen schoolloopbanen van mannen en vrouwen nogal onveranderd. De keuze van vakkenpakketten en studierichtingen verschilt tussen mannen en vrouwen nog bijna even sterk als vroeger. Deze zogenaamde horizontale segregatie tussen mannen en vrouwen is dus ondanks alle inlopen van onderwijsachterstanden niet veranderd. Deze stabiliteit is eigenlijk even raadselachtig als het inlopen van de achterstand.

Onderzoek naar de schoolloopbanen van etnische minderheden kwam in Nederland pas langzaam op gang, zowel door de ideologische geladenheid van het onderwerp als de privacy-gevoeligheid van de loopbaandata. De twee belangrijkste studies naar de schoolloopbanen van etnische minderheden zijn die van De Jong (1987) en Kerkhoff (1988). De centrale vraag lijkt hier de mate van overeenkomst tussen de patronen in schoolloopbanen van allochtonen en autochtonen. Naarmate deze patronen in de loopbanen meer op elkaar gelijken, zou de onderwijsachterstand van etnische minderheden minder te wijten zijn aan specifieke culturele barrières of aan discriminatie maar 'slechts' aan de lage sociaal-economische positie die de meeste allochtonen in Nederland innemen. Hoewel deze laatste factor niet licht onderschat kan worden, zijn er toch duidelijke aanwijzingen dat de schoolloopbanen tussen allochtonen en autochtonen van elkaar verschillen, ook na controle voor sociaal-economische positie, geslacht, leeftijd en CITO-score (Schouten, 1989).

10. CONCLUSIES

Het afgelopen decennium heeft het longitudinale schoolloopbaanonderzoek een snelle opgang doorgemaakt. Het heeft zich ontwikkeld van een 'one-issue' onderzoeksterrein naar een hulpmiddel voor een groot aantal uiteenlopende onderwerpen, zowel binnen de onderwijsresearch als daarbuiten. Deze groeiende diversiteit is van groot belang voor de verdere ontwikkeling van het longitudinale onderzoek. Het verzamelen van longitudinale data vereist veel geld, veel geduld en veel medewerking van derden. Bovendien wordt het verzamelen van schoolloopbaangegevens al dan niet terecht door leerlingen, ouders en instellingen als bedreigend ervaren voor hun privacy. Derhalve kan het verzamelen van longitudinale gegevens niet meer een individuele aangelegenheid zijn, maar behoort zij ten dienste te staan van de gehele wetenschappelijke gemeenschap. Dit betekent echter wel dat de verzamelde data een breed spectrum van mogelijke onderzoeksvragen moet kunnen bedienen. Uiteraard is het moeilijk vooraf de onderzoeksbehoeften van komende decennia nauwkeurig vast te stellen. Maar toch zijn er enkele richtlijnen op te stellen, die voortvloeien uit de ontwikkelingen van het afgelopen decennium. Allereerst moet de schoolloopbaan niet geïsoleerd in kaart gebracht worden, maar dient gezien te worden als onderdeel van de levensloop van de respondent. Dit betekent dat alle facetten van die levensloop in kaart gebracht moeten worden en niet alleen maar zijn of haar schoolloopbaan.

Ten tweede dient het proces-karakter van loopbanen uitgangspunt te zijn bij de dataverzameling. Alleen dan kan men voldoende inzicht krijgen in de (zelf-)selectie-processen.

Ten derde moeten de contexten waarbinnen het individu leeft completer en systematischer in kaart worden gebracht dan tot nu toe gebruikelijk is. Dus niet alleen schoolkenmerken, maar ook gegevens over het gezin, de broers en zussen, het vrienden en vriendinnen-netwerk, de woonbuurt, de alternatieven bij de verschillende keuze-momenten in de loopbaan zullen verzameld moeten worden.

Voorlopig ligt er dus nog genoeg werk voor de boeg. Gelukkig behoeven wij niet te wachten op nieuwe databestanden. Ook de thans reeds beschikbare bestanden zijn nog niet uitputtend gebruikt. Daarom beloven de komende jaren nog veel vruchtbaar schoolloopbaanonderzoek.

LITERATUUR

- Baeten, J.J.M., Cuyk-Remijssen, A. van, Dronkers, J. & 't Hof, L., (1989, in druk). Horizontale differentiatie in het onderwijs en ongelijkheid op de arbeidsmarkt. *Sociale Wetenschappen*.
- Bakker, B.F.M. (1987). Onderwijs en de kans op betaald werk. In G.W. Meijnen, A.F.M. Nieuwenhuis, & J.L. Peschar (Red.), *Selectie en kwalificatie in het onderwijs* (p. 149-164). Lisse: Swets en Zeitlinger.
- Bakker, B.F.M., Dronkers, J. & Ganzeboom, H.B.G. (1984). *Social stratification and mobility in the Netherlands*. Amsterdam: SISWO.
- Bakker, B.F.M., Dronkers, J. & Meijnen, G.W. (Eds.). (1989). *Educational opportunities in the welfare state; longitudinal studies in educational and occupational attainment in the Netherlands*. Nijmegen: I.T.S.
- Blau, P.M. & Duncan, O.D. (1967). *The American Occupational Structure*. New York: Wiley.
- Blossfeld, H.P. & Mayer, K.U. (1988). Labor market segmentation in the Federal Republic of Germany. *European Sociological Review*, 4, 123-140.
- Bosker, R.J. & Van der Velden, R.K.W. (1989). Schooleffects on educational perspectives. In B.F.M. Bakker, J. Dronkers & G.W. Meijnen (Eds.), *Educational opportunities in the welfare state* (p. 131-150). Nijmegen: I.T.S.
- Bosker, R., Van der Velden, R. & Otten, L. (1989). Social stratification and educational career. In W. Jansen, J. Dronkers & K. Verrips (Eds.), *Similar or different? Continuities in Dutch research on social stratification and social mobility* (p. 99-120), Amsterdam: SISWO.
- Bosman, R. & Louwes, W. (1988). Eenouder- en tweecoudergezinnen en schoolloopbanen. *Mens en Maatschappij*, 63, 5-23.
- Bosman, R., Louwes, W. & Van der Meer, A. (1979). Sexe, school, beroep. *Info*, 11, 217-343.
- Boudon, R. (1974). *Education, inequality and social opportunity*. New York: Wiley.
- Bourdieu, P. & Passeron, J.C. (1977). *La Réproduction*. Parijs: Minuit.
- Bowles, S. & Gintis, H. (1976). *Schooling in Capitalist America*. New York: Basic Books.
- Campbell, R.T. (1983). Status attainment research: end of the beginning or beginning of the end? *Sociology of Education*, 56, 47-62.
- CBS (1988). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs, deel 3; cohort 1977, schoolloopbanen en rendementen; stand na 7 jaar*. Den Haag: Staatsuitgeverij.
- CBS (1988). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs SLVO-cohort 1982; deel 1 schoolkeuze*. Den Haag: Staatsuitgeverij.
- Coleman, J.S. e.a. (1966). *Equality of Educational Opportunity* (vol. 1 en 2). Washington D.C.
- Coleman, J.S., Hoffer, T. & Kilgore, S. (1982). *High School Achievement: public, catholic and private schools compared*. New York: Basic Books.
- Coopmans, J. (1989). *De middenschool uit; eindverslag van onderzoek bij leerlingen die de middenschool afsluiten*. Nijmegen: ITS.

- Diederer, J. (1983). *Van Jaar tot Jaar. Derde fase: de keuze van een beroep*. Nijmegen: ITS.
- Diederer, J. (1983). *Meer of minder ongelijk: over sexeverschillen in schoolloopbaan en beroepskeuze en over de gedeeltelijke conditionering van die sexeverschillen door sociaal milieu en schoolprestatie*. Nijmegen: ITS.
- Dongen, D. van (1984). *Leesmoeilijkheden: naar diagnostiserend onderwijzen bij het leren lezen*. Tilburg: Zwijssen.
- Dronkers, J. (1978). Manipuleerbare variabelen in de schoolloopbaan. In J.L. Peschar en W.C. Ultee (Red.), *Sociale Stratificatie* (p. 116-143). Deventer: Van Loghum Slaterus.
- Dronkers, J. (1983). Have inequalities in educational opportunity changed in the Netherlands? A review of empirical evidence. *The Netherlands Journal of Sociology*, 19, 133-150.
- Dronkers, J. (1987). Ouders, liefde en geld; de relaties tussen ouderlijk milieu, schoolloopbaan, beroep, huwelijk en gezinsinkomen bij vrouwen. In G.W. Meijnen, A.F.M. Nieuwenhuis & J.L. Peschar (Red.), *Selectie en kwalificatie in het onderwijs* (p. 121-136). Lisse: Swets en Zeitlinger.
- Dronkers, J. (1987). *Het effect van de schoolkenmerken richting en coeducatie op de individuele prestaties en het vervolgdadvies in de jaren vijftig*. Paper zevende Onderwijssociologische conferentie, Lunteren.
- Dronkers, J. (1987). Een empirisch onderzoek naar de effecten van betaalde beroepsarbeid door vrouwen op de schoolloopbaan van hun kinderen. *Pedagogische Studiën*, 64, 277-284.
- Dronkers, J. (1989). The power of the family. In W. Jansen, J. Dronkers & K. Verrips (Eds.), *Similar or different; continuities in Dutch research on social stratification and social mobility* (p. 121-136). Amsterdam: SISWO.
- Dronkers, J. & Bakker, B. (1986). Van het dubbeltje en het kwartje; de relatie tussen milieu, onderwijs, partnerkeuze, beroep en gezinsinkomen. *Economisch-Statistische Berichten*, 71, 1184-1189.
- Dronkers, J. & Jong, U. de (1978). Jencks en Fägerlind op zijn Hollands, *Sociologische Gids*, 25, 4-30.
- Dijkstra, A. (1989). Schoolloopbanen, godsdienst en sociaal milieu. In J. van Damme & J. Dronkers (Red.), *Jongeren in school en beroep*. Lisse: Swets en Zeitlinger.
- Eeden, P. van den, & Koopman, P. (1989). The stability of school effects in a centralized educational system: the case of Amsterdam. In B.F.M. Bakker, J. Dronkers & G.W. Meijnen (Eds.), *Educational Opportunities in the welfare state* (p. 119-130). Nijmegen: ITS.
- Ellemers, J.E. (1976). Veel kunnen verklaren of iets kunnen veranderen: krachtige versus manipuleerbare variabelen. *Beleid en Maatschappij*, 3, 281-291.
- Faasse, J.H., Bakker, B., Dronkers J. & Schijf, H. (1986). Voor en na de mammoet. Vergelijking van individuele schoolloopbanen in het voortgezet onderwijs in Noord-Brabant voor en na de invoering van de mammoet-wet. *Mens en Maatschappij*, 61, 75-94.
- Graaf, P. de (1987). De invloed van financiële en culturele hulpbronnen in schoolloopbanen. Nijmegen: ITS.
- Grotenhuis, H. te, & Dronkers, J. (1989). Enkele gevolgen van werkloosheid en arbeidsongeschiktheid in de verzorgingsstaat; ongelijke onderwijskansen van kinderen. *Amsterdams Sociologisch Tijdschrift*, 15, 634-651.
- Heck, F. van, et al. (1968). *Het verborgen talent: milieu, schoolkeuze en schoolgeschiktheid*. Meppel: Boom.
- Jencks, C. et al. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.
- Jencks, C. et al. (1978). *Who gets ahead? The determinants of economic success in America*. New York: Basic Books.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen; allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs*. Lisse: Swets en Zeitlinger.
- De Jong, U. & Meyer, J. (1988). *Studie-uitval en -vertraging bij zes studierichtingen in het wetenschappelijk onderwijs*. Paper Onderwijsresearchdagen 1988, Leuven.
- Kemenade, J.A. van & Kropman, J.A. (1972). Verborgen talenten? Kritische kanttekeningen bij een onjuiste interpretatie. *Sociologische Gids*, 19, 219-229.
- Kerckhoff, A. (1988). *Taalvaardigheid en schoolsucces*. Lisse: Swets en Zeitlinger.
- Kodde, D.A. & Theunissen, M.A.M. (z.j.). *De schoolloopbaan van de snelle VWO-leerlingen, sexe-ongelijkheid en de keuze voor hoger onderwijs*. researchmemorandum 8410, Nijmegen.
- Kreft, G.G. (1987). *Models and methods for the measurement of schooleffects*. Proefschrift Universiteit van Amsterdam.

- Meesters, M.J. & Einerhand, M.G.K. (1986). LBO- en Mavoschoolverlaters op de arbeidsmarkt. *Tijdschrift voor arbeidsvraagstukken*, 2, 67-77.
- Meesters, M.J. & Pol, F.J.R. van de (1989). Once unemployed, always unemployed? In B.F.M. Bakker, J. Dronkers & G.W. Meijnen (Eds.), *Educational opportunities in the welfare state* (p. 167-182). Nijmegen: ITS.
- Meijnen, G.W. (1977). *Maatschappelijke achtergronden van intellectuele ontwikkeling*. Groningen: Wolters-Noordhoff.
- Meijnen, G.W. (1984). *Van zes tot twaalf, een longitudinaal onderzoek naar de milieu- en schooleffecten van loopbanen in het lager onderwijs*. Den Haag: SVO.
- Peschar, J.L. (1975). *Milieu, school, beroep*. Groningen: Tjeenk Willink.
- Peschar, J.L. (1978). Educational opportunity within and between Holland and Sweden; the semi-experimental approach. *Sociologische Gids*, 25, 273-296.
- Peschar, J.L. (1988). *Evaluatie van de basisvorming; kader voor het uitvoeringsplan*. Den Haag: SVO.
- Peschar, J.L. & Ultee, W.C. (Red.). (1978). *Sociale Stratificatie*. Deventer: Van Loghum Slaterus.
- Renique, A.J.E.G. (1987). Nabeschouwing bij het thema selectie en kwalificatie. In G.W. Meijnen, A.F.M. Nieuwenhuis & J.L. Peschar (Red.), *Selectie en kwalificatie in het onderwijs* (p. 173-182). Lisse: Swets en Zeitlinger.
- Rupp, J.C.C. & Wesselingh, A.A. (1989). Onderwijs en maatschappelijke ongelijkheid; een overzicht van 25 jaar theorievorming en onderzoek in Nederland. *Tijdschrift voor Onderwijsresearch*, 14, 237-258.
- Rutter, M. et al., (1979). *Fifteen Thousand Hours; secondary schools and their effects on children*. Somerset: Open books.
- Scheerens, J. & Stoel, W.G.R. (Red.). (1987). *Effectiviteit van onderwijsorganisaties*. Lisse: Swets en Zeitlinger, 1987.
- Scheerens, J. & Verhoeven, J.C. (Red.). (1989). *Schoolorganisatie, beleid en onderwijskwaliteit*. Lisse: Swets en Zeitlinger.
- Schouten, S.P. (1989). *Schoolloopbanen van Turken en Marokkanen in het SLVO-cohort*. Paper Onderwijsresearchdagen 1989, Leiden.
- Sewell, W.H. & Hauser, R.M. (1980). The Wisconsin longitudinal study of social and psychological factors in aspirations and achievements. In A.C. Kerckhoff (Red.), *Research in Sociology of Education and Socialisation*, vol. 1 (p. 59-100). Greenwich CT: JAI-Press.
- Tesser, P. (1981). *Schoolloopbanenonderzoek in Nederland; een trendstudie naar ontwikkelingen in het empirisch onderzoek tussen 1960 en 1980*. Nijmegen; ITS.
- Vervoort, C.E. (1968). *Gezin en schoolkeuze bij handarbeiders*. Proefschrift Universiteit van Leiden.
- Willemsse, P. (1987). Overleving in de bovenstroom van het voortgezet onderwijs: veranderde milieu-effecten voor jonge meisjes. *Tijdschrift voor Onderwijsresearch*, 12, 329-343.
- Wolf, J.C. van der (1984). *Schooluitval*. Lisse: Swets en Zeitlinger.
- Young, M. (1985). *The rise of the meritocracy, 1870-2033*. London: Thames and Hudson.

Manuscript ontvangen 11-9-1989

Definitieve versie ontvangen 2-11-1989

Het effect van leerkrachtverwachtingen en streefniveaus op schoolloopbanen in het basisonderwijs

Anna A. van der Hoeven-van Doornum*, Marinus J.M. Voeten** en Paul Jungbluth*

* *Instituut voor Toegepaste Sociale Wetenschappen, K.U. Nijmegen* en ** *Vakgroep Interdisciplinaire Onderwijskunde, K.U. Nijmegen*

ABSTRACT

The effect of teacher expectations and aspiration levels on school careers of pupils in primary education.

Aptitude and socio-economic status (SES) are the main factors influencing school careers. The aim of the study was to explain the relationship between SES and school careers. To this end, we investigated the effects of teacher expectations and aspiration levels teachers set for their pupils, on school careers. A longitudinal study was carried out with 11 and 12-years old pupils (grades 7 and 8) of 50 schools for primary education. Questionnaires about expectations and aspiration levels were filled out by the teacher for each individual pupil. In addition, there were achievement tests and an intelligence test for the pupils. Information about SES was directly collected from the parents. We hypothesized that the effects of SES on learning achievement and on school career are mediated by teacher expectations and aspiration levels. A LISREL-model was developed to test this hypothesis. It is shown that teacher expectations and aspiration levels, in grade 7 as well as grade 8, do partly explain the effects of SES on learning achievement and on recommended type for secondary education. It is our conclusion that teacher expectations and aspiration levels have longterm and sustaining effects on learning achievement, thereby influencing the school careers of pupils.

INLEIDING

In recent onderzoek naar schoolloopbanen wordt veelal gebruik gemaakt van structurele modellen (LISREL) om de relatie te analyseren tussen het sociale milieu van leerlingen en hun schoolloopbaan (toetsprestaties, adviezen voor voortgezet onderwijs, bereikt onderwijsniveau). De aandacht gaat hierbij vooral uit naar de overgang van basisonderwijs naar voortgezet onderwijs. Zoals bekend hangen verschillen in onderwijsposities van leerlingen in het voortgezet onderwijs samen met de sociale herkomst van de leerlingen. Volgens de genoemde structurele modellen (o.a. De Jong, Dronkers & Saris, 1982) kan deze samenhang grotendeels teruggevoerd worden op indirecte effecten van milieu via de schoolprestaties in het basisonderwijs. Zoals opgemerkt door Tesser (1986) dragen deze modellen echter weinig bij tot inzicht in de relatie tussen sociale herkomst en bereikt onderwijsniveau, omdat Cito-toetsscore, advies van het schoolhoofd en onderwijspositie na de brugklas opgevat kunnen worden als indicatoren van één latente variabele: schoolsucces.

Tesser introduceert de variabelen intellectueel thuis klimaat, faalangst en intelligentie als

Adres: Instituut voor Toegepaste Sociale Wetenschappen, Postbus 9048, 6500 KJ Nijmegen

Met dank aan Simone van Zolingen voor haar bijdrage aan de literatuurstudie

mediërende variabelen ter verklaring van de samenhang tussen sociale herkomst en schoolsucces. Alleen de variabele intelligentie bleek hierbij bruikbaar. Er blijven in zijn model echter significante directe effecten van milieuvariabelen op schoolsucces (Tesser, 1986, p. 59). Door anderen (Meijnen, 1984; De Graaf, 1987) wordt de verklaring vooral gezocht in gezinskenmerken. In dit artikel willen we de aandacht richten op binnen-schoolse processen als verklaringsbron voor effecten van sociaal milieu op schoolloopbanen van leerlingen.

Sociaal milieu en schoolloopbaan

Door o.m. Blok & Saris (1980) en De Graaf (1987) wordt een direct effect gevonden van sociaal milieu op het advies van het schoolhoofd, waarbij de leerprestaties in het laatste leerjaar constant gehouden zijn. Door Visser (1987) zijn de schoolprestaties opgesplitst in vier aspecten van het schoolse leren, te weten technische leesvaardigheid, begrijpend lezen, spelling- en rekenvaardigheid gemeten in leerjaar 5 en 8. In de analyse voor leerjaar 8 bleek een aanzienlijk direct effect van sociaal milieu op advies. Naast de genoemde schoolprestaties was daarbij ook de Citotoets constant gehouden. Dit directe effect was ongeveer de helft van het totale effect van sociaal milieu op het advies. Bij vervanging van de schoolprestaties van leerjaar 8 door die van leerjaar 5 kreeg sociaal milieu naast een direct effect op advies ook een direct effect op de Cito-score. Dit zou kunnen betekenen dat sociaal milieu ook positief samenhangt met relatieve veranderingen in schoolprestaties gedurende het basisonderwijs. Boland, Voeten & Visser (1989) ontwikkelden, op hetzelfde databestand, een longitudinaal model voor de leerjaren 4, 5 en 8 met lees- en spellingvaardigheid. Hieruit bleek dat ook wanneer rekening gehouden wordt met de ontwikkeling in schoolprestaties, een significant direct effect blijft bestaan van sociaal milieu op het advies voor vervolgonderwijs. Het indirecte effect van sociaal milieu via de schoolprestaties bleek m.n. te verlopen via begrijpend lezen. Waarschijnlijk wordt de ontwikkeling in spellingvaardigheid en in technisch lezen meer direct door het onderwijs beïnvloed, terwijl begrijpend lezen hoog gecorreleerd is met intelligentie. In ieder geval duiden deze resultaten erop dat het voor onderzoek naar milieuspecifieke schoolloopbanen van belang is te differentiëren naar de aard van de leerprestaties.

Wat betreft het effect van sociaal milieu op de ontwikkeling in de schoolvorderingen gedurende de basisschoolperiode lijken de resultaten van Boland et al. te kunnen worden verklaard uit de hypothese, dat de aan- of afwezigheid van een dergelijk effect vooral afhangt van de lengte van het tijdsinterval tussen de metingen van de schoolprestaties. Meet men met betrekkelijk korte tijdsintervallen (een leerjaar of minder) dan zal sociaal milieu alleen een effect hebben op de beginmeting en niet meer rechtstreeks op latere metingen. Het al dan niet vinden van evidentie voor een "cumulatief deficit" (De Graaf, 1987 op de data van Meijnen, 1984) op basis van padcoëfficiënten lijkt daarmee een functie van de lengte van het tijdsinterval tussen de prestatiemetingen.

Vraagstelling

In dit onderzoek kiezen we het advies voor vervolgonderwijs, gegeven aan het eind van de basisschool, als de te verklaren variabele. Gezien de hoge samenhang tussen advies en feitelijk gekozen type brugklas (Boland et al., 1989), is het echter redelijk te veronderstellen, dat de resultaten ook van toepassing geacht kunnen worden op de feitelijk gemaakte keuzes voor voortgezet onderwijs, aan het eind van de basisschool.

We maken hier onderscheid tussen enerzijds het effect van sociaal milieu op het prestatieniveau aan het eind van de basisschool en anderzijds het directe effect van sociaal milieu op advies, gegeven het prestatieniveau. Het belang van dit onderscheid ligt vooral in het gegeven dat het uitbrengen van een advies een beslissing impliceert van een leerkracht.

Dit onderscheid komt deels overeen met het door Boudon (1974) geïntroduceerde onderscheid tussen primaire en secundaire milieu effecten. We zullen in het onderstaande deze termen gebruiken.

Hoe kan nu de samenhang tussen sociaal milieu en het advies voor vervolgonderwijs bij gelijke schoolprestaties verklaard worden? Een mogelijkheid is dat de leerkracht bij zijn/haar advies mede rekening houdt met andere kenmerken van het kind, kenmerken die gecorreleerd zijn met sociaal milieu. Persoonlijkheidskenmerken van het kind (intelligentie, prestatiemotivatie, neurotische labiliteit, introversie-extraversie, testattitude) lijken slechts weinig bij te dragen aan de verklaring van verschillen in adviezen, als schoolprestaties constant gehouden worden (Groeneboom, Hoogstraten, Mellenbergh, & Van Santen, 1978). Saris & Blok (1982) vonden dat de relatie tussen schoolprestaties en advies verschillend is voor verschillende scholen. Hierbij werd ook een milieu-effect gevonden: naast een direct effect op het advies werd ook een indirect effect gevonden via de "adviseringsstrategie" van de leerkracht. Hierbij werd het oordeel van de leerkracht over de schoolprestaties van de leerlingen (een predictie van de score op de Cito-toets) constant gehouden.

Niet alleen de secundaire effecten van sociaal milieu vragen om een verklaring, ook de primaire effecten worden door de boven genoemde structurele modellen nog niet verklaard. Door o.m. Jungbluth (1985) is de aandacht gericht op de beelden die leerkrachten zich van hun leerlingen vormen en op de verwachtingspatronen van leerkrachten ten aanzien van de prestatiegeschiktheid van hun leerlingen. Deze leerkrachtverwachtingen zouden volgens Jungbluth leiden tot differentiatie in nagestreefde prestatieniveaus en daarmee tot verschillen in onderwijsaanbod. Volgens deze gedachtengang zouden leerlingen uit verschillende sociale milieus aan het eind van het basisonderwijs tot verschillende schoolprestaties komen, omdat ze verschillend onderwijs hebben ontvangen.

Het doel van dit onderzoek is om na te gaan of leerkrachtverwachtingen en differentiatie in nagestreefd prestatieniveau (streefniveaus) inderdaad bijdragen aan de verklaring van de samenhang van sociaal milieu met leerprestaties en adviezen voor vervolgonderwijs. Wij gaan dit na door leerkrachtverwachtingen en streefniveaus in een schoolloopbaanmodel op te nemen als mediërende variabelen tussen enerzijds sociaal milieu en anderzijds schoolprestaties en advies. Als de these van Jungbluth juist is, dan zouden de directe effecten van sociaal milieu op leerprestaties en vooral op advies moeten verdwijnen (of in ieder geval sterk moeten afnemen). Hierbij dienen intelligentie en eerdere leerprestaties constant gehouden te worden. Het feitelijk onderwijsaanbod wordt in dit onderzoek buiten beschouwing gelaten.

LEERKRACHTVERWACHTINGEN EN STREEFNIVEAUS

In onderzoek naar leerkrachtverwachtingen kan onderscheid worden gemaakt tussen twee benaderingen: experimenteel gemanipuleerde leerkrachtverwachtingen en natuurlijke leerkrachtverwachtingen. Bij experimenteel gemanipuleerde leerkrachtverwachtingen, zoals Rosenthal en Jacobson (1968) die toepasten, worden bij leerkrachten valse verwachtingen gewekt over de prestaties van hun leerlingen. De oorspronkelijke verkeerde verwachting leidt tot gedrag waardoor deze verwachting uitkomt: het "self-fulfilling prophecy" effect. Bij natuurlijke leerkrachtverwachtingen gaat het om verwachtingen van leerkrachten over hun leerlingen in natuurlijke situaties, zoals in de klas. In de natuurlijke situatie heeft de leerkracht, uit eigen waarneming en eventueel uit door anderen verstrekte gegevens, zich een beeld gevormd van elke individuele leerling in zijn of haar klas. Deze beeldvorming houdt de verwachting in dat het bestaande leerlingbeeld in de toekomst

bevestigd zal worden in het gedrag en de leerprestaties van de leerling. Dit wordt het "sustaining expectation" effect (Beez, 1968; Burnham, 1968; Dusek, 1975) genoemd.

In dit onderzoek gaat het om deze in de normale schoolpraktijk bij leerkrachten bestaande variatie in leerlingbeelden en verwachtingen. We veronderstellen, dat er een relatie bestaat tussen de verwachtingen die men van iemand heeft en het gedrag ten aanzien van die persoon. Men neigt ertoe zijn gedrag aan te passen aan de verwachting. Dit geldt zowel voor de leerkracht als voor de leerling. Het is bekend dat leerkrachtverwachtingen van invloed zijn op de schoolprestaties van leerlingen (zie bv. Dusek, 1985). Naarmate leerkrachten hogere verwachtingen hebben van leerlingen, streven zij hogere onderwijsdoelen bij hun leerlingen na. Hogere eisen van leerkrachten lokken bij de leerling, voorzover mogelijk, betere prestaties uit. Aangezien verwacht mag worden, dat het beeld en daarmee ook de verwachtingen die een leerkracht van een leerling heeft, deels bepaald zullen worden door de sociale achtergrond van de leerling, doet zich de vraag voor in hoeverre de relatie tussen de sociale achtergrond van leerlingen en hun schoolprestaties terug te voeren valt op leerkrachtverwachtingen en daaruit voortvloeiend leerkrachtgedrag. Differentiatie in streefniveaus leidt tot differentiatie in onderwijsaanbod, deels samenvallend met de indeling van leerlingen naar sociaal milieu. Leerlingen uit lagere sociale milieus zouden hierdoor kunnen worden benadeeld, doordat leerkrachten aan hen te lage eisen stellen. Jungbluth (1985) noemt dit verborgen differentiatie.

Wanneer de bovenstaande redenering juist is, dan is de samenhang tussen sociaal milieu en bereikt onderwijsniveau (deels) de resultante van een continu proces in de alledaagse schoolpraktijk, zich uitstrekkend over de totale basisschoolperiode. Onderzoek naar dit verschijnsel dient dan ook bij voorkeur longitudinaal te zijn. Wanneer het startpunt van het onderzoek niet samenvalt met de intrede in de basisschool, dan heeft het te onderzoeken proces reeds kans gehad in te werken op de schoolprestaties. In de samenhang van sociaal milieu met de beginmeting van de schoolprestaties kan dus reeds het "sustaining expectation effect" aanwezig zijn. Verschillen in schoolprestaties bij de start van het onderzoek moeten constant gehouden worden. Het effect van leerkrachtverwachtingen dat men dan nog kan waarnemen, is de resultante van de differentiatie in streefniveaus gedurende de onderzochte periode. Aangezien verondersteld wordt, dat hier sprake is van een continu proces dat cumulatief inwerkt op de samenhang tussen sociaal milieu en schoolprestaties, zal de omvang van het te vinden effect mede afhangen van de lengte van de onderzoeksperiode.

Leerlingbeeld

Het leerlingbeeld beperken we voor het doel van deze studie tot de inschatting van de leerkracht van de prestatiegeschiktheid van de leerling: het actuele beeld dat de leerkracht heeft van de prestaties en van de mogelijkheden van een leerling om prestaties te leveren. Voorbeelden hiervan zijn uitspraken als: "De leerling moet op zijn tenen lopen"; "heeft groot abstractievermogen", "doorziet de zaken goed".

Daarnaast onderscheiden we het beeld dat de leerkracht heeft van de sociale achtergrond van de leerling. Hierin worden betrokken de inschatting door de leerkracht van het opleidingsniveau van de vader en/of de moeder en de inschatting door de leerkracht van het onderwijsondersteunend thuisklimaat. Met dit laatste wordt bedoeld of naar het oordeel van de leerkracht bij de leerling thuis sprake is van een positief intellectueel klimaat, m.n. of de ouders hoge verwachtingen hebben van of een grote belangstelling hebben voor de schoolvorderingen van hun kinderen.

In onderzoek van Jungbluth (1985, p. 89 e.v.) bleek het oordeel over de prestatiegeschiktheid duidelijk samen te hangen met de perceptie van de leerkracht van de sociale achtergrond van de leerling. Er was geen samenhang met school- en leerkrachtkenmerken.

Streefniveau

Met de term 'streefniveau' doelt Jungbluth (o.c., p. 101 e.v.) zowel op een (vroegtijdig) anticiperen door leerkrachten op het niveau van voortgezet onderwijs dat een leerling kan bereiken, alsmede op het niveau van leerstofbeheersing dat daarvoor nodig zou zijn, bijv. wel of niet uitgebreid leren zinsontleden. In dit onderzoek gebruiken we de term 'absoluut streefniveau' voor het anticiperen op het niveau van voortgezet onderwijs. Daarnaast onderscheiden wij 'relatieve streefniveaus', waarvoor we de term 'cognitieve onderwijsdoelen' reserveren (ter onderscheiding van sociale doelen, die overigens in deze studie buiten beschouwing blijven). Bij cognitieve onderwijsdoelen zoals hier opgevat, gaat het niet om het aanbod van specifieke leerstofonderdelen, maar om de mate waarin de leerkracht het belangrijk vindt om aan een leerling hoge eisen te stellen, veel nadruk te leggen op taal en rekenen, of alleen die leerstof aan te bieden die het kind later, in het voortgezet onderwijs, nodig heeft. Cognitieve onderwijsdoelen hebben een pedagogisch-didactisch karakter, terwijl de verwachting van leerkrachten over het meest geschikte type voortgezet onderwijs aangeeft, welk niveau van voortgezet onderwijs voor een leerling haalbaar is. Het onderscheid tussen absolute en relatieve prestatieniveaus is van belang, omdat we veronderstellen dat er variatie kan bestaan in de mate waarin leerkrachten lage of hoge eisen stellen aan hun leerlingen, ongeacht het prestatieniveau van die leerlingen.

In deze studie leggen we een rechtstreeks verband tussen streefniveaus en schoolprestaties. Dit verband moet worden opgevat als een aggregaat van een onderliggend dynamisch proces dat zich afspeelt in de dagelijkse klaspraktijk. Over de wijze waarop verwachtingen van leerkrachten prestaties van hun leerlingen kunnen beïnvloeden is een model geformuleerd door Cooper (1979). In dit model staat centraal de differentiële feedback die leerlingen ontvangen op hun prestaties, afhankelijk van de verwachtingen die de leerkracht van een leerling heeft. De invloed hiervan op leerlingprestaties verloopt via cognitieve processen bij de leerlingen. Het hangt dus mede van de individuele leerling af, of leerkrachtverwachtingen doorwerken in schoolprestaties (Wittrock, 1986).

Modelspecificatie

In plaats van één model over beide leerjaren hebben we ervoor gekozen aparte modellen te maken voor de leerkrachtverwachtingen en streefniveaus in leerjaar 7 en 8. Deze keuze berust op de overweging dat de leerkrachtverwachtingen en streefniveaus in leerjaar 7 een andere status hebben dan in leerjaar 8. Beide modellen zijn longitudinaal, maar de periode die ligt tussen de meting van leerlingbeelden en streefniveaus enerzijds en de Cito Eindtoets en het advies anderzijds is voor leerjaar 7 twee keer zo lang als voor leerjaar 8. Verder is in de meeste gevallen de leerkracht van leerjaar 7 een andere dan die van leerjaar 8. De laatste is over het algemeen tevens degene die het advies voor voortgezet onderwijs uitbrengt. Het model voor leerjaar 7 levert dan ook een sterkere toetsing van het effect van leerkrachtverwachtingen en streefniveaus dan het model voor leerjaar 8. Een tweede reden is, dat er in een model voor beide leerjaren tegelijk een extra complicatie bijkomt. Er moeten dan immers veronderstellingen worden opgenomen over relaties tussen de leerlingbeelden en streefniveaus van de leerkrachten van leerjaar 7 en 8. Dit leidt tot ingewikkelde (kruis)relaties die niet relevant zijn voor de hier te beantwoorden vraagstelling.

Uitgangspunt voor de modelspecificatie is een schoolloopbaan model voor verklaring van het advies voor voortgezet onderwijs uit leerprestaties (taal en rekenen; Cito-score), intelligentie en sociaal milieu (Van der Hoeven-van Doornum, in voorbereiding). In dit model zijn sociaal milieu en intelligentie opgevat als gecorreleerde exogene variabelen. Dit heeft tot gevolg dat een deel van de correlatie van SES met schoolprestaties en advies niet geanalyseerd wordt. Dit deel van de correlatie wordt door het model noch aan SES noch aan intelligentie toegeschreven. Intelligentie is enkel gemeten aan het begin van leerjaar 8.

We veronderstellen over de periode van één leerjaar een dermate hoge stabiliteit van de intelligentie-verschillen, dat het gerechtvaardigd is deze meting te gebruiken als indicator voor de intelligentie begin leerjaar 7. Deze keuze is mede ingegeven door het feit dat we over slechts één IQ-meting konden beschikken.

In overeenstemming met de boven vermelde schoolloopbaan onderzoeken bleek sociaal milieu naast een indirecte invloed via de schoolprestaties ook een directe invloed te hebben op het advies voor voortgezet onderwijs. Bovendien bleek een direct effect van sociaal milieu op de score op de Cito-toets, waarbij intelligentie en eerdere leerprestaties constant gehouden zijn.

Aan het schoolloopbaan-model zijn de leerlingbeelden en streefniveaus toegevoegd, in een poging de gevonden milieu effecten te verklaren. Hiertoe hebben we zoals gezegd twee longitudinale modellen gespecificeerd, beide gericht op het verklaren van de score op de Cito Eindtoets en het vervolgadvis in leerjaar 8. Figuur 1a en 2a geven onze uitgangsmoellen weer. Deze modellen zijn gebaseerd op bovenstaande redenering inzake het effect van leerlingbeelden en streefniveaus. Daarnaast berusten de uitgangsmoellen deels op eerdere analyses, uitgevoerd op dezelfde data (Van der Hoeven-van Doornum, in voorbereiding), waarbij werd nagegaan hoe de leerlingbeelden en streefniveaus samenhangen met sociaal milieu, intelligentie en eerdere leerprestaties.

Het beeld dat de leerkracht heeft van de sociale achtergrond van een leerling, zal natuurlijk in de eerste plaats bepaald worden door het feitelijke sociale milieu van die leerling, maar gedeeltelijk ook via de omweg van het waargenomen leerlinggedrag zelf. Uit eerdere analyses is gebleken dat deze leerlingbeelden mede bepaald worden door de intelligentie en door de reken- en taalprestaties van de leerlingen. Naarmate het kind intelligenter is, c.q. hogere prestaties levert, wordt de ondersteuning door het thuisklimaat en de opleiding van de ouders hoger ingeschat.

Het oordeel van de leerkracht over de prestatiegeschiktheid van de leerling zal in onze opvatting zowel bepaald worden door de eerdere leerprestaties en de intelligentie alsmede door het beeld dat de leerkracht heeft van het sociale milieu van de leerling. De streefniveaus zullen voornamelijk bepaald worden door de leerlingbeelden en wat betreft het geanticipeerd type voortgezet onderwijs ook door SES en IQ.

We veronderstellen een directe relatie tussen de streefniveaus en de Cito-score. Tevens veronderstellen we een directe relatie tussen het geanticipeerd type voortgezet onderwijs en het uiteindelijke advies voor voortgezet onderwijs. Immers dezelfde overwegingen die naast de gemeten schoolprestaties het uiteindelijke advies bepalen, kunnen reeds een rol spelen in het eerdere oordeel met betrekking tot het type voortgezet onderwijs dat de leerkracht bij een bepaalde leerling voor ogen staat.

De kromme pijlen in de figuren tussen de beginprestaties voor rekenen en taal en tussen de leerkrachtoordelen met betrekking tot thuisklimaat en opleidingsniveau ouders geven correlaties aan, die we in eerdere analyses moesten toestaan tussen de residuen van de betrokken variabelen. Deze correlaties zijn overigens niet hoog, tussen 0.10 en 0.15.

Hypotheses

Willen leerkrachtverwachtingen en streefniveaus kunnen bijdragen aan de verklaring van de invloed van het sociale milieu op leerprestaties en vervolgadvis, dan dient aan twee voorwaarden te zijn voldaan. Er dient op de eerste plaats direct of indirect een relatie te zijn van sociaal milieu met de leerkrachtverwachtingen en streefniveaus. Deze relaties zijn vastgesteld in eerdere analyses (Jungbluth, 1985; van der Hoeven van Doornum, in voorbereiding). De tweede voorwaarde is dat er een effect is van streefniveaus op leerprestaties en vervolgadvis. Deze tweede voorwaarde leidt tot twee hypotheses die we in beide modellen zullen toetsen:

1. De prestatieniveaus die leerkrachten bij hun leerlingen nastreven, de cognitieve onderwijsdoelen en het geanticipeerd type voortgezet onderwijs, hebben een direct effect op de score op de Cito Eindtoets in leerjaar 8.
2. De streefniveaus, m.n. het geanticipeerd type voortgezet onderwijs, hebben een direct effect op het vervolgadvis.

Wanneer deze hypothesen niet worden verworpen, kunnen we vaststellen in hoeverre de primaire en secundaire milieu-effecten die in het schoolloopbaanmodel zijn geconstateerd, kunnen worden verklaard uit de leerlingbeelden en de streefniveaus. We verwachten dat het directe effect van het sociale milieu op de Cito Eindtoets en het vervolgadvis zal afnemen. In zijn meest stringente vorm leidt dit tot de volgende hypothese, die in beide modellen wordt getoetst:

3. Er bestaat geen direct effect van sociaal milieu op de Cito Eindtoets en het vervolgadvis in leerjaar 8.

Als deze hypothese moet worden verworpen, dan zullen we beschrijven in welke mate de directe effecten van sociaal milieu op de beide afhankelijke variabelen zijn gereduceerd als gevolg van de opname van leerlingbeelden en streefniveaus in het schoolloopbaanmodel. Tevens zullen we nagaan in welke mate de indirecte effecten van sociaal milieu verlopen via leerlingbeelden en nagestreefde prestatieniveaus.

METHODE

Proefpersonen

Om een evenwichtige verdeling van leerlingen uit verschillende sociaal-economische milieu's te bereiken, zijn op basis van het gemiddelde leerlinggewicht - bepaald volgens de normen van het Ministerie van Onderwijs en Wetenschappen - vijftig scholen geselecteerd uit een steekproef van 397 basisscholen die hebben deelgenomen aan een onderzoek naar schooleffectiviteit (Van der Hoeven-van Doornum & Jungbluth, 1987). Proefpersonen zijn alle leerlingen die per 1-9-1986 deel uitmaakten van groep 7 van de geselecteerde scholen, in totaal ruim 1000 leerlingen.

Dataverzameling

Voor de beantwoording van de vraagstelling is een longitudinale onderzoeksopzet met minstens een beginmeting en eindmeting van de leerprestaties noodzakelijk.

De leerprestaties van de leerlingen zijn gedurende twee jaar gevolgd. Aan het begin van leerjaar 7 zijn een taal- en een rekentoets afgenomen. De items voor deze toetsen zijn ontleend aan toetsen voor begrijpend lezen en rekenen van het Cito. Aan het begin van leerjaar 8 is de ISI (Van Boxtel, Snijders & Welten, 1982) afgenomen, zowel het schoolvorderingendeel als het intelligentiedeel. Aan het eind van leerjaar 8 zijn de scores beschikbaar op de Eindtoets Basisonderwijs van het Cito.

Daarnaast zijn de leerkrachten aan het begin van leerjaar 7 en leerjaar 8 geïnterviewd over hun leerlingen. Voor elke leerling hebben zij vragen beantwoord o.a. over de prestatiegeschiktheid van de leerling, het onderwijsondersteunend thuisklimaat en het opleidingsniveau van de ouders. Ook hebben de leerkrachten voor elke leerling aangegeven welke onderwijsdoelen zij bij hun leerlingen nastreven. Het feitelijk beroep en opleidingsniveau van de ouders is vastgesteld middels een vragenlijst aan de ouders. Aan het einde van leerjaar 8 is het

definitieve advies van de leerkracht voor het voortgezet onderwijs opgevraagd.

De variabelen

Tabel 1 en 2 bevatten descriptieve gegevens over de in dit onderzoek gebruikte variabelen. Bij enkele van deze variabelen is een nadere toelichting nodig.

Als index voor het feitelijke sociale milieu gebruiken we een middels HOMALS (Gifi, 1981) geconstrueerde score, gebaseerd op het beroep van de vader en het opleidingsniveau van beide ouders. De belangrijkste reden om niet te werken met de drie afzonderlijke milieuvariabelen is gelegen in het feit, dat we dan meer uitval van leerlingen zouden krijgen wegens ontbrekende scores. Het advies voor voortgezet onderwijs is als volgt geschaald: 5 = schakelschool / individueel onderwijs, 10 = lbo, 15 = lbo/mavo, 20 = mavo, 25 = mavo/havo, 30 = havo, 35 = havo/vwo, 40 = vwo.

Tabel 1. Verdelingskenmerken van de leerprestaties in leerjaar 7 en 8, het advies voor voortgezet onderwijs, het sociale milieu en de intelligentie van de leerlingen (N=701)

meetmoment	variabele	gemiddelde	sd	min-max
begin leerjaar 7	rekentoets	22.85	7.83	3-39
begin leerjaar 7	taaltoets	24.78	5.53	5-34
begin leerjaar 8	rekentoets	11.39	5.36	1-24
begin leerjaar 8	taaltoets	11.81	3.43	2-21
tweede helft leerj. 8	Cito Eindtoets	534.43	10.79	501-550
eind leerjaar 8	advies voortgezet onderwijs	22.95	10.20	5-40
tweede helft leerjaar 7	sociaal milieu	50.14	10.04	39-77
begin leerjaar 8	IQ	12.87	2.49	3-18

De leerkrachtverwachtingen en streefniveaus (Tabel 2) zijn gemeten door aan leerkrachten een aantal vragen voor te leggen over hun leerlingen. De inschatting van het opleidingsniveau van de vader en van de moeder is door de leerkrachten verricht op een schaal van 1-3 (laag, midden en hoog). Bij de hierna beschreven analyses is de inschatting van het opleidingsniveau van de vader gebruikt en in het geval van een éénoudergezin de inschatting van het opleidingsniveau van de moeder. Wij noemen deze variabele in het vervolg 'opleiding vader'. Het oordeel over het meest geschikte schooltype is op dezelfde wijze geschaald als het advies voor voortgezet onderwijs. De inschatting van het onderwijsondersteunend thuisklimaat en van de prestatiegeschiktheid zijn evenals de cognitieve onderwijsdoelen vastgesteld middels items met een 5-puntsschaal. De leerkrachten werd gevraagd om voor alle leerlingen van zijn of haar klas aan te geven in welke mate een uitspraak voor een leerling van toepassing is. De uitspraken zijn - soms in gewijzigde vorm grotendeels ontleend aan Jungbluth (1985). Hierboven zijn al voorbeelden genoemd van uitspraken over prestatiegeschiktheid. Voorbeelden van uitspraken over het onderwijsondersteunend thuisklimaat zijn: 'Thuis is er veel belangstelling voor lezen en algemene ontwikkeling' en 'Er is een grote kloof tussen de manier waarop er thuis met dit kind wordt omgegaan en wat er op school gebeurt.' Enkele typische items voor cognitieve onderwijsdoelen zijn: 'Uitgebreid leren zinsontleden' en 'Leren werken

met tabellen en grafieken'. Door middel van factoranalyse zijn (voor beide leerjaren apart) schaaltes geconstrueerd, met scores variërend van 1 (laag) tot 5 (hoog). In Tabel 2 zijn de interne consistenties (α -coëfficiënten) van deze schaaltes vermeld. Met uitzondering van thuisklimaat en vooral cognitieve onderwijsdoelen in leerjaar 7 zijn deze zeer bevredigend te noemen. Om de interne consistentie van de cognitieve onderwijsdoelen te verhogen zijn voor leerjaar 8 vragen toegevoegd zoals: 'Laten werken met materiaal uit lagere leerjaren' en 'verrijkingsstof aanbieden.'

Van 701 leerlingen zijn scores beschikbaar op alle variabelen. Voor een volledig overzicht van het databestand zie Van der Hoeven-van Doornum (in voorbereiding).

Tabel 2. Verdelingskenmerken en interne consistentie van de meting van leerkrachtverwachtingen en streefniveaus in leerjaar 7 en leerjaar 8

meetmoment	variabele	N items	α	gem.	sd
begin leerjaar 7	opleiding vader	1		1.76	0.82
	thuisclimaat	5	.76	3.75	0.72
	prestatie- geschiktheid	7	.92	3.50	0.90
	cognitieve doelen	5	.72	3.36	0.65
	geschikt type v.o.	1		21.00	9.50
begin leerjaar 8	opleiding vader			1.79	0.81
	thuisclimaat	5	.88	3.80	0.76
	prestatie- geschiktheid	7	.91	3.46	0.86
	cognitieve doelen	9	.86	3.61	0.71
	geschikt type v.o.	1		22.00	9.80

Data-analyse

De data worden geanalyseerd met behulp van de LISREL/VI procedure (Jöreskog & Sörbom, 1981) voor lineaire structurele vergelijkingen. De analyse wordt uitgevoerd voor de leerprestaties, leerkrachtverwachtingen en streefniveaus van leerjaar 7 en leerjaar 8 afzonderlijk. De score op de Citotoets en het advies voor voortgezet onderwijs aan het eind van leerjaar 8 vormen in beide gevallen de te verklaren variabelen. Er is geen gebruik gemaakt van de LISREL-mogelijkheid om latente variabelen in een model op te nemen. Voor elke variabele beschikken we over slechts één meetinstrument. De parameters worden geschat met de 'maximum likelihood'-methode.

Ter ontwikkeling van het model wordt de totale steekproef opgesplitst in twee helften: een exploratie-steekproef en een validatie-steekproef. De exploratie-steekproef wordt gebruikt om het uitgangsmodel op die onderdelen waarover vooraf geen hypothesen zijn geformuleerd, zodanig te wijzigen dat het model goed bij de data past. We streven daarbij naar een zo zuinig mogelijk model. Op grond van de chi-kwadraat, de 'adjusted goodness of fit' (AGF) en de residuele correlaties gaan we na hoe goed het model bij de data past. Indien nodig voor het verkrijgen van een betere passing van het model, worden stapsgewijs vrije parameters toegevoegd. Alleen die aanpassingen worden gemaakt waar een plausibele interpretatie voor bestaat. Wanneer een goed passend model is gevonden, gaan we aan de hand van de t-waarden van de parameters na welke van de parameters waarover geen hypothese is geformuleerd, alsnog kunnen worden gefixeerd. Als significantieniveau wordt steeds .05 gebruikt.

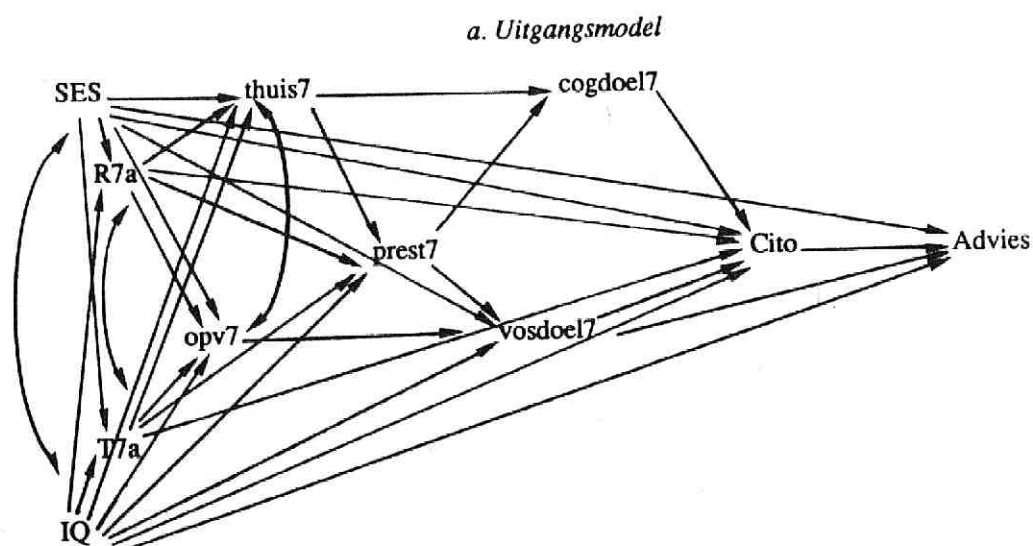
Het model dat is verkregen op de exploratie-steekproef, wordt vervolgens toegepast op de validatie-steekproef. Uiteindelijk worden de parameters geschat en de hypothesen getoetst op de totale steekproef. De hypothesen worden voor elke betrokken parameter afzonderlijk getoetst middels t-toetsen. De directe en de totale effecten van de diverse variabelen op de Cito-toets en het advies voor voortgezet onderwijs worden berekend en vergeleken met die uit het schoolloopbaanmodel zonder leerkrachtverwachtingen en prestatieniveaus. Het indirecte effect van SES op Cito Eindtoets en op advies voor voortgezet onderwijs wordt opgesplitst in specifieke effecten via (eerdere) leerprestaties enerzijds en via leerlingbeelden en streefniveaus van leerkrachten anderzijds. Met het specifieke indirecte effect van bijvoorbeeld leerkrachtverwachtingen op de Cito-toets bedoelen we het verschil in totale SES-effecten op de Cito-toets, dat resulteert wanneer alle directe paden van leerlingbeelden en streefniveaus van leerkrachten naar de Cito-toets op nul gesteld worden (vgl. Bollen, 1987, p. 55 e.v.).

RESULTATEN

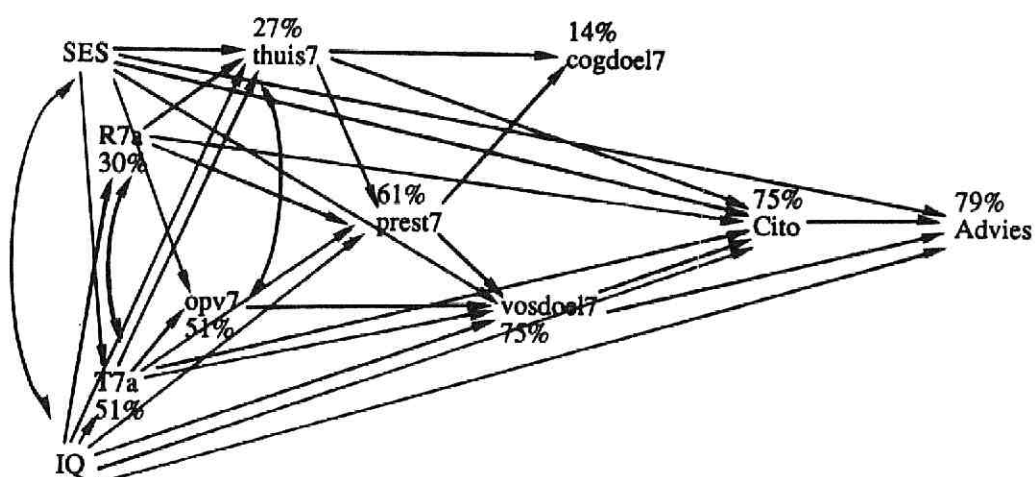
Modelontwikkeling

Uit toepassing van de beide uitgangsmodellen (Figuur 1a en 2a) op de exploratie-steekproef blijkt dat nog enkele vrije parameters aan het model moeten worden toegevoegd. Het geanticipeerd type voortgezet onderwijs blijkt zowel in leerjaar 7 als leerjaar 8 niet alleen bepaald te worden door de leerlingbeelden, SES en IQ. Er is tevens een effect van de schoolprestaties (taal in leerjaar 7 en 8; rekenen in leerjaar 8). Deze variabele blijkt zich daarmee in het model op ongeveer dezelfde wijze te gedragen als het advies voor voortgezet onderwijs. In leerjaar 8 moet ook een relatie van de taalprestatie naar de cognitieve doelen worden toegelaten, naast het in het uitgangsmodel reeds voorziene effect van intelligentie. De variabele cognitieve doelen zoals gemeten in leerjaar 7 en die gemeten in leerjaar 8 lijken zich niet geheel op dezelfde wijze te gedragen. In leerjaar 8 wordt in tegenstelling tot leerjaar 7 een direct effect toegestaan van SES op het oordeel over prestatiegeschiktheid. In leerjaar 7 lijkt in afwijking van onze uitgangspunten de inschatting van het onderwijsondersteunend thuisklimaat een direct effect te hebben op de Cito-score.

Na de genoemde toevoegingen van parameters wijkt het model niet meer significant af van de data (in leerjaar 7 $\chi^2=28.67$, $df=18$; in leerjaar 8 $\chi^2=16.29$, $df=12$). Bij toepassing op de validatie-steekproef, blijft de fit redelijk goed (resp. $\chi^2=34.65$, $df=18$ en $\chi^2=26.99$, $df=12$). In de totale steekproef blijken voor beide leerjaren enkele parameters niet significant



b. Eindmodel toegepast op de totale steekproef



Figuur 1. Invloed van Leerkrachtverwachtingen en Streefdoelen in leerjaar 7 op de Cito Eindtoets en het Advies in leerjaar 8

SES	= sociaal-economische status
IQ	= intelligentie (ISI)
R7a	= rekenprestatie, begin van leerjaar 7
T7a	= taalprestatie, begin van leerjaar 7
thuis7	= inschatting van het thuisklimaat, begin leerjaar 7
opv7	= inschatting van de opleiding van de vader, begin leerjaar 7
prest7	= inschatting van de prestatiegeschiktheid van de leerling, begin leerj. 7
cogdoel7	= cognitieve onderwijsdoelen, begin leerjaar 7
vosdoel7	= geschikt type voortgezet onderwijs, begin leerjaar 7
Cito	= standardscore Cito Eindtoets, leerjaar 8
Advies	= vervolgadvis voortgezet onderwijs, leerjaar 8

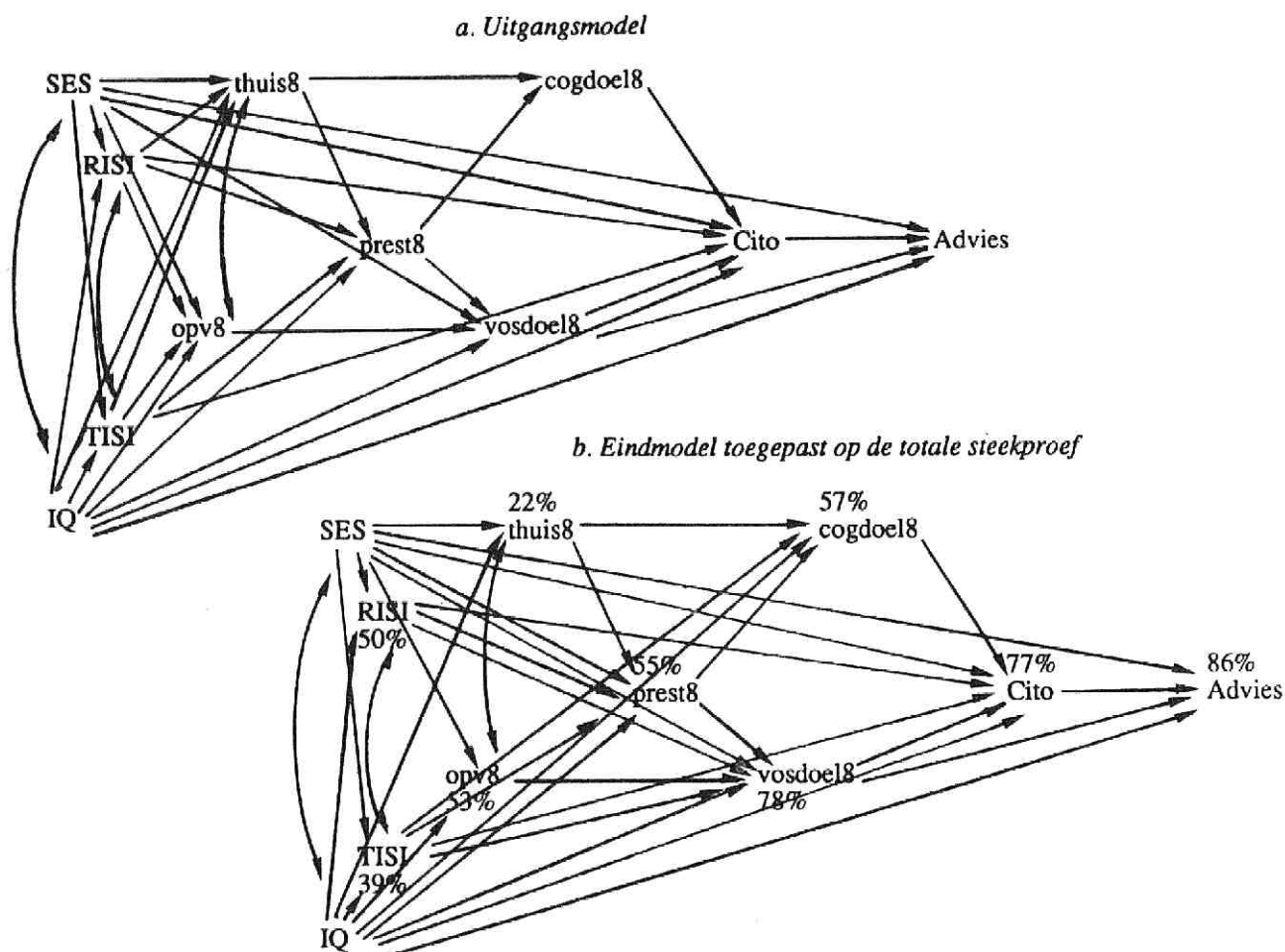
van 0 af te wijken. De niet-significante relaties waarover we in dit model geen specifieke hypothesen hebben, kunnen worden verwijderd. Het model dat uit deze vereenvoudigingen resulteert, heeft op de totale steekproef voor leerjaar 7 een chi-kwadraat van 44.77 bij 21 vrijheidsgraden ($AGF=.964$). Voor leerjaar 8 is de chi-kwadraat 32.77 bij 17 vrijheidsgraden ($AGF=.968$). Met behulp van deze modellen (zie Figuur 1b en 2b) worden de hypothesen getoetst.

Hypothesetoetsing

In deze paragraaf gaan we na, of onze veronderstellingen met betrekking tot de werking van de leerkrachtverwachtingen en de streefniveaus juist zijn. Tabel 3 bevat de resultaten van de hypothesetoetsingen voor beide modellen.

Ten aanzien van de eerste hypothese betreffende het effect van de streefniveaus op de score op de Cito toets blijkt uit Tabel 3 dat de hypothese bevestigd wordt, behalve voor de cognitieve doelen in leerjaar 7. Er is dus inderdaad sprake van een effect van de streefniveaus van de leerkrachten op de schoolprestaties, waarbij onder meer intelligentie en eerdere schoolprestaties constant zijn gehouden.

De tweede hypothese wordt in beide modellen bevestigd. Het oordeel van de leerkracht over het meest geschikte type van voortgezet onderwijs aan het begin van leerjaar 7 beïnvloedt niet alleen de score op de Cito Eindtoets, maar heeft ook nog een direct effect op



Figuur 2. De invloed van Leerkrachtverwachtingen en Streefdoelen in leerjaar 8 op de Cito Eindtoets en het Advies

SES	= sociaal-economische status
IQ	= intelligentie (ISI)
RISI	= rekenprestatie, begin van leerjaar 8
TISI	= taalprestatie, begin van leerjaar 8
thuis8	= inschatting van het thuisklimaat, begin leerjaar 8
opv8	= inschatting van de opleiding van de vader, begin leerjaar 8
prest8	= inschatting van de prestatiegeschiktheid van de leerling, begin leerj. 8
cogdoel8	= cognitieve onderwijsdoelen, begin leerjaar 8
vosdoel8	= geschikt type voortgezet onderwijs, begin leerjaar 8
Cito	= standardscore Cito Eindtoets, leerjaar 8
Advies	= vervolgadvis voortgezet onderwijs, leerjaar 8

het advies voor voortgezet onderwijs.

Tot slot de toetsing van de belangrijkste hypothese over het directe effect van het sociale milieu op de Cito-score en het vervolgadvis. Tabel 3 toont dat ook onafhankelijk van eerdere leerprestaties, leerkrachtverwachtingen en streefniveaus, het sociale milieu een direct effect heeft op de Cito Eindtoets en op het vervolgadvis. Onze derde hypothese wordt dus niet door de data ondersteund. In de volgende paragraaf gaan we na, in hoeverre de directe invloed van het sociale milieu is afgenomen door de leerkrachtverwachtingen en streefniveaus.

Tabel 3. Toetsing van de hypothesen. T-waarden van de regressiecoëfficiënten

Hypothese	leerjaar 7	leerjaar 8
1. cogdoel → Citotoets	-0.02	5.06
vosdoel → Citotoets	4.64	6.86
2. vosdoel → advies	4.67	29.05
3. SES → Citotoets	2.60	1.99
SES → advies	5.03	4.90

Parameterschattingen, directe en indirecte effecten

Op grond van de toetsing van de hypothesen dient nog één verandering te worden aangebracht in het model voor de leerkrachtverwachtingen en streefniveaus van leerjaar 7. Het niet-significante effect van de cognitieve onderwijsdoelen op de Cito Eindscore wordt uit het model verwijderd. Het resultaat is het model weergegeven in Figuur 1b ($\chi^2=44.77$, $df=22$, $p=.003$ en $AGF=.965$). Alle modelparameters zijn nu significant op 5%-niveau. Voor leerjaar 8 behoeven geen verdere wijzigingen in het model te worden aangebracht (zie Figuur 2b). De parameterschattingen (gestandaardiseerde padcoëfficiënten) zijn opgenomen in Tabel 4. Voor de gestandaardiseerde totale effecten zie Tabel 5.

Het model voor de leerkrachtverwachtingen in leerjaar 7 (Figuur 1b) verklaart 79% van de variantie van het advies eind leerjaar 8. Bij het model voor leerjaar 8 (Figuur 2b) is dit 86%.

Tabel 4. Gestandaardiseerde regressiegewichten (directe effecten)

LEERJAAR 7

	SES	IQ	onafhankelijke variabelen							
			R7a	T7a	thuis7	opv7	prest7	cog7	vos7	Cito
R7a		.55								
T7a	.09	.69								
Thuis7	.24	.15	.09	.20						
Opv7	.63			.18						
Prest7		.21	.31	.15	.33					
Cog7					.11		.29			
Vos7	.13	.15		.09		.13	.57			
Cito	.06	.32	.17	.31	.08				.14	
Advies	.10	.11							.48	.33

LEERJAAR 8

	SES	IQ	RISI	TISI	thuis8	opv8	prest8	cog8	vos8	Cito
RISI	.13	.65								
TISI	.08	.59								
Thuis8	.17	.26		.16						
Opv8	.63	.13	.08							
Prest8	-.06	.29	.24	.20	.24					
Cog8		.08		.17	.18		.48			
Vos8	.13	.09	.18	.14		.13	.47			
Cito	.04	.28	.31	.07				.12	.22	
Advies	.08	.07							.70	.16

Voor de betekenis van de afkortingen zie de toelichting bij Figuur 1 en 2.

Tabel 5. Gestandaardiseerde totale effecten

LEERJAAR 7										
	SES	IQ	onafhankelijke variabelen							
			R7a	T7a	thuis7	opv7	prest7	cog7	vos7	Cito
R7a		.55								
T7a	.09	.69								
Thuis7	.26	.33	.09	.20						
Opv7	.64	.13		.18						
Prest7	.10	.59	.34	.22	.33					
Cog7	.06	.21	.11	.09	.21		.30			
Vos7	.28	.56	.19	.22	.19	.13	.57			
Cito	.14	.73	.20	.36	.10	.02	.08		.14	
Advies	.28	.62	.16	.23	.13	.07	.30		.53	.33

LEERJAAR 8										
	SES	IQ	RISI	TISI	thuis8	opv8	prst8	cog8	vos8	Cito
RISI	.13	.65								
TISI	.08	.59								
Thuis8	.18	.36		.16						
Opv8	.64	.18	.08							
Prest8	.03	.65	.24	.23	.24					
Cog8	.06	.55	.12	.31	.29		.48			
Vos8	.25	.61	.30	.25	.11	.13	.47			
Cito	.15	.74	.40	.16	.06	.03	.11	.12	.22	
Advies	.27	.62	.28	.20	.09	.10	.35	.02	.74	.16

Voor de betekenis van de afkortingen zie de toelichting bij Figuur 1 en 2.

De variantie van de Cito-score en van het geanticipeerd type voortgezet onderwijs wordt in beide modellen eveneens in hoge mate verklaard. De inschatting van het thuisklimaat wordt in beide modellen slecht voorspeld (resp. 28% en 22%). In leerjaar 7 geldt dit ook voor de cognitieve onderwijsdoelen (14%). In leerjaar 8 worden deze veel beter door het model verklaard (57%).

Om de onderzoeksvraag te beantwoorden, gaan we eerst na hoe de leerkrachtverwachtingen en streefniveaus zijn gerelateerd aan SES. Daartoe splitsen we het totale effect van SES op leerkrachtverwachtingen en streefniveaus op in directe en indirecte effecten. Bij de laatste onderscheiden we indirecte effecten die lopen via de leerprestaties, via de leerlingbeelden (thuisklimaat en opleiding vader) en via de inschatting van de prestatiegeschiktheid van de leerling (zie Tabel 6). Uit de vergelijking van de geobserveerde correlaties tussen SES en de verwachtingen en streefniveaus met de totale effecten van SES, blijkt dat een aanzienlijk deel van de correlaties ongeanalyseerd blijft. Dit komt vooral op rekening van de correlatie tussen SES en IQ ($r=.37$), die in het model als exogeen is opgevat.

Het effect van SES is duidelijk het grootst bij het geanticipeerd type voortgezet onderwijs. Een aanzienlijk deel van dit effect, nl. ongeveer de helft, is een direct effect van SES. In onze theoretische veronderstellingen verwachten we een SES effect via de leerlingbeelden. In leerjaar 7 zien we dat ongeveer de helft van het totale effect van SES op geanticipeerd type

Tabel 6. Effecten van SES op leerkrachtverwachtingen en streefniveaus

	direct effect	indirecte effecten via			totaal effect	r
		leerprestaties	soc. achtergr.	prest.		
prest7	.0	.01	.09		.10	.31
prest8	-.06	.05	.04		.03	.26
cogdoel7	.0	.0	.03	.03	.06	.09
cogdoel8	.0	.01	.03	.01	.06	.24
vosdoel7	.13	.01	.08	.06	.28	.48
vosdoel8	.13	.03	.08	.01	.25	.48

soc. achtergr. = inschatting door leerkracht van opleidingsniveau vader en thuisklimaat

prest. = inschatting door leerkracht van de prestatiegeschiktheid (= prest7 of prest8).

r = correlatie met SES

voortgezet onderwijs voor rekening komt van de leerlingbeelden, in leerjaar 8 is dat 36%. Ten aanzien van de cognitieve onderwijsdoelen is, in overeenstemming met onze verwachtingen, alleen sprake van een indirect effect van SES, voornamelijk via de leerlingbeelden. De inschatting van de prestatiegeschiktheid gedraagt zich in leerjaar 7 ook overeenkomstig onze verwachtingen: een indirect SES-effect dat voor 90% voor rekening komt van het beeld dat de leerkracht heeft van de sociale herkomst van de leerling. In leerjaar 8 wordt het beeld echter vertroebeld door suppressor effecten. Tenslotte dient te worden opgemerkt dat de leerkrachtverwachtingen en streefniveaus veel sterker worden beïnvloed door IQ en schoolprestaties dan door SES (zie Tabel 5).

Voor het antwoord op de vraag in hoeverre de samenhang tussen sociaal milieu enerzijds en de Cito Eindtoets en het vervolgadvis anderzijds kan worden verklaard door leerkrachtverwachtingen en streefniveaus, hebben we een referentie-punt nodig. We gaan uit van het schoolloopbaanmodel van Van der Hoeven-van Doornum (in voorbereiding), dat gebaseerd is op dezelfde data. We kunnen dan de relatieve grootte van directe en indirecte effecten van het sociale milieu op de Cito Eindtoets en het vervolgadvis vergelijken voor een model zonder en een model met leerkrachtverwachtingen en streefniveaus (zie Tabel 7).

Tabel 7. Effecten van SES op Cito-score en advies voor voortgezet onderwijs.

Directe en indirecte effecten (als percentage van het totale effect) en geobserveerde correlaties

	zonder verwachtingen		met verwachtingen			r
	direct effect	indirect effect via leerprestaties	direct effect	indirecte effecten via leerprest. verwachtingen		
<i>leerj. 7</i>						
Cito.	.07 (47%)	.08 (53%)	.06 (43%)	.03 (20%)	.05 (37%)	.42
Advies	.19 (69%)	.09 (31%)	.10 (36%)	.05 (16%)	.13 (48%)	.51
<i>leerj. 8</i>						
Cito	.07 (47%)	.08 (53%)	.04 (27%)	.04 (27%)	.07 (46%)	.42
Advies	.19 (69%)	.09 (31%)	.08 (30%)	.02 (7%)	.18 (63%)	.51

verwachtingen = leerlingbeelden en streefniveaus van de leerkrachten

r = correlatie met SES

In het model zonder leerkrachtverwachtingen bestaat, rekening houdend met intelligentie en leerprestaties in de leerjaren 7 en 8, een direct effect van het sociale milieu op de Cito Eindtoets en op het vervolgadvis. Met betrekking tot de Cito-toets is dit directe effect 47% van het totale effect. Wanneer rekening gehouden wordt met de leerkrachtverwachtingen en de streefniveaus in leerjaar 7 dan zien we een lichte daling (4%) van dit directe effect, bij een gelijkblijvend totaal effect. Het indirecte effect van SES kunnen we opsplitsen in twee bronnen: een effect via de leerprestaties - zoals gefixeerd begin leerjaar 7, resp. begin leerjaar 8 - (20%) en een additioneel effect via de verwachtingen en streefniveaus van leerkrachten (37%). Bij het laatste wordt vanuit de belangrijkste bijdrage geleverd door het geanticipeerd type voortgezet onderwijs. De cognitieve onderwijsdoelen leveren hieraan geen bijdrage. Het indirecte SES-effect via de leerprestaties, bij constant houding van de intelligentie, kan binnen de theorie van de streefniveaus worden geïnterpreteerd als (deels) een effect van de werking van verwachtingen en streefniveaus in de leerjaren voorafgaande aan leerjaar 7.

In het model voor leerjaar 8 is de daling van het directe effect van SES, als gevolg van de introductie van de leerlingbeelden en de streefniveaus, veel groter, nl. van 47% naar 27%. Ook nu wordt naast een indirect effect via de leerprestaties, dat in overeenstemming met de idee van cumulatieve werking van de streefniveaus iets groter is dan in leerjaar 7, een aanzienlijk deel van het indirecte SES-effect toegeschreven aan de leerkrachtverwachtingen en streefniveaus (nl. 46% van het totale SES-effect). Het leeuwenaandeel wordt hierin weer geleverd door het geanticipeerd type voortgezet onderwijs. De bijdrage van de cognitieve onderwijsdoelen aan het totale SES-effect is iets minder dan 1%.

Het directe effect van sociaal milieu op het vervolgadvis daalt van 69% naar 36% als de leerkrachtverwachtingen en streefniveaus van leerjaar 7 in het model worden opgenomen, voor leerjaar 8 daalt dit percentage nog sterker, naar 30%. Bij het vervolgadvis verklaren de leerkrachtverwachtingen en streefniveaus 48% (leerjaar 7), resp. 63% (leerjaar 8) van het totale effect van het sociale milieu. Anders gezegd, bijna 75% (leerjaar 7) van het indirecte SES-effect kan worden toegeschreven aan leerkrachtverwachtingen en streefniveau. Ongeveer 25% (leerjaar 7) verloopt via het effect van de Cito-toets op het advies. Zoals boven al geconstateerd, komt dit laatste effect ook weer ten dele voor rekening van de leerkrachtverwachtingen en streefniveaus. Vergeleken met leerjaar 7 is de invloed van de streefniveaus in leerjaar 8 op de Cito Eindtoets en het vervolgadvis groter. Dit is vooral toe te schrijven aan het verschil in tijd dat tussen de meetmomenten ligt en aan het feit dat nu de leerkracht bij wie de streefniveaus zijn vastgesteld dezelfde persoon was als de leerkracht die het vervolgadvis uitbracht. Bij de streefniveaus voor leerjaar 7 was dat in het algemeen niet het geval.

De correlaties van SES met Cito-score en advies zijn veel hoger dan de totale effecten die uit de modellen kunnen worden berekend. Dit komt voornamelijk doordat de correlatie tussen SES en IQ in de modellen als exogeen is opgevat.

Tot slot vergelijken we de effecten van sociaal milieu en intelligentie op advies. De directe effecten van het sociale milieu en intelligentie op het advies zijn vrijwel gelijk (zie Tabel 4), maar het totale effect van IQ is veel groter dan dat van SES (zie Tabel 5). Het directe effect van intelligentie is 18%, respectievelijk 11% van het totale effect. Met uitzondering van het effect van intelligentie op de inschatting van de opleiding van de vader, heeft intelligentie op alle overige variabelen steeds hogere effecten dan het sociale milieu.

DISCUSSIE

Het doel van het onderzoek is een bijdrage te leveren aan de verklaring van ongelijke onderwijskansen, die niet kunnen worden toegeschreven aan verschillen in aanleg. Factoren

die samenhangen met het sociale milieu dragen bij aan verschillen in leerprestaties en het advies voor het voortgezet onderwijs. In het onderzoek dat hier is beschreven, is aangetoond dat rekening houdend met verschillen in aanleg en leerprestaties, de verwachtingen die leerkrachten van hun leerlingen hebben en de daarmee samenhangende prestatieniveaus die zij bij hun leerlingen nastreven, bijdragen aan de verklaring van de invloed van het sociale milieu op latere leerprestaties en het vervolgadvis. Dit betekent dat leerkrachten met hun verwachtingen en streefniveaus de schoolloopbanen van hun leerlingen beïnvloeden. Het is niet gelukt de milieu-effecten geheel te verklaren uit de verborgen differentiatie, zoals omschreven door Jungbluth (1985).

Bij deze conclusie zijn enkele opmerkingen te maken. In leerjaar 7 is de meting van de variabele 'cognitieve onderwijsdoelen' niet goed gelukt. De betrouwbaarheid van deze variabele is laag, evenals het percentage door het model verklaarde variantie. Dit is een mogelijke verklaring voor het ontbreken van een effect van de cognitieve onderwijsdoelen op de Cito Eindtoets en is aanleiding tot voorzichtigheid bij uitspraken over de invloed van cognitieve onderwijsdoelen. Op grond van onze bevindingen concluderen we, dat cognitieve onderwijsdoelen van leerjaar 7 geen invloed hebben op leerprestaties op een later tijdstip. In leerjaar 8 dragen cognitieve onderwijsdoelen wel bij aan de verklaring van de Cito Eindtoets en (indirect) aan het vervolgadvis. Dit wijst erop dat verder onderzoek naar cognitieve onderwijsdoelen de aandacht verdient.

Het oordeel van de leerkracht in leerjaar 7 over het meest geschikte type van voortgezet onderwijs heeft zowel een direct effect op de Cito Eindtoets als op het vervolgadvis. In engere zin betekent dit, dat bij de uitspraak van de leerkracht in leerjaar 7 over het type van voortgezet onderwijs de rangorde van de leerlingen in grote lijnen overeenkomt met de rangorde bij de leerprestaties en het vervolgadvis in leerjaar 8. Leerkrachten in leerjaar 7 geven met hun oordeel over het geanticipeerde schooltype aan, dat zij een goed beeld hebben van de mogelijkheden van hun leerlingen om een bepaald prestatieniveau te bereiken en dat zij uitstekend kunnen inschatten welk type voortgezet onderwijs aan het einde van het volgende schooljaar zal worden geadviseerd. Zowel bij het geanticipeerd schooltype in leerjaar 7 als in leerjaar 8 en bij het vervolgadvis vinden we effecten van het sociale milieu. Het geanticipeerd schooltype medieert een deel van de invloed van het sociale milieu op het vervolgadvis en ook op de Cito Eindtoets. Hierbij geldt dat het geanticipeerde schooltype en het vervolgadvis hoger zijn naarmate de leerling uit een hoger sociaal milieu afkomstig is. Dit komt overeen met onze veronderstelling over de rol die streefniveaus spelen bij de invloed van het sociaal milieu op schoolprestaties en vervolgadvis: Het geanticipeerde schooltype heeft een zichzelf bevestigend effect op schoolprestaties en het vervolgadvis. Deze betekenis van de leerkrachtoordelen komt vooral tot uiting in het sterke directe effect van geanticipeerd schooltype op vervolgadvis dat bestaat ondanks het feit dat de score op de Cito-toets constant is gehouden.

Het effect van leerkrachtverwachtingen en streefniveaus concentreert zich in onze modellen vooral in de variabele 'geanticipeerd schooltype voortgezet onderwijs'. Dit geldt vooral voor het effect van het geanticipeerd schooltype voortgezet onderwijs begin leerjaar 8 op het vervolgadvis. Enerzijds moeten we dit toeschrijven aan de grote overeenkomst tussen beide variabelen, anderzijds concluderen we dat aan het begin van leerjaar 8 de verwijzing naar het voortgezet onderwijs al grotendeels vastligt. Dat daar bovenop de cognitieve onderwijsdoelen ook nog een relatie vertonen met de Cito-score ondersteunt onze opvatting dat de cognitieve onderwijsdoelen vooral een pedagogisch-didactisch karakter hebben met een positief effect op leerprestaties in de toekomst. Het onderzoek laat zien dat cognitieve onderwijsdoelen en het geanticipeerd schooltype verschillen in de mate waarin zij onderhevig zijn aan sociale bias. Hoge cognitieve onderwijsdoelen, zoals hier bedoeld, hoeven niet noodzakelijk gepaard te gaan met het streven naar hoge vormen van voortgezet onderwijs.

Een andere kwestie, die we in nader multi-level onderzoek willen bestuderen, is of leerkrachten sterk verschillen in hun uitspraken over cognitieve onderwijsdoelen en geanticipeerd schooltype en of die verschillen ertoe leiden dat leerlingen binnen de ene klas gemiddeld hogere of lagere vervolgadvisen krijgen dan binnen een andere klas. Met andere woorden, maakt het uit of een leerling bij de ene leerkracht in de klas zit of bij een andere leerkracht?

Door de longitudinale opzet van het onderzoek was het mogelijk het effect van de leerkrachtverwachtingen en streefniveaus over een langere periode te bekijken. Het resultaat van deze opzet is dat de langdurige werking van leerkrachtverwachtingen op leerprestaties is aangetoond. Verschillen in leerkrachtverwachtingen berusten niet alleen op verschillen in intelligentie, maar ook op verschillen in het sociale milieu waaruit de leerlingen afkomstig zijn. Rekening houdend met aanvangsverschillen tussen leerlingen, wordt de invloed van het sociale milieu op latere leerprestaties en de verwijzing naar het voortgezet onderwijs deels verklaard door leerkrachtverwachtingen en streefniveaus. Naarmate de leerkrachtverwachtingen en streefniveaus hoger zijn hebben kinderen meer kans op een succesvolle schoolloopbaan.

LITERATUUR

- Beez, W. (1968). Influence of biased psychological reports on teacher behavior and pupil performance. *Proceedings of the Annual Convention of the American Psychological Association*, 3, 605-606.
- Blok, H. & Saris, W.E. (1980). Relevante variabelen bij het doorverwijzen na de lagere school: een structureel model. *Tijdschrift voor Onderwijsresearch*, 5, 63-79.
- Boland, Th., Voeten, M.J.M. & Visser, J.J.C.M. (1989). De invloed van de leesvaardigheid in het basisonderwijs op het niveau van voortgezet onderwijs: een longitudinale benadering. In J. van Damme & J. Dronkers (red.), *Jongeren in school en beroep* (pp. 11-23). Lisse: Swets & Zeitlinger.
- Bollen, K.A. (1987). Total, direct, and indirect effects in structural equation models. In C.C. Clogg (Ed.), *Sociological Methodology 1987, Vol. 17* (pp. 37-69). San Francisco: Jossey Bass.
- Boudon, R. (1974). *Education, inequality and social opportunity*. New York: Wiley.
- Boxtel, H. van, Snijders, J.Th. & Welten, V.J. (1982). *ISI Reeks: Interesse, Schoolvorderingen en Intelligentie*. Groningen: Wolters-Noordhoff.
- Burnham, J. (1968). *Effects of experimenter's expectancies on children's ability to learn to swim*. Unpublished master's thesis. West Lafayette, Purdue University.
- Cooper, H.M. (1979). Pygmalion grows up: A model for teacher expectation communication and performance influence. *Review of Educational Research*, 49, 389-410.
- Dusek, J. (1975). Do teachers bias children's learning? *Review of Educational Research*, 45, 661-684.
- Dusek, J.B. (1985). *Teacher expectancies*. Hillsdale, NJ: Erlbaum.
- Gifi, A. (1981). *HOMAL's User's Guide*. Leiden: University of Leiden.
- Graaf, P.M. de (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Groeneboom, P., Hoogstraten, J., Mellenbergh, G.J. & Santen, J.P.H. van (1978). Relevante variabelen bij het doorverwijzen na de lagere school: een correlationele analyse. *Tijdschrift voor Onderwijsresearch*, 3, 262-280.
- Hoeven-van Doornum, A.A. van der (in voorbereiding). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen*.
- Hoeven-van Doornum, A.A. van der & Jungbluth, P. (1987). De bijdrage van schoolkenmerken aan schooleffectiviteit. In J. Scheerens & W.G.R. Stoel (red.), *Effectiviteit van onderwijsorganisaties*. Lisse: Swets & Zeitlinger.
- Jong, U. de, Dronkers, J. & Saris, W.E. (1982). Veranderingen in de schoolloopbanen tussen 1965 en 1977. Ontwikkelingen in de Nederlandse samenleving en haar onderwijs. *Mens en Maatschappij*, 57, 26-54.

- Jungbluth, P. (1985). *Verborgene differentiatie: Leerlingbeeld en onderwijsaanbod op de basisschool*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Meijnen, G.W. (1984). *Van zes tot twaalf. Een longitudinaal onderzoek naar de milieu- en schooleffecten van loopbanen in het lager onderwijs*. 's-Gravenhage: SVO.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Saris, W.E. & Blok, H. (1982). Het onderwijzersadvies nader bekeken. *Tijdschrift voor Onderwijsresearch*, 7, 49-59.
- Tesser, P. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Visser, J.J.C.M. (1987). *Het advies voor voortgezet onderwijs: De invloed van leesvaardigheid in leerjaar drie en zes, sociaal milieu en geslacht* (SVO-project 6564, Deelrapport V). Nijmegen: Instituut voor Onderwijskunde.
- Wittrock, M.C. (1986). Students' thought processes. In M.C. Wittrock (Ed.), *Handbook of research on teaching (Third Edition)*. New York: MacMillan, 1986, p. 297-314.

Manuscript ontvangen 6-9-1989

Definitieve versie ontvangen 13-11-1989

Studie-uitval en vertraging bij zes studierichtingen uit het Wetenschappelijk Onderwijs

Uulkje de Jong en Joost Meijer

Stichting Centrum voor Onderwijsonderzoek, Universiteit van Amsterdam

ABSTRACT

In 1982 Dutch university education was restructured. Duration of courses was limited to four years. In the context of a research project aiming to evaluate the implementation of the new structure, data about students who started their studies in the course-year 1982/1983 were gathered. These data pertain to six courses, nine locations were involved, i.e. three courses were represented by two locations (universities), the remaining three courses were only represented by one location each. The problem of the study reported here is to what extent differences in study careers can be explained by characteristics of courses and locations, as well as background-characteristics of students. Several performance-indicators concerning study-results and progress were constructed on the basis of the data that were gathered. Analysis of these performance-indicators points out that there exist substantial differences between the courses involved in this research-project.

1. INLEIDING

In 1982 is de Wet Twee-Fasenstructuur voor het Wetenschappelijk Onderwijs ingevoerd. Eén van de doelstellingen van deze wet was het verhogen van de doelmatigheid van het Wetenschappelijk Onderwijs. Om dit te bereiken werd de studieduur verkort tot 4 jaar en de inschrijvingsduur op maximaal 6 jaar gesteld met de mogelijkheid tot het aanvragen van verlenging. Een ander middel ter verhoging van de doelmatigheid was de verplichte invoering van de propedeuse als afsluitend examen van het eerstejaarscurriculum. Aan de propedeuse werden drie functies toegekend namelijk een oriënterende -, een selecterende - en een verwijzende functie. Hiermee zou een verdere selectie in het doctoraal overbodig worden. In het Beleidskader Invoering Twee Fasenstructuur (1981) wordt dan ook gesteld dat alleen de propedeuse een selecterende functie krijgt. Al deze maatregelen behoeven geen afbreuk te doen aan de kwaliteit van de afgestudeerden indien de herprogrammering goed wordt uitgevoerd.

Om de uitkomsten van de wet te beoordelen zijn in het kader van het evaluatieonderzoek naar de invoering van de wet Twee Fasenstructuur (SCO-CSHOB, 1987) gegevens verzameld bij zes studierichtingen. Van studenten die zich in 1982/83 voor het eerst hebben laten inschrijven bij één of twee gekozen locaties van deze studierichtingen zijn gegevens verzameld over vooropleiding, leeftijd en geslacht en alle studieresultaten behaald in de eerste vier jaar. Hieruit zijn een aantal zogenaamde "performance-indicatoren" afgeleid. Deze indicatoren vormen een beeld van de efficiëntie en de effectiviteit van de studierichtingen (Dochy, Wijnen & Bouhuijs, 1987). Eén groep indicatoren heeft betrekking op geaggregeerde onderwijsuitkomsten zoals numeriek rendement van de propedeuse, verblijfsduur van niet propedeuse-geslaagden, gerealiseerde studielast na vier jaar en percentage uitval in de eerste drie jaren van het doctoraal. Een tweede groep

Adres: SCO, Grote Bickersstraat 72, 1013 KS Amsterdam

indicatoren heeft betrekking op de mate waarin toegelaten studenten verschillende kansen op een goed eindresultaat hebben al naar gelang hun achtergronden en hun studieresultaten in het eerste studiejaar.

Probleemstelling

Op basis van het bovenstaande komen wij tot de volgende probleemstelling:

- Zijn er verschillen in *onderwijsuitkomsten* tussen de studierichtingen en zo ja kunnen studierichtingskenmerken deze verschillen verklaren?
- Zijn er verschillen in *studieloopbanen* van studenten op de verschillende studierichtingen en zo ja kunnen deze verschillen worden verklaard door effecten van studierichtingskenmerken en/of achtergrondkenmerken van studenten?
- Kunnen de eventuele verschillen worden geïnterpreteerd als verschillen in efficiëntie en effectiviteit en kan er derhalve worden gesproken van een meer of minder geslaagde *implementatie van de wet Twee-Fasenstructuur* bij de verschillende studierichtingen?

2. DE SAMENSTELLING VAN HET COHORT 1982/1983

Om de onderwijsopbrengsten van studenten na invoering van de Twee-Fasenstructuur zo volledig mogelijk te kunnen onderzoeken is gekozen voor het bestuderen van het cohort 1982/1983. Bij de vaststelling van het cohort is uitgegaan van de definitie die in het Beleidskader Invoering Twee-Fasenstructuur (1981) wordt gegeven van de instroom van eerstejaars: "daaronder te begrijpen studenten die zich, eventueel na omzwaaien, voor de eerste maal in een bepaalde nieuwe studierichting laten inschrijven".

De gegevens die in dit deelonderzoek gebruikt worden zijn afkomstig van zes studierichtingen. Oorspronkelijk was het de bedoeling iedere studierichting op twee universiteiten te onderzoeken. Doordat twee studierichtingslocaties weigerden mee te werken en één locatie de gegevens niet tijdig ter beschikking kon stellen, zijn er uiteindelijk negen locaties onderzocht, verdeeld over vier universiteiten. (Onder studierichtingslocatie verstaan we een bepaalde studierichting op een bepaalde universiteit. In het vervolg wordt met locatie steeds studierichtinglocatie bedoeld.) De studierichtingen die zijn onderzocht zijn: biologie, scheikunde, geschiedenis, nederlands, psychologie en rechten. Deze studierichtingen kunnen verder worden ingedeeld in een bèta-cluster, een alpha-cluster en een gamma-cluster. (De indeling die we verder in deze publicatie zullen hanteren in bèta 1 en bèta 2, etc. komt niet overeen met de hierboven genoemde volgorde van studierichtingen. Dit is gebeurd om identificatie van afzonderlijke studierichtingslocaties tegen te gaan.) Van de studenten van cohort 1982/1983 zijn de gegevens verzameld uit bestanden van de bureau's centrale inschrijving en de studieadministraties.

Het samenvoegen van gegevens uit de twee genoemde bestanden bracht in sommige gevallen grote discrepanties aan het licht. De verschillen tussen beide bestanden variëren tussen de clusters. Het percentage studenten dat niet in beide bestanden voorkomt is in één studierichting zelfs groter dan 30%. Navraag bij studieadministraties en de centrale universitaire administraties leverde op dat de verschillen moeten worden toegeschreven aan:

- studenten die hun centrale inschrijving niet laten volgen door een inschrijving bij een studierichting (niet in de steekproef opgenomen);
- studenten die bij een verkeerd cohort zijn ingedeeld (niet in de steekproef opgenomen);
- studenten die bij de centrale administratie niet bij de juiste studierichting staan ingeschreven of helemaal niet centraal staan ingeschreven (wel in de steekproef opgenomen);

- studenten die (snel) de studie hebben gestaakt (wel in de steekproef opgenomen).

In totaal leverde deze procedure een steekproef op van 1799 studenten. Bovenstaande punten duiden op een gebrek aan éénduidigheid in het hanteren van criteria voor het vaststellen van de samenstelling van jaar-cohorten. De oorzaak hiervoor ligt in het feit dat verreweg de meeste administraties niet zijn ingericht op het doen van cohortstudies. Gevolg hiervan is dat de doorgaans gehanteerde rendementsgegevens zich laten kenmerken door een zekere mate van onbetrouwbaarheid. Alleen door een longitudinaal onderzoek naar studieprestaties van individuele studenten kunnen deze fouten worden voorkomen. Dit toont het belang aan van het verder ontwikkelen van systemen om student- en leerlingadministraties te kunnen gebruiken voor onderwijsbeoordelingen. Deze systemen dienen wel zodanig te zijn opgezet dat de privacy van de geregistreerden gewaarborgd blijft.

3. RESULTATEN

3.1 Het numeriek rendement

Als eerste performance indicator behandelen we het numeriek rendement van de propedeuse. Deze indicator is gekozen omdat het verhogen van het rendement één van de hoofddoelen was van de Wet Twee-Fasenstructuur. Vanouds is dit één van de heikele onderwerpen uit de Nederlandse Hoger Onderwijs discussie. In de literatuurstudie van Voorthuis en Wilbrink (1987) is dit uitvoerig behandeld. Zij stellen, dat ondanks het vele werk dat door commissies is besteed aan het te laag geachte rendement er weinig empirische onderzoeksgegevens bekend zijn over de oorzaken van studieuitval.

Toch is in de Wet Twee-Fasenstructuur een taakstellende rendementsverhoging van 10% voor het doctoraal opgenomen, waarmee het doctoraal rendement op 70% zou moeten komen. Bij een uitval in het doctoraal van 10%, zoals wordt verondersteld in de Wet Twee-Fasenstructuur is een propedeuserendement van minstens 78% nodig om tot een doctoraalrendement van 70% te kunnen komen. Het aggregatieniveau (universiteit, faculteit en studierichting) waarop dit rendement berekend moet worden, wordt niet aangegeven. Dit is voor de hoogte van het rendement wel van belang in verband met het al dan niet meetellen van omzwaaiers. Tegenover bovenstaande rendementsnorm kunnen we de uitkomsten plaatsen van twee oude cohortstudies van het CBS, betreffende de cohorten van 1961-1963 en het cohort 1970. Het CBS neemt als aggregatieniveau de faculteit en komt tot een numeriek rendement voor het kandidaatsexamen na 10 jaar van 74% respectievelijk 70%. Voor het afsluitend examen na 12 jaar bedraagt het numeriek rendement van beide cohorten 62% respectievelijk 56%. In beide gevallen heeft ongeveer 1 op de 10 studenten het examen afgelegd in een andere faculteit dan de faculteit waar de studie werd begonnen. Het rendement voor het kandidaatsexamen in de oorspronkelijke faculteit komt daarmee voor cohort 1961-63 op 67% en voor het cohort 1970 op 63%. Indien de oorspronkelijke studierichting als aggregatieniveau zou zijn genomen zouden de rendementen weer iets lager komen te liggen.

In Tabel 1 presenteren wij het numeriek rendement over de propedeuse van de zes studierichtingen. Het numeriek propedeuse-rendement over de totale groep is ongeveer 70%. Indien men bedenkt dat het rendement is berekend over een ruim gedefinieerd cohort (inclusief snelle studiestakers) en alleen de diploma's in de oorspronkelijke studierichting zijn gemeten, is dit geen slecht resultaat. Als we de snelle studiestakers verwijderen leidt dit tot een rendementsvermeerdering van circa 10%. De verschillen tussen de studierichtingen binnen clusters zijn aanzienlijk in het bèta- en alphacluster, terwijl binnen het gamma-

Tabel 1. Numeriek rendement propedeuse voor 6 studierichtingen, uitval doctoraal gemeten tot drie jaar na aanvang studie en het maximaal doctoraal rendement per studierichting

studierichting	propedeuse rendement	% doct. uit- val tot 3jr.*	maximaal te behalen docto- raal rendement
bèta 1	82.3	10.7	73.4
bèta 2	64.7	8.9	58.9
bèta tot.	73.2	10.0	65.9
gamma 1	70.1	2.9	68.0
gamma 2	70.8	9.5	64.1
gamma tot	70.6	8.0	65.0
alpha 1	70.5	5.4	66.8
alpha 2	58.7	25.2	43.9
alpha tot.	66.0	15.5	55.9
totaal (1799)	69.9	10.0	62.9

*) uitval = gedurende één jaar geen studiepunten meer behaald

cluster - dat het minst gezien kan worden als een echt cluster - weinig onderling verschil te zien is. Van de studenten van het cohort 1982 die het propedeusediploma hebben gehaald is 10% binnen drie jaar na aanvang van de studie gestopt. Dit betekent dat de totale doctoraal uitval na zes jaar hoger zal zijn dan de in het Beleidskader Invoering Twee-Fasenstructuur veronderstelde 10%. We zien in tabel 1 dat er opnieuw vrij grote verschillen zijn tussen de studierichtingen waarbij alpha 2 er bijzonder ongunstig uitspringt. De grote uitval bij deze studierichting kan deels worden toegeschreven aan het omzwaaien naar een andere doctoraalstudie binnen dezelfde faculteit waarvoor een alpha-propedeuse vereist is.

Indien we een vergelijking maken met de uitkomsten van het CBS over het cohort 1983/84 (CBS, 1989) zien we dat cohort 1983 in totaal iets betere resultaten behaalt (numeriek rendement 75%). Maar ook hier is weer de oorspronkelijke faculteit het aggregatieniveau, hetgeen het verschil deels verklaart. Het verschil kan verder worden verklaard uit de instabiliteit van de numerieke rendementen van verschillende jaar-cohorten en de andere wijze van dataverzameling (voor cohort 82 ook bij studieadministraties). De uitkomsten van beide cohorten komen met elkaar overeen indien voor cohort 82 de snelle studiestakers niet worden meegerekend.

3.2 Doorstroomsnelheid in de propedeuse en in het doctoraal

Als tweede performance-indicator nemen we de *doorstroomsnelheid* van studenten. In de Wet Twee-Fasenstructuur is de doelstelling opgenomen dat deze verbeterd moet worden. De middelen die genoemd werden om deze doelstelling te bereiken zijn beperking van de inschrijvingsduur en verbetering van de aansluiting van het curriculum op de belangstelling en de capaciteiten van de studenten door middel van horizontale differentiatie van het doctoraal-programma. De doorstroomsnelheid van studenten *in de propedeuse* meten we door middel van de vertraging die studenten oplopen bij het behalen van het diploma. In

Tabel 2 zien we dat slechts 50% van de studenten het diploma in de nominale studieduur van 1 jaar weet te behalen. Bij de studierichtingen uit het alpha-cluster ligt dit zelfs nog aanzienlijk lager. Opvallend is verder het resultaat van bèta 2 waar 76,1% van de studenten zonder vertraging het diploma behaalden.

Indien we deze uitkomsten combineren met die omtrent de gebruikmaking van een compensatieregeling voor onvoldoendes (Tabel 3) dan zien we dat bij de bèta-2 locaties verreweg het meeste gebruik gemaakt wordt van een dergelijke regeling. Hoewel hier geen conclusie in absolute zin aan kan worden verbonden lijkt het wel een ondersteuning voor de stelling van Wilbrink (1978) dat een compensatorische examenregeling in vergelijking met een conjunctieve regeling, waarbij elk onderdeel voldoende moet zijn, besparingen kan opleveren wat betreft de studietijd van studenten. Bij deze studierichting treedt tevens de minste vertraging in de propedeuse op. Uit Tabel 3 blijkt echter dat de mogelijkheid tot compensatie tijdens de propedeuse geen invloed heeft op het propedeuse-rendement na drie jaar. Ofschoon bèta-2 de meeste "compensatie-geslaagden" laat zien is het numeriek rendement van de propedeuse, zoals na drie jaar gemeten, niet groter dan dat van de overige studierichtingen, het is zelfs het op één na laagste. Dat betekent dat een compensatorische examenregeling voor de propedeuse aanvankelijk meer rendement oplevert, maar dat dit effect na enige jaren teniet wordt gedaan. Een compensatorische propedeuse-examen regeling bevordert het rendement na één jaar doordat meer studenten het propedeuseexamen zonder vertraging kunnen behalen maar een dergelijke regeling heeft geen effect op het aantal studenten dat uiteindelijk het propedeuse-examen behaalt.

Tabel 2. Vertraging in maanden bij het behalen van het propedeuse diploma per studierichting en cluster

s.r. cluster	geen vertr.	1 - 6 maanden	7 - 12 maanden	13 - 24 maanden	meer dan 24 maanden
bèta 1	45.6	35.4	17.1	1.9	.0
bèta 2	76.1	4.5	16.1	2.2	.7
bèta tot.	59.6	21.2	16.8	2.1	.3
gamma 1	56.6	17.6	8.1	17.6	.0
gamma 2	52.4	14.2	21.2	11.1	1.0
gamma tot.	53.3	15.0	18.4	12.6	.8
alpha 1	38.8	31.4	29.3	.5	.0
alpha 2	33.9	37.4	22.6	6.5	.0
alpha tot.	37.0	33.7	26.7	2.6	.0
totaal	50.7	21.1	20.1	7.6	.5

(1257)

Een tweede indicator van de doorstroomsnelheid die we behandelen is de verblijfsduur van de studenten die niet geslaagd zijn voor de propedeuse. Uitgangspunt bij het behandelen van deze indicator zijn twee van de drie functies van de propedeuse zoals genoemd in de Wet Twee-Fasenstructuur. Volgens de Wet Twee-Fasenstructuur moet de propedeuse een oriënterende, een selectieve en een verwijzende functie vervullen. In een deelonderzoek naar de implementatie van de Wet Twee-Fasenstructuur (Van Erp, Voorthuis en Tellegen, 1987) is aan de hier onderzochte studierichtingslocaties gevraagd in hoeverre de

Tabel 3. Percentage studenten dat geslaagd is voor de propedeuse door gebruikmaking van een compensatieregeling voor onvoldoende tentamenresultaten per studierichting en per cluster.

	PROPEDEUSE RENDEMENT	
	numeriek rendement	waarvan met compensatie (in procenten)
bèta 1	82.3	29.1
bèta 2	64.7	55.2
bèta tot.	73.2	41.2
gamma 1	70.1	26.5
gamma 2	70.8	10.7
gamma tot.	70.6	14.2
alpha 1	70.5	26.2
alpha 2	58.7	0
alpha tot.	66.0	17.4
totaal (1799)	69.9	21.3

propedeuse aan de genoemde drie functies voldoet. Hieruit blijkt dat de verwijzende functie slechts bij één locatie in praktijk wordt gebracht. De oriënterende functie daarentegen bij alle locaties. De selectieve functie wordt bij alle vervuld door middel van de zwaarte van het propedeuseprogramma. Per locatie varieert bovendien de vorm van het wettelijk verplichte advies dat de studierichting (meestal in de persoon van de studieadviseur) aan de student dient te verstrekken, de zogenaamde schriftelijke raad. Hierin adviseert de studierichting aan de student om de studie voort te zetten of te staken aan de hand van behaalde studieresultaten. De vraag is nu in hoeverre deze functies voor studenten vervuld worden.

Wij stellen dat de propedeuse voor studenten oriënterend en/of selecterend functioneert als een groot deel van de niet geslaagden stopt vóór het tweede studiejaar. Als een groot deel van de niet geslaagden stopt na het eerste en voor het derde studiejaar dan vervult de propedeuse van de betreffende studierichting geen duidelijke oriënterende maar wel een selecterende functie voor studenten. En als tenslotte een groot deel pas stopt na een verblijfsduur van meer dan twee jaar (de maximale inschrijvingsduur voor de propedeuse) dan werkt de propedeuse voor een belangrijk deel van de studenten noch oriënterend noch selecterend.

In Tabel 4 zien we dat veel van de studenten, die niet slagen, tijdens of direct na het eerste studiejaar stoppen. Alleen de studierichting bèta 1 heeft een opvallend laag aandeel in deze categorie. Bij deze studierichting vervult de propedeuse voor studenten geen duidelijk oriënterende maar wel een selecterende functie. Weinig studenten die de propedeuse niet binnen één jaar behalen zien redenen om met de studie te stoppen. Gezien het grote percentage niet propedeuse geslaagden dat in het tweede jaar afvalt lijkt de studie echter voor velen toch te zwaar te zijn. Bij bèta 2 en alpha 2 hebben veel studenten na twee jaar nog niet kunnen ontdekken of de studie hen al dan niet bevalt en of ze het aankunnen, waarna ze alsnog de studie zonder enig diploma verlaten.

Tabel 4. Verblijfsduur van de niet propedeuse geslaagden in maanden per studierichting per cluster

sr cluster	gestopt na 0 - 12 mnd.	gestopt na 13 - 24 mnd.	verblijfsduur meer dan 24 mnd.
bèta 1	20.6	67.6	11.8
bèta 2	60.3	13.7	26.0
bèta tot.	47.7	30.8	21.4
gamma 1	55.2	25.9	19.0
gamma 2	38.5	43.5	18.0
gamma tot.	42.2	39.5	18.2
alpha 1	47.3	36.4	16.4
alpha 2	61.7	12.3	25.9
alpha tot.	55.9	22.1	24.1
totaal	47.1	32.9	20.0

Als derde indicator voor doorstroomsnelheid nemen we de gerealiseerde studielast van de studenten die geslaagd zijn voor de propedeuse vier jaar na aanvang van de studie, uitgedrukt in een percentage van de totale studielast. Een probleem bij de berekening van deze indicator is dat veel studenten nog bezig zijn met stage en/of scriptie. Deze onderdelen vormen vaak een vrij groot deel van de doctoraal studielast zodat het veel uitmaakt of men net wel of nog net niet deze onderdelen heeft afgerond. Een tweede probleem is de uitval van studenten in het doctoraal, die bij sommige studierichtingen aanzienlijk is. Om het tweede probleem op te lossen hebben we een tweede berekening gemaakt voor gerealiseerde studielast waarbij de groep die gestopt is in het doctoraal buiten beschouwing is gelaten. De resultaten in tabel 5 geven een alarmerend beeld te zien. Bijna 90% van de studenten heeft op het moment dat de nominale studieduur is afgelopen meer dan 25% achterstand, hetgeen bij een vierjarige studieduur gelijk te stellen is aan meer dan één studiejaar achterstand. Meer dan 30% heeft zelfs meer dan twee jaar achterstand (minder dan 51% gerealiseerde studielast). Dit sombere beeld is niet bij alle studierichtingen even sterk terug te vinden. Gamma 2 heeft duidelijk de meeste studenten met een zeer grote achterstand (slechts een kwart of minder van de totale studielast gerealiseerd). Het is natuurlijk mogelijk dat het eerder genoemde probleem van de stage en scriptie hier sterk doorwerkt.

Bij alpha 2 is het grootste verschil te zien tussen de groep doctoraal studenten inclusief en exclusief gestopten. Dit is een logisch gevolg van het feit dat bij deze studierichting het hoogste percentage studiestakers voorkomt (Tabel 1).

De doelstelling 'verhoging van de doorstroomsnelheid' is mogelijk gerealiseerd ten opzichte van de oude structuur. Desalniettemin dreigt een grote groep studenten in de problemen te komen in verband met de inschrijvingsduurbepanking tot zes jaar.

Tabel 5. Gerealiseerde studielast - uitgedrukt in procenten van het totaal te behalen studiepunten - van de geslaagden voor de propedeuse, gemeten na vier jaar studie per studierichting en per cluster.

sr cluster	Gerealiseerde studielast							
	t/m 25%		26 - 50%		51 - 75%		75% en meer	
	incl.	excl.	incl.	excl.	incl.	excl.	incl.	excl.
	gestopten		gestopten		gestopten		gestopten	
bèta 1	7.7	.7	24.5	24.8	60.0	66.0	7.7	8.5
bèta 2	1.5	0	14.2	7.4	79.9	84.4	7.5	8.2
beta tot.	4.8	.4	19.7	16.7	67.8	74.5	7.6	8.4
gamma 1	1.5	0	17.6	16.7	50.0	51.5	30.9	31.8
gamma 2	15.8	12.5	33.8	34.6	39.0	41.0	11.3	11.8
gamma tot	12.6	9.6	30.2	30.5	41.5	43.4	15.7	16.5
alpha 1	7.5	1.0	33.9	26.7	49.3	67.6	9.2	4.8
alpha 2	16.7	0	39.5	41.9	40.4	53.5	3.5	4.7
alpha tot	10.6	.5	35.8	33.5	46.3	61.3	7.3	4.7
totaal	10.2	5.6	29.3	27.5	49.1	54.7	11.5	12.2

(incl. gestopten = alle doctoraal studenten; excl. gestopten = doctoraal studenten minus de groep die tijdens het doctoraal met de studie is gestopt.)

3.3 De studieloopbanen van studenten

Een andere manier om de resultaten van de Wet Twee-Fasenstructuur te beoordelen is om de uitkomsten te relateren aan de "gebruikte bronnen". De term gebruikte bronnen is afkomstig van Calvert (geciteerd in: Dochy, Wijnen en Bouhuys, 1987). De gangbare betekenis die aan dit begrip wordt gegeven heeft betrekking op inputkenmerken van instellingen ("resources", cf. Astin, 1985). Dergelijke hulpbronnen worden bijvoorbeeld samengesteld door financiële middelen die aan een onderwijsinstelling ter beschikking staan of bijvoorbeeld bibliothecaire voorzieningen, practicummaterialen enzovoorts. Ze kunnen echter ook betrekking hebben op de samenstelling van de studentenpopulatie, waarbij variabelen als gemiddeld niveau van vooropleiding of gemiddelde sociaaleconomische herkomst van belang kunnen zijn. Ofschoon dergelijke variabelen in deze optiek strikt genomen kenmerken van onderwijsinstellingen zijn en derhalve op dat aggregatie-niveau gemeten behoren te worden zullen we hier nagaan in hoeverre geslacht, leeftijd en vooropleiding op individueel niveau (dus als kenmerken van studenten) een verklaring kunnen bieden voor verschillen in het verloop van studieloopbanen.

Allereerst zijn we geïnteresseerd in de variabele herinschrijving in het tweede jaar. Van de studenten uit cohort 1982 komt 31% na één jaar niet terug bij dezelfde studierichting in het tweede studiejaar. De vraag is nu welke factoren een rol spelen in de beslissing al dan niet door te studeren na een jaar. In de Verenigde Staten is vrij veel onderzoek gedaan naar dit probleem (Noel e.a., 1981; Tinto, 1975; Aitken, 1982). In deze analyse sluiten we aan bij het structureel model van Aitken (1982). Aitken gaat na in hoeverre herinschrijving in het tweede jaar verklaard kan worden door studieprestaties en door een viertal factoren die betrekking hebben op objectieve en subjectieve leef- en studie-omstandigheden. In dit artikel nemen we

geen variabelen op over de leef- en studieomstandigheden van de studenten. Wel is het mogelijk dezelfde indicatoren voor studieprestaties te hanteren. Deze bleken ook in het model van Aitkin het meest van belang te zijn voor de verklaring van herinschrijving, ze verklaarden daar 19% van de variantie in herinschrijving op de betrokken Amerikaanse opleidingen. De indicatoren voor studieprestaties die we zullen gebruiken zijn:

- gerealiseerde studielast eerste jaar = STPT1
- gemiddeld behaald cijfer eerste jaar = GCIJF1
- niet/wel geslaagd voor propedeuse in het eerste jaar = PN
- al dan niet opnieuw ingeschreven als student na het eerste jaar = HERIN2
- gerealiseerde studielast na vier jaren = STPT4

Bovendien beschikken we over een drietal achtergrondkenmerken van studenten. Deze zijn:

- geslacht 0 = vrouw; 1 = man (SEXE)
- leeftijd (LEEFT)
- vooropleiding

Vooropleiding blijkt in geen van de gevallen enig effect te hebben op indicatoren voor studievoortgang. Voor wat betreft de achtergrondkenmerken zullen we ons dus beperken tot uitspraken over de invloed van leeftijd en geslacht. De regressies van herinschrijving op deze variabelen en de indicatoren voor studieprestaties zijn berekend voor twee groepen: inclusief en exclusief snelle studiestakers. Tot de snelle studiestakers worden gerekend studenten die minder dan 2% van de totale studielast hebben gerealiseerd. In tabel 6 zien we allereerst dat de verklaarde variantie van herinschrijving een stuk hoger is dan gerapporteerd door Aitken, namelijk 29% voor de totale groep exclusief snelle studiestakers en 52% voor het totale cohort. Hierbij moet wel direct aangetekend worden dat er duidelijk sprake lijkt te zijn van differentiële effecten voor de verschillende clusters. Het proces dat van invloed is op het terugkeren in het tweede studiejaar verloopt verschillend in elk van de drie clusters. Alleen bij alpha-studenten is er een positief effect van geslacht, wat betekent dat vrouwen bij gelijke studieprestaties minder terugkomen in het tweede jaar dan mannen. Bij de bèta-studenten wordt herinschrijving alleen beïnvloed door het aantal studiepunten dat is behaald en er is verder geen enkele factor uit ons model van belang.

Opvallend is dat het geslaagd zijn voor de propedeuse in de nominale studieduur (pn) geen enkel effect heeft op herinschrijving. Kennelijk haken niet alleen de studenten met de slecht(st)e studieprestaties af na het eerste jaar. We kunnen hieruit concluderen dat er van het propedeuse diploma als zodanig bij de overgang van het eerste naar het tweede studiejaar geen selecterende werking uitgaat. Ten opzichte van het Amerikaanse onderzoek van Aitken is dit een groot verschil. Daar had het geslaagd zijn verreweg het grootste effect op herinschrijving. In Nederland werkt de selectie veel meer als een proces van zelfselectie, maar binnen dit proces van zelfselectie vatten studenten het niet halen van de propedeuse niet op als een aanwijzing dat ze minder 'geschikt' zijn voor deze studie. Het is de vraag of studenten hiertoe aangespoord worden door hun studierichting. Een antwoord op deze vraag kan worden verkregen door na te gaan hoe studierichtingen vorm hebben gegeven aan de schriftelijke raad. De schriftelijke raad was namelijk expliciet bedoeld om de selectieve functie van de propedeuse inhoud te geven.

In Tabel 7 zien we dat bij slechts één bèta-studierichting vaak het advies wordt gegeven met de studie te stoppen. Bovendien, zo blijkt eveneens uit Tabel 7, is het bij veel studierichtingen mogelijk zonder (veel) beperkingen te beginnen aan het doctoraal-gedeelte van de studie als men nog niet geslaagd is voor de propedeuse. De inrichting van het onderwijs op de universiteiten is er sterker op gericht studenten met studieproblemen optimale kansen te bieden dan om ze weg te selecteren.

Tabel 6. Gestandaardiseerde coëfficiënten uit het studieloopbaanmodel voor de studenten van cohort 1982/83 uitgesplitst naar exclusief en inclusief snelle studiestakers per cluster en totaal (tussen haakjes de verklaarde variantie)

effect op:	excl. snelle studiestakers		incl. snelle studiestakers	
	HERIN2	STPT4	HERIN2	STPT4
effect van:				
STPT1				
bèta	.57	.37	.77	.44
gamma	.35	.52	.29	.51
alpha	.40	.33	.27	.32
TOT.	.38	.39	.43	.44
GCIJF1				
bèta		.13		
gamma	.21		.43	
alpha	.28	.09	.52	.10
TOT	.19	.08	.32	
PN				
bèta		.08		.12
gamma		.21		.20
alpha				
TOT		.19		.18
SEXE				
bèta				
gamma		-.05		-.04
alpha	.12		.09	
TOT				
LEEFT				
bèta				
gamma				
alpha		.07		.05
TOT				
HERIN2				
bèta	(32%)	.48	(59%)	.47
gamma	(28%)	.32	(48%)	.33
alpha	(37%)	.54	(61%)	.53
TOT	(29%)	.41	(52%)	.42
STPT4				
bèta		(78%)		(86%)
gamma		(79%)		(83%)
alpha		(71%)		(80%)
TOT		(75%)		(82%)

Tabel 7. Beantwoording vragen over schriftelijke raad*) en toelatingsregeling tot het doctoraal programma voor niet-propedeuse geslaagden per studierichting

sr	schriftelijke raad advies studie te stoppen	toelating tot doctoraal programma voor niet- propedeuse geslaagden
bèta 1	vaak	(heel) beperkt deel doctoraal programma
bèta 2	nooit/zelden	ja
gamma 1 gamma 2	nooit zelden	ja beperkt
alpha 1 alpha 2	nooit/zelden nooit/zelden	beperkt beperkt, maar niet strikt nageleefd

*) de antwoorden over de schriftelijke raad zijn afkomstig uit de enquête van het implementatie-onderzoek (Van Erp, Voorthuis en Tellegen, 1987)

Dit leidt dan wel tot de vraag of dit niet zal leiden tot een uitgestelde selectie in het doctoraal gedeelte van de studie. Dit willen we nagaan met behulp van een tweede studieresultaat namelijk de gerealiseerde studielast na vier jaar (studiepunten behaald na vier jaar). Wij verwachten dat het na vier jaar behaalde aantal studiepunten beïnvloed zal worden door dezelfde studieprestaties en achtergronden die herinschrijving beïnvloeden, maar niet in de laatste plaats natuurlijk ook door herinschrijving zelf.

In Tabel 6 zien we dat er inderdaad sprake is van een grotere selectiviteit na vier jaar dan bij de herinschrijving. De verklaarde variantie van zowel de totale groep als de groep exclusief de snelle studiestakers is bijzonder groot, resp. 82% en 75%. Opnieuw lijken er binnen de clusters verschillende mechanismen een rol te spelen in het studieloopbaanproces, maar bij alle drie de clusters is herinschrijving vanzelfsprekend een belangrijke factor in dit proces, evenals de studieprestaties in het eerste jaar.

In verband met de selectiviteit van het loopbaanproces is door Holleman (1984) de 'Wet van de remmende studieachterstand' ontwikkeld. Studenten die een studieachterstand hebben opgelopen in het begin van hun studie komen volgens deze 'wet' op een steeds grotere achterstand, doordat ze behalve met inhoudelijke studieproblemen ook nog eens te maken krijgen met organisatorische en sociale problemen. Het niet volgen van het standaard onderwijsaanbod leidt ertoe dat studenten zelf meer tijd moeten besteden aan het plannen van de studie. Bovendien kan het ertoe leiden dat ze minder sociale contacten onderhouden met medestudenten die in eenzelfde studiefase zitten.

Het proces van de remmende studieachterstand is in ons model te analyseren door te letten op de mate waarop de studieprestaties uit het eerste jaar van invloed zijn op die na vier jaar. Het sterkste zien we een groeiende studieachterstand optreden bij de gamma-studenten. Het geslaagd zijn voor de propedeuse in het eerste jaar leidt hier tot 11% meer studiepunten na vier jaar (zie tabel 6a: ongestandaardiseerde effecten). Het effect van gerealiseerde studielast in het eerste jaar (stpt1) op studiepunten 4 jaar is hier 1.81 hetgeen betekent dat bijvoorbeeld studenten die het eerste jaar 20% van de totale studielast af hadden na 4 jaar 36.2% af zullen

Tabel 6A. Ongestandaardiseerde coëfficiënten van Studiepunten na 4 jaar behorende bij Tabel 6 exclusief snelle studiestakers per cluster en totaal

effect op:	STPT4			
	excl. snelle studiestakers			
	bèta	gamma	alpha	TOT
effect van:				
STPT1	1.40	1.81	.95	1.24
GCIJF1	2.40		1.03	1.11
PN	3.60	11.23		9.63
SEXE		-.28		
LEEFT		3.14		
HERIN2	28.47	21.42	27.41	25.18
N excl.stakers	351	759	442	1552
N incl.stakers	399	879	521	1799

hebben, terwijl studenten die 10% in het eerste jaar hadden gerealiseerd na vier jaar op 18.1% zitten (*ceteris paribus*). De achterstand van deze studenten ten opzichte van de in het eerste jaar snellere groep loopt derhalve op van 10% naar 18%. In iets mindere mate zien we ditzelfde proces terug bij de bèta-studenten, maar hier speelt het gemiddelde cijfer uit het eerste jaar ook nog een rol. Bij de alpha-studenten is het ontbreken van het effect van propedeusediploma in de nominale studieduur zeer opvallend. Hier komt nog bij dat het effect van de studiepunten uit het eerste jaar kleiner is dan 1. Dit betekent dat achterstanden zelfs worden ingelopen. Hier lijkt opnieuw sprake van een effect van de hoge doctoraal uitval bij dit cluster: ook studenten met goede resultaten in het eerste jaar hebben een grote kans om later alsnog uit te vallen. Zoals reeds eerder opgemerkt is het mogelijk dat de hoge doctoraal-uitval in het alpha-cluster mede bepaald wordt door studenten die tijdens hun doctoraal-studie overstappen naar een andere alpha-kopstudie.

Van de achtergrondkenmerken zijn er twee die een effect hebben op studiepunten na vier jaar. Geslacht heeft een negatief effect bij gammastudenten, wat betekent dat vrouwen bij gelijke overige omstandigheden het iets beter doen. Het CBS (1983) vond een vergelijkbaar verschil tussen mannen en vrouwen bij rechten voor het cohort 1970 zonder controle voor eerdere studieprestaties. Leeftijd heeft een positief effect bij de alpha-studenten, oftewel oudere studenten doen het iets beter.

Al met al vinden we geen grote invloeden van de achtergrondkenmerken. Er bleken in het studieloopbaanproces grote verschillen te bestaan tussen clusters, hetgeen kan wijzen op het bestaan van studierichtingseffecten.

4. DISCUSSIE

Er zijn verschillen in onderwijsuitkomsten tussen verschillende (clusters van) studierichtingen geconstateerd. Voor wat betreft doorstromingssnelheid in de propedeuse kan geconcludeerd worden dat een compensatorische examenregeling het rendement van het eerste jaar gunstig beïnvloedt. Het uiteindelijke propedeuse-rendement hangt echter niet samen met het al dan niet bestaan van een dergelijke regeling. Het ligt voor de hand dat een dergelijke regeling alleen respijt biedt voor het onvertraagd (dat wil zeggen binnen één jaar) behalen van het propedeuse-diploma.

Voorts kan opgemerkt worden (zie tabel 5) dat er bij de bèta-studierichtingen de minste spreiding in studiesnelheid bestaat en dat de gamma-studierichtingen gekenmerkt worden door relatief veel studenten met zeer geringe studievorderingen maar ook opvallend veel studenten die na vier jaar meer dan driekwart van de totale studielast hebben gerealiseerd. Voor deze verschillen in onderwijsuitkomsten kunnen vooralsnog echter geen verklarende kenmerken van studierichtingen worden aangewezen.

De vraag of studieloopbanen in verschillende (clusters van) studierichtingen op andere wijze verlopen kan bevestigend beantwoord worden in die zin dat onafhankelijke variabelen (kenmerken van studenten en studieresultaten in het eerste jaar) per studierichting een verschillende mate van invloed hebben op afhankelijke variabelen (opnieuw ingeschreven na één jaar, studieresultaten na vier jaar). De "wet van de remmende studieachterstand" lijkt het sterkst voor gamma-studenten te gelden, terwijl bij de alpha's achterstanden juist kunnen worden ingelopen.

Gezien het uitgangspunt van dit onderzoek, namelijk evaluatie van de wet Twee-Fasenstructuur kunnen we concluderen dat binnen de clusters van studierichtingen de verschillende achtergrondkenmerken van studenten niet tot sterk wisselende kansen op studiesucces leiden. Wel zijn er verschillen in numeriek rendement en in doorstroomsnelheid, maar deze zijn niet of slechts zeer ten dele tot loopbaanverschillen te herleiden. Opgemerkt moet worden dat het nagestreefde rendement voor dit cohort waarschijnlijk niet gehaald zal worden, omdat de uitval in het doctoraal groter is dan de voorziene 10%. De grote verschillen in uitval tussen de studierichtingen kunnen ook te maken hebben met omzwaaiers. Het is goed mogelijk dat facultair berekende rendementen voor bepaalde clusters van studierichtingen hoger zullen uitvallen, omdat omzwaaiers binnen een faculteit dan niet tot de uitval worden gerekend.

Bij praktisch alle studierichtingen vervult de propedeuse een oriënterende functie. De grootste uitval van studenten die de propedeuse niet behaalden, vindt plaats onmiddellijk na het eerste studiejaar. De bedoeling van de wet Twee-Fasenstructuur om een propedeuse met een selecterende functie in te stellen lijkt echter niet erg geslaagd, gezien het feit dat het behalen van de propedeuse binnen één jaar geen waardevolle voorspeller is voor studievorderingen na vier jaar. Er wordt dan ook haast nooit het advies gegeven de studie na één jaar te staken, met uitzondering van één studierichting. Bij twee andere studierichtingen is nauwelijks enige sprake van een selecterende werking van de propedeuse; meer dan een kwart van de niet-propedeuse geslaagden staat daar twee jaar of langer ingeschreven.

Toch worden studieresultaten bij alle studierichtingen voor het overgrote deel verklaard door studieprestaties uit het eerste jaar. Het recht op een tweejarige inschrijving voor de propedeuse wordt algemeen zo opgevat dat het 'normaal' is om twee jaar over de propedeuse te doen. Gezien de doorwerking van (slechtere) studieprestaties aan het begin van de studie lijkt het ons dat dit 'normale' gedrag voor velen onder de huidige studieorganisatorische omstandigheden tot problemen kan leiden. Eerder constateerden wij al dat de doorstroomsnelheid van studenten laag te noemen is met het oog op de inschrijvingsduurbepanking. Een vrij grote groep studenten dreigt in de problemen te komen doordat ze een achterstand hebben opgelopen die binnen de hen resterende inschrijvingsduur waarschijnlijk niet meer in te lopen is. Het is de vraag of deze groep eigenlijk wel geschikt is voor de betreffende studie. Als dat niet het geval is was het eerder staken van de studie zowel voor de student als voor de universiteit beter geweest. Het universitair onderwijs zal zich opnieuw moeten bezinnen op de 'selecterende' verantwoordelijkheid die haar bij wet gegeven is.

LITERATUUR

- Aitken, N.D. (1982). College student performance, satisfaction and retention: Specification and estimation of a structural model. *Journal of Higher Education*, 53, 32-50.
- Astin, A.W. (1985). *Achieving educational excellence*. San Francisco, Jossey-Bass.
- Beleidskader invoer in tweefasen structuur (1981). 's-Gravenhage: Ministerie van Onderwijs en Wetenschappen.
- Centraal Bureau voor de Statistiek (1983). *Statistiek van het Wetenschappelijk onderwijs. Studieresultaten van de studentencohorten 1961-63 en 1970*. 's Gravenhage: Staatsuitgeverij.
- Centraal Bureau voor de Statistiek (1989). *Statistiek van het Wetenschappelijk Onderwijs. Studieresultaten en vertrek, cohort 1983/84 (stand na 3 jaar)*. 's Gravenhage: Staatsuitgeverij.
- Dochy, F.J.R.C., Wijnen, W.H.F.W. & Bouhuijs, P.A.J. (1987). *Kwaliteitsbewaking in het hoger onderwijs*. 's-Gravenhage: Staatsuitgeverij.
- Erp, M. van, Voorthuis, M. & Tellegen, B. (1987). *De implementatie van de Wet Twee-fasenstructuur bij studierichtingen*. Deelrapport 3: Evaluatieonderzoek Wet Twee-fasenstructuur. Amsterdam: SCO.
- Holleman, J.W. (1984). Handvatten voor de behandeling van de studieduur. *Tijdschrift voor Onderwijsresearch*, 9, 292-302.
- Noel, L., Levitz, R., Schurr, D., and associates (1985). *Increasing student retention*. Programs for reducing the dropout rate. San Francisco-London: Jossey-Bass Publishers.
- SCO/CSHOB (1987). *Evaluatieonderzoek Wet Tweefasen Structuur*. 's-Gravenhage: Staatsuitgeverij.
- Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-126.
- Tinto, V. (1985). Dropping out and other forms of withdrawal from college. In L. Noel, et al. (Eds.), *Increasing student retention*. San Francisco-London: Jossey Bass Publishers.
- Voorthuis, M., & Wilbrink, B. (1987). *Studielast, rendement en functies propedeuse. Relatie tussen wetgeving, theorie en empirie*. Deelrapport 2: Evaluatie-onderzoek Wet Twee-fasenstructuur. Amsterdam: SCO.
- Wilbrink, B. (1978). *Examenregeling. Deel A: studiestrategieën*. Amsterdam: Universiteit van Amsterdam.
- Wenno, M.O. (1987). *Werkloosheidscijfers hoger opgeleiden. HW/T-verkenning no. 28* 's-Gravenhage: Ministerie van Onderwijs en Wetenschappen.

Manuscript ontvangen 3-10-1989

Definitieve versie ontvangen 17-11-1989

Doet Godsdienst leren? Schoolloopbanen in de gereformeerde sector van het Nederlandse onderwijs*

Anne Bert Dijkstra**

*LVGS; Landelijk Verband van Gereformeerde Schoolverenigingen te Wezep/
RION; Instituut voor Onderwijsonderzoek Rijksuniversiteit Groningen*

ABSTRACT

The article reports on a longitudinal cohort study of schoolcareers of students in the first three years of secondary education in the Netherlands. The cohort consists of a cluster sample of 700 students who were in 1985/'86 in their last year of elementary education. These students are from 35 Calvinist schools, characterized by a strong religious involvement. The schoolcareers of these students are compared to the schoolcareers of students in a 1982/'83 cohort of 17000 students from a national sample of Dutch schools.

After adjusting for SES and academic ability some minor differences remain between schoolcareers of the two groups. Non-promotion occurred less frequently in the religious sample than in the national sample.

Through path analysis the same schoolcareer model is estimated in both samples. SES appears to be associated with differences in schoolcareers in both samples in the same way, though school achievement at the end of elementary education seems less strongly related to SES in the religious group than in the national sample.

1. INLEIDING

"Sommige scholen zijn beter dan andere in staat om de leerprestaties te beïnvloeden. Dat ligt aan allerlei factoren, maar onderzoek toont aan dat bijzondere scholen effectiever zijn dan openbare." (1)

"Uit onderzoek blijkt dat er geen verband bestaat tussen de levensbeschouwelijke kleur van de school en de kwaliteit van het onderwijs." (2)

Citaten uit twee onlangs verschenen artikelen waarin de resultaten van onderzoek naar (onder meer) de betekenis van de richting van de school voor onderwijskwaliteit onder de aandacht van een groter publiek worden gebracht. Echter, hoe is het nu? Hangen onderwijsprestaties samen met de richting van de school, of komen deze onafhankelijk van denominatie tot stand?

Beide uitspraken vormen een aardige illustratie van de nog bescheiden empirische en

* Deze bijdrage is geschreven in het kader van het onderzoek 'Schoolloopbanen in gereformeerd en nationaal onderwijs', een gezamenlijk project van het RION en LVGS, dat met steun van het Instituut voor Onderzoek van het Onderwijs wordt uitgevoerd (SVO 7018/8301)

** Met dank aan drs R.J. Bosker, drs P. van Laarhoven en prof dr J.L. Peschar voor hun bijdrage aan de totstandkoming van dit artikel

Adres: Postbus 151, 8090 AD Wezep

theoretische inzichten in de betekenis van de factor richting voor schoolsucces van leerlingen in het Nederlandse onderwijs.

Bescheiden empirische inzichten. Daarvan is ondanks het geleidelijk beschikbaar komen van meer onderzoeksresultaten (bijv. Dronkers, 1987; Blok & Eiting, 1988; De Jong & Roeleveld, 1988; Van Laarhoven & Dronkers, 1988; Costongs & Dronkers, 1989) toch nog steeds sprake.

Aangaande de lacunes in de kennis van richtingeffecten kan onder meer het beperkte aantal studies met recente informatie worden genoemd. Veel van de uitgevoerde analyses hebben betrekking op gegevens van wat oudere datum. Daarnaast springen de uiteenlopende resultaten in het oog: de uitkomsten van de verschillende onderzoeken lijken niet altijd consistent. Soms worden denominatiegewijze verschillen aangetroffen, soms ook niet, terwijl in de gevonden richtingeffecten bovendien maar weinig systematiek aanwezig lijkt. De rangorde tussen de denominaties verschilt en valt nu eens wel en dan weer niet samen met de grens tussen openbaar en bijzonder onderwijs.

Ook theoretisch laat de kennis van denominatie-effecten op de schoolloopbaan te wensen over. Een afdoende verklaring voor de relatie tussen de kleur van de school en verschillen in het loopbaanverloop is nog niet gevonden.

Wel zijn daartoe enkele aanzetten gegeven. Eén daarvan ziet per richting variërende kwaliteitsverschillen als een neven-effect van een verzuild opgezet onderwijsstelsel (zie bijv. Jungbluth & Breemans, 1984). Verzuild onderwijs zou impliciet ook standenonderwijs met zich brengen (vgl. Jungbluth, 1985; Jungbluth & Driessen, 1987).

De theorie van de regionale marktpositie zoekt de verklaring van richtingeffecten bij de secundaire functie van confessioneel onderwijs. In plaats van op de door secularisatie en ontkerkelijking afgenomen betekenis van religieuze socialisatie, oriënteren ouders zich meer en meer op de concurrentiepositie die scholen op regionale onderwijsmarkten innemen (Van Laarhoven & Bosker, 1988).

Een variant op deze theorie is de verklaring vanuit de marktstrategie van scholen, die zou bestaan uit het exploiteren van kwaliteits- en instroomverschillen teneinde leerlingen aan te trekken (Dronkers, 1987; De Jong & Roeleveld, 1988).

Kort samengevat: er zijn zeker aanwijzingen voor het bestaan van een verband tussen de richting van de school en leerlingprestaties. Echter, hoe dat verband precies verloopt, waar en onder welke omstandigheden verschillen optreden en welke mechanismen daar achter schuil gaan, zijn vragen die vooralsnog maar heel gedeeltelijk beantwoord kunnen worden.

Het is tegen deze achtergrond dat de resultaten van een vergelijking tussen een specifiek godsdienstige onderwijssector en het gehele Nederlandse onderwijs worden gepresenteerd. De aanleiding daartoe is de beschikbaarheid van een databestand van een groep leerlingen met een sterk godsdienstig geïnvolveerde achtergrond binnen een in hoge mate verzuilde sector van ons onderwijsbestel. Door de combinatie van een zowel op school- als gezinsniveau sterk doorwerkende godsdienstige overtuiging ontstaat de mogelijkheid het effect van denominatie op de schoolloopbaan na te gaan in een situatie waarin richting nog in belangrijke mate als een homogeen en inhoudelijk kenmerk kan worden opgevat. Een situatie die te midden van van levensbeschouwelijke kleur verblekende en verschietende scholen nog maar weinig frequent aangetroffen lijkt te kunnen worden.

Interessant is o.a. de vraag naar de onderwijskansen van leerlingen binnen zo'n specifieke sector ten opzichte van het overige Nederlandse onderwijs. Deze binnen het Nederlandse richtingen-onderzoek niet eerder gestelde vraag staat hier centraal. De beantwoording is echter gedeeltelijk. De analyse beperkt zich tot een vergelijking van het schoolloopbaanverloop binnen de specifiek godsdienstige sector en het overige onderwijs

zonder op de achtergronden van mogelijke verschillen in te gaan. Elders (o.a. Dijkstra & Van Laarhoven, 1989) wordt bijvoorbeeld nagegaan in hoeverre uiteenlopende gezinskenmerken als verklaring zouden kunnen dienen.

Als probleemstelling is gekozen: wijkt het schoolloopbaanverloop binnen een richting in het bijzonder onderwijs, die zich kenmerkt door een sterke godsdienstige geïnvolveerdheid, af van loopbaanpatronen zoals die voor het overige Nederlandse onderwijs kunnen worden vastgesteld? Deze probleemstelling wordt in drie deelvragen uitgewerkt: 1) worden in de twee onderscheiden groepen verschillende onderwijsresultaten behaald, 2) blijven eventuele verschillen ook na controle voor relevante instroomkenmerken bestaan, en 3) treden er verschillen op in de relaties tussen onafhankelijke, intermedierende en afhankelijke variabelen?

Nadat in paragraaf 2 de methoden en data zijn toegelicht, gaat de 3e paragraaf in op de in beide groepen behaalde resultaten. Het daarop volgende onderdeel bevat enige informatie over instroomkenmerken en corrigeert daarvoor, waarna voor beide groepen een schoolloopbaanmodel wordt geschat (par. 5). Paragraaf 6 vat de resultaten nog eens samen en plaatst er enige kanttekeningen bij.

2. DATA EN ANALYSE

De genoemde 'specifiek godsdienstige sector' betreft onderwijs van gereformeerd vrijgemaakte signatuur (verder aangeduid als 'gereformeerd'). Het gereformeerd onderwijs is een als aparte denominatie erkende, relatief kleine richting binnen het bijzonder onderwijs en functioneert als zelfstandige zuil met een eigen bestuurs- en verzorgingsstructuur in het Nederlandse onderwijssysteem. De gereformeerde richting omvat o.a. ruim 100 basisscholen en ca. 20 instellingen voor voortgezet onderwijs die begin 1989 door bijna 30.000 leerlingen werden bezocht (LVGS, 1989). De vestigingen van gereformeerde scholen zijn over het gehele land verspreid, gedeeltelijk geconcentreerd binnen de 'calvinistische gordel' (de band noord-oost - zuid-west Nederland; Van de Ven & Vink, 1986). Korthedshalve zij voor een inhoudelijke plaatsbepaling verwezen naar bijvoorbeeld Pelkmans (1983, p. 207, 208) of Kisjes & Te Velde (1983).

In verband met een vergelijkend onderzoek naar de relatie tussen sociale herkomst en onderwijskansen in het gereformeerd onderwijs wordt vanaf 1985/'86 de schoolloopbaan gevolgd van leerlingen die zich in het voorjaar van 1986 in achtste groepen van gereformeerde basisscholen bevonden. Daartoe is een a-selecte clustersteekproef van 35 scholen en ca. 700 leerlingen getrokken, het zgn. 'LVGS-cohort' (Dijkstra, 1988). In deze bijdrage worden de gegevens gebruikt van een (voor de gereformeerde richting representatieve) groep leerlingen waarover op alle meetmomenten informatie kon worden verzameld ($n=454$).

De gegevens voor het gehele Nederlandse onderwijs zijn afkomstig uit het CBS-cohortonderzoek 'Schoolkeuze Lager en Voortgezet Onderwijs' (het 'SLVO-cohort'). Het CBS brengt sinds 1982/'83 de schoolloopbaan van zo'n 17.000 leerlingen in kaart, die toen in het laatste leerjaar van de basisschool zaten (Blommers, 1983; CBS, 1988).

De opzet van het cohortonderzoek binnen de gereformeerde richting is op belangrijke onderdelen afgestemd op de bij het SLVO-onderzoek gevolgde werkwijze (zelfde meetinstrumenten, identieke operationalisering etc.). De data-sets zijn daardoor goed vergelijkbaar.

In deze bijdrage zijn uit beide onderzoeken de volgende variabelen gebruikt (zie ook noot 3). De gegevens over de loopbaan in het voortgezet onderwijs bestaan in beide cohorten uit informatie over leerjaar en schoolsoort, samengevat in de variabelen positie in leerjaar één, twee en drie. Het advies voor voortgezet onderwijs is verzameld via de basisscholen. De variabele geslacht is met een vragenlijst voor de ouders gemeten. Ook voor de factor sociale herkomst (beroep ouders) is dat het geval. Met twee non-verbale toetsen (PSB-3 en PSB-8) zijn aspecten van de cognitieve begaafdheid van leerlingen vastgesteld. In tegenstelling tot voorgaande variabelen is de prestatiescore in beide groepen niet zonder meer vergelijkbaar. Een verschillend afnametijdstip van de gebruikte schoolvorderingentoets, waardoor voor het LVGS-cohort een leerwinstperiode van acht maanden optreedt, is daarvan de oorzaak. Voor de LVGS-groep wordt daarom een voor leertijd gecorrigeerde variabele gebruikt, zodat de prestatiescore in beide cohorten beter vergelijkbaar wordt (4).

Voor beantwoording van de eerste vraag wordt volstaan met de presentatie van enkele frequentietabellen voor het LVGS- en SLVO-cohort. De instroomcontrole vindt plaats met eenvoudige tabel-elaboratie. Met multipele regressie- en padanalyse wordt het relatieve belang van de variabelen in het model nagegaan. Voor de regressieanalyse is het SPSS-X pakket (Regression) gebruikt (Norusis, 1985).

3. ONDERWIJSRESULTATEN

"Worden in beide groepen verschillende onderwijsresultaten behaald?", zo formuleerden we de eerste onderzoeksvraag. De Tabellen 1, 2 en 3 geven daarop het antwoord. Als indicatie voor behaalde resultaten fungeren de intermediaire variabele prestatiescore en de afhankelijke variabelen advies en de posities na één, twee en drie jaar verblijf in het voortgezet onderwijs.

Tabel 1 geeft de verdeling van de schoolkeuze-adviezen in het LVGS- en SLVO-cohort. Binnen de gereformeerde richting worden minder vaak adviezen voor LBO en VWO gegeven, terwijl leerlingen vaker wordt geadviseerd naar MAVO en HAVO door te stromen. De verschillen zijn niet significant.

Tabel 1. Verdeling leerlingen naar advies in LVGS- en SLVO*-cohort (%)

	LVGS	SLVO
LBO	26.0	29.7
MAVO	36.3	31.2
HAVO	25.3	21.4
VWO	12.3	14.4
overig	-	3.3
	100	100
	(454)	(16813)

* bron: CBS, 1985

De onderwijsposities in de eerste drie jaren voortgezet onderwijs zijn bij elkaar gezet in Tabel 2. Op scholen van gereformeerde signatuur lijken in het eerste leerjaar betere resultaten te worden behaald (verschillen significant, chi-kw 39.5; df 2; alfa .05 bij teruggewogen n SLVO). In hoeverre van blijvende en werkelijke verschillen sprake is, kan

pas later in de schoolloopbaan beoordeeld worden. De verdeling van de posities in het tweede leerjaar laat zien dat de gunstiger resultaten voor het LVGS-cohort terug komen in een minder aantal lage onderwijsposities en een geringer percentage doublures (chi-kw 26.8; df 4). In het SLVO-cohort blijven meer leerlingen zitten en is de categorie LBO-leerlingen groter, de doorstroming naar de MAVO is in het gereformeerd onderwijs beduidend hoger. De keuze voor hogere schooltypen blijft in het LVGS-cohort achter bij het overige Nederlandse onderwijs. De betere resultaten in het gereformeerd onderwijs moeten dus vooral bij de lage doorstromers worden gezocht.

Ook de laatste kolom van de tabel laat dat zien. Terwijl de percentages HAVO- en VWO-leerlingen in het derde leerjaar weinig verschillen, bezoeken LVGS-leerlingen minder vaak het LBO en kiezen ze vaker voor MAVO. Eveneens wordt binnen de gereformeerde denominatie minder frequent gedoubleerd (significante verschillen, chi-kw 18.2; df 4).

Tabel 2. Verdeling leerlingen naar positie in 1e, 2e en 3e leerjaar voortgezet onderwijs in LVGS- en SLVO*-cohort (%)

positie in	1e leerjaar VO		2e leerjaar VO		3e leerjaar VO	
	LVGS	SLVO	LVGS	SLVO	LVGS	SLVO
LBO-1	11.2	28.3	-	-	-	-
MAVO-1	32.4	23.8	-	-	-	-
brugklas-1	48.2	46.2	-	-	-	-
LBO-2	-	-	24.2	28.5	-	-
MAVO-2	-	-	42.7	28.1	-	-
HAVO-2	-	-	11.0	8.1	-	-
brugklas-2	-	-	0.2	15.7	-	-
VWO-2	-	-	16.7	10.6	-	-
LBO-3	-	-	-	-	24.9	28.0
MAVO-3	-	-	-	-	37.2	26.7
HAVO-3	-	-	-	-	11.0	10.7
brugklas-3	-	-	-	-	-	1.8
VWO-3	-	-	-	-	15.6	14.2
gedoubleerd	-	0.9	3.5	7.8	9.3	16.6
BAO-8	-	0.9	-	-	-	-
LBO-1	-	-	1.1	1.9	-	0.1
MAVO-1	-	-	0.9	2.8	-	0.1
brugklas-1	-	-	1.5	3.1	-	0.0
LBO-2	-	-	-	-	3.3	4.3
MAVO-2	-	-	-	-	4.0	7.4
HAVO-2	-	-	-	-	1.1	1.7
brugklas-2	-	-	-	-	-	2.2
VWO-2	-	-	-	-	0.9	0.8
overig	-	0.0	-	0.1	0.4	0.1
vertrokken/onbek.	8.1	0.3	1.5	1.2	1.5	2.0
	100	100	100	100	100	100
	(454)	(16813)	(454)	(16813)	(454)	(16813)

* bron: CBS, 1986; 1988

Tabel 3 geeft inzicht in de centrum- en spreidingsmaten van de ruwe somscores op de schoolvorderingentoets. De gemiddelden ($t = 5.1$; alfa .05; ongewogen n SLVO) verschillen significant. Alhoewel het om geringe afwijkingen gaat, wordt in het SLVO-cohort een hogere gemiddelde prestatiescore (en sd) gemeten.

De onderwijsprestaties van leerlingen op gereformeerde scholen wijken aanzienlijk af van die in het overige Nederlandse onderwijs, zo kunnen we samenvatten. Waar dat de lagere onderwijstypen betreft is dat in het voordeel van LVGS-leerlingen. Deze gunstige resultaten zetten niet door naar de hogere schooltypen. Geconcludeerd kan worden dat LVGS-leerlingen in vergelijking met leerlingen op andere scholen wat hun prestaties betreft veelal in de middenmoot gezocht moeten worden.

4. PRESTATIES EN INSTROOMKENMERKEN

Het voorgaande leidt niet als vanzelf tot de conclusie dat binnen de gereformeerde richting (op een aantal punten) ook betere onderwijsresultaten worden behaald. In de schoolkwaliteits- en effectiviteitsliteratuur wordt veelvuldig op het belang van controle voor instroomkenmerken gewezen (bijv. Creemers & Schaveling, 1985; Blom, Brandsma & Stoel, 1986; Scheerens & Stoel, 1987; Bosker & Van der Velden, 1988). Met name gaat het daarbij om variabelen als sociale herkomst en intelligentie- en prestatieverschillen bij intrede.

In het SLVO- en LVGS-cohort zijn als prestatiegeschiktheidsmaten twee cognitieve bekwaamheidsscores (reasoning en abstraction) en een prestatiescore opgenomen (tabel 3). De gemiddelde scores op de variabelen reasoning en abstraction wijken in beide groepen, evenals bij de prestatiescore het geval was, enigszins van elkaar af (verschillen significant: reasoning $t = -33.7$; abstraction $t = -7.6$; alfa .05; ongewogen n SLVO).

Een indicatie van het milieu van herkomst van de leerling geeft tabel 4, waar voor beide cohorten de verdeling van het sociaal milieu van man en hoofdkostwinner kan worden afgelezen.

Beide variabelen meten grotendeels hetzelfde. Het milieu van de hoofdkostwinner onderscheidt zich door de toevoeging van beroepsinformatie van de vrouw indien geen

Tabel 3. Gemiddelde prestatie- en bekwaamheidsscores in LVGS- en SLVO*-cohort

	LVGS	SLVO
prestatiescore:		
gemiddelde	32.7	33.1
standaarddeviatie	8.3	10.2
n	454	16588
bekwaamheidsscore, reasoning:		
gemiddelde	25.0	23.8
standaarddeviatie	4.8	4.6
n	454	16680
bekwaamheidsscore, abstraction:		
gemiddelde	23.9	23.5
standaarddeviatie	6.8	6.8
n	454	16679

* bron: CBS, 1988

gegevens van de man beschikbaar zijn. Verschillen treden vooral op bij de klassen overig en onbekend.

Vergelijking van beide cohorten levert (significante) verschillen op (hoofdkostwinner chi-kw 56.0; df 7; man chi-kw 59.3; df 7). De hogere milieucategorieën zijn in het LVGS-cohort aanzienlijk oververtegenwoordigd; het percentage leerlingen afkomstig uit lagere milieus blijft achter ten opzichte van de SLVO-groep.

Van slechts twee van de voor een instroomcontrole in aanmerking komende variabelen zijn ook bruikbare SLVO-gegevens beschikbaar: de prestatiescore en het sociaal milieu van herkomst. De keuze van de controlevariabelen staat daarmee vast. Al laat deze situatie te wensen over (vergelijkbaarheid van de prestatiescores, geen informatie over opleiding man/vrouw), de beschikbare variabelen lijken toereikend om eventuele instroomverschillen voldoende onder controle te krijgen.

Tabel 4. Verdeling leerlingen naar sociaal milieu hoofdkostwinner en sociaal milieu man in LVGS- en SLVO*-cohort (%)

	sociaal milieu hoofdkostwinner		sociaal milieu man	
	LVGS	SLVO	LVGS	SLVO
hoger employé	14.5	10.7	14.5	10.7
middelbaar employé	29.3	16.4	29.5	16.0
lager employé	8.4	10.8	7.9	10.2
zelfst. met personeel	5.9	4.4	5.9	4.3
zelfst. zonder personeel	9.0	5.7	8.8	5.6
arbeider	21.1	28.7	21.1	28.4
overig	9.9	14.0	5.3	17.1
onbekend	1.8	9.3	7.0	7.8
	100	100	100	100
	(454)	(16813)	(454)	(16813)

* bron: CBS, 1988 en CBS (ongepubliceerd materiaal)

Wanneer we, rekening houdend met instroomkenmerken, de onderwijsprestaties nogmaals bezien komen zo hier en daar opnieuw verschillen naar voren.

Allereerst weer de schoolkeuze-adviezen. Gunstige resultaten voor de LVGS-groep vinden we voor leerlingen afkomstig uit een hoog milieu met een laag of gemiddeld prestatieniveau (minder LBO-adviezen). Ook krijgen LVGS-leerlingen met een gemiddeld prestatieniveau vaker een VWO-advies en leerlingen met een laag prestatieniveau uit het hoge milieu vaker een HAVO-advies. Significante verschillen (na terugweging SLVO) doen zich echter niet voor.

De overgang naar het voortgezet onderwijs laat voor het LVGS-cohort in de lage en gemiddelde prestatiegroep enigszins betere resultaten zien. Leerlingen met een laag prestatieniveau kiezen minder vaak voor LBO en stromen vaker door naar MAVO en brugklas (hoog milieu n.s.). In de gemiddelde prestatiegroep wordt door LVGS-leerlingen eveneens vaker MAVO en minder frequent LBO gekozen, leerlingen in de SLVO-groep belanden vaker in de brugklas. Ook bij hoog presterende leerlingen uit hoge milieus is dat het geval, zij het dat -in tegenstelling tot de gemiddelde presteerders- nu van significante verschillen sprake is.

Tabel 5. Positie in derde leerjaar voortgezet onderwijs naar sociaal milieu hoofdkostwinner en prestatieniveau in LVGS- en SLVO* cohort (%)

prestatieniveau laag					
sociaal milieu					
laag			hoog		
positie VO-3					
	LVGS	SLVO	LVGS	SLVO	
LBO-3	64.0	58.7	28.9	36.1	
MAVO-3	23.3	19.3	51.1	33.9	
HAVO-3	1.2	1.5	-	3.4	
brugklas-3	-	0.4	-	0.6	
VWO-3	-	0.2	2.2	1.6	
gedoubleerd	11.6	19.6	17.8	24.4	
overig	-	0.3	-	0.1	
	100	100	100	100	n = 131
	(86)	(4014)	(45)	(1086)	n = 5100
prestatieniveau midden					
sociaal milieu					
laag			hoog		
positie VO-3					
	LVGS	SLVO	LVGS	SLVO	
LBO-3	25.4	29.3	11.8	10.9	
MAVO-3	57.1	40.1	57.9	40.7	
HAVO-3	7.9	7.4	14.5	17.4	
brugklas-3	-	1.0	-	2.9	
VWO-3	3.2	4.7	3.9	9.8	
gedoubleerd	6.3	17.4	11.8	18.2	
overig	-	0.1	-	0.1	
	100	100	100	100	n = 139
	(63)	(3265)	(76)	(1709)	n = 4974
prestatieniveau hoog					
sociaal milieu laag					
positie VO-3					
	LVGS	SLVO	LVGS	SLVO	
LBO-3	12.9	8.9	3.1	1.9	
MAVO-3	27.4	25.9	22.9	14.5	
HAVO-3	29.0	22.2	15.6	19.9	
brugklas-3	-	3.0	-	4.5	
VWO-3	27.4	29.7	49.0	48.4	
gedoubleerd	3.2	10.2	7.3	10.8	
overig	-	-	2.1	-	
	100	100	100	100	n = 158
	(62)	(2366)	(96)	(2378)	n = 4744

* bron: CBS

De na drie jaar verblijf in het voortgezet onderwijs bereikte positie is weergegeven in tabel 5. In het SLVO-cohort wordt vaker gedoubleerd, bevinden zich minder leerlingen in LBO en MAVO en kiezen doorgaans wat meer leerlingen voor HAVO en VWO. Een uitzondering vormen de leerlingen met een laag prestatieniveau uit een hoog milieu alsook de groep afkomstig uit de lagere milieus met een gemiddeld prestatieniveau, die in het LVGS-cohort minder vaak naar het LBO doorstromen (geen significante verschillen).

5. SCHOOLLOOPBAANMODELLEN

Voor beantwoording van de laatste deelvraag -is er verschil in de relaties tussen de variabelen in het SLVO- en LVGS-model?- wordt tenslotte voor beide cohorten een schoolloopbaanmodel (paddiagram) geschat.

Padanalyse kent enkele belangrijke beperkingen. Eén daarvan is dat het model niet op betrouwbaarheid kan worden getoetst maar theoretisch moet worden onderbouwd. Omdat het in deze bijdrage niet zozeer gaat om het vinden van een zo goed mogelijk passend model (5), alswel om de relatieve bijdrage van de variabelen aan de verklaring van de schoolloopbaan in beide cohorten, valt met deze beperking te leven. Een belangrijk punt is ook de causale ordening van de variabelen in het model. Gezien het geringe aantal variabelen en de chronologische volgorde daartussen levert die ordening hier weinig problemen op. Verder moet bedacht worden dat de resultaten afhankelijk zijn van de keuze van de in het model opgenomen variabelen: wijziging van het model leidt tot veranderde conclusies (bijv. De Leeuw, 1978).

Aan de toepassing van padanalyse worden verschillende voorwaarden verbonden. Aan één daarvan, de vereiste lage multicollineariteit, lijkt tot op zekere hoogte te zijn voldaan (vgl. tabel 6). Ook het meetniveau van de variabelen wordt voldoende geacht, al is strikt genomen niet steeds sprake van meting op intervalniveau (zie noot 6). Dat de relaties tussen de variabelen lineair zijn wordt aangenomen en evenmin staat vast dat alle van belang zijnde variabelen in het model zijn opgenomen.

Om te beginnen zijn in Tabel 6 de produkt-moment correlatie-coëfficiënten tussen de in de modellen opgenomen variabelen vermeld. De tabel laat zien dat in het SLVO-cohort de verbanden tussen het milieu van de vader enerzijds en anderzijds het advies en de posities in het voortgezet onderwijs iets sterker lijken dan in de LVGS-groep, uitgezonderd het derde leerjaar. Voor de samenhang van een bekwaamheidsindicator (reasoning) en de prestatie-

Tabel 6. Correlatiematrix LVGS- en SLVO*-cohort

	1	2	3	4	5	6	7	8
								LVGS
1. sociaal milieu		.08	.22	.23	.33	.20	.33	.30
2. bekwaamheid-reasoning	.11		.36	.45	.30	.05	.33	.37
3. bekwaamheid-abstraction	.09	.36		.36	.34	.07	.31	.35
4. prestatiescore	.30	.34	.27		.63	.18	.65	.66
5. advies	.35	.29	.24	.68		.49	.78	.75
6. positie VO-1	.35	.25	.19	.58	.78		.45	.40
7. positie VO-2	.35	.26	.21	.63	.76	.79		.92
8. positie VO-3 SLVO	.28	.23	.19	.57	.64	.57	.77	

SLVO: listwise deletion missing data, n=13744; LVGS: pairwise deletion missing data, min. n=388; coëfficiënten significant ($p \leq .01$)

* bron: CBS

score met het milieu is dat eveneens het geval, in tegenstelling tot de andere bekwaamheidsvariabele (abstraction). De correlaties tussen de prestatiescore en de schoolloopbaanvariabelen zijn vanaf het tweede verblijfsjaar in het voortgezet onderwijs wat sterker in het LVGS-cohort. Daarvòòr (advies en positie eerste leerjaar) geldt dat voor de samenhang met de prestatiescore in het SLVO-cohort. Ook het advies correleert na het eerste jaar voortgezet onderwijs in de LVGS-groep sterker met de loopbaan. Voor het eerste jaar is dat opnieuw niet het geval.

Voor de laatste onderzoeksvraag zijn vooral de Figuren 1 en 2 interessant. In de figuren zijn de ruwe regressie-coëfficiënten af te lezen. De gestandaardiseerde coëfficiënten (beta's) zijn tussen haken vermeld. Op basis daarvan is vervolgens de bijdrage van een variabele aan het loopbaanverloop in directe en indirecte effecten uiteengelegd (vgl. bijv. Dessens, Jansen & Swanborn, 1974). Voor beide groepen is eenzelfde model geschat (coëfficiënten vermeld indien $\geq .05$). (In Tabel 7 is voor de variabelen in het model de verklaarde variantie af te lezen, evenals het gemiddelde en de standaardfout.)

De voornaamste punten worden kort aangeduid; met name aan de loopbaanvariabelen, de prestatie-indicatoren en de sociale herkomst wordt aandacht geschonken.

Tabel 7. Gemiddelde, standaardafwijking en verklaarde variantie in LVGS- en SLVO*-cohort

	rg		sd		n		R ²	
	LVGS	SLVO	LVGS	SLVO	LVGS	SLVO	LVGS	SLVO
geslacht	1.50	1.50	0.50	0.50	448	13744		
sociaal milieu	4.45	4.14	2.01	2.08	422	13744		
bekwaamheid-reasoning	25.00	23.94	4.77	4.48	448	13744	.01	.02
bekwaamheid-abstract.	23.89	23.79	6.80	6.57	445	13744	.05	.01
prestatiescore	32.72	33.64	8.31	10.17	442	13744	.27	.21
advies	2.24	2.81	0.98	1.09	454	13744	.45	.50
positie VO-1	2.41	2.21	0.70	0.90	417	13744	.28	.61
positie VO-2	3.18	3.12	1.03	1.08	447	13744	.67	.70
positie VO-3	4.08	3.87	1.09	1.37	447	13744	.86	.61

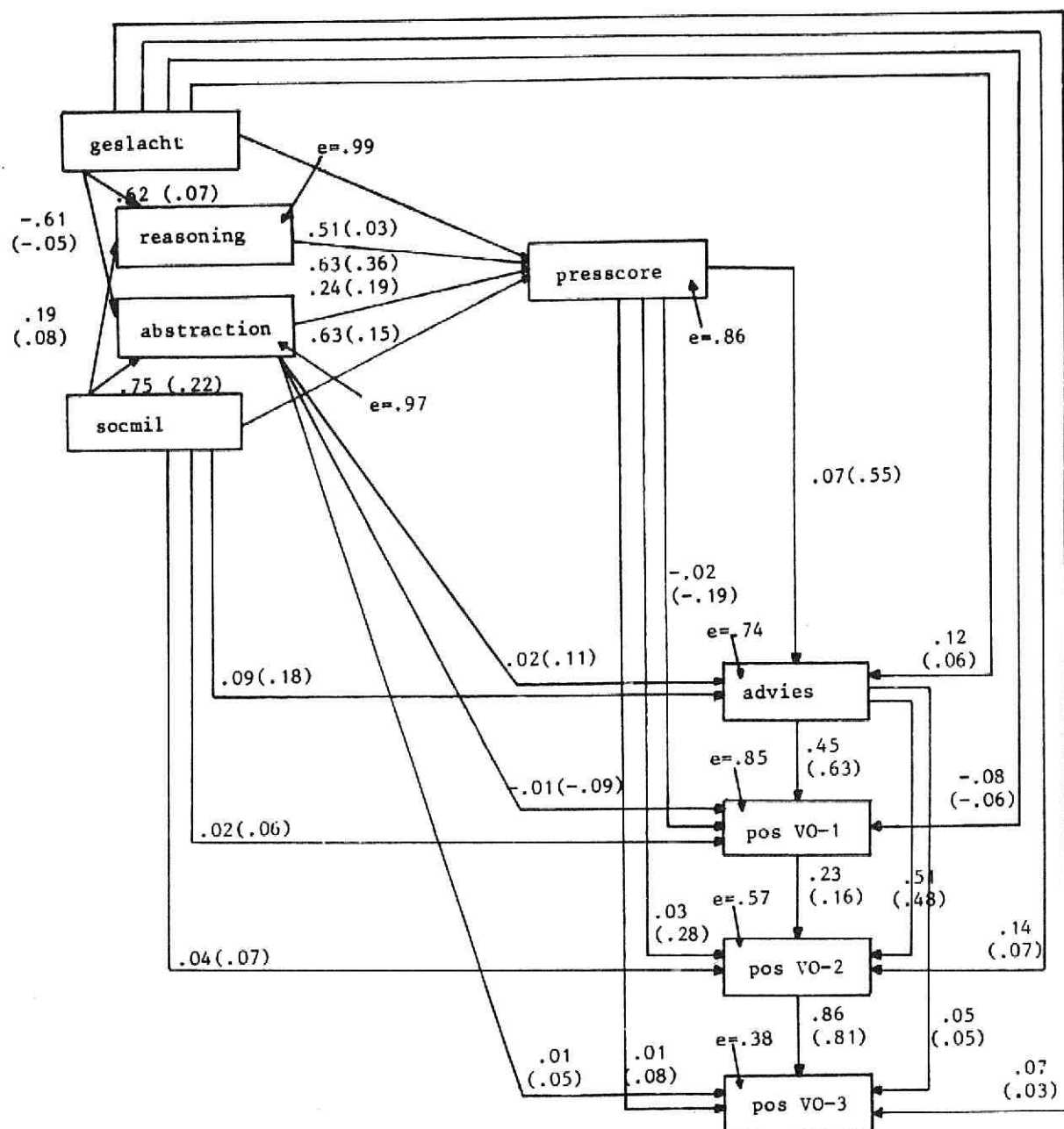
* bron: CBS

Allereerst valt op dat de totale verklaarde variantie in beide groepen uiteen loopt. De in het model opgenomen variabelen binden voor het LVGS-cohort meer van de variantie van de onderwijspositie drie jaar na het verlaten van de basisschool dan voor de SLVO-groep het geval is (R^2 .86 en .61). Ook voor de andere variabelen laat de tabel verschillen in verklaarde variantie zien, al ligt deze doorgaans in dezelfde orde van grootte. Een uitzondering daarop vormt met name de voor de positie in het eerste jaar voortgezet onderwijs verklaarde variantie, welke in het LVGS-cohort beduidend lager is dan het voor de SLVO-groep gevonden resultaat. Dit verschil lijkt grotendeels te worden verklaard door de aanwezigheid van een relatief groot aantal (potentiële) LBO-leerlingen in de brugklas-categorie in het LVGS-cohort.

De uiteenlopende residu-componenten ten aanzien van de variabele positie VO-3 laten zich minder gemakkelijk verklaren. Wellicht is het de-in vergelijking met de gereformeerde sector- minder homogene samenstelling van het gehele Nederlandse onderwijs, die z'n eigen verklarende variabelen behoeft?

Vergelijking van de verbanden in beide groepen vindt plaats met de ongestandaardiseerde coëfficiënten.

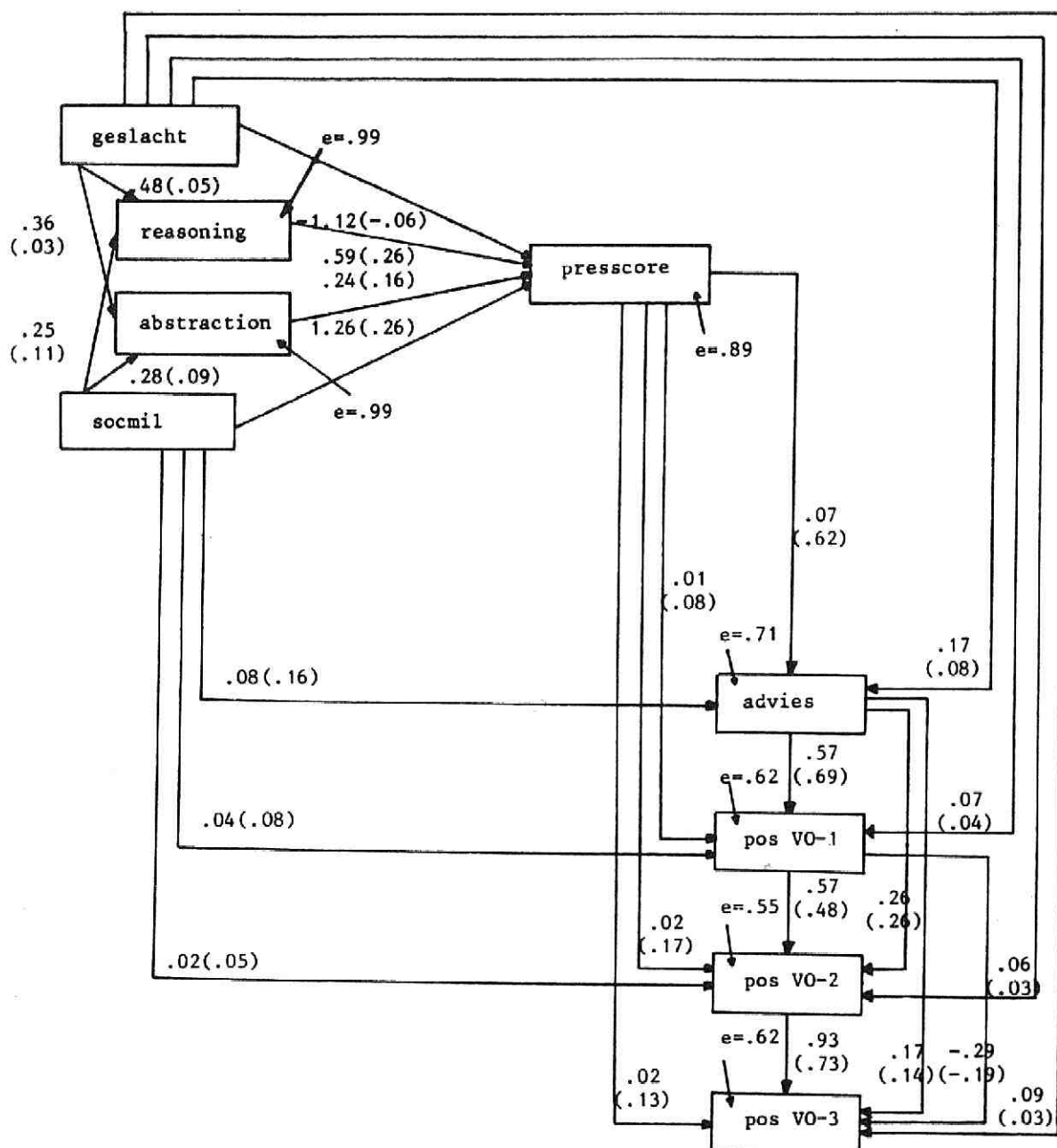
Het sociaal milieu van de vader is als exogene variabele in het model opgenomen. De



Figuur 1. Paddiagram LVGS-cohort met ruwe (en gestandaardiseerde) padcoëfficiënten ($p < .05$)

verschillen aangaande de invloed van het milieu op de loopbaan (advies, positie) tussen beide groepen zijn gering. Opvallend is wel het verschil in de samenhang met de prestatie-indicatoren. De samenhang tussen milieu en prestatiescore is in het SLVO-cohort sterker, in het LVGS-cohort vinden we een sterkere relatie met de bekwaamheid (-abstraction) van leerlingen.

Als prestatie-variabelen bevat het model twee exogene bekwaamheidsvariabelen en de prestatiescore, een intermediërende modelvariabele. Direct springen weer de verschillen rond de abstractionscore in het oog. In tegenstelling tot Figuur 2 laat Figuur 1 relaties met het schoolkeuze-advies en de positie in het eerste en derde leerjaar voortgezet onderwijs zien. De effecten van de prestatie-score op de schoolloopbaan verschillen weinig. Het belang van advies wijkt in de LVGS-groep af. Het schoolkeuze-advies beïnvloedt in het gereformeerd onderwijs het sterkst de positie in het tweede leerjaar. In het eerste en derde



Figuur 2. Paddiagram SLVO-cohort met ruwe (en gestandaardiseerde) padcoëfficiënten ($\geq .05$)

leerjaar vinden we voor het SLVO-cohort een sterker verband. Samenvattend kan worden geconstateerd dat de verbanden tussen de variabelen in het LVGS- en SLVO-model over het algemeen weinig verschil vertonen. Een uitzondering daarop vormt met name de invloed van de bekwaamheidsscore (abstraction) in het LVGS-cohort. Verder lijkt het milieu van herkomst een iets minder grote invloed op de prestatiescore van LVGS-leerlingen te hebben. In het tweede leerjaar voortgezet onderwijs hangt het schoolkeuzeadvies binnen de gereformeerde richting in wat sterkere mate samen met de onderwijspositie.

Tenslotte worden met de gestandaardiseerde regressie-coëfficiënten nog de directe en indirecte effecten voor beide modellen nagegaan.

Het sociaal milieu van herkomst is zowel in de LVGS- als SLVO-groep van belang voor

alle in het model opgenomen factoren. In beide cohorten oefent de sociale herkomst de sterkste invloed uit op het schoolkeuzeadvies: .29 in het LVGS- en .28 in het SLVO-cohort. In het SLVO-cohort is het milieu van vrijwel even grote betekenis voor de positie in het eerste, tweede en derde leerjaar. De invloed van milieu is vooral indirect, via prestaties en advies. Ook in het LVGS-cohort gaat het vooral om een indirecte invloed van het sociaal milieu. Resumerend: de sociale achtergrond van leerlingen beïnvloedt in beide sectoren vrijwel steeds de schoolloopbaan, zowel rechtstreeks als via andere factoren. In de modellen worden verder effecten van de prestatie- en bekwaamheidsscores op de loopbaanvariabelen verondersteld.

Zoals we zagen hangen de bekwaamheidsscores in de LVGS-groep sterker samen met het schoolkeuze-advies en de positie in VO-2 en VO-3. Het effect op advies verloopt indirect, via de prestatiescore. De invloed van bekwaamheid op VO-2 en VO-3 komt eveneens hoofdzakelijk indirect tot stand en speelt een rol onder meer door de variabelen advies en schoolprestaties. Bekwaamheid(-reasoning) neemt de grootste invloed voor z'n rekening. Het effect van abstraction is minder sterk. In tegenstelling tot VO-2 en VO-3 wordt de positie in het eerste leerjaar voortgezet onderwijs in het SLVO-cohort sterker door bekwaamheid bepaald. In de SLVO-groep komt dit effect op indirecte wijze via schoolprestaties en advies tot stand. In het LVGS-cohort is er bovendien een direct effect van abstraction op positie VO-1.

Tevens vonden we zowel in het LVGS- als SLVO-cohort een effect van de prestatiescore op de positie in het tweede en derde leerjaar voortgezet onderwijs. De invloed op positie VO-2 komt in het SLVO-cohort vooral indirect tot stand (via advies en eerste schoolkeuze), binnen het gereformeerd onderwijs speelt ook het directe effect een belangrijke rol. Omgekeerd draagt de samenhang tussen prestatiescore en positie VO-3 in het LVGS-cohort grotendeels een indirect en in het SLVO-cohort een direct karakter. Net als bij de bekwaamheidsscores wordt ook voor de prestatiescore in het SLVO-cohort een groter (vooral indirect) effect op de positie in het eerste leerjaar voortgezet onderwijs gevonden. Daarnaast is nu de invloed van schoolprestaties op advies in het SLVO-cohort wat sterker. Het schoolkeuze-advies wordt in beide groepen voor een belangrijk deel door de prestatiescore bepaald. De bekwaamheid is van minder betekenis. Beide bekwaamheidsvariabelen oefenen in het LVGS-cohort een wat grotere invloed uit. Vrijwel steeds gaat het om een indirect verband via de prestatiescore. Binnen de gereformeerde sector heeft abstraction ook een direct effect op het schoolkeuze-advies. De invloed van de sociale achtergrond van de leerling op advies is in het LVGS-cohort voor een groot deel direct, bij SLVO-leerlingen is het indirecte effect wat groter.

Dat de verbanden tussen de eerste schoolkeuze en de daaraan in het model voorafgaande variabelen in het gereformeerd onderwijs zwakker zijn dan in het SLVO-cohort is al eerder opgemerkt. In beide sectoren hangt de positie in het eerste jaar voortgezet onderwijs sterk samen met het schoolkeuze-advies. Via het advies spelen, althans in het SLVO-cohort, ook de prestaties in het basisonderwijs een belangrijke rol. De invloed van sociaal milieu is grotendeels indirect en loopt via advies en prestaties. Van enige indirecte betekenis zijn in het SLVO-cohort de bekwaamheidsscores.

Ook twee jaar na het verlaten van de basisschool zijn het vooral het advies en de prestatiescore die de weg van de leerling door het voortgezet onderwijs bepalen. In de SLVO-groep gebeurt dat grotendeels via andere variabelen, in het LVGS-cohort spelen directe effecten een meer belangrijke rol. Voor het SLVO-cohort is daarnaast ook de positie in VO-1 van relatief grote invloed op het in het tweede jaar bereikte onderwijsniveau.

Na drie jaar verblijf in het voortgezet onderwijs verandert het patroon weinig. Nog steeds zijn de prestatiescore en het advies van invloed op de gang door het onderwijs. In het

LVGS-cohort is die invloed wat kleiner dan in de SLVO-groep. Binnen de gereformeerde richting vinden we sterkere indirecte effecten. De directe effecten overtreffen in het SLVO-cohort de rechtstreekse invloed in de LVGS-groep. De beide bekwaamheidsvariabelen hangen eveneens samen met de positie in het derde jaar voortgezet onderwijs, met name in de gereformeerde sector. Drie jaar na de overgang naar het voortgezet onderwijs laat ook de sociale herkomst van leerlingen nog steeds z'n invloed gelden. Het gaat dan om een indirecte invloed op de positie VO-3 o.a. via schoolprestaties, advies en positie VO-1 en -2. Verreweg het belangrijkste is de positie in het tweede leerjaar voortgezet onderwijs. De eerste schoolkeuze is in beide cohorten van veel minder belang.

6. CONCLUSIE

Ter beantwoording van de onderzoeksvragen worden de conclusies uit de voorgaande paragrafen tenslotte kort gerecapituleerd.

Uitgaande van het door de leerlingen bereikte onderwijsniveau in het eerste, tweede en derde jaar in het voortgezet onderwijs lopen de in beide sectoren behaalde resultaten uiteen. Binnen de gereformeerde richting stromen in verhouding weinig leerlingen het LBO binnen en wordt een grotere MAVO-deelname geconstateerd. Tevens lopen relatief weinig leerlingen op gereformeerde scholen gedurende de eerste periode in het voortgezet onderwijs vertraging op.

Vervolgens vroegen we ons af of deze verschillen ook na correctie voor instroomkenmerken blijven bestaan. Dat is slechts gedeeltelijk het geval. Wanneer met de sociale herkomst en het prestatieniveau rekening is gehouden, worden in de gereformeerde sector geen betere resultaten aangetroffen, enkele deelgroepen uitgezonderd. De na de instroomcorrectie gevonden resultaatverschillen zijn vooral in het voordeel van het overige Nederlandse onderwijs, waar in het algemeen een hoger onderwijsniveau (HAVO, VWO) wordt bereikt. Wel blijft ook na controle voor instroomverschillen het grotere aantal onvertraagde schoolloopbanen binnen de gereformeerde denominatie opvallen.

De laatste onderzoeksvraag betrof mogelijke verschillen in de relaties tussen de variabelen. Deze zijn gevonden, zowel in de sterkte als het verloop van verbanden, maar vrijwel steeds gaat het slechts om betrekkelijk geringe verschillen.

Concluderend: de vergelijking van het schoolloopbaanverloop binnen een sterk godsdienstig geïnvolveerde richting in het onderwijs met loopbaanpatronen in het overige Nederlandse onderwijs laat zo hier en daar verschillen zien. Naast een verschillende invloed van een aantal loopbaanpredictoren op de onderwijs carrière, zijn ook de in beide groepen bereikte resultaten niet gelijk. Van duidelijk betere dan wel achterblijvende onderwijskansen binnen de specifiek godsdienstige onderwijssector lijkt geen sprake.

Bij deze conclusie zijn enkele kanttekeningen op z'n plaats.

Onder meer de beperkingen van padanalyse en de operationalisering van de variabele sociaal milieu maken verdere analyse wenselijk.

De bevindingen rond de positie in het eerste leerjaar voortgezet onderwijs doen vermoeden dat ook van het onderwijsaanbod binnen de gereformeerde denominatie en de bereikbaarheid daarvan een belangrijk effect op het verloop van LVGS-schoolloopbanen uitgaat. Nagegaan wordt of schoolloopbanen binnen de gereformeerde richting beter beschreven kunnen worden met de factoren onderwijsaanbod en reisafstand.

Met de uiteenlopende verdelingen van variabelen is in de uitgevoerde analyses maar weinig rekening gehouden. Bovendien zijn mogelijk intermediaire factoren als gezins-

grootte, etniciteit, werkloosheid en de regionale spreiding van de gereformeerde richting buiten beschouwing gebleven.

Ook de gevonden resultaten geven aanleiding voor een genuanceerde benadering. Onderzocht is het schoolloopbaanverloop tot ongeveer halverwege het voortgezet onderwijs, hetgeen aan de uitkomsten een voorlopig karakter geeft. Niet duidelijk is bijvoorbeeld welk effect zal uitgaan van doubleren of doorstroming naar hogere onderwijstypen op het uiteindelijk te bereiken onderwijsniveau.

Door de gereformeerde denominatie niet met andere richtingen te vergelijken maar het gehele Nederlandse onderwijs als referentiegroep te kiezen, is een relatief hoge drempel voor het vinden van richtingeffecten gehanteerd: mogelijk verdwijnen minder succesvolle sectoren in het algemene beeld.

NOTEN

1. Intermediair, jrg 25 (8), 24 februari 1989, p. 61; J. Dronkers: Verschillen tussen scholen.
2. Consumentengids, maart 1989, p. 163; Zo kiest u de juiste basisschool ('artikel () geschreven in samenwerking met het Rotterdams Instituut voor Sociologisch en Bestuurskundig Onderzoek van de Erasmusuniversiteit').
3. Uit het LVGS- en SLVO-onderzoek zijn de volgende variabelen beschikbaar:
 - POSITIE VO-1; positie van de leerling in het voortgezet onderwijs, één jaar na het verlaten van het basisonderwijs
 - POSITIE VO-2; idem na twee jaar
 - POSITIE VO-3; idem na drie jaar
 - SCHOOLKEUZE-ADVIES; (hoogste) advies voor voortgezet onderwijs
 - GESLACHT; geslacht van de cohortleerling
 - SOCIAAL MILIEU MAN; beroepsniveau vader naar CBS-classificatie
 - SOCIAAL MILIEU HOOFDKOSTWINNER; beroepsniveau hoofdkostwinner naar CBS-classificatie
 - BEKWAAMHEIDSSCORE -REASONING; somscore PSB-3 toets (denkvaardigheid gemeten met het 'Prüfsystem für Schul- und Bildungsberatung PSB'; Horn, 1968)
 - BEKWAAMHEIDSSCORE -ABSTRACTION; somscore PSB-8 toets (abstractievermogen gemeten met het 'Prüfsystem für Schul- und Bildungsberatung PSB'; Horn, 1968)
 - PRESTATIESCORE; gecorrigeerde somscore van toets met CITO eindtoetsitems (onderdelen taal, rekenen en informatieverwerking)
4. Omdat geen correctiefactor voor leertijdtoename beschikbaar was, is deze geschat. De in het SLVO- en LVGS-cohortonderzoek gebruikte schoolvorderingentoets is samengesteld uit een selectie items uit de CITO Eindtoets 1981. Daarmee zijn voor drie momenten gedurende het schooljaar toetsgegevens aanwezig (SLVO Schoolvorderingentoets, afgenomen in (cursus)maand 0/1; CITO Eindtoets, afgenomen in maand 5; LVGS Schoolvorderingentoets, afgenomen in maand 8/9). Voor de correctie is gebruik gemaakt van de p-waarden van de CITO- en SLVO-toets: de toename van de CITO p-waarden ten opzichte van de SLVO-waarden wordt opgevat als maat voor de opgetreden leerwinst. De gecorrigeerde prestatiescore voor de LVGS-groep is berekend door de per item gemeten score te vermenigvuldigen met het produkt van de LVGS p-waarde en het quotiënt van de SLVO en CITO p-waarden. Na correctie verschillen de, aanvankelijk hogere, prestatiescores in het LVGS-cohort nog maar weinig van het voor de SLVO-groep gevonden resultaat.
De correctie kent enkele tekorten, waarvan de voornaamste is dat de vergelijkbaarheid van de gebruikte data (zodanig dat verschillen uitsluitend kunnen worden toegeschreven aan ongelijke toetsafnametijdstippen) gedeeltelijk wordt verondersteld. De betrouwbaarheid wordt vergroot door getransformeerde (standard-five verdeelde) scores te gebruiken.

5. Met de door de vergelijkende doelstelling van de analyse ('verschillen de relaties tussen de variabelen?') ingegeven keuze voor het schatten van identieke padmodellen wordt voorbij gegaan aan de vraag naar de best passende modellen. Wellicht komen de optimaal fittende modellen voor beide cohorten niet geheel overeen. De gevolgde werkwijze wordt echter ondersteund door de resultaten van loglineaire analyse, waaruit blijkt dat beide groepen met identieke modellen goed beschreven kunnen worden (Dijkstra, 1989).
6. De drie variabelen positie voortgezet onderwijs zijn gehercodeerd, waarbij is gebruik gemaakt van de leerjarenladder van Bosker en Van der Velden (1986). De volgende categorieën worden onderscheiden:
 - POSITIE VO1/2/3: 1) LBO-1; 2) MAVO-1, MS-1, LBO-2; 3) HAVO-1, BKL-1, MAVO-2, MS-2, LBO-3; 4) VWO-1, HAVO-2, BKL-2, MAVO-3, MS-3; 5) VWO 2, HAVO-3, BKL-3, MAVO-4; 6) VWO-3.
 Overige variabelen in het model:
 - SCHOOLKEUZE-ADVIES: 1) LBO; 2) MAVO, MS; 3) HAVO; 4) VWO.
 - GESLACHT: 1) jongen; 2) meisje.
 - SOCIAAL MILIEU MAN: 1) overig; 2) arbeider; 3) lager employé; 4) zelfstandige zonder personeel; 5) zelfstandige met personeel; 6) middelbaar employé; 7) hoger employé.
 - BEKWAAMHEIDSSCORE -REASONING: somscore (max. 40)
 - BEKWAAMHEIDSSCORE -ABSTRACTION: somscore (max. 40)
 - PRESTATIESCORE: somscore (max. 60). Alhoewel de gecorrigeerde prestatievariabele is gebruikt, is de toegepaste correctie nu zonder betekenis (lineaire transformatie).

LITERATUUR

- Blok, H. & Eiting, M.H. (1988). Openbare en bijzondere lagere scholen vergeleken op de taalvaardigheid van hun zesdeklassers. *Mens en maatschappij* 63, 260-276.
- Blom, M.J.T., Brandsma, H. & Stoel, W.G.R. (1986). *Effectieve scholen in het basisonderwijs, een voorstudie. Basisschoolkenmerken als determinanten van het functioneren van de leerlingen*. Groningen: RION.
- Blommers, A.J. (1983). *Het onderzoek schoolloopbaan van leerlingen bij het voortgezet onderwijs (SLVO-1982)*. Voorburg: CBS.
- Bosker, R.J. & Velden, R.K.W. van de (1986). Onderwijspositie en selectie. Nijhof, W.J. & Warries, E. (red.): *De opbrengst van onderwijs en opleiding* (p. 109-123). Lisse: Swets en Zeitlinger.
- Bosker, R.J. & Velden, R.K.W. van der (1988). *Schooleffecten en rendementen*. Paper voor de Onderwijs Research Dagen. Groningen: RION.
- CBS (1985). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs (cohort 1982, SLVO, peildatum september 1983)*, mededeling nr. 7833.
- CBS (1986). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs (cohort 1982, SLVO, peildatum september 1984)*, mededeling nr. 7858.
- CBS (1988). *Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs. SLVO-cohort 1982, deel 1 schoolkeuze*. 's-Gravenhage: Staatsuitgeverij.
- Costongs, J. & Dronkers, J. (1989). *A bet on what works in your school*. Paper presented at the international congress for school effectiveness. Tilburg University.
- Creemers, B.P.M. & Schaveling, J. (1985). *Verhoging van onderwijs-effectiviteit*. Werkdocumenten Basisvorming in het onderwijs. 's-Gravenhage: WRR.
- Dessens, J., Jansen, W. & Swanborn, P.G. (1974). Padanalyse: uitgangspunten en basisbegrippen. *Mens en maatschappij* 49, 157-184.
- Dijkstra, A.B. (1988). *De overgang van basis- naar voortgezet onderwijs in het gereformeerd onderwijs. Tussenverslag schoolloopbanen in gereformeerd en nationaal onderwijs*. Wezep/Groningen: LVGS/RION.
- Dijkstra, A.B. & Laarhoven, P. van (1989). *Een religieuze factor in opvoedingsstijlen?* Paper voor de achtste Onderwijssociologische conferentie. Wezep/Groningen: LVGS/RION.
- Dijkstra, A.B. (1989). *Onderwijsresultaten binnen en buiten het gereformeerd onderwijs. Tweede*

- tussenverslag *Schoolloopbanen in gereformeerd en nationaal onderwijs*. Wezep/Groningen: LVGS/RION.
- Dronkers, J. (1987). *Het effect van de schoolkenmerken richting en coëductie op de individuele prestaties en het vervolg-advies in de jaren vijftig*. Paper voor de zevende Onderwijssociologische Conferentie. Tilburg: Katholieke Universiteit.
- Horn, W. (1968). *Prüfsystem für Schul- und Bildungsberatung PSB. Handanweisung für die Durchführung, Auswertung und Interpretation*. Göttingen: Verlag für Psychologie.
- Jong, U. de & Roeleveld, J. (1988). *Openbaar en bijzonder onderwijs: een constant verschil? Effectiviteit en selectiviteit van Amsterdamse scholen voor voortgezet onderwijs in de zeventiger jaren*. Paper voor de Vlaams-Nederlandse Studiedagen voor sociologen en antropologen. Amsterdam: Vrije Universiteit.
- Jungbluth, P. (1985). *Verborgene differentiatie. Leerlingbeeld en onderwijsaanbod op de basisschool*. Nijmegen: ITS.
- Jungbluth, P. & Breemans, A. (1984). Ongelijkheidsreproductie door differentiatie. Het zoekgeraakte ideaal van de eenheidsschool. *Comenius* 4, 188-211.
- Jungbluth, P. & Driessen, G. (1987). Eindtermen en verzuiling. Verkennende analyses rond school-effectiviteit. *Tijdschrift voor Onderwijswetenschappen* 18, 1-12.
- Kisjes, A.J. & Velde, P.A. te (red.) (1983). *Het gereformeerd onderwijs. Identiteitsbezinning*. Kampen: Kok.
- Laarhoven, P. van & Bosker, R.J. (1988). *Richtingseffecten op schoolloopbanen: een multilevel onderzoek (voorstel voor onderzoek)*. Groningen: RION.
- Laarhoven, P. van & Dronkers, J. (1988). Verschillen de onderwijsresultaten in openbaar en bijzonder onderwijs? Wieringen, A.M.L. van, Karsten, S. & Ax, J. (red.): *Gemeenten en onderwijs* (p. 75-88). Lisse: Swets en Zeitlinger.
- Leeuw, J. de (1978). De politieke relevantie van correlaties. *Sociologische Gids* 25, 31-39.
- LVGS (1989). Cijfers leerlingenbestand gereformeerd onderwijs 1988/'89. *LVGS-Infoblad*, nr. 62, 2-4.
- Norusis, M.J. (1985). *SPSS-X Advanced statistics guide*. Chicago: SPSS Inc.
- Pelkmans, A. (1983). Nieuwe richtingen en de planning van VO-scholen. Pelkmans, A., Kessel, N. van, Vierssen, E. van, Hahn, A. & Buurke, W.: *Wensen omtrent scholen en de onderwijsplanning. Vijf deelonderzoeken over kwalitatieve onderwijsbehoeften, 'verlangd onderwijs' (WVO art. 65) en de planning van het voortgezet onderwijs* (p. 185-227). Nijmegen: ITS.
- Scheerens, J. & Stoel, W.G.R. (1987). Conceptuele en methodologische problemen bij onderzoek naar de effectiviteit van schoolorganisaties. Scheerens, J. & Stoel, W.G.R. (red.): *Effectiviteit van onderwijsorganisaties* (p. 1-16). Lisse: Swets en Zeitlinger.
- Ven, W. van de & Vink, B. (1986). Het geloof der vaderen. *Intermediair*, jaargang 22 (20). 7-15.

Manuscript ontvangen 12-9-1989

Definitieve versie ontvangen 3-11-1989

Analyse van bijzondere gebeurtenissen in schoolloopbanen: kansen op zittenblijven

Marinus J.M. Voeten en Theo Boland

Vakgroep Interdisciplinaire Onderwijskunde, K.U. Nijmegen en Stichting voor de Leerplanontwikkeling, Enschede

ABSTRACT

Analysis of special events in school careers: failing a grade

A discrete-time method for the analysis of event histories is applied to study how the probability of failing a grade over time is affected by background variables. The method is implemented by logistic regression. Data are taken from a longitudinal cohort study of pupils entering Dutch elementary education in 1979. These pupils from 24 schools were followed during seven years, until they reached the end of their first year of secondary education. Variables used in the analysis include SES and sex of pupils, schoolreadiness, school characteristics, and reading achievement. Results include the following. The probability of failing a grade is negatively associated with SES, but only during the first year of elementary education. The teacher's judgment of schoolreadiness of a pupil is a firm predictor of failing a grade throughout the school career. In mixed-age classes the probability of retaining a pupil appears to be lower than in single-age classes.

INLEIDING

In longitudinaal onderzoek van schoolloopbanen doet zich veelal het probleem voor, dat in de loop van de tijd verschuivingen in de samenstelling van de onderzoeksgroep optreden als gevolg van een voortdurende uitstroom en instroom van leerlingen. De belangrijkste bronnen hiervan zijn: overgang naar een andere school (bijv. als gevolg van verhuizing), zittenblijven en verwijzing van het reguliere basisonderwijs naar het speciaal onderwijs. Doordat de dataverzameling om praktische redenen gewoonlijk gekoppeld wordt aan scholen of klassen, worden alleen die leerlingen bereikt die op het moment van meting tot de betreffende klassen behoren. Een gevolg hiervan is, dat men t.a.v. een aanzienlijk aantal leerlingen over de hele schoolloopbaan bezien onvolledige gegevens zal hebben. Longitudinale analyses, bijv. met behulp van LISREL-modellen, zullen dan veelal alleen betrekking hebben op die leerlingen waarvoor volledige gegevens aanwezig zijn. Dat zijn alleen leerlingen die over de hele periode normaal zijn doorgestroomd en die gedurende het tijdsbestek waarover het onderzoek zich uitstrekt op dezelfde school zijn gebleven. Aangezien de processen van instroom en uitstroom in de onderzoeksgroep gerelateerd kunnen zijn aan de variabelen die onderwerp van onderzoek zijn, kan dit leiden tot vertekening in de conclusies getrokken uit dergelijke longitudinale analyses, of minstens

Adres: Vakgroep Interdisciplinaire Onderwijskunde, K.U. Nijmegen, Erasmusplein 1-20, 6525 GG Nijmegen

Wij danken de beoordelaars voor hun opmerkingen, die hebben bijgedragen tot verduidelijking van de tekst

tot een beperking van de geldigheid van de conclusies tot een speciale groep: leerlingen die normaal zijn doorgestroomd en die niet van school zijn veranderd.

Boland, Voeten en Visser (1989) toetsten een longitudinaal LISREL-model op een cohort van leerlingen die in 1979 begonnen in wat toen nog heette het eerste leerjaar van de lagere school. De data waren ontleend aan het SVO-project 'Preventie van leesmoeilijkheden' (Van Dongen, 1984) en twee vervolgprojecten (SVO-1126, 1983-1985 en SVO-6564, 1985-1987). Van leerlingen die uitvielen (doubleren, terugplaatsing naar een lager leerjaar en verwijzing naar het speciaal onderwijs) of verhuisden waren na het moment van uitvallen of verhuizen geen verdere gegevens bekend. Aangezien de analyse zich uitstreckte tot het tweede jaar voortgezet onderwijs, konden alleen leerlingen in de analyse worden betrokken die gedurende hun hele basisschoolperiode normaal doorstroomden op dezelfde school. Een dergelijke beperking kan vermeden worden, wanneer niet de leerling als analyse-eenheid wordt genomen, maar wanneer elke tijdsperiode (bijv. een schooljaar) voor elke leerling als een afzonderlijke analyse-eenheid wordt geteld. Dit kan worden gerealiseerd bij toepassing van 'event history analysis' (Allison, 1982). In dat geval kan elke leerling in de analyse worden betrokken, ook leerlingen die later zijn ingestroomd of eerder zijn uitgestroomd. Dit is een groot voordeel van 'event history' analyses in vergelijking met LISREL.

Boland et al. (1989) gingen het effect na van sociaal milieu en van de ontwikkeling in lees- en spellingvaardigheid op het advies voor voortgezet onderwijs, gegeven door het hoofd van de school aan het eind van het basisonderwijs, en op de positie in de eerste twee leerjaren van het voortgezet onderwijs. In deze studie richtten wij ons op de kans op zittenblijven tijdens de schoolloopbaan als een functie van sociaal milieu, kenmerken van de leerling en kenmerken van het genoten onderwijs. We onderzoeken dit met behulp van een vorm van 'event history analysis' (Allison, 1984).

Doubleren komt tot stand als gevolg van een beslissing genomen door een leerkracht. Het is een maatregel die wordt toegepast, omdat men van oordeel is dat een leerling de leerdoelen van het betreffende leerjaar in onvoldoende mate heeft bereikt om met vrucht het onderwijs in een volgend leerjaar te kunnen volgen. In dit oordeel spelen zowel kenmerken van de leerling (behaalde schoolprestaties) als kenmerken van het curriculum en van de gevolgde didactiek (o.m. mogelijkheden van individualisering en remedial teaching) een rol. Scholen en leerkrachten kunnen verschillen in het beleid dat ze voeren met betrekking tot de maatregel van zittenblijven. Dit oordeels- en beslissingsproces zullen we niet nader trachten te modelleren. We richten ons enkel op de uitkomsten van dergelijke beslissingen. Voor de pedagogische, onderwijskundige en organisatorische implicaties van zittenblijven verwijzen we naar Doornbos (1969), Bos (1984) en Boekaerts (1987).

Het oorspronkelijke doel van het onderzoeksproject 'Preventie van leesmoeilijkheden' was om middelen te vinden voor een vroegtijdige herkenning van leerlingen met leesmoeilijkheden om vervolgens maatregelen te kunnen nemen ter preventie van leesmoeilijkheden (Van Dongen, 1984). In lijn met deze oorspronkelijke vraagstelling gaan wij hier voor een aantal variabelen na, in hoeverre deze samenhangen met de kans op zittenblijven gedurende de hele periode vanaf begin leerjaar 1 (thans groep 3 genoemd) tot aan het eind van de brugklas. De analyses hebben betrekking op data van hetzelfde cohort van leerlingen als in Boland et al. (1989). Aangezien het hier een secundaire analyse betreft, zijn wij bij de keuze van de variabelen beperkt tot variabelen aanwezig in het databestand van boven genoemd onderzoek, dat met name gericht was op de ontwikkeling van de leesvaardigheid. Een andere beperking bij de keuze van verklarende variabelen ligt in het gegeven dat doubleren een betrekkelijk zeldzame gebeurtenis is. Dit betekent dat slechts een beperkt aantal onafhankelijke variabelen in de analyse kan worden betrokken.

Bij de keuze van variabelen hebben we ons door de volgende overwegingen laten leiden.

Het is bekend dat de kans op zittenblijven in het basisonderwijs samenhangt met het sociaal milieu en het geslacht van leerlingen. Bij leerlingen uit lagere sociale milieus komt doubleren vaker voor dan bij leerlingen uit hogere milieus. Verhoudingsgewijs doubleren meer jongens dan meisjes. Deze samenhang vormt het uitgangspunt voor onze analyses. We zullen daarbij nagaan of de samenhangen met sociaal milieu en geslacht gedurende de hele schoolloopbaan tot en met de brugklas constant blijven. Als tweede stap in de analyse introduceren wij leerlingkenmerken gemeten bij de intree in de lagere school, m.n. school- en leesrijpheid. In hoeverre kan de kans op zittenblijven gedurende de onderzochte periode voorspeld worden uit deze aanvangsverschillen tussen leerlingen? Als derde stap worden kenmerken van het genoten onderwijs aan de analyse toegevoegd. In de onderzoeksopzet van het project 'Preventie van leesmoeilijkheden' was de methode van aanvankelijk lezen (leerjaar 1) een belangrijke variabele. Er was een gestratificeerde steekproef getrokken van scholen die de methode 'Veilig leren lezen' gebruiken en van scholen die de methode 'Letterstad' gebruiken. Als tweede onderwijskenmerk kiezen we het onderscheid tussen combinatieklas en enkelvoudige klas. Deze keuze is ingegeven door de zorg, die men hier en daar kan beluisteren, over de kwaliteit van het onderwijs in combinatieklassen (Veenman, Lcm, Voeten, Winkelmolen & Lassche, 1986). Door Roelofs en Voeten (1987) is de ontwikkeling in leesvaardigheid gedurende het basisonderwijs vergeleken tussen enkelvoudige klassen en combinatieklassen. Voor leerlingen die normaal zijn doorgestroomd, werden geen verschillen tussen beide klastypen aangetroffen. Nu willen we nagaan of de leesmethode en het klastype samenhangen met de kans op zittenblijven. Hierbij dienen de achtergrondkenmerken van de leerlingen en de aanvangsverschillen constant gehouden te worden. Tenslotte, als vierde stap, gaan we na of de kans op zittenblijven gedurende de schoolloopbaan tot en met de brugklas samenhangt met de ontwikkeling in de technische leesvaardigheid. Hiermee wordt bedoeld de vaardigheid in het decoderen van woorden (Brus & Voeten, 1973). De vraagstelling is beperkt tot de technische leesvaardigheid vanwege het probleem om voor de vaardigheid in begrijpend lezen een meetschaal te definiëren die gebruikt kan worden over de hele periode van de basisschool. Voor de meting van de technische leesvaardigheid kon gedurende de hele basisschoolperiode dezelfde meetschaal gebruikt worden.

Voor de 'doorstromers' concludeerden Boland et al. (1989), dat sociaal milieu zowel een indirect als een direct effect heeft op bereikte onderwijspositie, als gecontroleerd wordt voor lees- en spellingvaardigheid. Hier stellen we ons de vraag: Zijn er milieu- en sexe-effecten op de kans op zittenblijven, wanneer rekening gehouden wordt met verschillen in school- en leesrijpheid, met onderwijskenmerken en met leesprestaties?

Samengevat, valt de vraagstelling uiteen in de volgende deelvragen:

1. Hangt de kans op zittenblijven op verschillende momenten in de schoolloopbaan samen het geslacht en het sociaal milieu van de leerling? Blijft deze samenhang constant gedurende de hele onderzochte periode? Blijft deze samenhang bestaan als rekening gehouden wordt met verschillen in school- en leesrijpheid en leesprestaties?
2. Is het oordeel over schoolrijpheid van de kleuterleidster en van de leerkracht van het eerste leerjaar een goede voorspeller van de kans op zittenblijven? Dezelfde vraag wordt gesteld ten aanzien van metingen van de leesrijpheid bij de aanvang van het eerste leerjaar.
3. Hangt de kans op zittenblijven samen met kenmerken van het ontvangen onderwijs? Het gaat hierbij om twee kenmerken: de in het eerste leerjaar gebruikte methode voor aanvankelijk lezen en het zitten in een combinatieklas dan wel een enkelvoudige klas.
4. Hoe hangt de kans op zittenblijven samen met de ontwikkeling in de technische leesvaardigheid?

METHODE

Doubleren als een 'event-history'

In een toenemend aantal longitudinale onderzoeken wordt de aandacht gericht op het optreden van gebeurtenissen, of op overgangen tussen een discreet aantal toestanden en het tijdstip waarop zo'n overgang zich voordoet. Voorbeelden van dergelijk onderzoek zijn vooral te vinden in de medisch-biologische wetenschappen. In de sociologie vinden we voorbeelden bij onderzoek naar de duur van werkeloosheid (Meesters & Van de Pol, 1989) en naar andere belangrijke gebeurtenissen in het leven van een mens, zoals huwelijk of de geboorte van een kind. De data betreffen in dit geval gebeurtenissen die individuen overkomen binnen een bepaald tijdsbestek. Een gebeurtenis is een kwalitatieve verandering die zich op een bepaald moment in de tijd voordoet. In onderzoek van beroepsloopbanen kan dat betrekking hebben op bijv. verandering van baan, bevordering, werkeloos worden, pensionering enz. Bij schoolloopbanen kan men denken aan bijv. het bereiken van een volgend leerjaar of zittenblijven, intree in een ander schooltype, het behalen van een diploma enz. Voor een overzicht van 'event history analysis' raadplege men Allison (1984), Tuma en Hannan (1984), Hutchison (1988) en Kalbfleisch & Prentice (1980). In het onderstaande wordt toegelicht welke keuzes zijn gemaakt bij de toepassing van 'event history analysis' op de data t.a.v. zittenblijven.

Longitudinale registratie van gebeurtenissen houdt in, dat men voor een steekproef van individuen vastlegt op welk tijdstip een bepaalde gebeurtenis zich voordoet. Het resultaat hiervan noemt men een steekproef van 'event histories' (Tuma & Hannan, 1984). Zoals omschreven door Tuma en Hannan veronderstelt het registreren van een 'event-history' een continue observatie van de proefpersonen: aangegeven wordt op welk moment in de tijd een proefpersoon een bepaalde gebeurtenis overkomt. De data worden opgevat als zijnde gegenereerd door een onderliggend stochastisch proces met discrete toestanden, waarbij echter de tijd als continu wordt gezien. In de meeste longitudinale onderzoeksdesigns toegepast in de onderwijsresearch wordt de tijd echter als discreet opgevat. Er worden metingen verricht op enkele vooraf bepaalde tijdstippen. Dit geldt ook voor het databestand gebruikt in dit onderzoek.

In ons geval is een registratie in discrete tijd adequaat, aangezien zittenblijven optreedt op een vast punt in de tijd, nl. aan het einde van een schooljaar. Toch bestaat hier nog een complicatie. Tijdens een schooljaar kan een leerling worden teruggeplaatst naar een lager leerjaar. Zo'n terugplaatsing gebeurt echter niet op een vast punt in de tijd, al zal het steeds in de beginperiode van een schooljaar plaatsvinden. In het databestand kunnen we een terugplaatsing pas constateren bij het eerste meetmoment nadat de terugplaatsing feitelijk heeft plaatsgevonden. Het precieze tijdstip van terugplaatsing is dus niet bekend. Door als tijdsinterval een half schooljaar te nemen, kunnen we terugplaatsing coderen als 'zittenblijven gedurende het schooljaar'. Dus in de eerste helft van een schooljaar kan terugplaatsing voorkomen en in de tweede helft doubleren. Beide soorten gebeurtenissen behandelen we in het vervolg als doubleren. Met deze laatste aanpassing kunnen we doubleren opvatten als een proces dat verloopt in discrete tijd.

Men kan doubleren ook opvatten als bestaande uit meerdere soorten gebeurtenissen op grond van de reden waarom een leerkracht besluit om een leerling te laten doubleren. Aan het eind van het eerste leerjaar is de meest voorkomende reden een zwakke leesvaardigheid (Van Dongen, o.c., p. 44). In deze studie wordt geen onderscheid gemaakt naar de reden van doubleren, omdat we deze reden hoogstens langs indirecte weg zouden kunnen achterhalen en vooral omdat het om zeer kleine groepen leerlingen gaat. De verklarende variabelen die we in dit onderzoek gebruiken, zijn vooral gericht op de leesvaardigheid.

Wel maken we onderscheid tussen doubleren en verwijzing naar het speciaal onderwijs.

Hoewel het aantal verwijzingen naar het speciaal onderwijs in het beschikbare databestand klein is ($n=12$), wilden we deze toch niet zonder meer samenvoegen met de zittenblijvers. Van de andere kant zijn de data te beperkt om conclusies toe te staan over de verwijzing naar het speciaal onderwijs. Als oplossing hebben we gekozen om een *ordinale* variabele 'uitval' te construeren, met als categorieën 1 = zittenblijven en 2 = verwijzing naar speciaal onderwijs. Verwijzing is immers een meer extreme gebeurtenis dan doubleren.

Doubleren is een herhaalbare gebeurtenis, men kan meerdere malen blijven zitten. Van leerlingen die eenmaal hebben gedoubleerd, zijn echter geen verdere gegevens over hun schoolloopbaan beschikbaar (noot 1). Dit betekent dat we in deze studie doubleren moeten beperken tot 'voor de eerste maal doubleren', zodat het niet langer een herhaalbare gebeurtenis is. Verwijzing naar het speciaal onderwijs kan optreden nadat een leerling eerst heeft gedoubleerd. Deze mogelijkheid blijft hier buiten beschouwing.

Er zijn verschillende strategieën mogelijk voor de analyse van 'event-history' data (zie Tuma & Hannan, 1984, p. 28 e.v.). Hier volgen we een exploratieve aanpak zonder verdere assumpties over het stochastisch proces dat ten grondslag zou kunnen liggen aan de data. We richten ons op de vraag of bepaalde kenmerken van de leerlingen en hun omgeving het al dan niet 'uitvallen' (doubleren en verwijzing naar speciaal onderwijs) beïnvloeden. Dit is vergelijkbaar met gewone multiple regressie-analyse; echter de afhankelijke variabele is in dit geval discreet. Het stochastisch proces dat aan deze afhankelijke variabele ten grondslag ligt, kan gekarakteriseerd worden middels het begrip 'hazard' functie (Allison, 1982):

$$(1) \quad P_{it} = \text{Prob} [T_i = t \mid T_i \geq t, x_{it}].$$

De 'hazard' P_{it} is de kans dat een gebeurtenis (zittenblijven of verwijzing naar speciaal onderwijs) zal optreden op een tijdstip t bij een leerling i , op voorwaarde dat deze gebeurtenis vóór het tijdstip t nog niet is voorgekomen. De variabele T geeft aan het tijdstip van optreden van een gebeurtenis, gemeten in discrete eenheden vanaf de start van leerjaar 1. De 'hazard' zal afhangen van het tijdstip t (d.w.z. kansen op uitval variëren met het tijdstip in de schoolloopbaan) en van een k -dimensionele vector van verklarende variabelen x . In (1) is x voorzien van een subscript t om de mogelijkheid open te laten dat (sommige) variabelen op verschillende tijdstippen verschillende waarden kunnen aannemen bij eenzelfde leerling. De logit van de 'hazard' P_{it} schrijven we als een lineaire functie van de verklarende variabelen x , waarbij we de wijze waarop de 'hazard' van de tijd afhangt ongespecificeerd laten door toe te staan dat de intercept α varieert met het tijdstip (noot 2):

$$(2) \quad \log [P_{it} / (1 - P_{it})] = \alpha_t + \beta' x_{it},$$

ofwel

$$P_{it} = 1 / [1 + \exp (- \alpha_t - \beta' x_{it})].$$

Hierin is β een vector van regressiecoëfficiënten (noot 3). Het model kan verder gecompliceerd worden door toe te staan dat sommige van deze regressiecoëfficiënten afhangen van het tijdstip t .

Een belangrijk probleem dat zich bij de analyse van 'event history' data voordoet, is het volgende. De data zullen steeds betrekking hebben op een beperkt tijdsbestek. Gegevens ontbreken dus omtrent het optreden van gebeurtenissen voorafgaande aan het vroegste moment van registreren ('left censoring') en volgend op het gekozen eindtijdstip van

registreren ('right censoring'). De eerstgenoemde vorm van ontbrekende data sluiten we uit door de onderzoeksgroep te beperken tot het cohort van leerlingen die bij het begin van het schooljaar 1979/1980 begonnen met het eerste leerjaar van de lagere school op een van de in de steekproef opgenomen scholen. De informatie die we hierdoor missen betreft de mogelijkheid dat sommige leerlingen een jaar langer over hun kleuterschool deden. Leerlingen die, bijv. als gevolg van verhuizing, op een later tijdstip in de onderzoeksscholen zijn ingestroomd, worden niet in de analyse betrokken. Dit geldt ook voor leerlingen uit vroegere cohorten die middels doubleren in de onderzoeksgroep terecht zijn gekomen.

Rechter 'censoring' komt in onze data in twee vormen voor. Op de eerste plaats is er een (gelukkig grote) groep van leerlingen bij wie in de periode tot aan het einde van de brugklas de gebeurtenis 'zittenblijven' of 'verwijzing' niet wordt waargenomen. Bij deze leerlingen wordt de waarde van de variabele *T* dus niet waargenomen. Deze 'censoring' treedt op op een vooraf bekend tijdstip en kan dus worden aangeduid als gefixeerd. De tweede vorm van 'censoring' doet zich voor bij leerlingen die tijdens hun lagere school periode van school zijn gewisseld. Van deze leerlingen zijn na het tijdstip van verdwijnen uit de onderzoeksschool geen gegevens meer bekend. Voor deze leerlingen moeten we de assumptie maken, dat de tijdstippen waarop 'censoring' optreedt onafhankelijk zijn van de tijdstippen waarop de onderzochte gebeurtenissen optreden, controlerend voor de verklarende variabelen opgenomen in formule (2). Het mag dus niet zo zijn, dat leerlingen die een hoog risico lopen om te blijven zitten een grotere kans hebben om naar een andere school te gaan.

Het probleem van 'right censoring' wordt opgelost door de parameters van (2) te schatten via een 'maximum likelihood' procedure. De data van leerlingen dragen bij aan de 'likelihood' functie tot aan het tijdstip waarop ofwel een gebeurtenis dan wel 'censoring' door overgang naar een andere school optreedt. Elke discrete tijdseenheid voor elke leerling telt hierbij als een afzonderlijke observatie. Dit lost tevens het probleem op van variabelen met tijdafhankelijke waarden. In principe telt dus elke leerling per leerjaar mee voor twee analyseeenheden. Deze werkwijze levert de correcte maximum likelihood schatters onder de assumptie dat model (2) een adequaat model is voor de data.

Steekproef

De data betreffen een cohort van leerlingen die in het schooljaar 1979/80 hun intrede deden in het eerste leerjaar van de lagere school. De gegevens zijn afkomstig uit drie achtereenvolgende SVO-projecten (van Dongen, Hulsmans, Voeten, & Mommers, 1983; Boland, Mommers & Hulsmans, 1985; Mommers & Boland, 1987; Boland et al., 1989). Het gaat om leerlingen van 24 scholen voor gewoon lager onderwijs. Al deze scholen hanteerden het leerstof/jaarklassensysteem. De steekproefprocedure is beschreven in Van Dongen (1984, p. 136 e.v.).

Bij de aanvang van het onderzoek bestond het onderzoekscohort uit 582 leerlingen. Gedurende de lagere school periode zijn 99 leerlingen uit het bestand verdwenen wegens overgang naar een andere school; één leerling sloeg een klas over. Deze leerlingen worden in de analyses meegenomen tot aan het moment van 'verdwijnen'. Het laatste geldt ook voor de leerlingen die werden verwezen naar het speciaal onderwijs ($n=12$), op de lagere school bleven zitten ($n=72$), of tussentijds werden teruggeplaatst naar een lager leerjaar ($n=5$). Aan het einde van de lagere school bestond het cohort dus nog uit 393 leerlingen. In het eerste brugjaar zijn 13 leerlingen blijven zitten.

Variabelen, meetinstrumenten en meetmomenten

Het databestand bevat gegevens op een groot aantal variabelen op diverse meettijdstippen van het eerste leerjaar lagere school tot aan het tweede leerjaar van het voortgezet

onderwijs. Gedurende de basisschoolperiode zijn m.n. periodiek metingen verricht m.b.t. de leesprestaties. De volgende variabelen worden in de analyses opgenomen:

- (a) achtergrondkenmerken van de leerlingen: sexe en sociaal milieu. De variabele 'sociaal milieu' is gemeten via een index geconstrueerd m.b.v. HOMALS (Gifi, 1981), gebaseerd op beroepsniveau van de vader, opleidingsniveau van beide ouders, nationaliteit van het kind en aantal kinderen in het gezin. De index is geconstrueerd met een gemiddelde van 50 en een standaardafwijking van 10. De scores variëren van 19 tot 75 (van Dongen et al., 1983). De achtergrondkenmerken blijven constant over de hele onderzoeksperiode.
- (b) school- en leesrijpheid. De schoolrijpheid is op twee momenten vastgesteld via het oordeel van de kleuterleidster en via het oordeel van de leerkracht van de eerste klas (december leerjaar 1). Deze beide oordelen, verkregen op eenzelfde vijfpuntsschaal (Van Dongen, 1984, p. 145 e.v.), zijn opgevat als één enkele variabele 'schoolrijpheid'. Voor tijdstip 1 bestaat deze variabele uit het oordeel van de kleuterleidster en vanaf tijdstip 2 uit het oordeel van de leerkracht van de eerste klas. Vanaf tijdstip 2 blijft de waarde van schoolrijpheid dus constant over de hele onderzochte periode. De leesrijpheid is bij de aanvang van het eerste leerjaar vastgesteld middels vier toetsen: lettercluster identificatie, auditieve synthese, fonemische analyse en auditieve discriminatie (noot 4). De eenmaal vastgestelde leesrijpheidsvariabelen worden opgevat als constant in de tijd.
- (c) kenmerken van het onderwijs: leesmethode, combinatieklas. De scholen waren voor het onderzoek geselecteerd mede op basis van de gebruikte methode voor aanvankelijk lezen. In de helft van de scholen werd de methode 'Veilig Leren Lezen' gebruikt en in de andere helft de methode 'Letterstad'. Gedurende hun hele periode op de lagere school is van de leerlingen bekend, of ze onderwijs hebben genoten in een combinatieklas dan wel een enkelvoudige klas. De waarden op deze variabele zijn uiteraard tijdafhankelijk. Naast de registratie van het type klas in elk leerjaar, hebben we deze variabele nog op een andere manier meegenomen, nl. als het aantal jaren dat een leerling in een combinatieklas heeft doorgebracht. Voor een aantal leerlingen is het type klas onbekend; in deze gevallen hebben we aangenomen dat sprake was van een enkelvoudige klas.
- (d) technische leesvaardigheid, gemeten met twee verschillende toetsen: de Een Minuut Test (EMT) (Brus & Voeten, 1973), gemeten in de leerjaren 2 t/m 4 en 6 (september en april), en de AVI-toetskaarten (Van den Berg & Te Lintelo, 1977), afgenomen in leerjaar 1 (april) en leerjaar 2 en 3 (september en april). De leesvariabelen worden uiteraard opgevat als tijdafhankelijk. Voor elk tijdsinterval in de 'event history' is de waarde op de twee leesvariabelen gedefinieerd als de laatst gemeten waarde. Een leerling met een ontbrekende score krijgt voor zover mogelijk de score behaald op het vorige tijdstip. Gedurende leerjaar 1 zijn er geen waarden gedefinieerd op de EMT; voor de eerste helft van leerjaar 1 geldt dat ook voor de AVI. Voor deze tijdsintervallen waarop geen scores gedefinieerd zijn voor technische leesvaardigheid is de regressie-coëfficiënt van technische leesvaardigheid op 0 gefixeerd (noot 5).

In totaal zijn er dertien meettijdstippen onderscheiden, twee per leerjaar gedurende het basisonderwijs en één aan het eind van de brugklas. Registratie van de afhankelijke variabele vond plaats op deze dertien tijdstippen (noot 6). De meetmomenten zijn vermeld in Tabel 1. Hierbij zijn drie meetmomenten (april '80, leerjaar 1; april '82, leerjaar 3; april '84, leerjaar 5) weggelaten, aangezien op deze momenten geen zittenblijven noch verwijzing werd geconstateerd. Tijdsintervallen waarin geen enkele gebeurtenis plaatsvindt, kunnen niet in de analyse worden opgenomen. Zou men dit wel doen, dan resulteren schattingen voor de betreffende intercepten die naar $-\infty$ tenderen.

Data-analyses

De data-analyses bestaan uit een reeks logistische regressies overeenkomstig formule (2), uitgevoerd met de SAS-procedure LOGIST (Harrell, 1983). Dit programma berekent 'maximum likelihood' schatters van de parameters. Om een intercept voor elk van de tien tijdstippen te kunnen bepalen, werden negen dummy variabelen (0-1 scores) geconstrueerd. Om tijdafhankelijke β 's te kunnen schatten, werden productvariabelen gevormd door vermenigvuldiging van de betreffende onafhankelijke variabele met een geschikte combinatie van dummy variabelen.

Het onderscheid tussen doubleren en verwijzing naar het speciaal onderwijs werd, zoals gezegd, ingevoerd door de afhankelijke variabele op te vatten als ordinaal. Dit kan door vergelijking (2) te wijzigen in

$$(3) \quad \log [P_{it}(j) / (1 - P_{it}(j))] = \alpha_{jt} + \beta' x_{it},$$

met $j = 1$ (doubleren) of 2 (verwijzing). De $P_{it}(j)$ zijn de cumulatieve kansen op doubleren en verwijzing, m.a.w. $P_{it}(1)$ is de kans op zittenblijven of verwijzing en $P_{it}(2)$ is de kans op verwijzing naar het speciaal onderwijs. De $P_{it}(j)$ vormen twee parallelle lijnen op een logistische schaal (Cox, 1970, p. 104). Het verschil tussen doubleren en verwijzing komt alleen tot uiting in de intercepts. Het model kan dan ook geen verklaring bieden voor de beslissing 'verwijzing' in vergelijking met de beslissing 'doubleren'. Voor deze oplossing is gekozen, omdat het data-materiaal te beperkt is om beide typen gebeurtenissen afzonderlijk te analyseren. Voor een nadere uiteenzetting van deze procedure van logistische regressie met een ordinale afhankelijke variabele verwijzen wij naar Agresti (1989).

Een probleem van analyse in discrete tijd is, dat de conclusies mogelijk gevoelig zijn voor de keuze van de lengte van de tijdsintervallen (Singer & Spilerman, 1976). Deze tekortkoming geldt niet voor discrete-tijd modellen die afgeleid zijn uit continue-tijd modellen. De discrete-tijd versie van het 'proportional hazards' model van Cox (1972) is de complementaire log-log functie $\log [-\log \{1 - P_{it}(j)\}]$ (Kalbfleisch & Prentice, 1980; Allison, 1982). Voor deze laatste zijn de schattingen van de coëfficiënten invariant onder de lengte van het tijdsinterval. Hoewel deze eigenschap niet opgaat voor het logistische model (3), zijn de verschillen tussen de logistische en de complementaire log-log functie in de praktijk niet groot (Allison, 1982, p. 73). Omdat het doel van de analyses beperkt is tot het identificeren van het relatieve belang van verschillende onafhankelijke variabelen bij één enkel cohort, hoeven we niet te verwachten, dat gebruik van de complementaire log-log functie tot andere conclusies zou leiden.

In plaats van de geschatte β 's van (3) zullen we exponentieel getransformeerde β 's presenteren. Deze zijn te interpreteren als multiplicatieve effecten op de 'odds' op doubleren en verwijzing:

$$(4) \quad P_{it}(j) / (1 - P_{it}(j)) = \exp(\alpha_{jt}) [\exp(\beta_1)]^{x_{1it}} [\exp(\beta_2)]^{x_{2it}} \dots [\exp(\beta_k)]^{x_{kit}}.$$

Als criterium voor de selectie van modellen gebruiken we de waarde van de logaritme van de 'likelihood' functie L , vermenigvuldigd met -2 . Gezien de lage frequentie van voorkomen van gebeurtenissen in de steekproef en omdat continue verklarende variabelen worden gebruikt, kunnen we niet aannemen, dat de verdeling van $G^2 = -2\log L$ adequaat benaderd wordt door een chi-kwadraat verdeling. We kunnen echter wel de verschillen in de likelihood-ratio chikwadraat G^2 voor hiërarchisch geordende modellen gebruiken om

een keuze te maken uit de diverse modellen (Fienberg, 1980; Agresti, 1989, p. 297). De onafhankelijke variabelen worden in vier blokken (achtergrondkenmerken, individuele verschillen bij het begin van de lagere school, onderwijskenmerken, leesprestaties) achtereenvolgens aan de regressievergelijking toegevoegd. Variabelen die geen significante bijdrage leveren worden verwijderd.

RESULTATEN

De 'hazard' functie kan worden geschat door voor elk van de tien tijdstippen het aantal gebeurtenissen te delen door het aantal leerlingen dat op het betreffende tijdstip het 'risico' loopt om met de gebeurtenis te worden geconfronteerd. Dit laatste aantal betreft alle leerlingen uit het onderzochte cohort die vóór het betreffende tijdstip nog niet zijn blijven zitten, noch naar het speciaal onderwijs zijn verwezen en die nog steeds aanwezig zijn op de onderzoeksscholen. Tabel 1 geeft een overzicht van de tien tijdstippen met de aantallen leerlingen (N) die tot de risico-verzameling behoren; dit zijn dus de leerlingen die tot aan het betreffende tijdstip normaal zijn doorgestroomd. Tevens bevat deze tabel de percentages uitvallers per tijdstip (= de geschatte 'hazard rate'), als een percentage van het aantal tot op dat moment normaal doorgestroomde leerlingen (noot 7).

Tabel 1. Percentages zittenblijvers en verwijzing naar het speciaal onderwijs (s.o.)

Meettijdstip	Leerjaar	N	zittenblijven	s.o.
1 dec. '79	1	572	0.17	0.35
2 sept. '80	1	550	6.73	1.27
3 april '81	2	499	0.40	0.00
4 sept. '81	2	492	2.85	0.20
5 sept. '82	3	456	1.32	0.22
6 april '83	4	436	0.46	0.23
7 sept. '83	4	422	1.42	0.00
8 okt. '84	5	403	1.24	0.00
9 maart '85	6	397	1.01	0.00
10 eind brugklas		393	3.31	0.00
Totaal		4620	90	12

Leerjaar is het leerjaar waarin het doubleren of de verwijzing plaatsvond.

Zittenblijven (inclusief terugplaatsing) komt in de steekproef in totaal 90 maal voor (77 gedurende het basisonderwijs en 13 in de brugklas) en verwijzing naar het speciaal onderwijs 12 maal (noot 8). Het totaal van alle N's (4620) geeft het maximum aantal observaties waarop de analyses zijn gebaseerd. Naarmate meer variabelen in de analyse worden betrokken, wordt dit aantal lager als gevolg van ontbrekende scores. Bij gebruik van alle onafhankelijke variabelen wordt het aantal observaties gereduceerd tot 4261. Hierbij komt zittenblijven 82 keer voor en verwijzing 8 maal.

Zoals te verwachten liggen de kansen op zittenblijven het hoogst aan het eind van de eerste klas. Gedurende het basisonderwijs neemt de kans op zittenblijven af; aan het eind van de brugklas is er weer een toename te zien. Verwijzing naar het speciaal onderwijs doet zich vooral in het eerste leerjaar voor; in de leerjaren 2 t/m 4 komt het ook nog voor, zij het sporadisch.

Het doel is nu regressiemodellen te schatten waarin de kansen op doubleren en verwijzing per tijdperiode afhangen van een aantal verklarende variabelen genoemd in de vraagstellingen. Allereerst kijken we naar de achtergrondkenmerken sexe en sociaal milieu.

In vergelijking met een model waarin alleen intercepten zijn opgenomen voor de diverse tijdstippen ($df = 11$) daalt het likelihood-ratio criterium van 962.16 naar 866.83 als sexe en sociaal milieu worden opgenomen als verklarende variabelen ($df = 13$). Een verdere daling wordt bereikt door toe te staan dat de effecten van sexe en sociaal milieu gedurende leerjaar 1 anders zijn dan in de hogere leerjaren ($df = 15$, $G^2 = 858.12$). We noemen dit model 1; de parameterschattingen zijn weergegeven in Tabel 2. Aan de hand van deze tabel zullen we de betekenis van de diverse parameters toelichten.

Tabel 2. Effecten van geslacht en sociaal milieu (model 1)

VARIABELE	EXP (BETA)	
<i>Intercepten</i>		
Doubleren of verwijzing	0.65	
Verwijzing	0.07***	
1 eerste helft leerj. 1	1.81	
2 tweede helft leerj. 1	98.52***	
3 eerste helft leerj. 2	0.12***	
4 tweede helft leerj. 2	0.91	
5 tweede helft leerj. 3	0.45*	
6 eerste helft leerj. 4	0.20**	
7 tweede helft leerj. 4	0.42*	
8 tweede helft leerj. 5	0.36*	
9 tweede helft leerj. 6	0.30**	
<i>Regressiecoëfficiënten</i>		
SEXE:		
Leerjaar 1	0.23***	(Meisjes: 0.05)
Overige tijdstippen	0.59*	(Meisjes: 0.35)
SOCIAAL MILIEU:		
Leerjaar 1	0.90***	(0.13 - 0.0003)
Overige tijdstippen	0.96***	(0.41 - 0.030)
$G^2 = 858.12$ ($df = 15$)		

* = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$

Effecten zijn sterker naarmate de exponentieel getransformeerde β 's meer van 1.0 afwijken. Hoe hoger de parameterwaarden hoe groter de 'odds' op uitvallen of achterblijven. De intercepten α_{jt} van model (3) zijn in de regressie-analyse vertegenwoordigd door elf parameters, waaronder de regressiecoëfficiënten voor de negen dummy variabelen voor het verschil tussen de tijdsintervallen (de coëfficiënt voor eind brugklas is 22.77). De schattingen voor de intercepten $\exp(\alpha_{jt})$ kunnen uit Tabel 2 worden verkregen door het intercept voor 'doubleren of verwijzing', resp. 'verwijzing', te vermenigvuldigen met de geschatte coëfficiënt voor het betreffende tijdstip t . De significantie-aanduidingen bij de parameters zijn gebaseerd op de 'Wald statistiek' (Harrell, 1983).

Tabel 2 geeft de samenhangen tussen enerzijds de kans op uitval en anderzijds geslacht en sociaal milieu van de leerling, zonder rekening te houden met de covariabelen. Zowel voor geslacht als voor sociaal milieu is het effect in leerjaar 1 groter dan in de hogere leerjaren. Meisjes hebben een kleinere kans om te blijven zitten (of te worden verwezen naar het speciaal onderwijs) dan jongens. De 'odds' voor de jongens zijn in leerjaar 1 meer

dan vier keer slechter dan voor meisjes; in de hogere leerjaren is dat nog een factor 1.7.

Voor sociaal milieu zijn in Tabel 2 tevens de waarden vermeld voor de laagste en de hoogste milieu-index in de steekproef. Deze waarden zijn verkregen door toepassing van formule (4) voor de laagste en de hoogste score. Voor hogere sociale milieu's is de kans op uitvallen geringer dan voor de lagere sociale milieus. Dit effect is in leerjaar 1 duidelijk sterker dan in de hogere leerjaren. We hebben ook getoetst of het effect van sociaal milieu rond de overgang naar het voortgezet onderwijs en aan het einde van de brugklas weer veranderd is. Hiervoor is geen indicatie gevonden (chi-kwadraat is 0.29 bij $df=1$).

De volgende analysestap is om de variabelen m.b.t. school- en leesrijpheid aan model 1 toe te voegen. Dit leidt tot een aanzienlijke daling van G^2 naar 681.79 ($df = 20$) door toevoeging van 5 parameters. De kansen op uitvallen dalen met toenemende score op schoolrijpheid en op de leesrijpheidstoetsen. Uitgaande van dit model is getoetst of de effecten van geslacht en sociaal milieu, zoals weergegeven in Tabel 2, verklaard kunnen worden uit verschillen in school- en leesrijpheid. Verwijdering van de sexe-effecten geeft slechts een stijging van 2.21 in de G^2 bij 2 vrijheidsgraden. Als rekening gehouden wordt met verschillen tussen leerlingen in school- en leesrijpheid, dan zijn er dus geen sexe-effecten meer waarneembaar. Het effect van sociaal milieu blijft echter significant, zij het alleen in het eerste leerjaar. Het model waaruit het sexe-effect en het milieueffect voor de hogere leerjaren zijn verwijderd, noemen we model 2 ($G^2 = 685.16$, $df = 17$).

Model 2 is het uitgangsmodel voor het toetsen van effecten van onderwijskenmerken (leesmethode, combinatieklas versus enkelvoudige klas). Het verschil in methode van aanvankelijk lezen blijkt geen effect te hebben, ook niet in het eerste leerjaar. Dit geldt echter wel voor het type klas, wanneer dit gedefinieerd wordt als het type klas (combinatieklas of enkelvoudige klas) waarin een leerling op tijdstip t vertoeft. Het aantal jaren doorgebracht in een combinatieklas heeft echter geen significant effect. Het resulterende model 3 heeft een G^2 van 672.53 bij 18 vrijheidsgraden. Volgens dit model hebben leerlingen in combinatieklassen een kleinere kans op uitvallen dan leerlingen in enkelvoudige klassen.

Tenslotte zijn aan model 3 de twee variabelen m.b.t. technische leesvaardigheid toegevoegd (AVI en EMT). Deze variabelen kunnen geen bijdrage leveren aan de verklaring in leerjaar 1, omdat ze pas vanaf leerjaar 2 zijn gemeten. We zijn ervan uitgegaan dat de effecten van AVI en EMT constant zijn gedurende de hele periode vanaf leerjaar 2 tot en met leerjaar 5, maar dat vanaf leerjaar 6 de effecten zullen zijn afgezwakt. Dit betekent dat aan model 3 vier parameters worden toegevoegd. De G^2 daalde naar 610.17 bij 22 vrijheidsgraden.

Vervolgens is getoetst of het effect van de technische leesvaardigheid vanaf leerjaar 6 nul is. Dit leidde tot een G^2 van 612.41 bij 20 vrijheidsgraden, zodat we dus inderdaad mogen aannemen dat er vanaf leerjaar 6 geen effect meer is van technische leesvaardigheid. Van dit model 4 zijn de parameterschattingen opgenomen in Tabel 3. De intercepten zijn hierbij weggelaten. Naast de parameterschatting worden de laagste en hoogste waargenomen score vermeld op elke variabele en de effecten op de 'odds' op uitvallen die corresponderen met deze laagste en hoogste score, zie formule (4).

Sociaal milieu heeft in leerjaar 1 nog steeds een significant effect op de kans op uitvallen, wanneer gecontroleerd wordt voor verschillen in schoolrijpheid (leerkrachtoordelen) en leesrijpheid. Het maximale geschatte effect is een factor 37 in de 'odds' op zittenblijven. Voor hogere leerjaren is er geen effect van sociaal milieu meer waarneembaar.

De oordelen over de schoolrijpheid van de leerlingen hebben een aanzienlijk effect op de kansen op zittenblijven of verwijzing gezien over de hele onderzochte periode. Het opmerkelijke is dat dit geldt onder constant houding van leesrijpheid en leesprestaties vanaf leerjaar 2. Nul stellen van het effect van school- en leesrijpheid vanaf leerjaar 2 gaf

Tabel 3. Parameterschattingen eindmodel (model 4)

VARIABELE	EXP (β_r)	RANGE	
		X_r	$[\exp(\beta_r)]^{x_r}$
Sociaal Milieu (leerj. 1)	0.94**	17-75	0.37- 0.01
Schoolrijpheid	0.51***	1-5	0.51- 0.03
Lettercluster identificatie	0.93***	2-38	0.87- 0.07
Auditieve Synthese	0.95***	2-30	0.90- 0.20
Fonemische Analyse	0.97	1-29	0.97- 0.45
Auditieve Discriminatie	0.98	0-38	1.00- 0.44
Combinatieklas	3.53***	1-2	3.53-12.47
AVI (leerj. 1 t/m 5)	0.59***	0-9	1.00- 0.01
EMT (leerj. 2 t/m 5)	0.96**	0-116	1.00- 0.01

$G^2 = 612.41$ (df = 20)

* = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$

een significante verslechtering van de fit te zien. De voorspellende waarde van de leerkrachtoordelen en van de leesrijpheidstoetsen strekt zich dus verder uit dan het eerste leerjaar. Afgaande op de 'Wald statistic' leveren twee van de vier leesrijpheidstoetsen geen significante bijdrage. Bij toetsing via de conditionele 'likelihood ratio' chi-kwadraat blijkt echter wel een significant effect van deze toetsen.

Van de onderwijskenmerken die in de analyse zijn betrokken, speelt alleen het gegeven of de leerling op het betreffende tijdstip in een combinatieklas zit een rol. In enkelvoudige klassen is de kans op zittenblijven of verwijzing groter dan in combinatieklassen.

De prestaties in de technische leesvaardigheid tenslotte vormen in de leerjaren 2 tot en met 5 een belangrijke verklarende factor. Beide toetsen leveren een significante bijdrage. Het verschil in de 'odds' tussen de laagste en de hoogste score bedraagt voor beide toetsen een factor 100.

CONCLUSIES EN DISCUSSIE

Jongens hebben een groter kans op zittenblijven en verwijzing naar het speciaal onderwijs dan meisjes. Deze samenhang kan volledig verklaard kunnen worden uit verschillen in school- en leesrijpheid. Meisjes hebben bij de aanvang van het basisonderwijs een voorsprong op jongens, waardoor ze een kleiner risico lopen op uitval, althans in de periode tot aan het einde van de brugklas.

De samenhang met sociaal milieu echter kan niet helemaal worden wegverklaard door verschillen in aanvangsniveau van de diverse milieugroepen. In leerjaar 1 blijft een statistisch significant en aanzienlijk milieu-effect bestaan ten nadele van de lagere sociale milieus, ook als gecontroleerd wordt voor aanvangsverschillen in school- en leesrijpheid. Voor de hogere leerjaren echter zijn er geen milieu-effecten meer wat betreft uitval door doubleren of verwijzing naar het speciaal onderwijs. In Boland et al. (1989) werden op hetzelfde databestand wel milieu-effecten gevonden op het advies voor voortgezet onderwijs, onder controle voor lees- en taalprestaties, in de groep leerlingen die zonder enige vertraging de lagere school heeft doorlopen.

Het oordeel van de kleuterleidster en van de leerkracht van de eerste klas over de schoolrijpheid van leerlingen levert een aanzienlijke bijdrage aan de voorspelling van uitval bovenop wat verklaard kan worden uit leesrijpheidstoetsen. Leerkrachten zijn dus

al in een vroeg stadium in staat leerproblemen te voorzien.

Gezien de zorgen over combinatieklassen (Veenman et al., 1986) is het op het eerste gezicht verrassend te constateren dat onderwijs in een combinatieklas een kleinere kans op zittenblijven of verwijzing oplevert dan onderwijs in een enkelvoudige klas, onder controle voor sociaal milieu, school- en leesrijpheid en technische leesvaardigheid. Een gedeeltelijke verklaring hiervoor zou kunnen zijn, dat leerkrachten van combinatieklassen een zwakke leerling eerder zullen laten overgaan, als ze deze leerling het volgend jaar toch weer onder hun eigen hoede kunnen houden. Het lijkt niet aannemelijk dat het gevonden verband iets van doen zou hebben met het verschil in onderwijs in combinatieklassen versus enkelvoudige klassen (combinatieklassen worden wellicht toegewezen aan de betere leerkrachten). Het blijkt immers, dat de variabele 'aantal jaren doorgebracht in een combinatieklas' geen significante bijdrage levert. De gebruikte methode van aanvankelijk lezen ('Veilig leren lezen' of 'Letterstad') vertoont geen samenhang met de kans op uitvallen.

De prestaties met betrekking tot technische leesvaardigheid blijken een krachtige voorspeller van kansen op vertraging in de schoolloopbaan tot en met het vijfde leerjaar. Ook na de periode van het aanvankelijk lezen blijkt de technische leesvaardigheid een belangrijke correlaat van de kans op vertraging in de schoolloopbaan. Overigens moet hierbij worden opgemerkt, dat technische leesvaardigheid hier kan functioneren als een substituut voor lees- en taalvaardigheid in het algemeen, omdat geen andere lees- of taalvaardigheden in de analyse waren betrokken.

De uitgevoerde analyses kennen een aantal beperkingen. Op sommige daarvan is in het voorgaande reeds gewezen. De gegevens over zittenblijven zijn beperkt tot de eerste keer dat een leerling is blijven zitten, terwijl ook geen rekening is gehouden met de reden waarom een leerling is blijven zitten. De omvang van het databestand is te beperkt om afzonderlijke regressievergelijkingen toe te staan voor doubleren en voor verwijzing naar het speciaal onderwijs. In onze analyses hebben we daardoor de assumptie moeten maken dat aan beide beslissingen hetzelfde mechanisme ten grondslag ligt, hetgeen zeker valt te betwijfelen. De beperktheid van het databestand liet ook niet toe om een model te ontwikkelen voor het beslissingsproces dat aan het besluit om een leerling niet te bevorderen of te verwijzen ten grondslag ligt. Daardoor bleef de rol van leerkrachten in dit proces onderbelicht.

In de toepassing van de 'event history' analyse hebben we een aantal keuzes gemaakt, die gevolgen kunnen hebben voor de resultaten. Omdat het hier ging om de analyse van zeldzame gebeurtenissen, moesten we ons beperken in het aantal variabelen dat tegelijkertijd in de analyse wordt betrokken. Toevoeging en weglating van variabelen kan effect hebben op de schatting van parameters voor andere variabelen. In de diverse analyses die wij hebben uitgevoerd met (subsets van) de gekozen onafhankelijke variabelen, bleken de resultaten vrij stabiel, behalve t.a.v. de 'verdeling' van de regressiegewichten over de vier leesrijpheidstoetsen. We hebben de analyse ook uitgevoerd met zittenblijven en doubleren samengevoegd tot één categorie en met tijdsintervallen van een heel schooljaar i.p.v. een half jaar. De resultaten weken nauwelijks af van de resultaten gerapporteerd in Tabel 3.

Tenslotte herinneren we eraan dat aan de analyses de assumptie ten grondslag ligt dat overgang naar een andere school onafhankelijk is van zittenblijven of verwijzing. De data staan geen rechtstreekse toetsing van deze assumptie toe (cf. Kalbfleisch & Prentice, 1980, p. 172-175). Wel kan vastgesteld worden, dat de leerlingen die naar een andere school zijn gegaan, niet kunnen worden beschouwd als een random steekproef uit het cohort. Leerlingen die gedurende het basisonderwijs naar een andere school zijn verhuisd behalen een significant hogere gemiddelde score op de milieu-index dan de leerlingen die op

dezelfde school bleven. Bovendien hadden deze leerlingen hogere gemiddelde prestaties op leesrijpheid en technische leesvaardigheid in leerjaar 1 en 2. Het is dus niet waarschijnlijk, dat juist leerlingen met een hoog risico op doubleren of verwijzing naar een andere school zijn gegaan. Een gedeeltelijke controle op de assumptie kan worden verkregen door een omkering van rollen: een logistische regressie als in Tabel 3 ter verklaring van de kans op verhuizen. Twee van de variabelen opgenomen in het eindmodel blijken ook samen te hangen met de kans op verhuizing naar een andere school, nl. sociaal milieu ($e^{\beta} = 1.025$) en het zitten in een combinatieklas ($e^{\beta} = 2.35$). Leerlingen in enkelvoudige klassen verhuizen vaker naar een andere school. Het laatste doet twijfel rijzen aan het verband gevonden tussen het zitten in een combinatieklas en de kans op zittenblijven.

NOTEN

1. Van de zittenblijvers die op de onderzoeksscholen bleven, zijn wel enkele gegevens bekend. Zie Mommers & Van Hooft-Aarnoutse (1987) en Boland & Mommers (1987).
2. Het is bekend dat de kans op uitval in lagere leerjaren groter is dan in hogere leerjaren. Voor de brugklas kan weer een hogere kans op zittenblijven verwacht worden. We gaan er niet vanuit dat deze temporele variatie in de kansen op uitval volledig verklaard zou kunnen worden uit verschillen tussen de leerlingen op de verklarende variabelen. Het onderwijs kent een eigen dynamiek, zodanig dat men in vroegere fases eerder geneigd zal zijn om leerlingen te laten uitvallen dan in latere fases. Deze autonome variatie met de tijd laten we tot uiting komen door een afzonderlijk intercept toe te staan per tijdstip. Hiermee zeggen we dus dat de 'hazard', gegeven de verklarende variabelen, een arbitraire functie is van de tijd, een functie die niet nader wordt gespecificeerd (analoog aan het 'proportional hazards' model (Cox, 1972) voor 'event histories' in continue tijd; zie bijv. Allison, 1982).
3. Voor logit-modellen zie bijvoorbeeld Cox (1970) of Fienberg (1980).
4. Nadere informatie over deze toetsen is te vinden in Van Dongen (1984, p. 145 e.v.) en in Van Dongen & Mommers (1984). Dr. M.J.C. Mommers maakte ons erop attent, dat in twee gevallen in deze publicaties een inadequate benaming voor de toetsen is gebruikt. 'Lettercluster identificatie' werd eerder aangeduid met de naam 'visuele letterperceptie', terwijl 'fonemische analyse' ten onrechte werd aangeduid als 'fonetische analyse'.
5. Technische leesvaardigheid werd voor leerjaar 1 gefixeerd door alle leerlingen gedurende leerjaar 1 de score 0 te geven op deze variabele. Anders gezegd technische leesvaardigheid wordt vemenigvuldigd met een dummy variabele met de waarde 0 voor de twee tijdsintervallen binnen leerjaar 1 en de waarde 1 voor de latere tijdstippen.
6. Bij het begin van een leerjaar kon worden geregistreerd welke leerlingen in het voorgaande leerjaar zijn blijven zitten en welke leerlingen in de tweede helft van het voorgaande schooljaar naar het speciaal onderwijs werden verwezen. Op het tweede meettijdstip binnen een schooljaar is vastgesteld welke leerlingen in de eerste helft van het betreffende schooljaar werden teruggeplaatst of werden verwezen. 'Zittenblijven' vastgesteld december leerjaar 1 betreft terugplaatsing naar de kleuterschool.
7. Hierbij wordt verondersteld, dat de 'hazard rate' kan verschillen per tijdsperiode maar constant is voor alle individuen binnen een gegeven periode.
8. De werkelijke frequentie van doubleren en verwijzing naar het speciaal onderwijs in het onderzochte cohort ligt hoger dan de in de tekst gegeven aantallen. Immers verwijzing naar het speciaal onderwijs kan plaatsvinden nadat een leerling eerst heeft gedoubleerd. Verder kan een leerling meerdere malen doubleren. En tenslotte hebben we geen verdere gegevens, nadat een leerling naar een andere school is gegaan.

LITERATUUR

- Agresti, A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin*, 105, 290-301.
- Allison, P.D. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological Methodology 1982*. San Francisco: Jossey-Bass (p. 61-98).
- Allison, P.D. (1984). *Event history analysis: Regression for longitudinal event data*. London: Sage.
- Berg, R.M. van den, & Te Lintelo, H.G. (1977). *Individualisering van het leesonderwijs*. Den Bosch: Katholiek Pedagogisch Centrum.
- Boekaerts, M. (1987). *Voorlopen op achterstand: Voorkomen van zitten blijven*. Den Haag: Ministerie van Onderwijs en Wetenschappen.
- Boland, Th., & Mommers, M.J.C. (1987). *Verschillen tussen doorstromers en achterblijvers aan het eind van het basisonderwijs*. (SVO-project 6564, deelrapport 1). Nijmegen: Instituut voor Onderwijskunde, K.U. Nijmegen.
- Boland, Th., Mommers, M.J.C., & Hulsmans, H. (1985). *Lees- en leerprestaties in het zesde leerjaar van het lager onderwijs* (SVO-project 1126, Deelrapport 2). Nijmegen: Instituut voor Onderwijskunde, K.U. Nijmegen.
- Boland, Th., Voeten, M.J.M., & Visser, J.J.C.M. (1989). De invloed van leesvaardigheid in het basisonderwijs op het niveau van voortgezet onderwijs: Een longitudinale benadering. In J. van Damme & J. Dronkers (red.), *Jongeren in school en beroep*. Lisse: Swets & Zeitlinger.
- Bos, D.J. (1984). *Blijven zitten met zittenblijven*. Harlingen: Flevodruk.
- Brus, B.Th., & Voeten, M.J.M. (1973). *Een Minuut Test*. Nijmegen: Berkhout.
- Cox, D.R. (1970). *Analysis of binary data*. London: Methuen.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34, 187-202.
- Dongen, D. van (1984). *Leesmoeilijkheden: Naar diagnostiserend onderwijzen bij het leren lezen*. Tilburg: Zwijsen.
- Dongen, D. van, Hulsmans, H., Voeten, M.J.M., & Mommers, M.J.C. (1983). *Preventie van leesmoeilijkheden. Eindverslag van het longitudinale onderzoek*. (SVO-project 492, Deelrapport 13). Nijmegen: Instituut voor Onderwijskunde, K.U. Nijmegen.
- Dongen, D. van, & Mommers, M.J.C. (1984). *Lees- en spellingtoetsen voor leerjaar 1 tot 4 van de lagere school*. (Mededelingen nr. 12). Nijmegen: Vakgroep Interdisciplinaire Onderwijskunde.
- Doornbos, K. (1969). *Opstaan tegen zittenblijven*. Den Haag: Staatsuitgeverij.
- Fienberg, S.E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, MA: MIT Press.
- Gifi, A. (1981). *HOMAL's user's guide*. Leiden: University of Leiden, Department of Data Theory.
- Harrell, F.E. (1983). The LOGIST procedure. In *SUGI supplemental library user's guide* (pp. 181-202). Cary, NC: SAS Institute.
- Hutchison, D. (1988). Event history and survival analysis in the social sciences. *Quality and Quantity*, 22, 203-219 (part 1) en 255-278 (part 2).
- Kalbfleisch, J.D., & Prentice, R.L. (1980). *The statistical analysis of failure time data*. New York: Wiley.
- Meesters, M.J., & Pol, F.J.R. van de (1989). Once unemployed, always unemployed? An analysis of the duration of unemployment and the stability of jobs of Dutch schoolleavers in the eighties. In B.F.M. Bakker, J. Dronkers, & G.W. Meijnen (Eds.), *Educational opportunities in the welfare state: Longitudinal studies in educational and occupational attainment in the Netherlands*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.
- Mommers, M.J.C., & Boland, Th. (red.) (1987). *De ontwikkeling van de leesvaardigheid in longitudinaal perspectief*. Nijmegen: Instituut voor Onderwijskunde, K.U. Nijmegen.
- Roelofs, E.C., & Voeten, M.J.M. (1987). *De leesontwikkeling van 7 tot 12: Invloeden van sekse, sociaal milieu en combinatieklassen*. (SVO-project 6564, Deelrapport IV). Nijmegen: Instituut voor Onderwijskunde, K.U. Nijmegen.
- Singer, B., & Spilerman, S. (1976). Some methodological issues in the analysis of longitudinal surveys. *Annals of Economic and Social Measurement*, 5, 447-474.

- Tuma, N.B., & Hannan, M.T. (1984). *Social dynamics: Models and methods*. Orlando: Academic Press.
- Veenman, S.A.M., Lem, P., Voeten, M.J.M., Winkelmolen, G., & Lassche, H. (1986). *Onderwijs in combinatieklassen*. 's Gravenhage: SVO (Selecta reeks).

Manuscript ontvangen 4-10-1989

Definitieve versie ontvangen 24-11-1989

De doelmatigheid van het algemeen voortgezet onderwijs

A.G.J. van der Torre en B. Kuhry
Sociaal en Cultureel Planbureau

ABSTRACT

This paper deals with the production structure of secondary general education in the Netherlands. Output is measured as the number of pupils corrected for drop-out and repeat, or more precisely as the number of successfully completed school years. Input is measured as the salary costs, representing 88% of total costs. Apart from a time-series analysis at the national level for the period 1969-1985, a micro-analysis is performed for individual schools in the school year 1985/'86 involving data for more than 1000 schools. A descriptive analysis of data and an analysis using a translog cost function confirm the presence of important positive scale effects, which are not only due to current regulations favouring small schools but also to a better performance of pupils in larger schools. Important differences in productivity appear to exist between schools of similar size. Important cost reductions can be realised by increasing the scale of production.

1. INLEIDING

1.1 Opzet en doelstelling

In Nederland is tot dusverre weinig onderzoek gedaan naar de produktiestructuur in de kwartaire sector. Problemen om de produktie te meten zijn daarvan de voornaamste oorzaak. In de marktsector worden subtiliteiten in verband met de meting van prestaties of effecten gewoonlijk omzeild door de waardering van het fysieke produkt aan de betaalde prijs af te meten. Er wordt stilzwijgend verondersteld dat eventuele effecten op de consument via de prijsstelling tot uiting komen in de produktiewaarde en daarmee in de omzet. Het budgetmechanisme, dat in de kwartaire sector de rol van allocatiemechanisme vervult, laat een dergelijke bepaling van de marktwaarde niet toe. Als besluitvormers bij de toewijzing van middelen onvoldoende zicht hebben op de bereikte effecten van voorzieningen, kan de produktie van minder zinvolle of te kostbare activiteiten doorgang blijven vinden. Anders dan in de marktsector kan men in de kwartaire sector niet uitgaan van de omzet als maat voor het produktievolume. Men zal derhalve gebruik moeten maken van fysieke produktindicatoren. Voor zover daarbij gebruik gemaakt wordt van meer dan één indicator wordt men daarbij geconfronteerd met een wegingsprobleem: men kan immers de verschillende produkten niet via de prijzen bij elkaar optellen.

Onlangs heeft het Sociaal en Cultureel Bureau een onderzoek gepubliceerd naar de produktiestructuur van vier voorzieningen in de kwartaire sector, te weten bejaardenoordelen, politie, openbare bibliotheken en algemeen voortgezet onderwijs (Goudriaan et al. 1989a en 1989b). In dit artikel wordt ingegaan op de uitkomsten met betrekking tot de laatstgenoemde voorziening.

Door het ontbreken van inzicht in de produktiviteit van kwartaire dienstverlening is de schaal van produktie in de regel niet aangepast aan daarvoor geldende optima. Door het

goeddeels ontbreken van concurrentie is het voorts meer dan in de marktsector mogelijk, dat verschillende instellingen van min of meer vergelijkbare grootte worden gekenmerkt door een uiteenlopende produktiviteit. Deze studie toont aan dat het mogelijk is om door een doelmatiger organisatie van het produktieproces bezuinigingen op de collectieve uitgaven te realiseren zonder dat de dienstverlening daaronder hoeft te leiden. Omgekeerd is het ook mogelijk om de dienstverlening uit te breiden zonder dat daarvoor extra middelen noodzakelijk zijn. In een tijdvak waarin een toenemende behoefte bestaat aan kwartaire diensten van zo hoog mogelijke kwaliteit, maar waarin anderzijds macro-economische argumenten pleiten voor een beheersing van collectieve uitgaven, is het nut van dit soort analyses onmiskenbaar. In dit verband mag er op gewezen worden, dat het Sociaal en Cultureel Planbureau in opdracht van het Ministerie van Onderwijs en Wetenschappen thans een vervolgonderzoek uitvoert, waarin het basisonderwijs, het speciaal onderwijs en het voortgezet onderwijs zijn betrokken (SCP 1989).

In paragraaf 1.2 wordt ingegaan op buitenlands onderzoek op het terrein van de produktiviteit in het onderwijs. Paragraaf 1.3 heeft betrekking op de meting van het produktievolume. In paragraaf 2 wordt ingegaan op de produktiviteitsontwikkeling van het avo in de periode 1969-1985. In paragraaf 3 wordt een beschrijvende analyse gegeven van produktiviteitsverschillen tussen scholen in het schooljaar 1985/'86, die in paragraaf 4 wordt gevolgd door een econometrische analyse. In paragraaf 5, tenslotte, wordt ingegaan op besparingen die kunnen worden bereikt door ingrepen in de produktiestructuur in het algemeen en in de schaal van produktie in het bijzonder.

1.2 Eerder onderzoek

Het meeste onderzoek over de produktiviteit van het onderwijs heeft betrekking op de situatie in de Verenigde Staten. Nederlands onderzoek op dit terrein is er nauwelijks.

Een klassieke studie op dit terrein is Coleman et al. (1966). De afhankelijke variabele in dit onderzoek wordt gevormd door de testscore. Onafhankelijke variabelen zijn karakteristieken van de leerlingen, de kwaliteit van de leerkrachten en schoolspecifieke variabelen als de groepsgrootte en het aantal leerlingen op een school. De afhankelijke variabele is de output, de onafhankelijke variabelen beschrijven de input. Daarom kan deze specificatie gezien worden als een produktiefunctie. Summers en Wolfe (1977) adopteren dit schema maar passen het toe op individuele leerlinggegevens in plaats van op gegevens over scholen. Een recente toepassing is te vinden in Sengupta en Sfeir (1986). Deze auteurs gebruiken een beperkt aantal variabelen maar gaan vrij diep in op de schattingsmethodiek, waarbij zij gebruik maken van een translogproduktiefunctie.

Een geheel andere benadering is gebruikt door Riew (1966). Hier zijn de kosten per leerling gekozen als afhankelijke variabele, terwijl het aantal leerlingen, het gemiddeld salaris van leerkrachten, het aantal contacturen per leerling en leerkracht en de ouderdom van het schoolgebouw gekozen zijn als onafhankelijke variabelen. Een dergelijke specificatie kan worden gezien als een kostenfunctie. De analyse-eenheid wordt hier gevormd door de school. Een nadere uitwerking is gegeven in Cohn en Riew (1974). Een vergelijkbare specificatie is ook gebruikt door Glennerster en Wilson (1970), Wales (1973), Watt (1980), Kumar (1983) en Bee en Dolton (1985). Kumar gaat expliciet in op de transportkosten voor leerlingen. Bee en Dolton meten de kwaliteit op basis van gegevens over examenresultaten en uitval. Verschillende auteurs introduceren een kwadratische term voor het leerlingenaantal om een schatting van een optimale schoolgrootte mogelijk te maken. Watt gebruikt een translog-specificatie van de kostenfunctie.

Butler en Monk (1985) en Jimenez (1986) kiezen als afhankelijke variabele de ongestandaardiseerde kosten. De eersten voeren een analyse uit op het niveau van het schooldistrict, waarbij als onafhankelijke variabelen gebruikt worden: het aantal leerlingen, het

gemiddelde salaris van leerkrachten, gegevens over de leerlingenpopulatie en examenresultaten. Jimenez kiest als onafhankelijke variabelen onder meer het aantal leerlingen in lager en voortgezet onderwijs, en de salariskosten. Hij besteedt aandacht aan de "economies of scope", de effecten van het geven van verschillende soorten onderwijs binnen één school. Beiden maken gebruik van een translogkostenfunctie.

De resultaten van de analyses lopen sterk uiteen, zoals ondermeer blijkt uit de literatuurstudie van Hanushek (1986), waarin 147 onderzoeken op dit terrein betrokken werden. Zo worden zelden significant positieve relaties gevonden tussen de verhouding leerkrachten/leerlingen en uitgaven per leerling enerzijds en schoolresultaten anderzijds. In een aantal gevallen worden zelfs significant negatieve relaties aangetroffen. Hanushek gaat niet in op het effect van schoolgrootte. Van de hier geciteerde onderzoeken wijzen de uitkomsten van Riew (1966), Wales (1973), Watt (1980), Kumar (1983), Bee en Dolton (1985) en Sengupta en Sfeir (1986) op schaalvoordelen of op een zeer hoge optimale schoolgrootte. Summers en Wolfe (1977) vonden als enigen (met een geheel andere benaderingswijze) een negatief effect van schoolgrootte op prestaties. De resultaten van Jimenez (1986) zijn tegenstrijdig, die van Glennerster en Wilson (1970) duiden niet op significante schaafeffecten.

1.3 Meting van het produktievolume

Een eenvoudige en voor de hand liggende indicator voor het produktievolume is het aantal leerlingen, dat in een bepaald tijdvak onderwijs bij de betreffende instelling volgt. Deze indicator geeft echter geen goed beeld van de produktie van de school; de gemiddelde jaarlijks per leerling toegevoegde waarde hoeft niet voor elke school gelijk te zijn. De meest ideale indicator voor het produktievolume van een school is de som van de door de school toegevoegde vaardigheden van alle leerlingen in een bepaald jaar. Daartoe moeten de vaardigheden via een testresultaat onder één noemer worden gebracht. In de internationale literatuur (Summers en Wolfe 1977) wordt aanbevolen bij elke leerling zowel aan het begin als aan het eind van de analyseperiode de vaardigheden (in de praktijk meestal alleen de kennis) te meten. Het produktievolume is dan de som over alle leerlingen van de toegevoegde vaardigheden. Dergelijke gegevens zijn echter niet beschikbaar.

Daarom worden de toegevoegde vaardigheden hier op een meer globale manier gemeten, die aansluit bij de beschikbare informatie. Met behulp van beschikbare overgangsmatrices is het jaarlijkse produktievolume bij benadering te berekenen als de som over leerlingen van de resterende nominale studieduur gedeeld door de verwachte resterende verblijfsduur. Deze berekeningswijze doet recht aan tijdens de studie opgelopen vertraging. Het zuiverst is deze maat voor leerlingen die het eindexamen behalen, omdat het niveau daarvan wordt gegarandeerd door een landelijke toetsing. Tussentijdse uitstroom van leerlingen (door verhuizing, uitval of verwijzing naar een ander onderwijs-element) wordt verdisconteerd door aan dergelijke leerlingen een produkt toe te kennen, dat in principe is gebaseerd op het aantal voltooide cursusjaren. De beschikbare gegevens geven echter geen uitsluitsel omtrent de vraag of tussentijds uitgestroomde leerlingen het laatst gevolgde cursusjaar al of niet succesvol hebben afgerond. Daarom wordt het laatst gevolgde cursusjaar maar voor de helft geteld.

Deze indicator kan ook worden gezien als het aantal leerlingen gecorrigeerd voor een kwaliteitsfactor. In plaats van de algemene term produktievolume wordt in het vervolg ook wel gesproken van het aantal leerlingequivalenten.

Bij de hier gehanteerde methode wordt geen rekening gehouden met verschillen in de kwaliteit van de instroom tussen scholen die tot hetzelfde onderwijs-element behoren. Het produktievolume van een probleemschool in een oude wijk van een grote stad kan daardoor lager uitvallen dan dat van een in een «betere» buurt gelegen school.

Kwaliteitsverschillen in de instroom kunnen ook een rol spelen in landelijke gebieden, waar de afstand tot beschikbare scholen een rol kan spelen. In het algemeen kan echter worden gesteld, dat kwaliteitsverschillen van de instroom in het avo veel minder groot zijn dan bij het basisonderwijs, omdat al een zekere voorselectie heeft plaatsgevonden. Dit geldt in nog sterkere mate voor de instroom in categoriale avo-scholen.

Verwachtingen voor het uitstroomniveau en de resterende verblijfsduur kunnen worden gebaseerd op de overgangsmatrices, die het CBS jaarlijks voor het gehele avo publiceert (bijvoorbeeld CBS 1986a), maar die voor dit onderzoek per afzonderlijke school ter beschikking stonden. Een dergelijke matrix geeft het verband aan tussen leerlingenbestanden naar onderwijselement en leerjaar in twee opeenvolgende schooljaren. Processen als overgang, zittenblijven, maar ook in- en uitstroom, worden in een dergelijke matrix vastgelegd. Onder de veronderstelling dat de doorstroom- en uitvalpercentages door de jaren heen constant blijven, kan men uitrekenen hoeveel procent van de leerlingen uiteindelijk zal uitstromen op de verschillende niveaus en hoelang zij daar gemiddeld over zullen doen (Blank 1986a en 1986b).

Bij wijze van voorbeeld is in Tabel 1 het produktievolume (het aantal leerlingequivalenten) van het tweede leerjaar van een denkbeeldige categoriale mavo berekend. In het voorbeeld wordt uitgegaan van 100 leerlingen, waarvan er 85 mettertijd het einddiploma zullen behalen. De overige 15 leerlingen zullen zonder diploma uitstromen uit diverse leerjaren. Afhankelijk van het niveau waarop zij uitstromen zullen de leerlingen een zekere gemiddelde vertraging oplopen.

Tabel 1. Berekening produktievolume tweede leerjaar categoriale mavo

	uitstroomniveau				totaal
	mavo2	mavo3	mavo4	diploma	
1 theoretisch verwachte uitstroom a)	5	5	5	85	100
2 nominale studieduur b)	0,5	1,5	2,5	3,0	
3 verwachte gemiddelde verblijfsduur b)	1,1	2,2	3,3	3,3	
4 produkt (1*2/3)	2,3	3,4	3,8	77,3	86,8 c)

a) Verdeling van 100 leerlingen over diverse uitstroomniveaus.

b) Gemeten in jaren.

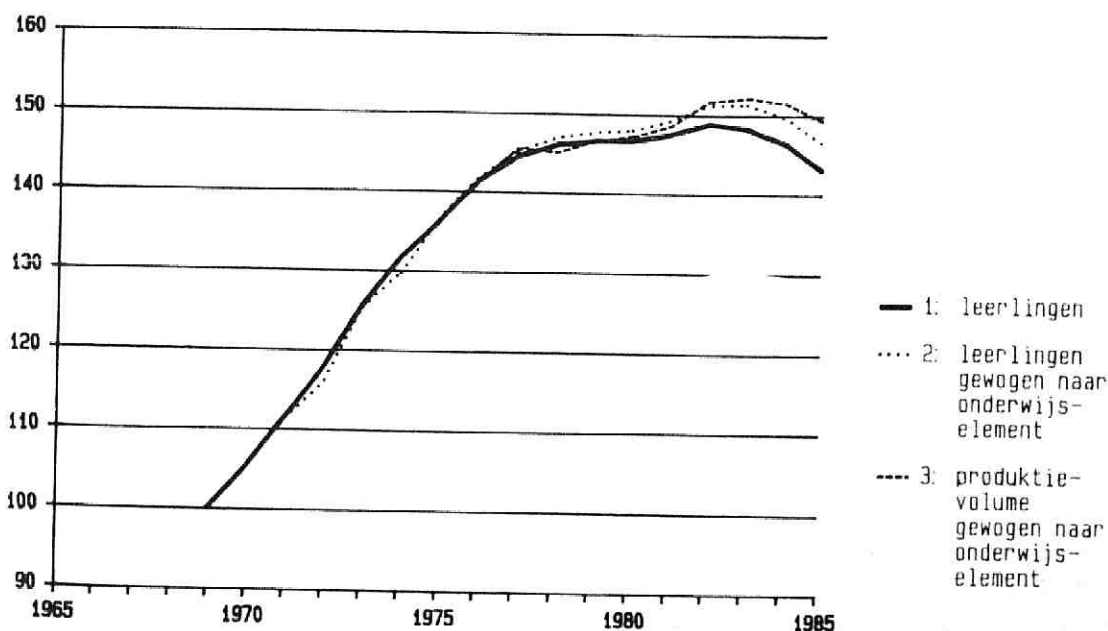
c) De 100 leerlingen corresponderen met een effectief jaarlijks produkt van 86,8 leerlingequivalenten.

Het produktievolume van categoriale scholen voor mavo is de som van het produktievolume voor de vier leerjaren. Het totale produkt omvat dus de som van de deelprodukten per leerjaar en uitstroomniveau. Voor de andere categoriale scholen en de onderwijselementen van de scholengemeenschappen is de berekening van het produktievolume op analoge wijze geschied. Doordat de indicator is gebaseerd op verwachtingen met betrekking tot de verdere schoolloopbaan, is het mogelijk om leerlingen in brugklassen toe te delen aan de onderwijselementen, waarin zij later terecht zullen komen. Dit is een belangrijk nevenargument om bij de keuze van de indicator uit te gaan van verwachtingen met betrekking tot de verdere schoolloopbaan. Om de vergelijkbaarheid met categoriale scholen te handhaven, wordt bij scholengemeenschappen de cursusduur en verblijfsduur geschat tot het behalen van het diploma in het betreffende onderwijselement. Bij uitstroom in de brugjaren wordt de leerling toegerekend aan het laagste onderwijselement (mavo voor mavo/havo en mavo/havo/vwo en havo voor havo/vwo). Het gaat daarbij overigens om relatief kleine aantallen leerlingen.

Het produktievolume van twee verschillende onderwijselementen is niet zonder meer vergelijkbaar. Uitgaande van de regelgeving voor overgangen tussen onderwijselementen, die inhoudt dat mavo-gediplomeerden door kunnen stromen naar leerjaar 4 van het havo en havo-gediplomeerden naar leerjaar 5 van het vwo, kan worden gepostuleerd dat de verhouding in kennisvermeerdering tussen een cursusjaar mavo, een cursusjaar havo en een cursusjaar vwo 3:4:5 bedraagt. Met een dergelijke verdeelsleutel kan de produktie van de drie onderwijselementen onder één noemer worden gebracht. Hiermee begeeft men zich echter op glad ijs, omdat het aanleren van bepaalde vaardigheden bij zwakkere leerlingen veel meer inspanning vergt dan bij betere leerlingen. Daarnaast kan de aggregatie van deze produktindicatoren tot één samengestelde indicator ook plaatsvinden op basis van empirische gewichten, die worden afgeleid uit de gemiddelde inzet van middelen die nodig is om één extra eenheid produkt te leveren in de afzonderlijke onderwijselementen. Op dit onderwerp wordt in paragraaf 4 nader ingegaan. Bij de presentatie van de uitkomsten zal tevens worden ingegaan op de gevoeligheid van de uitkomsten voor uiteenlopende aannames met betrekking tot de te hanteren gewichten.

2. PRODUKTIVITEITSONTWIKKELING, 1969-1985

In Figuur 1 wordt het verloop van het produktievolume (in leerlingequivalenten) over de periode 1969/'70-1985/'86 weergegeven. Het produktievolume is berekend met behulp van de jaarlijks door het CBS gepubliceerde avo-matrices (bijvoorbeeld CBS 1986a). Bij de weging van de afzonderlijke produktievolumina voor mavo, havo en vwo wordt daarbij uitgegaan van de in paragraaf 1.3 vermelde a priori gewichten (lijn 3). Om een interpretatie mogelijk te maken is tevens het verloop van het aantal leerlingen (lijn 1) en het verloop van het naar onderwijselement gewogen aantal leerlingen in Figuur 1 opgenomen (lijn 2).



(a) Indexcijfers met 1969/'70 : 100.

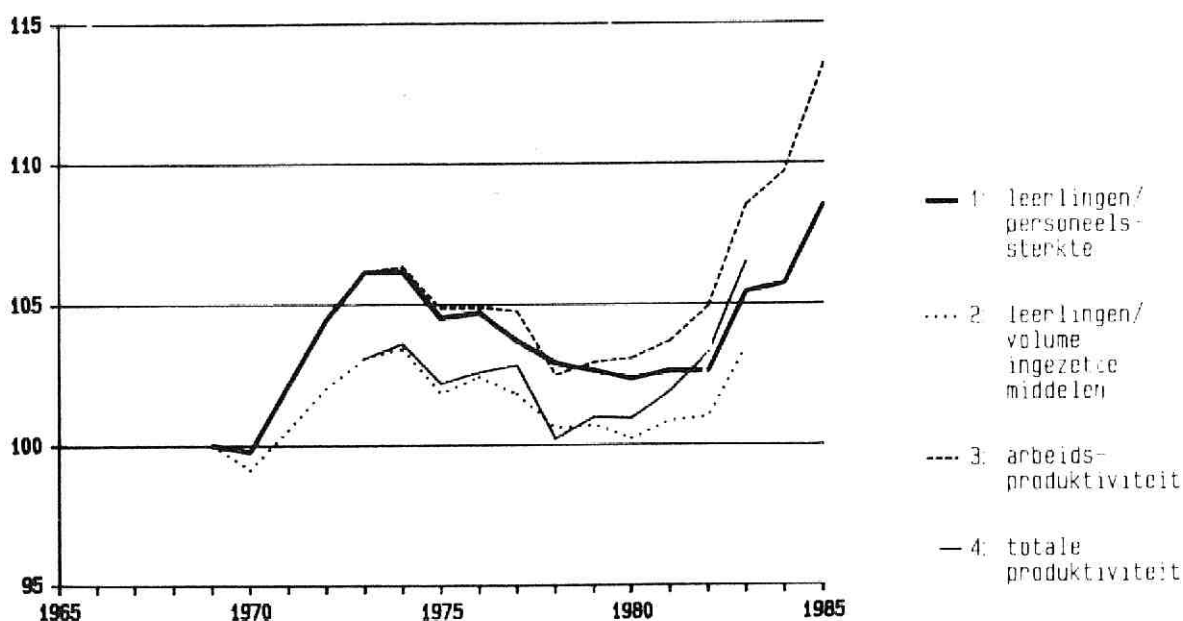
Figuur 1. Avo: leerlingen en produktievolume, 1969-1985 (a)

Lijn 1 blijft achter bij lijn 2 door een toenemend aandeel van vwo-leerlingen. Lijn 3 blijft rond 1977/'78 achter bij lijn 2 doordat in verhouding meer leerlingen het onderwijs zonder diploma verlaten. In latere jaren neemt zowel het zittenblijven als het ongediplomeerd vertrek af, waardoor lijn 3 uitstijgt boven lijn 2.

De produktiviteit is gedefinieerd als het quotiënt tussen produktievolume en ingezette middelen. Het eerste kan gemeten worden als het aantal leerlingen of de in paragraaf 1.3 gedefinieerde leerlingequivalenten. De ingezette middelen zijn uitgedrukt als het (voor arbeidsduur gecorrigeerde) personeelsvolume of als een gewogen gemiddelde van personeelsvolume en gedefleerde materiële kosten, waarbij de kostenaandelen in 1980 als uitgangspunt gekozen zijn. Gegevens met betrekking tot de laatstgenoemde onderwerpen zijn afkomstig uit CBS-bronnen (bijvoorbeeld CBS 1986b en CBS 1987).

Figuur 2 geeft het verloop van de produktiviteit van het avo weer. Daartoe worden achtereenvolgens gepresenteerd:

- (1) de leerlingenaantallen gedeeld door het onderwijzend personeel (in voltijdequivalenten);
- (2) de leerlingenaantallen gedeeld door het volume van de ingezette middelen;
- (3) de arbeidsproduktiviteit, dat wil zeggen, het produktievolume uitgedrukt in leerlingequivalenten gedeeld door het onderwijzend personeel (in voltijdequivalenten);
- (4) de totale produktiviteit, dat wil zeggen, het produktievolume uitgedrukt in leerlingequivalenten gedeeld door het volume van de ingezette middelen.



(a) Indexcijfers met 1969/'70 : 100.

Figuur 2. Avo: produktiviteit, 1969-1985 (a)

Zowel in de beginperiode als in de eindperiode stijgt de leerling/leraarverhouding (lijn 1) met circa 5% door veranderingen in de regelgeving (Staatsblad 1972 en 1983). In de tussenliggende periode treedt een geringe, maar moeilijk verklaarbare daling op. Vanaf 1979/'80 stijgt de arbeidsproduktiviteit (lijn 3) sneller dan lijn 1. Dit is te verklaren door de bij de bespreking van Figuur 1 genoemde factoren: het toegenomen aandeel van vwo-leerlingen en een vermindering van de percentages zittenblijvers en uitvallers. Kennelijk heeft de verhoging van de leerling/leraarverhouding in het begin van de jaren tachtig niet geleid tot een daling van de prestaties. Lijnen 2 en 4 blijven in de beginperiode enigszins

achter bij de corresponderende lijnen 1 en 3, doordat in die periode de materiële kosten sneller zijn gestegen dan de personeelskosten.

3. PRODUKTIVITEIT OP MICRONIVEAU, 1985/'86

Het produktievolumen wordt gemeten met de in paragraaf 1.3 geïntroduceerde produktindicator. De benodigde onderwijsmatrices op schoolniveau zijn ter beschikking gesteld door het CBS. Directe kostengegevens op schoolniveau waren niet beschikbaar. Wel beschikt het CBS over gegevens op schoolniveau over de personeelssterkte naar leeftijd en categorie. Deze zijn met behulp van de regelgeving met betrekking tot de salariering (O&W 1985) omgerekend naar personeelskosten. De materiële kosten, die circa 12% van de totale kosten uitmaken, zijn dus niet in de analyse betrokken. De analyses zijn gebaseerd op een deelbestand van 1039 scholen. Dit is 75% van het totaal aantal scholen. Buiten beschouwing zijn onder meer gelaten scholen met een driejarige brugperiode, scholen in opbouw en middenscholen.

Tabel 2 brengt de arbeidsproductiviteit in beeld. Het produktievolumen is voor de onderwijselementen mavo, havo en vwo gewogen met behulp van de standaardgewichten 3:4:5, die zijn gebaseerd op de relatieve kennisvermeerdering per leerjaar in de onderscheiden schoolstructuren.

Tabel 2. Arbeidsproductiviteit, 1985/'86 a) b) c)

	categoriale scholen			scholengemeenschappen		
	mavo	havo	vwo	m/h	h/v	m/h/v
kleine scholen	71		107			
middelkleine scholen	79	82	110	70		
middelgrote scholen	86	93	129	85	109	92
grote scholen				96	121	104
totaal	78	91	125	89	118	102

a) Indexcijfers met gemiddelde totaal avo=100.

b) Wegens het kleine aantal waarnemingen (N=7) en de daarmee gepaard gaande relatief grote onzekerheidsmarges mag aan de uitkomsten voor categoriale scholen voor havo niet te veel waarde worden gehecht.

c) Kleine scholen: 259 leerlingen of minder; middelkleine scholen: tussen 259 en 388 leerlingen; middelgrote scholen: tussen 388 en 921 leerlingen; grote scholen: meer dan 921 leerlingen.

De arbeidsproductiviteit loopt in de verschillende schoolstructuren sterk uiteen. De indexcijfers voor de categoriale mavo-, havo- en vwo-scholen bedragen 78, 91 en 125. De cijfers voor de verschillende soorten scholengemeenschappen stroken hier goed mee: 89 voor mavo/havo, 102 voor mavo/havo/vwo en 118 voor havo/vwo. Deze uitkomsten hangen samen met de gekozen a priori gewichten voor de produktievolumina van de afzonderlijke schooltypen. Schaafeffecten zijn in Tabel 2 duidelijk zichtbaar. Bij alle afzonderlijke schoolstructuren treedt een aanzienlijke stijging van de arbeidsproductiviteit op naarmate de schoolgrootte oploopt. Verschillen van 20 procentpunten of meer zijn geen uitzondering.

Het inzicht kan worden vergroot door een decompositie van de arbeidsproductiviteit in een component geassocieerd met de regelgeving en een component geassocieerd met het doorstroompatroon van leerlingen in termen van studievertraging en uitval.

Tabel 3 geeft kengetallen voor de verhouding tussen leerlingenaantallen en personeelssterkte, die een resultante is van de vigerende regelgeving.

Tabel 3. Verhouding tussen leerlingen en personeelssterkte, 1985/'86 a)

	mavo	havo	vwo	m/h	h/v	m/h/v
kleine scholen	88		82			
middelkleine scholen	98	88	85	85		
middelgrote scholen	104	92	97	100	96	92
grote scholen				106	104	103
totaal	97	91	95	102	102	101

a) Leerlingen per arbeidsjaar. Indexcijfers met gemiddelde totaal avo=100.

Zie voorts voetnoten b) en c) bij Tabel 2.

Aangezien de leerling/leraar-verhouding voor alle schoolstructuren met de grootte toeneemt, illustreert Tabel 3 de in de regelgeving vervatte schaaffecten. Bij gelijke schoolgrootte blijkt bij mavo-scholen een hogere leerling/leerkracht verhouding op te treden dan bij de andere schoolstructuren. Dit komt doordat een mavo-school slechts vier leerjaren kent, waardoor deze qua schaaffecten vergelijkbaar is met een havo-school die 1,25 maal groter is en een vwo-school die 1,5 maal groter is. Het is overigens niet juist om te stellen dat de regelgeving de oorzaak is van een deel van de schaaffecten. De regelgeving is gebaseerd op de notie dat onderwijs in klassen wordt gegeven, dat klassen een bepaalde optimale grootte kennen, en dat per klas een aantal leraarlessen moet worden gegeven. De regelgeving speelt daarmee al in op schaaffecten. De mate waarin dat het geval is, kan worden getoetst door een analyse van het doorstroompatroon van leerlingen.

Een karakteristiek voor het doorstroompatroon kan worden verkregen door het produktievolume per element te delen door het corresponderende aantal leerlingen. Kengetallen voor het produktievolume per leerling zijn opgenomen in Tabel 4.

Tabel 4. Produktievolume per leerling a), 1985/'86

	categoriaal			scholengemeenschappen						
				mavo/havo		havo/vwo		havo/havo/vwo		
	mavo	havo	vwo	mavo	havo	havo	vwo	mavo	havo	vwo
kleine scholen	89		84							
middelkleine scholen	90	78	84	85	74					
middelgrote scholen	92	85	86	88	76	80	86	90	75	83
grote scholen				88	84	82	87	90	77	86
openbare scholen	87	64	86	88	71	80	85	87	75	83
bijzondere scholen	91	85	85	88	81	82	87	93	78	87
totaal	90	84	86	88	80	81	86	90	77	85

a) Gedefinieerd als ongewogen produktievolume per leerling (x 100).

Zie voorts voetnoten b) en c) bij Tabel 2.

Een waarde van 100 voor het produktievolume per leerling in Tabel 4 impliceert dat er geen leerlingen uitvallen en dat er geen studievertraging optreedt. Naarmate deze waarde lager is treedt er meer uitval en/of studievertraging op. Dit volgt uit de in paragraaf 1.3 besproken berekeningswijze van deze indicator. Uit de tabel blijkt dat het doorstroompatroon gunstiger is naarmate de school groter is. Op het mavo komen uitval en/of studievertraging minder voor dan bij andere onderwijs-elementen. De tabel levert geen aanwijzing op voor een gunstiger doorstroompatroon bij scholengemeenschappen dan bij categoriale scholen van vergelijkbare grootte. Hierbij moet de kanttekening worden

gemaakt, dat doorstroom binnen scholengemeenschappen van het ene naar het andere onderwijselement hier ook als uitstroom wordt gekenschetst om een vergelijking met categoriale scholen mogelijk te maken. Voor het havo levert de tabel zelfs de suggestie op dat de doorstroom binnen het havo-element in scholengemeenschappen ongunstiger is dan die bij categoriale havo's. Het trekken van een conclusie is echter gevaarlijk, omdat de steekproef slechts zeven categoriale havo's omvat. Bij wijze van toegift wordt ook enige informatie verstrekt over bijzondere versus openbare scholen. Bijzondere scholen blijken te worden gekenmerkt door een gunstiger doorstroompatroon. Aangezien bijzondere scholen gemiddeld kleiner van omvang zijn dan openbare scholen, kunnen deze verschillen niet worden veroorzaakt door positieve schaafeffecten.

4. ECONOMETRISCHE ANALYSE OP MICRONIVEAU, 1985/'86

4.1 Opzet

Op schoolniveau zijn voor het schooljaar 1985/'86 econometrische analyses uitgevoerd die een nader inzicht kunnen geven in de produktiestructuur van het avo. Daarbij is uitgegaan van een benadering via een kostenfunctie. De salariskosten zijn hierbij gekozen als afhankelijke variabele, het produktievolume in leerlingequivalenten als onafhankelijke variabele. De gebruikte gegevens zijn toegelicht in de eerste alinea van paragraaf 3. Er is gekozen voor een translog-specificatie. De variabelen ondergaan hierbij een log-transformatie, terwijl naast lineaire termen ook alle kwadratische en tweedegraads kruistermen gespecificeerd worden. Een dergelijke specificatie, die gezien kan worden als een tweedegraads-benadering van een willekeurig functioneel verband, laat het toe om grootte-afhankelijke schaaelasticiteiten te schatten, die gekenmerkt kunnen worden door een optimum (zie verder Goudriaan et al. 1989a, hoofdstuk 2). In een eerste publikatie over dit onderzoek (Van der Torre en Blank 1989) werd overigens een translogproduktiefunctie toegepast. Een voordeel van de in het huidige artikel gehanteerde benadering via een kostenfunctie is dat meerdere produktindicatoren gehanteerd kunnen worden.

Het lag in de bedoeling om één simultane schatting te maken voor alle avo-scholen en -scholengemeenschappen ongeacht hun schoolstructuur, waarbij de gewichten voor de produktie van de onderscheiden onderwijselementen (mavo, havo en vwo) empirisch zouden worden bepaald. Een dergelijke analyse zou tevens aanknopingspunten bieden voor een evaluatie van synergie-effecten (voor- en/of nadelen van gemengde scholengemeenschappen). In de praktijk bleek deze ambitieuze opzet helaas niet uitvoerbaar te zijn. Bij de simultane schatting zijn namelijk merkwaardig lage coëfficiënten en marginale kosten gevonden voor het havo. Mogelijkerwijs zijn de uitkomsten vertekend door multicollineariteit. Daarom is gekozen voor een opzet waarbij afzonderlijke schattingen zijn uitgevoerd voor de zes onderscheiden schoolstructuren en waarbij bovendien de produktievolumina voor havo en vwo bij elkaar zijn opgeteld. Daarbij is gekozen voor ongewogen optelling op grond van de overeenkomstige marginale kosten voor categoriale havo- en vwo-scholen (zie Tabel 6).

De gehanteerde specificatie van de translogkostenfunctie voor een 'categoriale school' (mavo, havo, vwo of havo+vwo) ziet er als volgt uit:

$$\log C = a + b \log Q + c (\log Q)^2$$

waarin C staat voor de kosten en Q voor het produktievolume. Zoals in de econometrische literatuur gebruikelijk is, zijn C en Q gedeeld door hun gemiddelde waarden om de interpreteerbaarheid van de uitkomsten te vergroten. De schaaelasticiteit voor de

gemiddelde instelling kan nu berekend worden als $v = 1/b$.

In het geval van een 'scholengemeenschap' (mavo/havo of mavo/havo+vwo) is de volgende specificatie van toepassing:

$$\log C = a + b_1 \log Q_1 + b_2 \log Q_2 + c_1 (\log Q_1)^2 + c_2 (\log Q_2)^2 + e \log Q_1 \log Q_2$$

waarin C staat voor de kosten en Q_1 en Q_2 voor het gereconstrueerde produktievolume volgens de twee onderscheiden indicatoren, waarbij C , Q_1 en Q_2 weer gedeeld zijn door hun gemiddelde waarden. De schaaelasticiteit voor de gemiddelde instelling kan nu worden berekend als: $v = 1 / (b_1 + b_2)$. De marginale kosten van produkttypen Q_1 en Q_2 zijn te berekenen uit de partiële afgeleide van de kostenfunctie naar de hoeveelheid van het betreffende produkttype. Voor produkttype Q_1 , bijvoorbeeld, kunnen de marginale kosten uitgedrukt worden als $M_1 = b_1 C_1 / Q_1$.

4.2 Uitkomsten

Tabel 5 geeft een overzicht van de schattingsresultaten.

Tabel 5. Avo: Personeelskosten in relatie tot het produktievolume, 1985/'86 a) b)

	mavo	havo	vwo	m/h	h/v	m/h/v
mavo	0,79 *			0,39 *		0,18 *
havo/vwo		0,81 *	0,82 *	0,41 *	0,80 *	0,55 *
kostenflexibiliteit c)	0,79 d)	0,81	0,82 d)	0,80 d)	0,80 d)	0,73 d)
schaaelasticiteit e)	1,27 d)	1,23	1,22 d)	1,25 d)	1,25 d)	1,38 d)

a) De met * aangegeven coëfficiënten zijn op 5%-niveau significant afwijkend van 0. De resultaten zijn gebaseerd op OLS-schattingen van translogkostenfuncties. De variabelen zijn berekend rond de gemiddelde waarden.

b) Wegens het kleine aantal waarnemingen ($N=7$) en de daarmee gepaard gaande relatief grote onzekerheidsmarges mag aan de uitkomsten voor categoriale scholen voor havo niet te veel waarde worden gehecht.

c) Dit is de som van de coëfficiënten van de produkttypen.

d) Verschilt significant van 1 (bij een significantieniveau van 5%).

e) Dit is het omgekeerde van de kostenflexibiliteit.

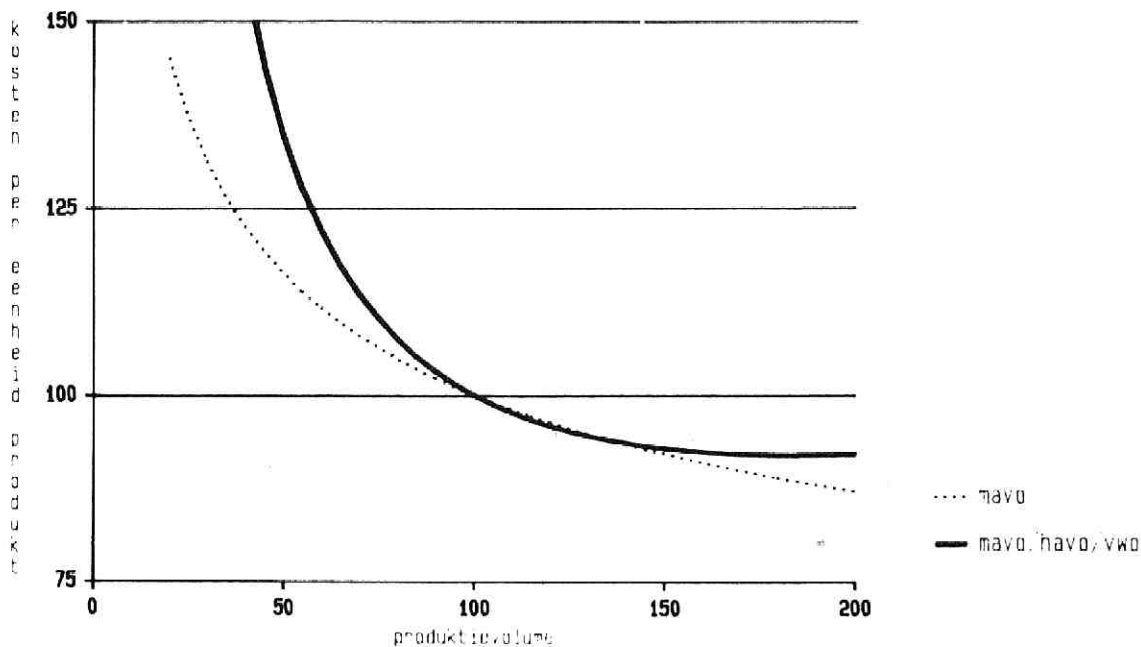
Uit de schattingen valt af te leiden, dat bij de gemiddelde school in de onderscheiden schoolstructuren een stijging van het produktievolume met 1% gepaard gaat met een stijging van de kosten, die varieert tussen 0,73 en 0,82%. Dit komt overeen met schaaeffecten van 1,2 à 1,4, die significant van één verschillen. Er is dus bij alle schoolstructuren in het avo sprake van schaalvoordelen.

In paragraaf 3 is opgemerkt, dat de regelgeving al inspeelt op het bestaan van schaaeffecten. Als illustratie van de in de regelgeving vervatte schaaeffecten zijn eveneens kostenfuncties geschat waarbij het produktievolume is vervangen door het aantal leerlingen. De op deze wijze bepaalde schaaeffecten zijn significant groter dan één, maar kleiner dan de corresponderende schaaeffecten in Tabel 5. Ondanks de door de regelgeving toegevoegde extra middelen is het doorstroompatroon bij kleinere scholen toch minder gunstig dan bij grotere scholen, in de zin dat er meer vertraging en/of uitval plaatsvindt. Dit vormt een bevestiging van de conclusies die reeds bij de globale analyse op instellingsniveau (paragraaf 3) zijn getrokken.

Er is ook getoetst op het bestaan van synergie-effecten, dat wil zeggen van voor- of nadelen voor scholengemeenschappen die meerdere onderwijselementen omvatten. In de Angelsaksische literatuur worden deze aangeduid als "economies of scope". Voor mavo en

havo/vwo blijken synergie-effecten nauwelijks waarneembaar te zijn.

De schattingsresultaten maken het mogelijk een beeld te geven van de relatie tussen de schaal van de produktie en de kosten per eenheid produkt. In Figuur 3 wordt deze relatie voor categoriale scholen voor mavo en voor scholengemeenschappen voor mavo/havo/vwo in beeld gebracht. Voor andere schoolstructuren zijn overeenkomstige curves gevonden. De figuur schetst het effect van verschillende produktieniveaus (horizontale as) op de kosten per eenheid produkt (verticale as). De index van het produktievolume van de gemiddelde school is in beide gevallen op 100 gesteld. Bij scholengemeenschappen wordt van een constante verhouding tussen de produkttypen uitgegaan, die is gebaseerd op het gemiddelde.



Figuur 3. Scholengemeenschappen voor mavo/havo/vwo en categoriale scholen voor mavo: produktievolume en kosten per eenheid produkt, 1985/'86

Opmerking: gemiddelde school : 100.

Uit Figuur 3 blijkt dat over vrijwel het gehele bereik een verhoging van het produktievolume gepaard gaat met een daling van de kosten per eenheid produkt. Over nagenoeg de gehele linie is dus sprake van schaalvoordelen. De meest produktieve scholen zijn de scholen met het grootste absolute produktievolume. In de praktijk zijn dit tevens de scholen met het grootste aantal leerlingen. Het optimum voor categoriale mavo-scholen ligt ver boven de in de praktijk optredende maximale schoolgrootte; het optimum voor scholengemeenschappen mavo/havo/vwo ligt daar net iets onder.

De in Tabel 5 gepresenteerde schattingen geven uitsluitend informatie over de procentuele toeneming van de kosten bij een stijging van het produktievolume van een onderwijselement met 1%. Daardoor hangt het effect van de verandering van de produktievolumina op de kosten af van het niveau van het produktievolume. De marginale kosten vormen een manier om het absolute effect van een toeneming van de produktievolumina op de kosten te meten. Deze zijn gedefinieerd als de kostenstijging bij een groei van het produktievolume van een onderwijselement met één eenheid. In Tabel 6 worden de geschatte marginale kosten gepresenteerd voor de gemiddelde school van elke schoolstructuur. Deze kosten hebben alleen betrekking op het ingezette personeel.

Daarom is bij de berekening van de marginale kosten een opslagfactor (1.14) gehanteerd voor materiële kosten.

Tabel 6. Marginale kosten a)

	marginale kosten mavo		marginale kosten havo/vwo	
	absoluut (x gld)	index b)	absoluut (x gld)	index b)
categoriaal mavo	4.400	100		
categoriaal havo			5.300	119
categoriaal vwo			5.100	116
scholengemeenschap mavo/havo	4.600	103	4.700	107
scholengemeenschap havo/vwo			5.000	114
scholengemeenschap mavo/havo/vwo	3.600	81	4.900	110

a) Berekend voor de gemiddelde school in de betreffende schoolstructuur.

Zie voorts voetnoot b) bij Tabel 5.

b) De marginale kosten van categoriale scholen voor mavo zijn op 100 gesteld.

Bij de interpretatie van de in de Tabel 6 aangegeven marginale kosten moet rekening worden gehouden met de in de regel vrij ruime onzekerheidsmarges rondom de aangegeven bedragen. Over de gehele linie liggen de marginale kosten voor havo/vwo echter boven die voor het mavo. Voorts liggen de marginale kosten bij scholengemeenschappen in de regel onder die voor vergelijkbare elementen bij categoriale scholen. De voor het mavo gevonden marginale kosten per eenheid produkt liggen tussen 3.600 en 4.600 gulden; die voor havo/vwo tussen 4.900 en 5.300 gulden. De schattingen voor het categoriale havo en vwo ontlopen elkaar nauwelijks.

De empirisch gevonden gewichten liggen veel dichter bij de verhouding 1:1:1 dan de op basis van kennisvermeerdering geconstrueerde a priori gewichten (3:4:5). Zoals ook al op grond van de regelgeving duidelijk is, worden de volgens deze laatste maatstaf geconstateerde produktiviteitsverschillen niet bereikt door een grotere inzet van mankracht en maar zeer ten dele door een grotere kwaliteit (als men het opleidingsniveau van docenten en daarmee samenhangende salarisverschillen als zodanig mag interpreteren) van het onderwijzend personeel in havo en vwo. Evenmin worden de op grond van de a priori gewichten verkregen produktiviteitsverschillen teniet gedaan door een systematisch hogere doubleer- of uitvalkans van leerlingen in havo en vwo. De verschillen ontstaan door het stellen van hogere eisen aan de leerlingen, in combinatie met een voorselectie op vaardigheden.

5. SIMULATIES

Op basis van de in de micro-analyse geschatte kostenfuncties kunnen de effecten van wijzigingen in de produktiestructuur worden gesimuleerd. Hier worden de uitkomsten van twee simulaties gepresenteerd, te weten:

- (1) de grensvariant
- (2) de schaalvariant.

In de grensvariant wordt bekeken welke kostenbesparingen kunnen worden gerealiseerd als de kostenfunctie van de 50% meest efficiënte scholen als uitgangspunt wordt genomen. De bijbehorende kosten worden «haalbare kosten» genoemd. Instellingen die hun produktievolume voortbrengen tegen hogere kosten dan deze haalbare kosten, worden

ertoe gebracht om het bestaande produktievolume tegen de haalbare kosten te realiseren. De schaal van de produktie en de schoolstructuur van de scholen blijven in deze variant ongewijzigd. Overigens moet hierbij worden bedacht, dat deze analyse maar beperkte indicaties oplevert over de oorzaken van dergelijke produktiviteitsverschillen. Verschillen in de capaciteiten van instromende leerlingen kunnen hierbij een rol spelen, maar ook finesses in de regelgeving die kunnen leiden tot vrij grote verschillen in de toewijzing van leerkrachten bij min of meer gelijke aantallen leerlingen. Het zal daarom niet eenvoudig zijn om de in de grensvariant aangegeven besparingsmogelijkheden te «verzilveren».

Uit de dwarsdoorsnede-analyse bleek dat de gemiddelde kosten per eenheid produkt met de schoolgrootte afnemen (zie Figuur 3). De grootste scholen zijn het meest produktief. In de schaalvarianten worden scholen gefuseerd tot eenheden met een zo doelmatig mogelijke omvang. Fusies worden alleen overwogen voor scholen binnen hetzelfde nodale gebied om de bereikbaarheid niet te veel geweld aan te doen. Daarbij is aangesloten bij een gangbare CBS-indeling van Nederland in 80 nodale gebieden. Er zijn vier varianten uitgewerkt. In subvariant 2a worden scholen binnen een nodaal gebied zodanig samengevoegd dat hun omvang de optimale omvang zo dicht mogelijk benadert. Bij het avo is de optimale omvang binnen de in de praktijk optredende spreiding die van de grootste bestaande school voor elke onderscheiden schoolstructuur. De schoolstructuur wordt in deze subvariant in stand gehouden. Slechts scholen met eenzelfde schoolstructuur worden gesplitst of samengevoegd. Voorts is als subvariant 2b een versie van de eerste schaalvariant doorgerekend, waarbij de samenvoeging is beperkt tot scholen met overeenkomstige denominatie (openbaar versus bijzonder).

De analyses wijzen uit dat de produktiviteit van scholengemeenschappen ongeveer gelijk is aan die van categoriale scholen met dezelfde omvang. De voordelen van scholengemeenschappen boven categoriale scholen zijn daardoor in feite terug te voeren op schaalvoordelen. In subvariant 2c zijn per nodaal gebied alle scholen samengevoegd tot scholen die in omvang zo veel mogelijk overeenstemmen met de grootste bestaande scholengemeenschap mavo/havo/vwo. Voorts is als subvariant 2d een versie van de voorafgaande subvariant doorgerekend, waarbij de samenvoeging is beperkt tot scholen met een overeenkomstige denominatie (openbaar versus bijzonder).

Tabel 7 geeft een overzicht van de resultaten van de verschillende simulaties.

Tabel 7. Resultaten simulaties a)

variant		effect in % op	
		aantal scholen	kosten
1 grensvariant		0	-12
2 schaalvariant	a	-26	- 4
	b	-22	- 3
	c	-66	-10
	d	-62	-10

a) In 1985/'86 waren er circa 1.380 scholen voor avo; de totale kosten kunnen voor dat schooljaar worden geschat op circa 4,5 miljard gulden.

Volgens de grensvariant kunnen aanzienlijke besparingen worden gerealiseerd (12%, corresponderend met een bedrag van 540 miljoen gulden) als aangenomen wordt dat minder efficiënte scholen hun produktievolume tegen de haalbare kosten gaan realiseren.

Als scholen volgens schaalvariant 2a op een meer optimale omvang worden gebracht,

wordt een besparing gerealiseerd van 4% (180 miljoen gulden). Dit gaat echter gepaard met een aanzienlijke inkrimping van het aantal scholen, namelijk met 26%. Indien bij de samenvoeging rekening wordt gehouden met het onderscheid tussen openbare en bijzondere scholen (subvariant 2b) is het effect iets geringer (een besparing van 3%, corresponderend met een bedrag van 135 miljoen gulden en een daling van het aantal scholen met 22%). Blijkens deze berekeningen zijn de kosten van het handhaven van een apart openbaar en bijzonder onderwijs althans voor het avo beperkt. Het verschil in besparing tussen subvarianten 2a en 2b zal iets groter zijn, indien rekening wordt gehouden met het bestaan van verschillende soorten bijzonder onderwijs. Meer dan enkele procentpunten kan dit verschil echter onmogelijk bedragen.

Als echter volgens subvariant 2c alle scholen binnen een nodaal gebied zodanig fuseren dat alleen scholengemeenschappen voor mavo, havo en vwo overblijven, die bovendien de meest optimale omvang zo veel mogelijk benaderen, dan zijn bij gelijkblijvend produktievolume besparingen van 10% (450 miljoen gulden) mogelijk. Het aantal scholen neemt in deze subvariant af met 66%. Als hierbij de restrictie geldt dat uitsluitend scholen met dezelfde denominatie (openbaar versus bijzonder) mogen fuseren, is de resulterende bezuiniging eveneens 10%, terwijl het aantal scholen met 62% afneemt (subvariant 2d). In dit geval zijn de kosten van handhaving van een apart openbaar en bijzonder onderwijs verwaarloosbaar.

LITERATUUR

- Bee M., en P.J. Dolton (1985). Costs and economies of scale in UK private schools. *Applied Economics* 17 (3), p.281-290.
- Blank J.L.T. (1986a). *Een stationair model voor de berekening van onderwijsindicatoren*. Voorburg: Centraal Bureau voor de Statistiek, notitie 3068-86-S3/intern.
- Blank J.L.T. (1986b). Verblijfsduur-, uitstroom- en snelheidsindices voor de jaren 1981 en 1982. In: *Zakboek onderwijsstatistiek 1986*. 's-Gravenhage: Staatsuitgeverij.
- Centraal Bureau voor de Statistiek (1986a). *Statistiek van het vwo, havo en mavo 1985/'86. Scholen en leerlingen*. 's-Gravenhage: Staatsuitgeverij.
- Centraal Bureau voor de Statistiek (1986b). *Statistiek van het vwo, havo en mavo 1985/'86. Leraren en bevoegdheidssituatie*. 's-Gravenhage: Staatsuitgeverij.
- Centraal Bureau voor de Statistiek (1987). *Statistiek van de uitgaven der overheid voor onderwijs 1984*. 's-Gravenhage: Staatsuitgeverij.
- Cohn, E., en J. Riew (1974). Cost functions in public schools. *Journal of Human Resources* 9 (zomer), p.408-414.
- Coleman, J.S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F. Weinfeld, and R.L. York (1966). *Equality of educational opportunity*. Washington D.C.: Government Printing Office.
- Glennerster, H., en G. Wilson (1970). *Paying for private schools*. Londen: Allen Lane.
- Goudriaan, R., F. van Tulder, J. Blank, A. van der Torre en B.Kuhry (1989a). Doelmatig Dienstverleners. Rijswijk: Sociaal en Cultureel Planbureau, *Sociale en Culturele Studie 11*.
- Goudriaan, R., F. van Tulder, J. Blank, A. van der Torre en B.Kuhry (1989b). Doelmatig Dienstverleners. *Economisch Statistische Berichten* 74 (3703), p.380-384.
- Hanushek, E.A (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature* 24 (3), p.1141-1177.
- Jimenez, E (1986). The structure of educational costs: multiproduct cost functions for primary and secondary schools in Latin America. *Economics of Education Review* 5 (1), p.25-39.
- Kumar, R. C. (1983). Economies of scale in school operation: evidence from Canada. *Applied Economics* 15 (3), p.323-340.
- Ministerie van O&W (1985). *Handleiding Onderwijs Salarisstructuur* (HOS-circulaire). Zoetermeer.
- Riew, J. (1966). Economies of scale in high school operation. *Review of Economics and Statistics* 48, p.280-287.

- SCP (1989). Naar een schaalvergrotingsbeleid in het onderwijs? (Een probleemverkenning). *Document 4, Sociaal en Cultureel Planbureau, Rijswijk*.
- Sengupta, Jati K., en Raymond E. Sfeir (1986). Production frontier estimates of scale in public schools in California. *Economics of Education Review* 5 (3), p.297-307.
- Staatsblad van het Koninkrijk der Nederlanden (1972). Besluit van 25 augustus 1972, houdende wijziging van het Besluit vwo-havo-mavo (Stb. 1967, 526) en het besluit avondscholen vwo-havo-mavo (Stb. 1968, 440). *Staatsblad*, (1972) 457.
- Staatsblad van het Koninkrijk der Nederlanden (1983). Besluit van 25 juli 1983, houdende wijziging van een aantal inrichtings- en eindexamenbesluiten (o.m. inzake de leraarlessen, een certificatieregeling en een vrijstellingsregeling). *Staatsblad*, (1983) 372.
- Summers, A.A., en B.L. Wolfe (1977). Do schools make a difference? *American Economic Review* 67 (4), p.639-652.
- Torre, A.G.J. van der, en J.L.T. Blank (1989). Economies of scale in het algemeen voortgezet onderwijs in Nederland. In: J. Scheerens en J.C. Verhoeven (ed.), *Schoolorganisatie, beleid en onderwijskwaliteit*. Lisse: Swets en Zeitlinger.
- Wales, T.J (1973). The effect of school and district size on education costs in British Columbia. *International Economic Review* 14, p.710-720.
- Watt, P.A. (1980). Economies of scale in schools: some evidence from the private sector. *Applied Economics* 12 (2), p.235-242.

Manuscript ontvangen 28-6-1989

Definitieve versie ontvangen 21-12-1989

The threat of scale drift in item banks

Dato N.M. de Gruijter

Educational Research Center, University of Leiden

ABSTRACT

Item banks contain a large number of items with item parameters on a common scale. The parameters are obtained from test forms administered at different occasions. Generally it is assumed that the item characteristics do not change with time. In this contribution the possibility of a systematic change in item difficulties is considered. Such a change poses a threat to the integrity of an item bank. A model with an estimable shift parameter is formulated for illustrative purposes.

INTRODUCTION

Items can be calibrated to a common latent scale, even items which have never been administered in one and the same test. In order to realize the calibration it is necessary that the test forms have enough items in common. New items are scaled to the same scale as the old items through these linking items, and old items can be updated simultaneously. In this way an item bank is created.

In principle two different methods for calibration are available. The first is to estimate the item parameters for each test separately and, next, to link the obtained scales through the common items. The second, statistically more satisfactory method is to estimate all item parameters simultaneously.

The construction of a coherent item bank based on IRT-methodology rests on the assumption that the item characteristics do not change with time. This assumption is not always a realistic one. Goldstein (1983) argued that item characteristics can change with a shifting instructional emphasis: some items may become outdated, whereas the subject matter relevant to other items gains importance. Bock, Muraki and Pfeifferberger (1988) convincingly demonstrated differential changes with time within the context of IRT-modelling. Results like theirs are relevant to all item bank constructors: the fit of a particular IRT-model at each test occasion does not imply the overall fit of the model with fixed item parameters.

In the Bock et al. model it is not possible to estimate average item drift over time. Such changes are incorporated in the average performance levels of examinee groups. The model is satisfactory when differential item functioning is the issue. However when performance standards must be maintained over time, detection of average item drift is of crucial importance. Item drift might result from test disclosure. In a different context, linear equating, possible effects of test disclosure were examined in a simulation study by Gilmer (1989).

This contribution demonstrates the potential threat of item drift to the integrity of an item bank. For the purpose of the demonstration it is assumed that the Rasch model fits at each test administration. Item difficulties change after the first administration. The different Rasch scales are linked by a least squares method (Engelhard & Osberg, 1983).

First, the linking method is discussed. Next the impact of the change in item characteristics is discussed. Finally, it is demonstrated that in this particular case adequate test linking remains possible.

LINKING RASCH SCALES WITH LEAST SQUARES

In the Rasch model the probability correct on item i given latent ability θ is given by

$$P_i(\theta) = \exp(\theta - b_i) / [1 + \exp(\theta - b_i)], \quad (1)$$

where b_i is the item difficulty parameter. The latent scale is determined up to an additive constant: substitution of b by $b+a$ and of θ by $\theta+a$ gives the same probability correct. One restriction is needed in order to fix the latent scale.

The latent scale for a particular test can be fixed by setting the average difficulty in the test equal to zero. Two tests j and k are scaled differently with this procedure: the item parameters of the common items in the more difficult test, the test with on the average higher b s on a common scale, are lower on the average than the corresponding items in the easy test. The average difference equals

$$y_{jk} = \sum (b_{ij} - b_{ik}) / n_{jk}, \quad (2)$$

where n_{jk} is the number of common items, and b_{ij} and b_{ik} refer to the item parameter estimates of the common items in test j and test k respectively. The shift y_{jk} estimates $\delta_j - \delta_k$, the difference between the easiness of test j and the easiness of test k , i.e. the difference between the average b of test k and the average b of test j on a common scale. In order to obtain estimates of all item parameters on the scale defined by test j , y_{jk} must be added to all item parameter estimates for test k . A shift with weighted item difficulty estimates was proposed by Wright and Bell (1984).

With $m > 2$ test forms there are at most $m(m-1)/2$ links between pairs of tests. Engelhard and Osberg (1983) proposed least squares estimation of the δ s based on values y_{jk} in Equation 2. In order to obtain a solution, the m -th δ was set equal to zero. In the following the first δ will be set equal to zero instead.

The least squares procedure can be described as follows. Assume that all $m' = m(m-1)/2$ links are present. The observed shifts are collected in an m' -element vector $\underline{y}$. This vector can be expressed as

$$\underline{y} = \underline{A} \underline{\delta} + \underline{e}, \quad (3)$$

where $\underline{\delta}$ is the $(m-1)$ -element vector with δ s for tests 2 through m , $\underline{A}$ is the $m' \times (m-1)$ matrix

$$\underline{A} = \begin{bmatrix} -1 & 0 & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & 0 & -1 \\ 1 & -1 & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 1 & 0 & . & . & . & 0 & -1 \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 & -1 \end{bmatrix}, \quad (4)$$

and $\underline{e}$ is an m' -element error component with variance-covariance matrix $\underline{C}$, the structure of which will be discussed later on.

The solution for δ is

$$\underline{\hat{\delta}} = (\underline{A}'\underline{A})^{-1}\underline{A}'\underline{y}. \quad (5)$$

With missing links m' must be adapted and the corresponding rows from $\underline{A}$ eliminated.

Next residuals

$$\hat{e}_{jk} = y_{jk} - (\hat{\delta}_j - \hat{\delta}_k) \quad (6)$$

can be obtained. An informal inspection of residuals might reveal some problems. A complication is that the residuals have different error variances, depending among others on the strength of the links. In order to make inspection easier residuals might be divided by the standard deviations associated with the links. Take for example the link between test j and test k . The variance of y_{jk} equals

$$\begin{aligned} \text{var}(y_{jk}) = n^{-2}{}_{jk} & \left[\sum_i \text{var}(\hat{b}_{ij}) + \sum_{i \neq i'} \sum \text{cov}(\hat{b}_{ij}, \hat{b}_{i'j}) + \sum_i \text{var}(\hat{b}_{ik}) \right. \\ & \left. + \sum_{i \neq i'} \sum \text{cov}(\hat{b}_{ik}, \hat{b}_{i'k}) \right], \end{aligned} \quad (7)$$

with summation over the n_{jk} common items. In contrast to a suggestion by Wright and Bell (1984) the covariances between item parameters within a test are not set equal zero in this equation. Due to the restriction $\sum_i b_i = 0$ applied to each test form, item variances and covariances add to zero within a test. The variance-covariance matrix of item parameter estimates for a particular test is obtained as the inverse of the item information matrix, derived under the restriction $\sum_i b_i = 0$. For CML this matrix is easily obtained, for UML an approximation is given by De Gruijter (1985).

The variance - covariance matrix of $\hat{\delta}$ equals

$$\underline{\underline{S}}_{\hat{\delta}} = (\underline{\underline{A}}' \underline{\underline{A}})^{-1} \underline{\underline{A}}' \underline{\underline{C}} \underline{\underline{A}} (\underline{\underline{A}}' \underline{\underline{A}})^{-1} , \quad (8)$$

where $\underline{\underline{C}}$ is the variance - covariance matrix of the m' variables y_{jk} . The diagonal elements of $\underline{\underline{C}}$ are given in Equation 7, the off-diagonal elements can be obtained in a similar way.

A weighted solution seems more appropriate when the links vary considerably in strength: when the number of common items varies and/or the number of examinees. A WLS-solution is

$$\hat{\delta} = (\underline{\underline{A}}' \underline{\underline{C}}^{-1} \underline{\underline{A}})^{-1} \underline{\underline{A}}' \underline{\underline{C}}^{-1} \underline{\underline{y}} \quad (9)$$

with

$$\underline{\underline{S}}_{\hat{\delta}} = (\underline{\underline{A}}' \underline{\underline{C}}^{-1} \underline{\underline{A}})^{-1} . \quad (10)$$

ITEM PARAMETER CHANGE

Items may become easier after the first administration, as some knowledge about the items becomes available to potential examinees. Items might be disclosed, for example. The effects might be disastrous, especially when the number of old, disclosed items is small.

On the other hand it might be possible that the effects are less dramatic in as far as the IRT-model - the Rasch model in our case - remains valid at each occasion, and that only item difficulties are effected. The simplest model in this case is a model in which all items become easier by the same amount after the first administration, i.e. $b_i^* = b_i + \beta$ for all i (in this case $\beta < 0$) and no further change thereafter. Such a change can be detected, but first let us investigate the disturbing effects of such a change on the outcome of the linking procedure.

Let us consider three thirty-item tests T_1 , T_2 and T_3 , administered in that order. T_2 has six items in common with T_1 . T_3 has three items in common with T_1 and T_2 , and three other items with only T_2 . All three tests have the same difficulty. Neglecting sampling fluctuations $y_{12} = y_{13} = y_{23} = 0$ under the restriction $\Sigma b_i = 0$ for each test.

Next, a shift $\beta = -.2$ is introduced. As a consequence T_2 and T_3 , which have six previously used items, become easier; the overall change in difficulty for both tests equals $-.04$ ($6\beta/30$). The changes could be detected from the values y_{jk} if $y_{12} = -.04$, $y_{13} = -.04$, $y_{23} = 0.0$.

Actually, y_{12} equals .16, y_{13} .16, and y_{23} .1. Consider y_{12} . Let the average difficulty of the common items in T_1 and T_2 before the introduction of shift β be denoted b_c . After introduction of the shift and renormalization $\Sigma b_i = 0$ for T_2 , the average difficulty of the common items in T_2 equals $b_c + \beta - (6\beta/30) = b_c - .16$, i.e. $y_{12} = .16$. The actual values y give us the impression that T_2 and T_3 are more difficult than T_1 , while they are slightly easier. Use of Equation 5 results in a relative easiness equal to $-.127$ for T_2 , and a relative easiness of $-.193$ for T_3 .

When the above facts remain unnoticed, scale drift causes overestimation of abilities at later test occasions. It is easy to detect the shift. The values y are inconsistent: y_{13} is smaller than $y_{12} + y_{23}$. Unfortunately, inconsistencies can also arise from sampling fluctuations and model misspecification. In the present case the inconsistencies are large. The residuals $y - \hat{y}$, obtained from $\hat{y} = \underline{\underline{A}} \hat{\delta}$, are .33, $-.33$, .33. The misspecification of the model affects all

links; detection and elimination of weak links, as suggested by Engelhard and Osberg (1983), clearly is not appropriate in this situation.

Wright and Bell (1984) proposed to check the item-between-link fit for deviant items. This approach can be used in the present case, but a more direct check on shift is possible. Let j and k be two tests, k being the more recent one. Divide the common items in two groups A and B . A contains common items new in test j , B contains the other items. In the A -items shift β occurs between the administration of j and k , in the B -items the item difficulties do not change. When A and B are not empty β is estimated as

$$\hat{\beta} = \bar{b}_{k(A)} - \bar{b}_{j(A)} - (\bar{b}_{k(B)} - \bar{b}_{j(B)}). \quad (11)$$

Each test link results in a different estimate. These have to be combined in order to obtain one final estimate. When this estimate is close to zero, one should verify whether the effect is real. These two problems can be solved, using least squares.

In the presence of shifts the links can be written as

$$y_{jk} = \delta_j - \delta_k - (n'_{jk} / n_{jk}) \beta + e, \quad (12)$$

where n'_{jk} is the number of common items, new in test j . The model can be written in matrix notation, as in Equation 3. The difference is that δ becomes an m -element vector with β as last element. Further, a column is added to A with elements $-n'_{jk}/n_{jk}$. Using the standard deviation of $\hat{\beta}$, the square root of the last diagonal element of $\hat{\Sigma}_{\hat{\beta}}$, one can decide on the presence of shift.

DISCUSSION

Item parameters are not fixed entities. Within the context of IRT-modelling it has been shown that item characteristics can change with time (Bock et al., 1988). The possibility of unidirectional changes has not yet been explored in the context of IRT-based item banking, notwithstanding the fact that such changes might pose a large threat to the integrity of an item bank system.

The possibility of item characteristics which are not invariant, has implications for item parameter estimation. Estimation procedures such as the minimum-chi-square procedure for the Rasch model (Fischer, 1974) and a similar procedure proposed by Van der Linden and Eggen (1986) are not acceptable when parameter change is possible because information on test occasion is eliminated from the data. In other approaches more attention should be given to the detection of item shifts. One possibility based on calibrating Rasch homogeneous tests by a linking procedure, has been described. For simultaneous calibration procedures it is also possible to develop detection techniques.

REFERENCES

- Bock, R.D., Muraki, E., & Pfeifferberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement*, 25, 275-285.
- De Gruijter, D.N.M. (1985). A note on the asymptotic variance-covariance matrix of item parameter estimates in the Rasch model. *Psychometrika*, 50, 247-249.
- Engelhard, G. Jr., & Osberg, D.W. (1983). Constructing a test network with a Rasch measurement model. *Applied Psychological Measurement*, 7, 283-294.

- Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber
- Gilmer, J.S. (1989). The effects of test disclosure on equated scores and pass rates. *Applied Psychological Measurement*, 13, 245-255.
- Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement*, 20, 369 - 377.
- Van der Linden, W.J., & Eggen, T.J.H.M. (1986). An empirical bayesian approach to item banking. *Applied Psychological Measurement*, 10, 345 - 354.
- Wright, B.D., & Bell, S.R. (1984). Item banks: what, why, how. *Journal of Educational Measurement*, 21, 331 - 345.

Manuscript ontvangen 16-10-1989

Definitieve versie ontvangen 4-1-1990

Interne structuur en constructvaliditeit van belevingsschalen voor wiskunde en Engels

Hans Kuhlemeier¹, Huub van den Bergh² en Franke Teunisse¹

¹*Afdeling Beginfase Voortgezet Onderwijs, Cito, Arnhem*

²*Afdeling Taalbeheersing, Rijksuniversiteit Utrecht*

ABSTRACT

This article describes an empirical study on the reliability and validity of the Attitude Scale towards Mathematics (BSW) and the Attitude Scale towards English as a Foreign Language (BSE). The a priori classification of the items in subscales is confirmed by means of confirmative factor analysis. Also significant relations with some external variables (grades, time spent on homework and appreciation of the school subject) are assessed.

1. INLEIDING

In 1987 kwam de 'Belevingsschaal voor Wiskunde (BSW)' ter beschikking (Cito, 1987). Deze schaal stelt wiskundedocenten en onderwijsonderzoekers in staat een belangrijk aspect van het affectieve domein te meten, namelijk de mate waarin leerlingen hun wiskunde-onderwijs als plezierig, moeilijk, interessant en nuttig beleven. Aan de uitgave ging een uitgebreid onderzoek naar de onderwijskundige en psychometrische kwaliteiten van de BSW vooraf. Ter verhoging van de bruikbaarheid zijn de meningen over en ervaringen met de schaal in het onderwijs systematisch geïnventariseerd (Martinot, 1986; Kuhlemeier en Martinot, 1987). Daarnaast is een normeringsonderzoek uitgevoerd (Martinot, Kuhlemeier & Feenstra, 1988). Op de verantwoording van de keuze van de subschalen wordt hier niet nader ingegaan, aangezien daaraan in de hiervoor genoemde publikaties reeds ruime aandacht is besteed.

Naar aanleiding van de relatief goede ontvangst van de BSW in het onderwijs is op het Cito een begin gemaakt met de constructie en validatie van belevingsinstrumenten voor andere vakken dan wiskunde. Op basis van bestudering van relevante literatuur over attitudemeting bij talen is een belevingsschaal ter meting van de beleving van het vak Engels (BSE) ontwikkeld en op bescheiden schaal in het onderwijs uitgeprobeerd (Teunisse, 1990). Naar de relatie tussen de attitudes en motivaties van leerlingen en hun leerprestaties in de moderne vreemde talen is reeds veel onderzoek gedaan. In de diverse overzichtsartikelen op dit terrein worden vrijwel altijd positieve, maar doorgaans niet bijzonder sterke verbanden met leerprestaties gerapporteerd (o.a. Spolsky, 1969; Gardner & Lambert, 1972; Randhawa & Korpan, 1973; Burstall, 1975; Gardner, 1980; Brown, 1981; Krashen, 1981a, 1981b, 1985; Bogaards, 1982; Laine, 1988). Natuurlijk vormt de relatie tussen attitudes en leerprestaties niet de enige legitimering voor de constructie van een attitudeschaal. Voor veel docenten is attitudevorming een doelstelling die alleen al omwille van zichzelf het nastreven waard is (Van den Bergh & Rijlaarsdam, 1988). Wellicht is dit de reden voor het verschijnen van steeds maar weer nieuwe instrumenten ter meting van de attitudes en motivaties van leerlingen ten opzichte van moderne vreemde talen (zie

¹ Adres: Afdeling Beginfase Voortgezet Onderwijs, Cito, Postbus 1034, 6801 MG Arnhem

² Adres: Afdeling Taalbeheersing, Rijksuniversiteit Utrecht, Trans 10, 3512 JK Utrecht

o.a.: Spoelders, 1979; Gardner & Smythe, 1981; Roger, Bull & Fletcher, 1981; Gliksmann, Gardner & Smythe, 1982; Horwitz, 1985; Horwitz, Horwitz & Cope, 1986; Lee, 1986; Pritchard, 1987; Laine, 1988). Tot de onderscheiden attitude-aspecten behoren onder meer de attitude ten opzichte van de docent, de lessen, het leerboek, de doeltaal, de doelgroep en de met het vak geassocieerde angstgevoelens. Naast deze attitudes is de integratieve en instrumentele motivatie om de doeltaal te leren van belang. Met integratieve motivatie wordt bedoeld op de interesse voor het land, de cultuur en het volk van de doeltaal; onder instrumentele motivatie verstaan we de mate waarin het leren van de doeltaal als nuttig ervaren wordt. Evenals in de BSW zijn in de BSE schalen opgenomen over de beleving van plezier, angst en inzet. Daarnaast bevat de BSE schalen ter meting van de instrumentele en integratieve motivatie. In tegenstelling tot de BSW bevat de BSE ook een schaal om de houding ten opzichte van de docent te meten. De veronderstelling is dat een positieve attitude ten opzichte van de docent als persoon, maar ook ten opzichte van zijn/haar didactische kwaliteiten een hogere beleving van het schoolvak Engels tot gevolg zal hebben.

In de BSW en de BSE zijn de onderscheiden belevingsaspecten door middel van subschalen geoperationaliseerd. Onderzoek naar de interne structuur van een belevingsschaal kan meer uitsluitsel verschaffen over de validiteit van het instrument. Daarbij zijn diverse onderzoeksvragen van belang: hoeveel aspecten meet de schaal nu eigenlijk? Wat is de aard van de aspecten en hoe betrouwbaar worden ze gemeten? Hoe is het gesteld met hun onderlinge relaties? De vraag naar het aantal aspecten en hun empirische onderscheidbaarheid is niet alleen een academische kwestie. Immers, naarmate we meer aspecten onderscheiden, neemt de informatieve waarde toe. Bovendien kunnen remediërende maatregelen meer gericht genomen worden wanneer meer gedifferentieerde informatie over de beleving beschikbaar is. Aan de andere kant neemt de bewerkelijkheid van de afname toe naarmate een docent voor meer schalen leerling- of klassescores moet berekenen (vgl. Martinot, 1986). Vanuit overwegingen van spaarzaamheid is terughoudendheid geboden. Zouden de verschillende aspecten van de beleving van wiskunde of Engels psychometrisch niet onderscheidbaar blijken te zijn, dan is één kortere schaal toereikend. Voor de samenstelling van belevingsschalen mogen praktische overwegingen uiteraard slechts een ondergeschikte rol spelen. Van veel groter belang is de mate waarin een veronderstelde interne structuur empirisch onderbouwd kan worden. Daarbij doelen we niet alleen op het aantal en de aard van de gepostuleerde aspecten, maar ook op hun onderlinge relaties.

De validiteit van een belevingsschaal moet niet alleen blijken uit de interne structuur, maar ook uit de relaties met externe variabelen. Daarmee kan inzicht worden verkregen in de convergente en discriminante validiteit als onderdeel van de constructvaliditeit (Cook & Campbell, 1979). Deze komt tot uiting in hoge correlaties tussen belevingsaspecten en variabelen die een vergelijkbaar begrip meten en lage correlaties met variabelen die een niet verwant begrip meten.

In het onderhavige artikel wordt eerst een summierende beschrijving van de belevingsschalen voor wiskunde en Engels gegeven (paragraaf 2). Daarna volgt een analyse van de interne structuur van de beide belevingsschalen (paragraaf 3). Vervolgens worden de relaties met rapportcijfers, tijdbesteding aan huiswerk en de waardering voor de vakken wiskunde, Engels en Nederlands gerapporteerd (paragraaf 4). Het artikel besluit met enkele conclusies en aanbevelingen (paragraaf 5).

2. INSTRUMENTATIE

De Belevingsschaal voor Wiskunde

De BSW bestaat uit vier schalen die elk uit acht uitspraken van het Likerttype zijn opgebouwd. Ter illustratie is hieronder voor iedere schaal een korte omschrijving en een karakteristieke uitspraak weergegeven.

1. *Plezier in wiskunde*. Hierbij gaat het om de vraag of de leerlingen het vak leuk vinden, of de lestijd snel voorbij gaat, of de leerling zelf voor wiskunde zou kiezen, etcetera. Een kenmerkende uitspraak is:

‘Ik vind wiskunde een leuk vak’.

2. *Angst en moeilijkheid*. Deze schaal heeft betrekking op de ervaren moeilijkheid van wiskunde en de ermee samenhangende angstgevoelens. Een typerende uitspraak is:

‘Bij wiskunde ben ik bang om fouten te maken dan bij andere vakken’.

3. *Inzet*. Hier gaat het erom of de leerlingen zich aangesproken voelen door het vak wiskunde en bereid zijn om er tijd aan te besteden. Een karakteristieke uitspraak is:

‘Ik maak wel eens meer huiswerk voor wiskunde dan ik heb opgekregen’.

4. *Nut*. In deze laatste schaal staat het praktisch nut van wiskunde centraal. Gevraagd wordt naar de bruikbaarheid van wiskunde binnen en buiten de school, voor nu en later. Een typerende uitspraak is:

‘Buiten de school heb je weinig aan wat je in de wiskundelessen leert’.

De Belevingsschaal voor Engels

De Belevingsschaal voor Engels bestaat uit zes schalen die ieder vijf à elf uitspraken van het Likerttype bevatten. Hieronder is voor elke schaal een korte omschrijving alsmede een kenmerkende uitspraak weergegeven.

1. *Plezier*. Hier gaat het om het plezier in het vak, de lessen en de leerboeken. Een karakteristieke uitspraak is:

‘Ik heb een hekel aan Engels’.

2. *Angst*. Deze schaal heeft betrekking op gevoelens van nervositeit en onzekerheid die met Engels geassocieerd zijn. Eén van de uitspraken luidt:

‘Tijdens de lessen Engels voel ik me meestal op mijn gemak’.

3. *Inzet*. In deze schaal is de mate waarin de leerling zich voor het vak Engels inzet geoperationaliseerd. Een voor deze schaal typische uitspraak luidt:

‘Ik doe erg mijn best voor het vak Engels’.

4. *Integratieve motivatie (interesse)*. Met deze schaal is getracht de interesse voor het land, de cultuur, het volk en de taal te meten. Een kenmerkende uitspraak luidt:

‘Het is fijn om in de Engelse les iets over Engeland te weten te komen’.

5. *Instrumentele motivatie (nut)*. In deze schaal wordt gevraagd naar het algemene nut van Engels, het nut voor een vervolgopleiding en het nut voor het toekomstige beroep. Een karakteristieke uitspraak luidt:

‘Engels is nodig om later een baan te krijgen’.

6. *Docent*. In deze schaal zijn uitspraken opgenomen over de docent als persoon en als vakdocent. Eén van de kenmerkende uitspraken luidt:

‘De docent(e) Engels legt duidelijk uit’.

In het kader van het vooronderzoek naar de onderwijskundige en psychometrische kwaliteit is de BSE aan ongeveer 550 leerlingen uit het tweede leerjaar voorgelegd. In de steekproef zijn alle getalsmatig belangrijke schooltypen voor voortgezet onderwijs vertegenwoordigd.

3 INTERNE STRUCTUUR

3.1 Methode

In het in 1988 uitgevoerde onderzoek met de BSW is getracht na te gaan in hoeverre de gekozen indeling in schalen op empirische gronden te rechtvaardigen is (Martinot, Kuhlemeier & Feenstra, 1988). In dit kader is de schaal afgenomen aan ruim 8000 leerlingen uit het eerste, tweede en derde leerjaar voortgezet onderwijs. Ter analyse van de interne structuur zijn destijds twee principale componentenanalyses met varimax rotatie uitgevoerd. De eerste is uitgevoerd op de correlaties tussen alle 32 uitspraken van de BSW. Nagenoeg iedere uitspraak laadde substantieel op de eerste factor. Daaruit werd geconcludeerd dat aan de BSW één algemene factor ten grondslag ligt. Vooral de uitspraken uit de schaal 'plezier' bleken in hoge mate representatief voor deze factor. De conclusie lag voor de hand dat de beleving van het vak wiskunde voor de leerlingen vooral een kwestie is van 'leuk' of 'niet leuk'. Omdat de gevonden één-factorstructuur in tegenspraak is met de theoretische notities die tijdens de constructie een rol speelden (Kremers, 1981) is een tweede factoranalyse uitgevoerd. In deze tweede factoranalyse zijn alleen de 24 uitspraken van de schalen angst, inzet en nut in de analyse betrokken. Na varimax rotatie bleken de uitspraken van de drie schalen ieder op een afzonderlijke factor (hoog) te laden. Dit gegeven werd geïnterpreteerd als een rechtvaardiging voor het handhaven van de schalen angst, inzet en nut. Deze analyse (ver)leidde (ons) tot de conclusie dat er naast het algemene plezier in wiskunde sprake is van duidelijk te onderscheiden specifieke aspecten, ondanks de bevindingen van de eerste factoranalyse. Nu lijkt exploratieve factoranalyse niet de meest geëigende techniek om uitspraken te doen over de interne structuur van een psychologisch construct, zeker niet wanneer men op grond van theoretische overwegingen een bepaalde structuur verwacht en wanneer men deze structuur wil vergelijken met alternatieve structuren. Gezien de veelheid van statistisch equivalente oplossingen, kunnen moeilijk 'harde' uitspraken worden gedaan over het aantal subconstructen en hun onderlinge relaties. Ook kunnen vraagtekens worden geplaatst bij de gehanteerde varimax rotatie: de aanname van ongecorrleerde aspecten is waarschijnlijk minder aannemelijk. Een rotatie met oblique factoren had wellicht meer recht gedaan aan de structuur van de data en de theorie omtrent attitudes. Kortom: er resteren twijfels omtrent het aantal te onderscheiden constructen en de relaties ertussen. Vandaar dat wij de BSW-data nogmaals geanalyseerd hebben, maar nu met behulp van covariantie structuuranalyse (Jöreskog & Sörbom, 1986). De LISREL-analyse is niet op de inter-itemcorrelatiematrix uitgevoerd. Om modeltechnische redenen zijn per schaal twee parallelle subtests geconstrueerd, volgens de methode van 'matched random subtests' (Gulliksen, 1950). Immers, dan kunnen de betrouwbaarheden per aspect en de correlaties tussen de latente factoren simultaan geschat worden (zonder de toevoeging van relatief arbitraire restricties). De analyse is uitgevoerd op de matrix met de correlaties tussen de somscores op de parallelle subtests. Omdat het parallelle subtests betreft, zijn in alle analyses twee restricties aan de data opgelegd: gelijke regressies van de beide subtestscores op de latente factor en gelijke errorvarianties. Voor de BSE zijn de analyses op analoge wijze uitgevoerd.

3.2 Te toetsen modellen

Met betrekking tot de interne structuur van de BSW en de BSE zijn vier modellen geformuleerd.

Model 1. Het eerste model postuleert uitsluitend één algemene, ondeelbare factor en kan derhalve als het meest spaarzaam worden beschouwd.

Model 2. In dit model is de beleving van het vak opgebouwd uit de vier (wiskunde) c.q. de

zes (Engels) belevingsaspecten die in de instrumenten vertegenwoordigd zijn. Model 2 veronderstelt dat de aspecten orthogonaal zijn.

Model 3. Dit model is analoog aan model 2, zij het dat de restrictie van orthogonaliteit is opgeheven. In model 3 mogen de correlaties tussen de factoren vrij geschat worden.

Model 4. Het laatste model sluit nog het meest aan bij de eerder gerapporteerde factoranalyses. Dit model omvat één algemene belevingsfactor en drie (wiskunde) c.q. vijf (Engels) meer specifieke factoren; correlaties tussen de latente factoren zijn in dit model niet toegestaan.

3.3 Resultaten

Alvorens in te gaan op de analyseresultaten presenteren we eerst enkele beschrijvende gegevens over de BSW en de BSE (Tabel 1). De uitspraken zijn zo gecodeerd dat een hogere score overeenkomt met een positievere beleving.

Tabel 1. Gemiddelde ($\bar{x}$), laagst haalbare score (min), hoogst haalbare score (max), standaardafwijking (sd), coëfficiënt alpha (alpha), aantal leerlingen (N) en aantal uitspraken (U)

Subschaal	$\bar{X}$	min	max	sd	alpha	N	U
Belevingsschaal voor wiskunde (BSW):							
Plezier	18.62	0	32	8.04	.88	7717	8
Angst	21.09	0	32	6.59	.80	7724	8
Inzet	15.02	0	32	6.36	.74	7707	8
Nut	21.64	0	32	6.02	.80	7666	8
Belevingsschaal voor Engels (BSE):							
Plezier	28.41	8	40	6.41	.78	577	8
Angst	13.43	4	20	3.82	.58	581	4
Inzet	16.83	5	25	4.34	.65	576	5
Nut	41.04	11	55	6.45	.68	573	11
Interesse	40.45	10	50	5.85	.66	579	10
Docent	26.44	7	35	5.50	.74	575	7

Uit de gegevens in Tabel 1 blijkt ondermeer dat er zich bij geen van de subschalen bodem- of plafondeffecten voordoen. Wel zijn de geschatte betrouwbaarheden voor de subschalen van de BSE aan de lage kant. De betrouwbaarheidsschattingen zijn zeker nog te laag voor op grotere schaal in het onderwijs te verspreiden instrumenten. Zeker wanneer we de NIP-code van .90 als maatstaf hanteren.

De vier modellen zijn getoetst onder de aanname van multivariate normaliteit van de verdeling van de geobserveerde scores. Er is gebruik gemaakt van de maximum-likelihood methode, zodat een toetsingsgrootheid verkregen is die bij benadering Chi-kwadraat verdeeld is. De gegevens over de passing van de vier modellen zijn weergegeven in Tabel 2. Bij de interpretatie van de gerapporteerde passingsmaten moeten we bedenken dat het aantal waarnemingen nogal groot is. Vooral bij de BSW zijn extreem veel waarnemingen. Kleine verschillen tussen de geobserveerde en de geschatte correlaties resulteren derhalve al snel in een verhoudingsgewijs grote toename van de toetsingsgrootheid Chi-kwadraat. Daarom zijn ook andere passingsmaten gerapporteerd: de 'goodness of fit' (gfi) en de wortel uit de gemiddelde gekwadraterde residuensom (RMR). Ter controle is een drie-groepenanalyse uitgevoerd. In deze analyse is nagegaan of de interne structuur van de BSW in de drie leerjaren gelijk is. Er zijn drie drie-groepenmodellen getoetst: een model voor parallelle metingen in de drie leerjaren ($X^2 = 1086.07$; $df = 94$; $gfi = .961$; $RMR =$

Tabel 2. De passing van vier modellen ter beschrijving van de interne structuur van de BSW (N = 7170) en de BSE (N = 556)

Model	X ²	df	p	gfi	RMR
Belevingsschaal voor wiskunde (BSW):					
1 Een algemene factor	7991.97	28	<.001	.786	.122
2 Vier orthogonale factoren	10552.62	28	<.001	.732	.331
3 Vier oblique factoren	871.16	22	<.001	.973	.047
4 Een algemene en drie (orthogonale) specifieke factoren	1632.21	25	<.001	.945	.069
Belevingsschaal voor Engels (BSE):					
1 Eén algemene factor	602.17	66	<.001	.849	.096
2 Zes orthogonale factoren	730.04	66	<.001	.774	.214
3 Zes oblique factoren	151.86	51	<.001	.955	.052
4 Eén algemene en vijf (orthogonale) specifieke factoren	195.72	61	<.001	.943	.062

.055), een model voor tau-equivalente metingen ($X^2 = 939.46$; $df = 86$; $gfi = .968$; $RMR = .056$) en een model voor congenerieke metingen ($X^2 = 931.86$; $df = 78$; $gfi = .968$; $RMR = .052$). Gegeven de minieme passingsverschillen tussen deze drie modellen spreken wij een voorkeur uit voor het restrictieve parallelle model. De drie leerjaren verschillen dus niet in de regressies van de waargenomen subtestscores op de latente factoren, noch in de residuele varianties. Opmerkelijk is overigens het (geringe) verschil in passing tussen het tau equivalente en het parallelle model. Omdat de analyse op de correlatiematrix is uitgevoerd zijn de beide modellen niet onderscheidbaar en behoren ze dezelfde resultaten op te leveren. De geobserveerde varianties van de schaalscores zijn immers gelijk aan 1. Het hier geconstateerde verschil zal derhalve te wijten zijn aan afrondingsverschillen.

Voor zowel de BSW als de BSE blijkt het één-factormodel (model 1) minder goed te passen dan de modellen 3 en 4; alleen het orthogonale vier-factorenmodel (model 2) past slechter. Op welke passingsmaat we ons ook baseren (ratio Chi-kwadraat/ df , gfi of RMR), de modellen 1 en 2 mogen verworpen worden. Het derde model past beter dan model 1 (BSW: Chi-kwadraat = 7120.81, $df = 6$; BSE: Chi-kwadraat = 450.31, $df = 15$). Model 3 valt daarom te verkiezen boven model 1. Rest ons de vergelijking van de passing van de modellen 3 en 4. Hoewel deze modellen niet in elkaar genest zijn, geven wij gezien de passingsmaten voor zowel de BSW als de BSE de voorkeur aan model 3.

In Tabel 3 zijn de parameterschattingen, voor zover deze relevant zijn voor de betrouwbaarheidsbepaling, weergegeven (onder de restricties van model 3). De betrouwbaarheid van een subschaal is berekend door de formule van Spearman Brown (voor homogene testverlenging) toe te passen op de geschatte regressiegewichten van de geobserveerde scores op de latente factor. In de analyse is immers uitgegaan van parallelle subtests met ieder een gelijk aantal uitspraken. In Tabel 4 zijn de geschatte correlaties tussen de factoren weergegeven (eveneens geschat onder de restricties van model 3).

Uit Tabel 3 kunnen we concluderen dat de onderscheiden aspecten van de BSW en de BSE redelijk betrouwbaar gemeten kunnen worden. Opmerkelijk is dat de verschillen tussen de in Tabel 1 gepresenteerde betrouwbaarheidsschattingen (Cronbachs α) en de uit theoretische overwegingen te prefereren split-half betrouwbaarheden miniem zijn.

Uit Tabel 4 blijkt dat de factor plezier voor beide belevingsschalen relatief hoog samenhangt met elk van de overige factoren; de angst voor het vak lijkt een tamelijk uniek belevingsaspect te representeren.

Tabel 3. ML-betrouwbaarheidsschattingen voor de schalen van de BSW en de BSE (standaardfouten tussen haakjes)

	Plezier	Angst	Inzet	Nut	Interesse	Docent
Belevingsschaal voor wiskunde (BSW):						
λ	.900 (.008)	.816 (.009)	.749 (.009)	.833 (.008)		
θ	.189 (.003)	.334 (.005)	.440 (.007)	.306 (.005)		
$p^2 = 2\lambda^2 / (2\lambda^2 + \theta_i)$	.896	.799	.718	.819		
Belevingsschaal voor Engels (BSE):						
λ	.771 (.032)	.638 (.036)	.710 (.033)	.663 (.035)	.714 (.033)	.764 (.032)
θ	.405 (.024)	.593 (.036)	.496 (.030)	.561 (.034)	.491 (.029)	.417 (.025)
$p^2 = 2\lambda^2 / (2\lambda^2 + \theta_i)$	.746	.579	.670	.610	.675	.737

Tabel 4. Geschatte correlaties tussen de schalen van de BSW en de BSE (standaardfouten tussen haakjes)

	Plezier	Angst	Inzet	Nut	Interesse	Docent
Belevingsschaal voor wiskunde (BSW):						
Plezier	1.000					
Angst	.613 (.010)	1.000				
Inzet	.818 (.008)	.294 (.014)	1.000			
Nut	.613 (.009)	.308 (.013)	.665 (.011)	1.000		
Belevingsschaal voor Engels (BSE):						
Plezier	1.000					
Angst	.478 (.057)	1.000				
Inzet	.615 (.047)	.178 (.067)	1.000			
Nut	.465 (.056)	.245 (.070)	.470 (.059)	1.000		
Interesse	.624 (.046)	.237 (.066)	.510 (.054)	.576 (.056)	1.000	
Docent	.624 (.043)	.348 (.061)	.416 (.054)	.276 (.061)	.379 (.055)	1.000

4. CONSTRUCTVALIDITEIT

Nu voor interne validiteit van de belevingsschalen voor wiskunde en Engels ondersteuning is gevonden kunnen relaties met externe variabelen worden bestudeerd. Het gaat daarbij om de volgende gegevens, in kaart gebracht met behulp van een achtergrondvragenlijst:

1. De waardering voor de vakken wiskunde, Engels en Nederlands.
2. De tijd die gemiddeld besteed wordt aan het huiswerk voor een wiskundeles, een les Engels en een les Nederlands.
3. Het rapportcijfer voor wiskunde, Engels en Nederlands. Bij de BSW is de leerlingen gevraagd naar hun verwachting omtrent het cijfer voor het eerstvolgende rapport; bij de BSE is het feitelijke cijfer op het vorige rapport opgevraagd.

Met betrekking tot de relaties met deze variabelen zijn, mede op basis van beschikbare literatuur (Kremers, 1981; Teunisse, in druk), enkele verwachtingen geformuleerd:

1. *Vakspecificiteit.* De BSW en de BSE zijn niet bedoeld om de beleving in zijn algemeenheid te meten. De BSW-schalen dienen vooral de beleving van wiskunde te meten en de BSE-schalen vooral die van Engels. Daarom verwachten wij dat de externe relaties

met rapportcijfer, huiswerk en waardering eveneens vakspecifiek zullen zijn. Voor de BSW zou deze verwachting bevestigd worden indien elk van de schalen hoger zou correleren met het rapportcijfer voor wiskunde dan met het rapportcijfer voor Engels en Nederlands. Evenzo dienen de schalen hoger te correleren met de tijdbesteding aan huiswerk voor wiskunde dan met de overeenkomstige tijdbesteding aan Engels en Nederlands en dienen ze hoger te correleren met de waardering voor wiskunde dan met de waardering voor Engels en Nederlands. Voor de BSE geldt een en ander op overeenkomstige wijze.

2. *Plezier en waardering.* Het plezier in een vak, zoals gemeten met een belevingsschaal, en de waardering voor het vak, zoals gemeten met de achtergrondvragenlijst, zijn conceptueel nauw verwante begrippen. We veronderstellen dan ook een relatief hoge, positieve correlatie tussen deze twee factoren. Bovendien dient de waardering voor het vak hoger te correleren met de factor plezier dan met de overige belevingsfactoren.

3. *Inzet en huiswerk.* De tijd die een leerling aan huiswerk zegt te besteden kan worden opgevat als een indicator van de mate waarin de leerling zich voor het vak inspant. Een positieve correlatie tussen de schaal 'inzet' en de tijdbesteding aan huiswerk zou deze aanname bevestigen.

4. *Angst en rapportcijfer.* Laag presterende leerlingen zullen het vak over het algemeen als moeilijk ervaren en mede in verband hiermee relatief veel angst beleven. Verwacht wordt dan ook dat het rapportcijfer voor het vak een positieve correlatie met de belevingsfactor 'angst' zal vertonen. Dit wil zeggen: hoe hoger de prestaties, hoe minder angst voor het vak en vice versa.

5. *Docent.* Het staat buiten kijf dat de docent als persoon en als vakdocent voor het onderwijsleerproces van doorslaggevende betekenis is. We veronderstellen dan ook een positieve relatie tussen de belevingsschaal 'docent' en het rapportcijfer voor Engels, de tijdbesteding aan huiswerk en de waardering van het vak Engels.

4.1 Methode

De samenhang tussen de belevingsfactoren en de externe variabelen is geanalyseerd met behulp van LISREL (Jöreskog & Sörbom, 1986). De analyse is uitgevoerd op de correlatiematrix van de negen externe variabelen en de acht (wiskunde) c.q. twaalf (Engels) subtestscores. Er zijn voor beide belevingsschalen drie modellen getoetst. In al deze drie modellen zijn de gespecificeerde relaties tussen de factoren van de BSW en de BSE conform de restricties van model 3 uit de analyse van de interne structuur. Een verdere invariante restrictie over modellen is de aanname dat de waardering voor de vakken wiskunde, Nederlands en Engels, alsmede de tijd die aan het huiswerk voor deze vakken besteed wordt en het rapportcijfer voor genoemde vakken, volledig betrouwbaar gemeten zijn. De te schatten relaties tussen de waardering, tijd besteed aan huiswerk, rapportcijfer en de factoren van de BSW respectievelijk BSE zullen derhalve onderschattingen van de werkelijke samenhang zijn.

4.2 Te toetsen modellen

Ten aanzien van de constructvaliditeit zijn drie modellen geformuleerd. We werken de modellen alleen uit voor de BSW; voor de BSE geldt een en ander op analoge wijze. Een eerste algemeen punt betreft de richting van de te schatten effecten. Sommigen zijn geneigd om de relaties tussen belevingen en bijvoorbeeld leerprestaties of gedragingen een richting te geven. De een veronderstelt dat beleving een positieve invloed heeft op het gedrag, anderen zijn juist een tegenovergestelde mening toegedaan: gedrag beïnvloedt de beleving. Wij hebben echter geen richting aan deze relatie willen meegeven. Wij beschouwen de relaties tussen belevingen en gedrag als reciproke: gedrag beïnvloedt beleving en beleving beïnvloedt gedrag. Omdat er geen a priori gronden zijn om aan te nemen dat gedrag meer

beïnvloed wordt door beleving dan vice versa, hebben wij in alle modellen deze effecten aan elkaar gelijk gesteld. De drie te toetsen modellen worden hieronder kort toegelicht.

Model 1. In het eerste model worden alleen de relaties tussen de vier BSW-factoren en de drie wiskundefactoren (rapportcijfer, huiswerk en waardering) vrij geschat; de relaties met de overeenkomstige factoren voor Engels en Nederlands zijn op nul gefixeerd.

Model 2. In het tweede model worden alleen de relaties met de drie externe factoren voor Nederlands vrij geschat en zijn geen relaties met de externe factoren voor wiskunde en Engels toegestaan.

Model 3. In het derde model worden alleen de relaties voor het vak Engels vrij geschat; de overige relaties tussen de vier BSW-factoren en de externe factoren voor wiskunde en Nederlands zijn gefixeerd op nul.

Vergelijking van de passing van de drie modellen geeft inzicht in de mate waarin de belevingsschalen vakspecifiek zijn. Overeenkomstig onze verwachtingen zou model 1 beter moeten passen dan de modellen 2 en 3.

Hetzelfde geldt voor de BSE, alleen verwachten we hier dat model 3, waarin de relaties met rapportcijfer, huiswerk en waardering voor het vak Engels vrij geschat mogen worden, beter zal passen dan de modellen 1 en 2.

4.3 Resultaten

De passingsmaten voor de drie getoetste modellen zijn weergegeven in Tabel 5.

Tabel 5. De passing van drie modellen ter beschrijving van de constructvaliditeit van BSW (N = 7170) en de BSE (N = 493).

Model	X ²	df	p	gfi	RMR
Belevingsschaal voor wiskunde (BSW):					
Model 1 (wiskunde)	1484.85	92	<.001	.976	.044
Model 2 (Nederlands)	9177.66	92	<.001	.894	.152
Model 3 (Engels)	9129.57	92	<.001	.895	.151
Belevingsschaal voor Engels (BSE):					
Model 1 (wiskunde)	568.47	153	<.001	.907	.104
Model 2 (Nederlands)	578.78	153	<.001	.906	.102
Model 3 (Engels)	279.13	153	<.001	.921	.053

Uit Tabel 5 blijkt dat conform de verwachtingen model 1 voor de BSW beter bij de data past dan de beide andere modellen, terwijl voor de BSE model 3 beter bij de geobserveerde intercorrelaties past. Hieruit mogen we concluderen dat de BSW-schalen specifiek zijn voor het vak wiskunde en dat de BSE-schalen specifiek zijn voor het vak Engels. In de volgende tabel is het percentage variantie in de externe factoren weergegeven dat niet door de belevingsschalen verklaard wordt.

Tabel 6 laat zien dat de BSW-schalen het rapportcijfer en de waardering voor wiskunde alleszins redelijk voorspellen. Zo kan 40 procent van de variantie in het rapportcijfer door de vier BSW-schalen worden verklaard; voor de waardering van het vak wiskunde bedraagt dit percentage 51 procent. De tijd die de leerling aan wiskundehuiswerk zegt te besteden kan niet of nauwelijks worden verklaard. De onder de aannamen van de modellen 2 en 3 geschatte percentages verklaarde variantie voor Nederlands en Engels kunnen als verwaarloosbaar gering worden aangemerkt. Voor de BSE zijn de percentages

Tabel 6. Percentage niet-verklaarde variantie in de externe factoren (tussen haakjes: standaardfouten)

	Belevingsschaal voor wiskunde (BSW)	Belevingsschaal voor Engels (BSE)
Wiskunde:		
1 Rapportcijfer	59.7 (1.1)	98.1 (6.3)
2 Huiswerk	91.3 (1.6)	96.4 (5.9)
3 Waardering	48.7 (.9)	98.7 (6.3)
Nederlands:		
4 Rapportcijfer	99.5 (1.6)	97.9 (6.1)
5 Huiswerk	98.7 (1.6)	99.1 (5.9)
6 Waardering	99.1 (1.7)	97.8 (6.3)
Engels:		
7 Rapportcijfer	99.8 (1.7)	73.9 (5.5)
8 Huiswerk	99.7 (1.6)	90.0 (5.8)
9 Waardering	98.5 (1.6)	63.6 (4.5)

Tabel 7. Schattingen van de relaties tussen de belevingsfactoren en de externe factoren voor de BSW en de BSE, onder de restricties van model 3 (tussen haakjes: standaardfouten).

	Plezier	Angst	Inzet	Nut	Interesse	Docent
Belevingsschaal voor wiskunde (BSW):						
1 Rapportcijfer	.091 (.006)	.341 (.006)	.033 (.007)	.026 (.006)		
2 Huiswerk	-.005 (.005)	-.131 (.006)	.104 (.007)	.016 (.006)		
3 Waardering	.274 (.006)	.145 (.006)	.157 (.007)	.104 (.006)		
Belevingsschaal voor Engels (BSE):						
1 Rapportcijfer	.045 (.027)	.277 (.030)	.001 (.028)	.046 (.029)	.021 (.028)	-.013 (.026)
2 Huiswerk	-.002 (.026)	-.139 (.029)	.108 (.027)	.037 (.029)	.016 (.028)	.008 (.025)
3 Waardering	.201 (.028)	.129 (.030)	.063 (.028)	.055 (.030)	.101 (.029)	.075 (.026)

aanmerkelijk lager: met behulp van de zes BSE-schalen kan een kwart van de variantie in het rapportcijfer voor Engels worden verklaard; de tijdbesteding aan huiswerk wordt voor ruim éénderde voorspeld.

In de bovenstaande tabel zijn de schattingen van de relaties tussen de belevingsschalen en de externe factoren weergegeven, onder de restricties van het best passende model.

De gegevens in Tabel 7 laten zich als volgt beschrijven:

1. Over het algemeen zijn de reciproke effecten tussen de beleving van het vak (wiskunde of Engels) en de externe variabelen (rapportcijfer, huiswerk en waardering) relatief klein.
2. Overeenkomstig onze verwachting hangt de waardering voor het vak positief samen met het beleefde plezier. Bovendien is dit verband sterker dan dat met de overige belevingsschalen.
3. Zoals verwacht bestaat er een positieve relatie tussen de tijd die een leerling aan huiswerk besteedt en de inzet voor het vak zoals gemeten met de BSW en de BSE.
4. In overeenstemming met onze verwachting vertoont het rapportcijfer voor het vak een positieve correlatie met de belevingsfactor 'angst'. Dit wil zeggen: naarmate een leerling minder angst ervaart en het vak als minder moeilijk percipieert, zijn de leerprestaties hoger en vice versa.
5. In afwijking van hetgeen wij verwachtten, blijkt de BSE-schaal 'docent' geen relatie te

vertonen met de leerprestaties voor het vak Engels en de tijd die de leerling aan zijn huiswerk zegt te besteden. Alleen de relatie met de waardering voor het vak verschilt significant van nul; absoluut gezien is het effect evenwel verwaarloosbaar gering.

5. CONCLUSIES EN AANBEVELINGEN

Op basis van de in dit artikel gepresenteerde analyses van de interne structuur van de belevingsschalen voor wiskunde en Engels kunnen drie conclusies worden getrokken. De eerste conclusie betreft het aantal te onderscheiden belevingsaspecten. Zowel in het geval van de BSW als de BSE bleek het éénfactormodel de oorspronkelijke correlatiematrix maar matig te kunnen reproduceren. De beleving van een schoolvak als wiskunde of Engels kan eerder als een meerdimensioneel dan als een ééndimensioneel construct worden beschouwd. Een tweede conclusie betreft de relaties tussen de factoren. Op grond van de eerder uitgevoerde factoranalyse veronderstelden we één centrale, door het plezier in het vak gedomineerde belevingsfactor en daarnaast een aantal minder belangrijke specifieke factoren als angst, nut en inzet. De thans gepresenteerde gegevens doen vermoeden dat alle onderscheiden factoren voor de beleving van ongeveer even groot belang zijn. De oblique meerfactorenmodellen bleken duidelijk beter te passen dan de orthogonale. Hieruit kunnen we concluderen dat de onderscheiden factoren in de beleving van de leerlingen onderling samenhangen. Wel bleek de factor angst zowel bij de BSW als de BSE relatief laag met de overige factoren te correleren. De met het vak geassocieerde angst kan dus als een relatief onafhankelijk belevingsaspect worden beschouwd. Een derde conclusie betreft het onderscheid tussen de integratieve en de instrumentele motivatie om Engels te leren. De interesse voor het land, cultuur en taal en het gepercipieerde nut kunnen als vrij duidelijk te onderscheiden aspecten van de beleving van een taal als Engels worden beschouwd.

Onze verwachtingen ten aanzien van de relaties van de belevingsschalen met de externe variabelen rapportcijfer, huiswerk en waardering werden op één uitzondering na bevestigd. Alleen de BSE-schaal 'docent' bleek onverwacht niet substantieel aan de externe variabelen gelieerd. Wel bleken de vier BSW-schalen aanzienlijke percentages variantie in het rapportcijfer en de waardering voor het vak te verklaren. Bij het hoge percentage verklaarde variantie in het rapportcijfer voor wiskunde en de geconstateerde reciproke relaties met de beleving van dit vak dient echter een kritische kanttekening geplaatst te worden. In de analyse van de BSW is gebruik gemaakt van het cijfer dat de leerling voor het eerstvolgende rapport verwacht. De geobserveerde correlaties tussen de BSW-schalen en het rapportcijfer voor wiskunde zouden waarschijnlijk lager zijn geweest indien wij evenals bij de BSE 'echte' rapportcijfers hadden verzameld. Zo zouden leerlingen met een negatieve beleving van het vak hun toekomstige rapportcijfer kunnen onderschatten; leerlingen met een uiterst positieve beleving zouden tot een te optimistische inschatting kunnen zijn gekomen. Door deze veronderstelde systematische over- en onderschatting worden de reciproke effecten tussen belevingsschaal en rapportcijfer waarschijnlijk geflatteerd, waardoor ook het percentage onverklaarde variantie onderschat wordt. Het wekt dan ook geen verbazing dat dit percentage voor de BSE beduidend hoger is.

De tijd die een leerling aan huiswerk zegt te besteden kan nauwelijks met de belevingsschalen verklaard worden. Ook de relatie tussen de tijdbesteding aan huiswerk en de mate van inzet zoals gemeten met de belevingsschalen is uiterst zwak. Achteraf gezien wekt deze uitkomst niet zoveel verwondering. Er zijn vakken die meer huiswerk meebrengen dan andere en er zijn leraren die meer huiswerk geven dan anderen, zonder dat

dit van invloed hoeft te zijn op de beleving. Er zijn leerlingen die zich extra inspannen voor een vak dat ze moeilijk vinden en het vak wellicht daardoor minder leuk vinden. En hoeveel leerlingen hebben een hekel aan een vak juist omdat ze er veel huiswerk voor moeten maken? Voor dit soort variabelen is niet gecontroleerd en valt wellicht niet goed te controleren.

Al met al geven de resultaten vertrouwen in de interne structuur en de constructvaliditeit van de BSW en de BSE. Op basis van de resultaten van het vooronderzoek met de BSE verdient het aanbeveling door te gaan met de constructie van instrumenten ter meting van de beleving van Engels en andere talen. Vooralsnog kan daarbij worden uitgegaan van een zestal schalen.

LITERATUUR

- Bogaards, P. (1982). *Moderne vreemde talen op school. Een studie betreffende de determinanten van succes bij het leren van Frans in de brugklas*. Amsterdam: Vrije Universiteit.
- Bergh, H. van den & Rijlaarsdam, G. (1988). De relatie tussen lees- en schrijfattituden en lees- en schrijffprestaties. In: R. Jaarsma (red.). *Taalvorming, taaldrukken en kinderboeken: verslag van het congres 'Ieder kind is een kinderboekenschrijver', gehouden op 12 en 13 november 1987*. Den Haag: NBLC, 17-29.
- Brown, H.D. (1981). Affective factors in second language learning. In: Alatis, J.E., Altman, H.B. & Alatis, P.M. (Eds.). *The second language classroom: Directions for the eighties*. New York: Oxford University Press, 113-129.
- Burstall, C. (1975). Factors affecting foreign-language learning: a consideration of some recent research findings. *Language Teaching & Linguistics: Abstracts*, 8, 1, 5-25.
- Cito (1987). *Belevingsschaal voor wiskunde: Handleiding*. Arnhem: Cito.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally college publishing company.
- Gardner, R.C. (1980). On the validity of affective variables in second language acquisition: conceptual, contextual and statistical considerations. *Language Learning*, 30, 255-271.
- Gardner, R.C. & Lambert, W.E. (1972). *Attitudes and motivation in second language learning*. Rowley, Mass.: Newbury.
- Gardner, R.C. & Smythe, P.C. (1981). The development of the Attitude/Motivation Test Battery. *The Canadian Modern Language Review*, 37, 510-525.
- Gliksmann, L., Gardner, R.C. & Smythe, P.C. (1982). The role of integrative motive on students' participation in the French classroom. *The Canadian Modern Language Review*, 38, 625-649.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Horwitz, E.K. (1982). Using student beliefs about foreign language learning and teaching in the foreign language methods course. *Foreign Language Annals*, 18, 333-338.
- Horwitz, E.K., Horwitz, M.B. & Cope J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70, 125-132.
- Jöreskog, K.G. & Sörbom, D. (1986). *LISREL: Analysis of linear structural relationships by the method of maximum likelihood (Version VI)*. Uppsala: University of Uppsala.
- Krashen, S.D. (1981a). Effective second language acquisition: insights from research. In: Alatis, J.E., Altman, H.B. & Alatis, P.M. (Eds.). *The second language classroom: Directions for the eighties*. New York: Oxford University Press, 97-109.
- Krashen, S.D. (1981b). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Krashen, S.D. (1985). *The input hypothesis: Issues and Implications*. New York: Longman.
- Kremers, E.J.J. (1981). *Affectieve doelstellingen in het onderwijs en een toepassing voor het vak wiskunde*. Cito-memo Nr. 277. Arnhem: Cito.
- Kuhlemeier, J.B. & Martinot, M.J. (1987). *Verslag van het enquête-onderzoek met de Wiskundebelevingsschaal*. Interne Documentatie Nr. 240. Arnhem: Cito.
- Laine, E.J. (1988). *The affective filter in foreign language learning and teaching*. Jyväskylä Cross-Language Studies No.15. Jyväskylä: Department of English.

- Lee, B. (1986). 'French is hard because it's a different language'. *British Journal of Language Teaching*, 24, 71-76.
- Martinot, M.J. (1986). *Verslag van het vooronderzoek met de Wiskundebelevingsschaal: De interviewgegevens*. Interne Documentatie Nr. 206. Arnhem: Cito.
- Martinot, M.J., Kuhlemeier, J.B. en Feenstra, H.J.M. (1988). Het meten van affectieve doelen: De validering en normering van de Belevingsschaal voor Wiskunde (BSW). *Tijdschrift voor Onderwijsresearch*, 13, 65-76.
- Pritchard, R.M.O. (1987). Boys' and girls' attitudes towards French and German. *Educational Research*, 1, 65-72.
- Randhawa, B.S. & Korpan, S.M. (1973). Assessment of some significant affective variables and the prediction of achievement in French. *Canadian Journal of Behavioral Science*, 5, 24-33.
- Roger, D., Bull, P. & Fletcher, R. (1981). The construction and validation of a questionnaire for measuring attitude towards learning foreign languages. *Educational Researcher*, 33, 223-230.
- Spoelders, M. (1979). Taalattitudemeting in theorie en praktijk: TAS. *Pedagogisch Tijdschrift*, 4, 76-85.
- Spolsky, B. (1969). Attitudinal aspects of second language learning. *Language Learning*, 29, 271-283.
- Teunisse, F. (1990). *De beleving van Engels: verslag van een vooronderzoek naar de psychometrische en onderwijskundige kwaliteit van de Belevingsschaal voor Engels (BSE)*. Arnhem: Cito.

Manuscript ontvangen 6-11-1989

Definitieve versie ontvangen 2-1-1990

Intuïtieve, empirische en theoretische redeneringen

H. Vos¹

Faculteit der Elektrotechniek, Universiteit Twente

ABSTRACT

The subject-matter of our empirical research is content effects in three-valued students reasoning. We compare reasoning in a simple case to reasoning in a complex case. We find that students reasoning is intuitive and that the argumentation of students reasoning correctly differs significantly in both cases. We distinguish intuitive, empirical and theoretical reasoning. We discuss the consequences of our findings for problem solving and for teaching empirical reasonings as a heuristic. Our conclusion is that teaching problem solving can be effective only in a complex learning situation. At last we review and compare the characteristics of intuitive, empirical and theoretical reasoning.

1. INLEIDING

We gaan uit van twee observaties in het onderwijs. De eerste is deze. Docenten hoor je soms zeggen dat studenten niet kunnen redeneren. We zullen ons hier concentreren op die gevallen waarin bedoeld wordt dat studenten niet logisch met de vakinhoud omspringen in een complexe probleemsituatie. Ze komen er niet uit en redeneren daarbij niet correct. We willen hier onderzoeken of ze er niet uitkomen *doordat* ze niet kunnen redeneren.

Het gaat er dan enerzijds om te bepalen hoe studenten in het algemeen redeneren en of dat correct gaat, anderzijds hoe ze redeneren met betrekking tot de vakinhoud in die probleemsituaties waarin ze er niet uit komen. We onderzoeken dus effecten van de inhoud van het redeneren op de redeneringen zelf. Tevens moet onderzocht worden wat de relatie is tussen redeneren en het oplossen van problemen.

We onderzoeken het redeneren in een driewaardige logika. Hierin zijn er drie mogelijkheden waarmee rekening gehouden moet worden, namelijk iets is waar (symboolisch aangegeven met 1), iets is niet waar (0), of het is niet te zeggen of iets waar is of niet, het is onzeker (?). Dit in tegenstelling tot de twee-waardige logika, waarin iets alleen maar waar of niet waar kan zijn. De reden dat we een driewaardige logika onderzoeken is deze, dat het in probleemsituaties van groot belang is om te kunnen omgaan met uitspraken waarvan het nog niet te zeggen is of ze waar zijn of niet.

In probleemsituaties zijn er onbekende elementen in het spel. De eerste fase van het oplossen van het probleem bestaat hieruit dat geanalyseerd wordt wat het probleem eigenlijk is, dus wat je wilt bereiken en nog niet weet, en wat je uitgangsgegevens zijn die je wel weet (vergelijk Mettes en Pilot, 1980). Vervolgens vindt het zoeken naar de ontbrekende elementen daartussenin plaats. Uitgaande van de gegevens kan men verder redeneren. Er zijn meestal veel bruikbare gegevens en daarmee verbonden andere kennis om vanuit te gaan. Ook zijn vele redeneringen mogelijk. De meeste daarvan zullen niet tot het gewenste doel leiden.

Je weet van tevoren dus niet of je het probleem wel kunt oplossen. Je weet ook nog niet of dat doel wel bereikt kan worden en toch moet je daarop uitkomen. Je moet dus vanuit

een onzekere eindsituatie kunnen nagaan wat je moet weten om die situatie zeker te maken (wel oplosbaar, niet oplosbaar). We noemen dit terugredeneren (zie verder par. 5). Om op het doel uit te komen moet men ook terug kunnen redeneren.

Om te kunnen concluderen of op dit moment de zon in Enschede schijnt moet ik weten of er geen wolk voor de zon zit en of de zon op is. (We zien hier even af van de mogelijkheid om direct waar te nemen of de zon schijnt). De uitgangssituatie is hier dat ik niet weet of de zon schijnt of niet; het doel is om te weten te komen of hij wel schijnt of niet. Het gaat er om een conclusie die onzeker is om te zetten in een conclusie die wel zeker is (positief dan wel negatief). Om het doel te bereiken moeten dan eerst de premissen onderzocht worden. Op welke wijze dat gaat, zullen we verderop zien. De derde mogelijkheid van de driewaardige logika (het onzekere element ?) is dus van groot belang bij het terugredeneren dat nodig is om op het doel (of een tussendoel) uit te komen in een probleemoplossingsproces. Dit alles lijkt triviaal, en houdt verband met de volgende observatie.

De tweede observatie is de volgende. Docenten proberen soms fouten die studenten in een complexe situatie maken, te voorkomen door eenvoudige voorbereidende oefeningen aan te bieden. Het blijkt dan vaak dat studenten iets geïsoleerd wel kunnen, en zelfs triviaal vinden, wat in een complexe situatie niet lukt. Wij vermoeden dat dit komt doordat studenten een *intuïtieve redenering* hebben ontwikkeld, die wel werkt in eenvoudige situaties (en daar triviaal wordt gevonden) maar niet of nog niet in complexere probleemsituaties. Hierbij verstaan wij onder een intuïtieve redenering een succesvolle (dus correcte) redenering die gebaseerd is op ervaring in het redeneren maar die niet expliciet geleerd is door de proefpersonen en ook niet consistent onder woorden gebracht kan worden (vergelijk de Groot, 1985).

Een intuïtieve redenering moet dus geleerd zijn door een onbewust leerproces: Hij is het resultaat van wat Reber (1989) 'implicit learning' noemt. Volgens Reber leidt dit tot 'tacit knowledge' die impliciet gebruikt kan worden om problemen op te lossen. Wij zullen zien in hoeverre dit waar is.

In dit verband is er ook een andere interessante kwestie. Het is bekend dat ook het redeneren van experts voor een gedeelte intuïtief is. Voor het construeren van onderwijs en van expertsystemen probeert men de redeneringen van de expert te expliciteren. De meest voorkomende, correcte, consistente, geëxpliciteerde redenering noemen we hier een *empirische redenering*. Zo'n empirische redenering is op te vatten als een gewenst handelingsverloop in de zin van Mettes en Pilot (1980). We kunnen deze empirische redenering als een heuristiek gebruiken om op een systematische manier problemen op te lossen.

In probleemoplosonderwijs wordt nu van deze empirische redeneringen gebruik gemaakt om studenten (of leerlingen) te leren systematisch problemen aan te pakken. We zien dan twee problemen optreden. Ten eerste zijn de studenten moeilijk te motiveren om heuristieken te oefenen. Ze willen het op hun eigen manier doen. Zoals bekend is uit het probleemoplosonderwijs (zie bijvoorbeeld Van den Berg, 1983) gebruiken studenten de geleerde heuristiek in veel gevallen niet. Ten tweede is uit het probleemoplosonderwijs bekend dat de transfer van heuristieken naar andere domeinen van toepassing, gering is (vergelijk Mettes en Gerritsma, 1986). Waarschijnlijk hangt dit samen met het eerste punt. Soms is er transfer zonder dat de prestaties in het vak toenemen (loc.cit.). De vraag is nu waarom het onderwijs van Mettes en Pilot wel succesvol is gebleken.

We hebben tot nu toe gesproken over intuïtieve en empirische redeneringen. Zo'n intuïtieve of empirische redenering hoeft niet noodzakelijkerwijs overeen te komen met de theoretische redenering zoals die in de logika bestudeerd wordt. Ons onderzoek is erop gericht te achterhalen hoe de studenten redeneren, hoe deze redeneringen zich verhouden tot theoretische redeneringen, en wat de verklaring is voor de genoemde observaties en de besproken kwesties.

Ons onderzoek beperkt zich tot de redeneringen die te maken hebben met het trekken van een conclusie gebaseerd op een conjunctie. In par. 2 geven we de theoretische redeneringen die een rol spelen bij het trekken van een conclusie. Daar wordt ook beschreven op welke wijze wij de redeneringen van studenten onderzoeken. In par.3 wordt summier beschreven op welke wijze het empirisch onderzoek is uitgevoerd om na te gaan hoe studenten in een simpel geval (in alledaagse gevallen die het weer betreffen) redeneren, en wat de resultaten van dat onderzoek zijn. In par. 4 wordt beschreven hoe deze redenering verloopt in een complexer geval in een elektrotechnisch vak, namelijk bij vragen over het functioneren van de oscilloscoop. Omdat voor ons inhoudseffecten op de redeneringen van belang zijn, en vanwege de verstrekkende consequenties van ons onderzoek, worden de redeneringwijzen tamelijk uitgebreid besproken. In par. 5 volgen onze conclusies. In par. 6 bespreken we de relatie tussen probleemoplossen en redeneren nader, en geven een verklaring voor de observaties op basis van de resultaten van ons onderzoek. Tevens leggen we daar een relatie tussen de ervaringen in probleemoplosonderwijs en het onderwijzen van empirische redeneringen. Tenslotte volgt in par. 7 een vergelijking van de verschillende redeneringen ten behoeve van verder onderzoek.

2. THEORETISCHE REDENERINGEN
BIJ HET TREKKEN VAN EEN CONCLUSIE

Wij hebben onderzocht hoe studenten met een drie-waardige logika omgaan in simpele redeneringen die het weer betreffen. Er zijn verschillende driewaardige logika's mogelijk (Rescher, 1969). Het gaat ons om de theoretische redeneringen die tot een conclusie leiden wanneer we een bewering doen die gebaseerd is op twee premissen die samen een conjunctie vormen. Het trekken van een conclusie in de hier van toepassing zijnde driewaardige logika verloopt volgens de kolommen van de driewaardige waarheidstabel van Tabel 1.

Tabel 1. De conclusie in de drie-waardige waarheidstabel

Elke kolom stelt een theoretische redenering voor.
1 betekent: is waar
0 betekent: is niet waar
? betekent: het is niet te zeggen of het waar is, het is onzeker

premissie 1:	1	1	0	0	0	?	1	?	?
premissie 2:	1	0	1	0	?	0	?	1	?
conclusie:	1	0	0	0	0	0	?	?	?

Elke kolom stelt een theoretische redenering voor die als volgt te verwoorden is. We nemen als voorbeeld de vijfde kolom: Als de eerste premissie onjuist is en van de tweede premissie is het onzeker of hij juist is of niet, dan is de conclusie onjuist.

De vragen die aan de studenten gesteld werden waren van het volgende type. Eerst wordt er een bewering gedaan waarvan gegeven wordt dat hij juist is, namelijk: 'De zon schijnt als er geen wolk voor de zon zit en de zon op is.' Daarna wordt gevraagd of dan een andere bewering ook juist is, bijvoorbeeld: 'De zon schijnt want het is acht uur en volgens het KNMI zijn er wolkenvelden.' Tevens wordt gevraagd het antwoord te motiveren.

De student moet hier de gegeven argumenten (gegevens genoemd) voor de tweede bewering in verband brengen met de premissen van de eerste bewering. Een correct antwoord met een volledige argumentatie gaat in dit geval als volgt. Het is acht uur dus je

weet niet of de zon op is of niet, dit hangt van het jaargetijde af. Premisse 2 (!) is dus onzeker (premissie 2: ?). Er zijn wel wolkenvelden maar je weet niet of er een wolk precies voor de zon zit. Premisse 1 is dus ook onzeker (premissie 1: ?). Volgens de waarheidstabel is de conclusie nu ook onzeker (conclusie: ?): Je weet dus niet of de zon schijnt.

Hier wordt de argumentatie voor de tweede bewering als gegeven genomen, deze wordt in verband gebracht met de premissen van de eerste bewering, en tenslotte wordt een conclusie getrokken over de tweede bewering zelf ('de zon schijnt'), die als conclusie in de eerste bewering was opgenomen.

In dit geval komen de gegevens in de tweede bewering in omgekeerde volgorde voor als de premissen in de eerste redenering. We duiden dit aan met: invers. We noteren deze theoretische redenering nu als volgt: $\text{invers}?? \rightarrow ?$. Hij komt overeen met de laatste kolom uit Tabel 1. Theoretisch bestaat concluderen uit de 9 typen van redeneringen volgens de negen kolommen van Tabel 1, waarbij de premissen ook nog verwisseld mogen worden, zoals blijkt door nadere analyse van Tabel 1.

Tenslotte nog een opmerking over de bovenstaande redenering. Dit was een redenering uitgaande van de gegevens. Er werd echter niet aangegeven hoe de redenering daarmee uitkomt op de tweede bewering. Dit lijkt 'vanzelf' te gaan. Een volledige theoretische redenering die tot de oplossing van het gestelde probleem leidt, moet echter een element van terugredeneren bevatten (zie verder par. 5) en ziet er zo uit: Om te weten of de zon schijnt (tweede bewering), moet je weten of er geen wolk voor de zon zit en of de zon op is. Om te weten of er geen wolk voor de zon is, moet je naar de gegevens kijken, en op basis daarvan eveneens een conclusie trekken, nu over de juistheid van de eerste premisse. Evenzo voor de tweede premisse. Beide uitspraken kunnen nu volgens Tabel 1 gecombineerd worden tot een conclusie over de juistheid van de tweede bewering.

We onderzoeken alleen de wijze waarop de laatste conclusie wordt getrokken. Als de conclusie over de juistheid van een premisse niet goed wordt uitgevoerd, leidt dit tot een irrelevant of onjuist argument van de student. Op grond daarvan kan echter wel een juiste redenering worden uitgevoerd.

We vragen dus naar de juistheid van een bewering en naar de argumenten daarvoor in een driewaardige logika. We zijn van mening dat op deze wijze aanwijzingen verkregen worden over de manier waarop studenten redeneren, en met name over hun capaciteiten om terug te redeneren vanuit het doel (zie verder par. 5).

3. REDENERINGEN IN EEN EENVOUDIGE SITUATIE

De onderzoeksopzet was als volgt. Eerstejaars studenten werden op een praktikum getoetst over de werking van een meetinstrument, de oscilloscoop. Dit is een soort beeldscherm waarop elektrische signalen zichtbaar worden gemaakt. De werking van de oscilloscoop was aan de orde gekomen in een vorig praktikum, en nodig in het praktikum waarin aan het begin de toets werd afgenomen. Aangezien deze toetsresultaten (zie verder par. 4) in eerste instantie door redeneerproblemen niet te interpreteren waren, is na enkele weken een vervolgonderzoek gedaan over redeneren aan de hand van het weer.

We hebben 42 studenten de opgave gegeven om voor 5 verschillende beweringen de juistheid na te gaan, en de argumenten daarvoor te geven (vergelijk Vos, 1989). De beweringen, de gegevens waarop ze gebaseerd zijn en de daarbij behorende theoretische redenering staan in Tabel 2 weergegeven.

De manier waarop studenten blijken te redeneren kunnen we indelen in een aantal klassen (vergelijk het aanhangsel). Het antwoord kan correct zijn (A+) of onjuist (A-). Verder kan de motivatie van het antwoord volledig zijn, met beide argumenten, of

Tabel 2. Beweringen, gegevens en bijbehorende theoretische redeneringen

1. De zon schijnt want er is één wolk aan de hemel en het is een kwartier voor zonsopgang.		70→0
2. De zon schijnt want het weer is helder en het is zes uur 's avonds.		17→?
3. De zon schijnt want het is acht uur en volgens het KNMI zijn er wolkenvelden.	invers	??→?
4. De zon schijnt want het is wolkenloos en de zon is net onder.		10→0
5. De zon schijnt want het regent een beetje en het is juli.		??→?

onvolledig, met slechts één argument of geheel zonder. Een argument kan op grond van de gegevens juist, onjuist of irrelevant zijn. Ten slotte kan het dan nog zijn dat de redenering op grond van die argumenten juist is (L+) of onjuist (L—).

Een juiste redenering hoeft niet noodzakelijk tot een juist antwoord te leiden. Immers, op grond van een onjuist argument kan door een juiste redenering een onjuist antwoord verkregen worden. Een voorbeeld hiervan vinden we bij de student die als antwoord geeft: De bewering (nummer 3) is onjuist want bij wolkenvelden zijn er wolken voor de zon. Deze redenering valt in de klasse A—L+.

Onder een onjuiste redenering verstaan we hier een redenering waarbij iemand op grond van irrelevante argumenten (juist dan wel onjuist) tot een antwoord komt (juist dan wel onjuist), of waarbij op grond van een juist argument iemand tot een onjuist antwoord komt dan wel via een onjuist argument tot een juist antwoord dat niet door de waarheidstabel is toegestaan.

Dat het niet eenvoudig is tot een consistente indeling van redeneringen te komen blijkt uit het volgende voorbeeld. Een student geeft (weer bij bewering 3) het antwoord: Onjuist want er kan een wolk voor de zon hangen. Hier is het argument juist en het antwoord fout. Uit vergelijking met de andere nummers blijkt deze student de hele vraagstelling op een andere wijze te interpreteren. Hij gaat niet na of de bewering juist is maar of de redenering die achter de bewering steekt, juist is.

De redeneringen betreffende het weer zijn door drie personen ingedeeld in klassen (bij de oscilloscoop door twee personen). Door de ingewikkelde classificaties en het geringe aantal beoordeelaars is er geen goede maat voor de interbeoordelaarsovereenstemming te rapporteren. De verschillen tussen de classificaties waren zeer gering. Over minder dan 4 % van het totale aantal classificaties bestond verschil van mening. De verschillen zijn te herleiden tot interpretatieverschillen met betrekking tot de vermoedelijke redenering van de student die niet altijd met zekerheid achterhaald kan worden.

De verdeling van de redeneringen en antwoorden in klassen ziet er als volgt uit (zie Tabel 3). Slechts 7 % van de redeneringen is onjuist, en slechts 9% van de antwoorden fout.

Tabel 3. Verdeling van de redeneringen en antwoorden in klassen

L+ = juiste redeneringen.					
L— = onjuiste redeneringen.					
A+ = juiste antwoorden.					
A— = onjuiste antwoorden.					
klassen			aantallen (het weer)		
L+A+	L+A—	L+	183	13	196
L—A+	L—A—	L—	8	6	14
A+	A—	N	191	19	210

Wanneer we de klasse L+ van correcte redeneringen bezien vinden we de volgende soorten argumenten (zie Tabel 4 voor de verdeling in procenten. De absolute aantallen zijn in een aanhangsel vermeld).

Tabel 4. Verdeling van de soorten argumenten bij correcte redeneringen betreffende het weer (in %)

theor.redenering	juiste argumenten		onjuist/ irrel.arg.
	1 arg.	2 arg.	
$?0 \rightarrow 0$	83	13	5
$1? \rightarrow ?$	83	15	3
inv $?? \rightarrow ?$	22	69	8
$10 \rightarrow 0$	83	12	5
$?? \rightarrow ?$	26	62	13

Het blijkt uit deze Tabel dat het aantal argumenten dat de meeste studenten geven afhangt van de redenering die nodig is. We kunnen nog wat gedetailleerder kijken naar de argumenten die de meeste studenten geven bij hun redenering. Bij de eerste redenering bijvoorbeeld wordt als argumentatie meestal gegeven: ‘De zon is niet op’, dat is de onware premisse (0). In Tabel 5 zijn deze meest gebruikte argumentatiewijzen weergegeven.

Tabel 5. De meest gebruikte argumentiewijzen bij correcte redeneringen met juiste argumenten. (Het weer.)

theor.redenering	argumentatie		percentage
$?0 \rightarrow 0$	alleen onware premisse	0	(83%)
$1? \rightarrow ?$	alleen onzekere prem.	?	(83%)
inv $?? \rightarrow ?$	beide premissen	??	(69%)
$10 \rightarrow 0$	alleen onware premisse	0	(83%)
$?? \rightarrow ?$	beide premissen	??	(62%)

In de gevallen ?? hanteren de meeste studenten beide argumenten expliciet en bewust. Blijkbaar hebben ze twee argumenten ter beschikking en hanteren ze die ook als dat nodig is. In 3 van de 5 gevallen wordt maar één argument gegeven. In deze gevallen moet dus het tweede argument onbewust functioneren. Al naar gelang het geval kan dit tweede argument een onzekere (?), of een ware (1) premisse zijn.

Op grond van deze gegevens zouden de (hypothetische) redeneringen van de meeste studenten als volgt verlopen (zie Tabel 6). We hebben niet onderzocht of de volgorde in deze redeneringen empirisch is terug te vinden. De aangegeven volgorde is van structurele aard, dat wil zeggen hij volgt uit de gegeven argumentaties. Een andere volgorde is niet consistent met de gegeven argumentaties.

Tabel 6. Redeneringen van de meeste studenten in een bekende situatie

- | |
|--|
| 1. De premissen die waar zijn (1) worden onbewust opgenomen. |
| 2. Is één van de premissen niet waar (0), dan is de bewering onjuist (0), met één argument. |
| 3. Als beide premissen onzeker zijn (??), dan is niet te zeggen of de bewering juist is (?), met twee argumenten. |
| 4. Is het in de overblijvende gevallen voor één van de premissen niet te zeggen of hij waar is of niet (?), dan is het niet te zeggen of de bewering juist is (?), met één argument. |

We mogen aannemen dat deze eerstejaars studenten, die het normale VWO hebben gevolgd, geen drie-waardige logika hebben geleerd. Bij navraag achteraf bleek ook dat ze de drie-waardige waarheidstabel voor een conclusie niet kenden en niet konden genereren. Alle redeneringen van Tabel 4 zijn succesvol, en gebaseerd op ervaring. Daarom mogen we de conclusie trekken dat deze redeneringen intuïtief zijn. De meest gebruikte intuïtieve redeneringen vinden we dan in Tabel 6.

Deze intuïtieve redeneringen zijn niet verkortingen van de uitgebreide, volledige

redeneringen gebaseerd op Tabel 1. Het argument hiervoor is dat de studenten de theoretische waarheidstabel nooit geleerd hebben. Een verkorting door automatisering of een verkorting door inzicht is hierdoor uitgesloten.

Ook hebben de studenten de redeneringen nooit in verkorte vorm expliciet geleerd. Op grond hiervan kunnen we stellen dat bovenstaande intuïtieve redeneringen niet bewust of niet volledig bewust aanwezig zijn bij de studenten, maar geheel of gedeeltelijk onbewust functioneren. De aanwezige variatie in argumentatiewijze duidt hier eveneens op: Bij ongeveer 15 % van de antwoorden argumenteren de studenten op een alternatieve manier. Er zijn tien studenten die consequent op de meerderheidsmanier redeneren (slechts een student geeft consequent twee argumenten).

De intuïtieve redeneringen zijn door voorbeelden van redeneringen geleerd. Zij lijken wel erg op een verkorte vorm van bewust theoretisch redeneren, maar ze zijn niet goed bestand tegen veranderingen in de situatie zoals we in de volgende paragraaf zullen zien.

4. REDENERINGEN IN EEN COMPLEXE SITUATIE

Aan dezelfde studenten hebben we een zevental soortgelijke beweringen als hiervoor voorgelegd, maar nu betreffende de oscilloscoop.

De opgave was als volgt. Eerst werd weer een juiste uitspraak gegeven: ‘Voor de goede werking van de interne triggering is het nodig dat het ingangssignaal voldoende groot is en de oscilloscoop correct staat ingesteld’. (De interne triggering zorgt ervoor dat het beeld niet gaat lopen, maar stil staat.) Dit is zeker een complexe situatie omdat beide premissen complex zijn in die zin dat ze zijn samengesteld uit diverse onderdelen. Gevraagd wordt nu bijvoorbeeld of de volgende uitspraak juist is: “De trigger werkt goed want het ingangssignaal is 1 V en het beeldpunt is slecht gefocuseerd”. Dit is duidelijk vaktaal en een complexe probleemsituatie. Zeven van dit soort vragen werden aan de studenten voorgelegd. In Tabel 7 is de verdeling van redeneringen en antwoorden weergegeven. Het percentage onjuiste redeneringen is nu 51%, het percentage onjuiste antwoorden 49%. Deze cijfers kunnen vergeleken worden met die behorende bij Tabel 3.

Tabel 7. Verdeling van de redeneringen en antwoorden in klassen

klassen			aantallen (de oscilloscoop)		
L+A+	L+A—	L+	99	51	150
L—A+	L—A—	L—	57	101	158
A+	A—	N	156	152	308

We zien dat de helft van de antwoorden fout is. Als de studenten puur zouden gokken zou ongeveer 2/3 van de antwoorden fout moeten zijn. Er is dus wel enige vakkennis aanwezig, maar niet voldoende om de vragen even goed te kunnen beantwoorden als bij het weer.

We zien ook dat bij de helft van de studenten de logika hen in de steek laat bij het beantwoorden van vragen over de oscilloscoop. Deze studenten zijn blijkbaar in de war gebracht. De verdeling over het aantal goede en foute antwoorden binnen deze klasse is ongeveer 1:2 en zou dus geheel door het toeval verklaard kunnen worden (chi-squared 1 dof = 0,5).

De nieuwe verdeling van de soorten argumenten bij correcte redenering, te vergelijken met Tabel 4, is weergegeven in Tabel 8 (in procenten; voor de absolute aantallen, zie aanhangsel).

Tabel 8. Verdeling van de soorten argumenten bij correcte redeneringen betreffende de oscilloscoop (in %)

THEORET. REDENERING	OSCILLOSCOOP		
	juist arg. 1 arg.	2 arg.	onjuist/ irr.arg.
?0→0	83	2	15
1?→?	50	0	47
inv??→?	64	14	23
10→0	33	0	67
??→?	45	10	45

We zien dat correct redeneren met 2 argumenten praktisch niet meer voorkomt. (In totaal, dus over alle klassen, neemt het aantal dubbele argumenten af van 35% naar 8% (in aantallen: 72/210→24/308).)

De student heeft minder ervaring met de oscilloscoop dan met het weer. Bij het weer weet hij al welke aspecten relevant zijn en welke niet, en hoe hij de probleemsituatie moet interpreteren. Bij de oscilloscoop is dat minder het geval. Daarom is de situatie voor hem complexer.

De toename van het percentage onjuiste en irrelevante argumenten bij overigens correcte redeneringen, reflecteert onzes inziens de complexiteit van de nieuwe probleemsituatie en de geringere hoeveelheid domeinkennis (in aantallen: 13/196→51/150).

Wij nemen aan dat onjuiste of irrelevante argumenten bij correcte redeneringen door een afwijkende interpretatie van de probleemstelling tot stand komen (verg. par. 3). Deze aanname maakt het mogelijk een test uit te voeren op de nulhypothese dat bij correcte redeneringen op grond van een juiste interpretatie van de probleemstelling het aantal argumenten niet afhangt van het toepassingsgebied van de redenering. In Tabel 9 zijn de relevante aantallen en de waarde van chi-kwadraat aangegeven.

Tabel 9. Verschuiving in de argumentatiewijze bij correcte redeneringen met juiste argumenten (aantallen) en de waarde van chi-kwadraat voor het geval de verschuiving op toeval berust

REDENERING	WEER		OSCILLOSCOOP		CHI-SQUARED (1 dof)
	juist argument 1 arg.	2 arg.	juist argument 1 arg.	2 arg.	
?0→0	33	5	45	1	2,6
1?→?	33	6	19	0	1,8
inv??→?	8	25	14	3	13,4
10→0	35	5	6	0	0,04
??→?	10	24	9	2	7,4
	119	65	93	6	

Bij drie klassen van redeneringen is de verschuiving nihil of niet significant. Voor de redeneringen ??→? en inv??→? is de nulhypothese verworpen met een significantiegrens van 1%. Bij correcte redeneringen op grond van juiste argumenten met twee onzekere

Tabel 10. Intuïtieve redeneringen van de meeste studenten in een complexe situatie (correcte redeneringen op grond van juiste argumenten)

1. De premissen die waar zijn (1) worden onbewust opgenomen.
2. Is één van de premissen niet waar (0), dan is de bewering onjuist (0), met één argument.
3. Is het in de overblijvende gevallen voor minstens één van de premissen niet te zeggen of hij waar is of niet (?), dan is het niet te zeggen of de bewering juist is (?), met één argument.

premissen verandert dus de argumentatiewijze significant wanneer men overgaat van het weer naar de oscilloscoop. De argumentatiewijze is weergegeven in Tabel 10.

5. CONCLUSIES

Onze conclusies zijn dat de onderzochte soort redeneringen intuïtief verloopt en dat intuïtieve redeneringen van studenten beïnvloed worden door de hoeveelheid domeinkennis die zij ter beschikking hebben. Intuïtieve redeneringen zijn niet bestand tegen de overgang van een bekende naar een minder bekende, complexere situatie. Afgezien van het feit dat er dan veel meer onjuiste en irrelevante argumenten worden gehanteerd, neemt het aantal foute redeneringen sterk toe en verschuift de argumentatiewijze bij overigens correcte redeneringen. Het belangrijkste effect van overgang naar een minder bekende situatie is te merken bij correcte redeneringen op grond van twee onzekere premissen: in plaats van twee premissen wordt slechts één premisse expliciet als argument gehanteerd.

De conclusie van Reber (l.c.) is dus in zijn algemeenheid onjuist en moet genuanceerd worden. 'Tacit knowledge' kan wel gebruikt worden om problemen op te lossen, maar in een beperkt domein. In een ander domein is de 'tacit knowledge' anders. De vraag rijst dan of er wel van 'knowledge' sprake kan zijn. In elk geval vindt het redeneren hier niet plaats via domein-onafhankelijke formele regels. De inhoud heeft effect op het redeneren. Een dergelijk effect is al eerder aangetoond (Wason 1977, 1983). Tot nu toe werden deze effecten geanalyseerd met 2-waardige redeneringen. Verklaringen, bijvoorbeeld in termen van pragmatische redeneerschema's (zie Cheng, 1989, en de referenties daarin), bieden daarbij een alternatief voor de theorie dat redeneren plaatsvindt louter op grond van de herinnering van domein-specifieke ervaring. Op deze plaats zullen we niet verder gaan dan te stellen dat ons onderzoeksresultaat een alternatief analysekader biedt voor de geciteerde effecten.

Ons onderzoek levert materiaal voor een oud probleem. Uit Tabel 7 blijkt bij studenten met goede antwoorden 37% (in aantallen: $L-A+/A+=57/156$) een foute logika te hanteren, terwijl bij de studenten met foute antwoorden 34% (in aantallen: $L+A-/A-=51/152$) een juiste redenering toepast. We vinden dus dat bij 1/3 van de foute antwoorden de redenering correct is en dat bij 1/3 van de juiste antwoorden de redenering fout is. Een ervaren docent heeft hier gevoel voor: Hij kijkt niet alleen naar de antwoorden van studenten, maar ook naar hun redeneringen.

Overigens moeten we nog opmerken dat één van de studenten in de eenvoudige situatie consequent twee argumenten hanteerde. In de complexe situatie bleek dat ook deze student intuïtief redeneert. Toch kunnen we hieruit een belangrijke conclusie trekken: de theoretische redenering is te vinden door zorgvuldig alle (correcte) redeneringen van studenten te onderzoeken.

We zullen in de volgende paragraaf preciezer bekijken welke consequenties onze conclusies hebben voor het vermogen problemen op te lossen.

6. RELATIE TUSSEN REDENEREN EN PROBLEEMOPLOSSEN.

Probleemoplossen is niet hetzelfde als redeneren: men kan redeneren zonder problemen op te lossen, maar ook is het mogelijk de oplossing van een probleem te 'zien' zonder dat er op dat moment geredeneerd wordt (zie verder). Redeneren speelt, naast symboliseren en schematiseren, wel een belangrijke rol bij het voorbereiden van het 'zien' van de oplossing en het uitwerken van de oplossing.

We willen hier bespreken hoe de door ons onderzochte redeneringen van belang zijn voor *probleemoplossen* en hoe onze resultaten algemener van belang zijn voor het *leren* oplossen van problemen (probleemoplosonderwijs).

De redeneringen die wij hebben onderzocht zijn in tweeërlei opzicht van belang voor probleemoplossen. Ten eerste kan een conclusie gebruikt worden bij het identificeren van objecten. Een oscilloscoop kan ik bijvoorbeeld opvatten als een elektronisch beeldscherm met een instelbare tijdbasis. Deze twee kenmerken (in het algemeen zijn er meer) kunnen nu als premissen dienen voor de conclusie: Dit is (wel/ niet/ misschien) een oscilloscoop.

Ten tweede wordt de conclusie gebruikt bij redeneringen over eigenschappen, werking en functie van een object, zoals in de gegeven voorbeelden over de zon en de oscilloscoop. Bij het oplossen van problemen worden conclusies in beide opzichten gebruikt.

Zoals we boven gezien hebben veranderen sommige intuïtieve redeneringen bij conclusies wanneer overgegaan wordt van eenvoudige naar complexe situaties. Het betreft het aantal argumenten dat aangedragen wordt voor een conclusie die nog onzeker is. In het eenvoudige geval wordt meestal onderscheid gemaakt tussen $?1 \rightarrow ?$ (één argument) en $?? \rightarrow ?$ (twee argumenten). In het complexe geval wordt bij beide redeneringen meestal slechts een argument gebruikt.

Bij het redeneren vanuit het gegeven heeft dit nauwelijks gevolgen. Immers, bij deze redeneringen gaat het er om te bepalen wat wel en wat niet juist is op grond van de gegevens. Dit zou ook korrekt kunnen gebeuren aan de hand van de 2-waardige logika.

De driewaardige logika is echter van groot belang om de redeneringen zo te sturen dat het doel (de oplossing van het probleem) bereikt wordt. Om te weten of de interne triggering goed werkt, moet ik weten of het ingangssignaal voldoende groot is en of de oscilloscoop correct staat ingesteld. Dit is een omkering van de in par.4 gegeven uitspraak. Deze omkering is nodig om uit te komen op een uitspraak betreffende de goede werking van de trigger. We kunnen immers deze terugredenering verder voortzetten. Hoe kom ik te weten of het ingangssignaal voldoende groot is? Daartoe moet ik zoeken naar gegevens op grond waarvan ik een conclusie kan trekken over het ingangssignaal.

De consequentie van het bovenstaande is dat de student op grond van intuïtieve redeneringen in eenvoudige gevallen zal onderscheiden of hij één of dat hij twee argumenten zeker (waar of onwaar) moet maken. In complexe gevallen laat zijn redeneervermogen hem in de steek en/of denkt hij dat het voldoende is om één argument zeker te maken. Hij komt er in echte probleemsituaties dan vaak niet uit.

Probleemoplossen is in onze visie – even afgezien van het gebruik van formules, schema's e.d. – een 2-waardig redeneren op grond van gegevens en andere aanwezige kennis, een 3-waardig terugredeneren vanuit het gevraagde, en een 'zien' van een mogelijke structuur die heen- en terugredeneren verbindt. De oplossing komt tot stand door heen- en terugredenering uit te werken tot een logisch geheel. De zoekruimte voor oplossingen wordt bepaald door de zeer vele 2-waardige redeneringen die op grond van gegevens en aanwezige kennis mogelijk zijn. Als bij het terugredeneren te weinig premissen worden ingeschakeld, kan de verbinding niet tot stand gebracht worden.

Het afgaan op te weinig gegevens is een bekend verschijnsel in het probleemoplosonderwijs. Dit leidt tot het in het wilde weg te werk gaan van studenten (zie bijvoorbeeld Mettes en Pilot, 1980). De naam 'kick and rush' strategie duidt ook op het boven gesignaleerde afgaan op een te beperkt aantal kenmerken. We zouden graag willen dat studenten leren op alle relevante kenmerken te letten. Daartoe zouden ze moeten leren om correct en consistent te redeneren.

De eerste stap om hier iets aan te doen is dan om de redenering te expliciteren, zoals in ons onderzoek gebeurd is. Het blijkt dan dat deze redeneringen intuïtief zijn en dat er

verschillende typen correcte redeneringen zijn (met 1 of met 2 argumenten). Vervolgens wordt een keus gemaakt voor een gewenste redeneerwijze, bijvoorbeeld de correcte redenering die de meeste studenten in een eenvoudige situatie volgen (Tabel 6). We hebben deze meest gebruikte correcte, geëxpliciteerde redenering de empirische redenering genoemd. De gevonden (correcte) redeneringen vormen dan feitelijke (correcte) handelingsverlopen, de empirische redenering vormt een gewenst handelingsverloop in de zin van Mettes en Pilot. We kunnen deze empirische redenering als een heuristiek opvatten, die we kunnen toevoegen aan de andere heuristieken voor een systematische wijze van probleemoplossen.

De gesignaleerde problemen in het onderwijs zijn nu begrijpelijk. De studenten zijn moeilijk te motiveren om heuristieken te oefenen omdat de oefensituatie vaak eenvoudig is en studenten de aangeboden heuristiek dan intuïtief al toepassen en daarom triviaal vinden. In feite werkt dit onderwijs dan averechts omdat de studenten erop leren te vertrouwen dat in een complexe situatie de intuïtieve redenering ook wel zal voldoen. Het succes van de methode van Mettes en Pilot met probleemoplosonderwijs moet dan verklaard worden door de koppeling van heuristieken aan complexe probleemsituaties: de complexe probleemsituatie dwingt tot gebruik van de heuristiek - intuïtief werken is niet meer voldoende.

Het gebrek aan transfer van de heuristiek naar andere domeinen hangt dan inderdaad samen met het eerste punt. In die gevallen waarin wel transfer optreedt maar de prestaties in het vak niet toenemen wordt vermoedelijk de heuristiek louter formeel gehanteerd. We kunnen dit vergelijken met het toepassen van de juiste empirische redenering op onjuiste argumenten. De heuristiek laat het dan waarschijnlijk afweten bij het trekken van conclusies over de juistheid van de premissen.

7. VERGELIJKING VAN REDENERINGEN

In eerdere paragrafen hebben we al de verklaringen gegeven voor de beide observaties in het onderwijs. Resumerend: Studenten kunnen in een complexe situatie inderdaad niet zo redeneren als ze in een eenvoudige situatie doen, en je kunt ze de empirische redeneringen moeilijk leren omdat ze die in een eenvoudige oefensituatie vervangen door hun intuïtieve redeneringen. Dit duidt erop dat het van groot belang is verder onderzoek te doen naar onderwijs in de theoretische redeneringen. Dit onderwijs is nauw verwant aan het onderwijs in theoretische begrippen zoals gepropageerd door Davydov (vergelijk Haenen en Van Oers, 1983). Op dit moment kunnen we een vergelijkend overzicht geven over de theoretische, de empirische en de intuïtieve wijze van redeneren gebaseerd op conjunctie.

Allereerst een opmerking over de onderzochte driewaardige logika. Bij het vooruit redeneren vanuit ware of onware gegevens leiden redeneringen in een driewaardige logika tot dezelfde conclusies als in een tweewaardige logika. Het verschil zit in het omgaan met onzekere (doel of begin-) situaties.

De feitelijke driewaardige redenering blijkt intuïtief te zijn en niet stand te houden in een complexe situatie, en we kunnen ons dus met recht afvragen of hij wel echt als redenering plaatsvindt in het hoofd van de studenten. We zullen ons daarom verder beperken tot een vergelijking van de empirische en de theoretische redenering.

1. De theoretische redenering is *volledig*, in die zin dat alle argumenten in de betreffende redeneringen zijn opgenomen. Er zitten geen 'gaten' in. In alle negen redeneringen worden beide argumenten gegeven. Bij de empirische ontbreken sommige argumenten.

2. De theoretische redenering is *symmetrisch*. Dit kan duidelijker gezien worden in een andere representatie van Tabel 1 (zie Tabel 11).

Tabel 11. De theoretische drie-waardige conclusie in een symmetrische representatie.

		p1→		
		1	0	?
p2 ↓	1	1	0	?
	0	0	0	0
	?	?	0	?

De symmetrie houdt in dat verwisseling van premissen dezelfde redeneringen oplevert. Bij de theoretische conclusie is de symmetrie structureel en expliciet aanwezig, bij de empirische redenering is hij impliciet en moet hij als een aparte regel toegevoegd worden.

3. De theoretische redenering is *schematisch* weer te geven zonder volgorde van afhandeling, en geeft zo een totaal-overzicht van het concluderen. Als de empirische redenering schematisch wordt weergegeven, bevat zo'n schema altijd sequenties.

4. De theoretische redenering kan eenvoudig *omgekeerd* worden gebruikt om terug te redeneren vanuit het doel. Zowel in het schema van Tabel 1 als in dat van Tabel 11 kan in een oogopslag worden nagegaan welke conclusies en argumenten bij elkaar horen. De empirische conclusie brengt bij omkering moeilijke gedachtengangen mee die niet altijd worden vastgehouden.

5. De theoretische conclusie is *iteratief* te gebruiken: Zij kan herhaaldelijk toegepast worden. Vooral bij het terugredeneren is dit van belang. Immers, hoe bepalen we voor elk van de premissen of hij waar is of niet? Ook hiervoor is weer een conclusie nodig. Er vindt dan een soort nesting van conclusies plaats.

6. De theoretische redenering is *generaliseerbaar*, niet alleen naar andere domeinen (transfer) maar ook naar andere conclusies in dezelfde drie-waardige logika, bijvoorbeeld een conclusie die gebaseerd is op disjunctie in plaats van de hier gebruikte conjunctie. Zelfs is hij generaliseerbaar naar 'conclusies' in andere drie-waardige logika's die even consistent zijn maar nog geen bruikbare toepassing kennen (Rescher, 1969). De empirische redenering leidt niet tot verdere theorievorming maar geeft een (empirisch) model dat op zich zelf staat.

NOOT

1. Met dank aan Wiebe van der Veen voor de nauwgezette wijze waarop hij de data heeft verwerkt. Tevens vermelden we de Begripsontwikkelingsgroep van het Onderwijskundig Centrum (A. Pilot, F. Pothof, C. Pouw, C. Terlouw en de auteur) vanwege de vruchtbare discussies en het kritisch commentaar.

LITERATUUR

- Berg, J.S. van den (1983). *Natuurkunde-vraagstukken-oplossen*. Eindhoven: Proefschrift, Technische Hogeschool Eindhoven.
- Cheng, P.W. & Holyoak, K.J. (1989). On the natural selection of reasoning theories. *Cognition* 33, p. 285-313.
- Groot, A.D. de (1985). Over intuïtie. *Heijmans Bulletins Psych. Instituten*. Groningen: Rijksuniversiteit Groningen.
- Haenen, J. & Oers, B. van (1983). *Begrippen in het onderwijs: De theorie van Davydov*. Amsterdam: Pegasus.
- Mettes, C.T.C.W. & Pilot, A. (1980). *Over het leren oplossen van natuurwetenschappelijke problemen*. Enschede: Proefschrift, Technische Hogeschool Twente.

- Mettes, C.T.C.W. & Gerritsma, J.G.M. (1986). *Probleem oplossen*. Utrecht: Onderwijskundige informatie over het Hoger Onderwijs, Aula 819, Het Spectrum.
- Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Exper. Psych.: General* 118 , p.219.
- Rescher, N. (1969). *Many-valued logic*. New York: McGraw-Hill.
- Vos, H. (1989). *Redeneringen van studenten*. Enschede: OC-doc 89-26, Onderwijskundig Centrum, Universiteit Twente.
- Wason, P.C. (1977). Self-contradictions. In: P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking : Readings in cognitive science*. Cambridge: Cambridge University Press.
- Wason, P.C. (1983). Realism and rationality in the selection task. In: J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

Manuscript ontvangen 23-11-1989

Definitieve versie ontvangen 11-1-1990

AANHANGSEL

Tabel 12a. Absolute verdeling van de soorten van argumenten bij correcte redeneringen (klasse L+)

REDENERING	WEER					OSCILLOSCOOP				
	N	juist	arg.	onjuist/ irr.arg.		N	juist	arg.	onjuist/ irr.arg.	
		1	2	1	2		1	2	1	2
20→0	40	33	5	2	-	54	45	1	7	1
1?→?	40	33	6	1	-	36	19	0	10	7
inv??→?	36	8	25	3	-	22	14	3	3	2
10→0	42	35	5	2	-	18	6	0	10	2
??→?	38	10	24	4	-	20	9	2	9	-
	196	119	65	12		150	93	6	39	12

Tabel 12b. Absolute verdeling van de soorten van argumenten bij onjuiste redeneringen (klasse L—)

REDENERING	WEER					OSCILLOSCOOP					
	N	juist	arg.	onjuist/ irr.arg.		N	juist	arg.	onjuist/ irr.arg.		
		1	2	1	2		1	2	0	1	2
?0→0	2	-	-	2	-	34	3	-	20	10	1
1?→?	3	1	-	1	1	52	4	-	27	18	3
inv??→?	6	-	2	-	4	22	-	-	13	7	2
10→0	-	-	-	-	-	26	1	-	17	8	-
??→?	3	1	2	-	-	24	-	-	19	5	-
	14	2	2	5	5	158	8	0	96	48	6

Classificering van de argumenten en de redeneringen

De gegeven juiste bewering (de eerste bewering) bestaat uit een conclusie gebaseerd op conjunctie van twee premissen.

De tweede bewering, waarvan nagegaan moet worden of hij juist is, is voorzien van twee argumenten, verder gegevens genoemd.

Het antwoord van de student is gebaseerd op argumenten.

Klassen van argumenten

A. Een juist argument brengt een gegeven in verband met een premisse waardoor de waarheidswaarde van die premisse kan worden vastgesteld.

B. Een irrelevant juist argument brengt geen verbinding aan tussen premisse en gegevens, maar is op zich wel juist.

Voorbeeld: - Het KNMI kan het goed, maar ook fout hebben.

- Er is een wolk is in strijd met de eis 'er is geen wolk'.

C. Een irrelevant onjuist argument brengt geen verbinding aan tussen premisse en gegevens, en is op zich onjuist.

Voorbeeld: (bij bewering 5) de zon schijnt niet omdat het een beetje regent.

D. Een onjuist argument is wel relevant (brengt een verbinding aan tussen een premisse en een gegeven) maar volgt niet uit de gegevens. Hij is dus gebaseerd op een onjuiste interpretatie van de gegevens.

Voorbeeld: - Om 8 uur is de zon onder ('s morgens en 's avonds).

E. Geen argument.

Klassen van redeneringen

Om de juistheid van een redenering na te gaan, vragen we ons af: Wat voor antwoord zou logischerwijs op de door de student gegeven argumentatie moeten volgen?

A. Een juiste redenering (L+) wordt gevolgd als

1. een juiste argumentatie leidt tot een juist antwoord, en
2. een onjuist argument leidt tot een (volgens de waarheidstabel) mogelijk antwoord (voordeel van de twijfel).

B. Een onjuiste redenering (L—) wordt gevolgd als

1. een irrelevant argument (juist of onjuist) leidt tot een antwoord, juist dan wel onjuist.
2. een juist argument leidt tot een onjuist antwoord.
Voorbeeld: (bij bewering 3) bewering onjuist want er zijn wolken.
3. een onjuist argument leidt tot een (volgens de waarheidstabel) onmogelijk antwoord (juist dan wel onjuist).

Een juist argument kan tot een onjuist antwoord leiden doordat geen gebruik wordt gemaakt van de premissen in de gegeven juiste bewering. Het voorbeeld bij B2 is zo'n redenering. Een ander voorbeeld van zo'n redenering is deze: (bij bewering 1) het is niet te zeggen, want omdat de zon niet op is, is niet te zien of hij schijnt. Een andere mogelijkheid is dat de hele vraagstelling onjuist wordt geïnterpreteerd (zie het voorbeeld in par. 3).

Boekbesprekingen

Anne Kerkhoff, *Taalvaardigheid en schoolsucces*. Amsterdam, Swets & Zeitlinger, 1988. ISBN 90 265 0959 6, 237 pagina's (ook verschenen als proefschrift KUB, Tilburg).

Voorafgegaan door een literatuuroverzicht over ongelijke onderwijskansen en de rol van taalvaardigheid daarin, doet Kerkhoff verslag van een onderzoek over 'de relatie tussen taalvaardigheid Nederlands en schoolsucces van allochtone en autochtone leerlingen aan het einde van de basisschool', zoals de ondertitel van haar proefschrift luidt. Het onderzoek is uitgevoerd bij leerlingen die in het schooljaar 1983/84 in klas 5 van in totaal vijf Tilburgse en Bredase scholen zaten. Deze leerlingen werden tot het einde van de lagere school gevolgd, terwijl ook na één jaar voortgezet onderwijs nog doorstroomgegevens werden verzameld. De resultaten vormen grotendeels een bevestiging van eerder onderzoek.

Kerkhoff probeert in haar onderzoek de achterstandsproblematiek van allochtone en autochtone minderheidsgroepen in het onderwijs in één kader te vangen door zowel sociaal milieu als taalachtergrond mee te nemen. Jammer genoeg slaagt zij daarin niet erg goed. Op het geheel is zowel qua opzet als uitvoering nogal wat aan te merken, zoals in het volgende zal worden beargumenteerd.

De aanleiding tot het onderzoek is de achterstandspositie van etnische minderheidsgroepen in het onderwijs, waarbij wordt verondersteld dat een onvoldoende beheersing van de Nederlandse taal daarvan de belangrijkste oorzaak vormt. In de eerste twee hoofdstukken documenteert Kerkhoff de aanwezige kennis op dit gebied, waarbij ook autochtone minderheidsgroepen ruim aandacht krijgen. In dit deel, en ook op andere plaatsen waar aan literatuur wordt gerefereerd, loopt het niet over van kritische zin en de auteur is hier en daar bepaald slordig, met name wat betreft het toeschrijven van conclusies aan anderen. Op p. 12 wordt bijvoorbeeld gesteld: 'Zowel Wijnstra (1976) als Stijnen en Vallen (1981) komen tot de conclusie dat er sprake is van 'divergentie': de verschillen in het schoolsucces van leerlingen met verschillende taalachtergronden nemen in de loop van hun schoolloopbaan eerder toe dan af.' De aangehaalde publikatie laten dergelijke algemene conclusies geenszins toe en zijn daarin ook niet te vinden (opmerkelijk hierbij is dat Vallen als assistent-promotor optrad bij de totstandkoming van deze dissertatie). Op p. 18 wordt Stijnen en Vallen een soortgelijke conclusie met betrekking tot taalvaardigheid toegeschreven. Stijnen en Vallen (1981, p. 237) trekken zelf een tegenovergestelde conclusie. In hun slothoofdstuk vermelden Stijnen en Vallen (1981, p. 241) als elfde conclusie: 'Dialectsprekende leerlingen worden door hun leerkrachten op een aantal aspecten minder gunstig beoordeeld dan hun standaardtaalsprekende leeftijdgenoten', waarbij in de toelichting wordt vermeld dat de verschillen in de meeste gevallen niet significant zijn. Kerkhoff bombardeert deze conclusie op p. 94 tot de 'voornaamste conclusie' in de volgende formulering: 'Daarbij is naar voren gekomen (...) dat de wat minder gunstige onderwijsresultaten van dialectsprekende leerlingen in Kerkrade samenhangen met de wijze waarop leerkrachten de prestaties van dialectsprekers beoordelen.'

Deze voorbeelden geven te denken met betrekking tot de rest en het is niet moeilijk een aantal feitelijke onjuistheden aan te wijzen. Op p. 21 wordt bijvoorbeeld de formatieregeling voor het basisonderwijs niet correct weergegeven. De uit een onderzoek van Van Esch (1983) overgenomen gegevens over doorstroomadviezen in twee tabellen op p. 74 en 75 zijn niet consistent. Bij de tweede tabel wordt ten onrechte niet vermeld dat de daarin

opgenomen gegevens van Van Esch niet betrekking hebben op de volledige oorspronkelijke steekproef: leerlingen die in het basisonderwijs zijn gebleven of naar een internationale schakelklas zijn vertrokken (13.3%), blijven buiten beschouwing. In de eerste tabel heeft Kerkhoff deze kinderen in de IBO-categorie opgenomen.

Het ontwerp van het uitgevoerde onderzoek wordt nogal gebrekkig gemotiveerd. Dit geldt zowel voor de verantwoording van de opzet van de gelegenheidssteekproef als voor de instrumentatie van schoolsucces en taalvaardigheid. Waarom bijvoorbeeld de relatie tussen taalattitudes en taalvaardigheid van belang is voor de hoofdvraagstelling, wordt niet beargumenteerd. Ook de invulling van het instrumentarium voor taalvaardigheid, op zichzelf het meest interessante deel van het onderzoek, is niet erg planmatig. Het lijkt erop dat het principe 'van alles wat' heeft gegolden. De relevantie van een woordassociatietaak bijvoorbeeld is niet erg overtuigend. Het zwakste punt evenwel is de operationalisatie van de centrale variabele schoolsucces, die volledig wordt gedefinieerd in termen van leerkrachtoordelen (rapportcijfers, vragenlijsten en doorstroomadviezen). Met name de doorstroomadviezen krijgen een zwaar gewicht in het geheel. Afgezien van verschillen in adviseringsstrategie tussen leerkrachten (vgl. Saris & Blok, 1982), lenen deze gegevens zich minder goed voor vergelijking van groepen allochtone en autochtone leerlingen, omdat allochtone leerlingen vaak het voordeel van de twijfel krijgen (vgl. Wijnstra, 1984; De Jong, 1987), dat wil zeggen dat allochtone leerlingen gemiddeld hogere adviezen krijgen dan autochtone leerlingen met vergelijkbare niveaus van schoolvorderingen. Hoewel Kerkhoff dit verschijnsel kent, motiveert zij haar keuze door te wijzen op de relatie met scores op schoolvorderingentoetsen en het feit 'dat de oordelen van leerkrachten doorgaans doorslaggevend zijn bij de keuze van het type voortgezet onderwijs waar leerlingen naar doorstromen' (p. 57), waarbij zij aanneemt dat er zeer weinig 'verticale doorstroming' plaatsvindt (p. 3). Zij veronachtzaamt daarbij echter mogelijke verschillen in het vervolg van de schoolloopbaan tussen autochtone en allochtone leerlingen. Uit een nog te rapporteren onderzoek in verband met de Eindtoets Basisonderwijs blijkt dat het afstroompercentage (doublures en doorstroming naar een lager geclassificeerd schooltype) onder allochtone leerlingen twee tot vier keer zo hoog is als bij autochtone leerlingen (Uiterwijk, in voorbereiding; Ahlers, 1989). Daarom is het te betreuren dat Kerkhoff in de analyses niets doet met de verzamelde gegevens over de bereikte onderwijspositie na één jaar voortgezet onderwijs, ondanks de aankondiging dat dit wel zal gebeuren (p. 57).

Ook op de weergave van de resultaten en met name het uitvoerige, vaak nogal speculatieve commentaar daarbij, valt wel het een en ander aan te merken. Naar aanleiding van een berekende correlatiecoëfficiënt van .52 tussen sociaal-economisch milieu en schoolkeuze-advies (hoe?) wordt bijvoorbeeld vermeld: 'Dan nog blijft het echter opmerkelijk dat de in dit onderzoek gevonden correlatie tussen sociaal-economisch milieu en schoolkeuzeadvies hoger is dan de correlaties die in de meeste andere - doorgaans grootschaligere - onderzoeken worden gepresenteerd met betrekking tot autochtone leerlingen.' (p. 77, 78). Nog afgezien van verschillen in operationalisatie houdt de suggestie omtrent een relatie tussen de schaal van een onderzoek en de hoogte van een correlatiecoëfficiënt weinig rekening met steekproeffluctuaties, die in kleinschaliger onderzoeken nu eenmaal groter zijn dan in grootschaliger.

Illustratief in dit verband is hoe een van de hoofdconclusies tot stand komt. Naar aanleiding van het schoolsucces van zes allochtone leerlingen uit hoger sociaal-economisch milieu wordt het vermoeden geuit 'dat de factor sociaal-economisch milieu het effect van de variabele 'allochtoon' 'compenseert' (p. 78). Dit 'vermoeden' wordt op p. 92/93 tot conclusie verheven en leidt in het slothoofdstuk tot de stelling dat niet geconcludeerd kan worden 'dat de variabele 'taalachtergrond' bepalend is voor de onderwijskansen van kinderen uit etnische minderheidsgroepen. De gesignaleerde discrepanties tussen de

verschillende groepen leerlingen lijken eerder samen te hangen met verschillen in het sociaal-economisch milieu waaruit de kinderen afkomstig zijn.' (p. 190). Om te verklaren waarom allochtone kinderen uit lagere sociaal-economische milieus dan toch een ongunstiger onderwijspositie innemen dan autochtone leerlingen uit vergelijkbare milieus, wordt als verklaring van 'het schijnbare effect van de variabele taalachtergrond' (p. 191) de gebrekkige operationalisatie van de factor milieu opgevoerd. Afgezien van de strijdigheid van deze stelling met het commentaar bij bijvoorbeeld de tabellen 5.4 en 6.2 (resp. p. 105 en 130) leent de onderzoeksopzet zich niet voor dergelijke conclusies. In de genoemde tabellen onderscheidt Kerkhoff vier categorieën leerlingen: Turken en Marokkanen, overige allochtonen, dialectsprekers en standaardtaalsprekers. Onder de Turkse en Marokkaanse leerlingen komen echter in het geheel geen kinderen uit hoger sociaal-economisch milieu voor. De zes allochtone kinderen uit hoger milieu vallen allen in de categorie overige allochtonen. Het design is dus niet volledig gekruist zodat over Turken en Marokkanen - nota bene de grootste groep allochtone leerlingen - in dit opzicht in het geheel geen uitspraken mogelijk zijn.

Kortom, dit proefschrift kan nauwelijks als een aanwinst worden betiteld en biedt ook verder weinig nieuws.

REFERENTIES

- Ahlers, J. (1989). Overgang naar voortgezet onderwijs verloopt niet probleemloos. *School*, 17, nr. 3, 22-24.
- Esch, W. van (1983). *Toetsprestaties en doorstroomadviezen van allochtone leerlingen in de zesde klas van lagere scholen*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen*. Lisse: Swets & Zeitlinger.
- Saris, W.E. & Blok, H. (1982). Het onderwijzersadvies nader bekeken. *Tijdschrift voor Onderwijsresearch*, 7, 49-59.
- Stijnen, S. & Vallen, T. (1981). *Dialect als onderwijsprobleem*. 's-Gravenhage: Staatsuitgeverij (SVO-reeks nr. 50).
- Uiterwijk, J.H. (in voorbereiding). *Onderzoek naar item- en testbias in de Eindtoets Basisonderwijs*. Arnhem: Instituut voor Toetsontwikkeling.
- Wijnstra, J. M. (1976). *Het onderwijs aan van huis uit friestalige kinderen*. 's-Gravenhage: Staatsuitgeverij (SVO-reeks nr. 2).
- Wijnstra, J.M. (1984). *Verantwoording Eindtoets Basisonderwijs 1982*. Arnhem: Centraal Instituut voor Toetsontwikkeling.

Johan M. Wijnstra
Instituut voor Toetsontwikkeling (Cito)
Postbus 1034
6801 MG Arnhem

Anne van Streun (1989). *Heuristisch Wiskunde Onderwijs; verslag van een onderwijsexperiment*. (Proefschrift Universiteit Groningen) Uitgave van de auteur.

Een doelstelling van onderwijs die algemeen geaccepteerd zou dienen te worden is deze: het tegelijkertijd opschuiven naar boven van de grens waar standaardopgaven subjectief tot problemen worden en het verbreden van de zone, na die grens, waarbinnen problemen door de leerling opgelost kunnen worden door een beroep te doen op heuristische methoden. Volgens veel critici schieten de meeste methoden van wiskundeonderwijs

gemeten aan deze dubbele doelstelling tekort. Typische punten van kritiek zijn de volgende drie. Veel leerlingen passen algoritmen op een louter mechanische manier toe. Voorbeeld: op een opgave als 'indien in een taxi maximaal drie personen kunnen worden vervoerd, hoeveel taxi's zijn er dan nodig om zeven personen te vervoeren?' geven veel leerlingen antwoorden als '2,33'. De geleerde algoritmen worden 'mindless' toegepast. Een tweede punt van kritiek is dat voor velen zelfs de geringste transformatie van een bekend type opgave er subjectief een probleem van maakt. Als voorbeeld kan dienen de berekening van de rente op een hypotheek na 'n jaar bij een vaste rente en een vaste jaarlijkse aflossing. Eerste jaars economiestudenten leren daar een formule voor. In de gedaante gebracht van een persoon die jaarlijks de (vaste) rente ontvangt van een geërfd kapitaal, waarvan hij ook jaarlijks een vast bedrag opsoupeert, daalt de p-waarde met de helft. De contextafhankelijkheid van het geleerde is dus te hoog. Een derde punt van kritiek is dat leerlingen niet leren, op hun niveau, wiskunde te bedrijven. Als, bijvoorbeeld, uitgebreid is behandeld op hoeveel manieren n onderscheidbare objecten in een rij kunnen worden geplaatst ($n!$) dan is daarmee slechts voor een begaafde enkeling de basis gelegd voor de oplossing van het probleem op hoeveel verschillende manieren n kinderen in een kring kunnen staan. Wiskunde (met een hoofdletter) is inderdaad de bestemming van slechts enkelen, maar met een kleine letter geschreven, op het eigen niveau dus, is wiskunde kunnen bedrijven toch een aantrekkelijke doelstelling, onder andere omdat het mogelijk een goed uitgangspunt is voor een volgende, aansluitende leercyclus.

Dit soort kritiek is niet gebonden aan de Nederlandse situatie, zij wordt internationaal gehoord. Zij is ook zeker niet nieuw. Wel betrekkelijk recent is dat geprobeerd wordt onderwijs en wat daar eventueel aan schort te beschrijven in algemene cognitief psychologische termen. Het proefschrift van Van Streun past in deze nieuwe lijn, al moet het vooral nog als een poging worden gezien, en dat is één van de dingen die deze recensent erin waardeert. Nog niet zo lang geleden begon iedere zichzelf respecterende onderwijskundige onderzoeker zijn rapportage met een hoofdstuk, waarin omstandig de keuze van het 'theoretische kader' werd verantwoord. Aan zo'n kader was vaak een grote naam verbonden, bijvoorbeeld die van Selz of Vygotski of Anderson. Meestal functioneert het 'gekozen kader' in zulke gevallen slechts als een geordende lijst namen voor begrippen en verschijnselen. Dat is niet onnuttig, het dient de communicatie, maar het hindert de communicatie evenzeer (denk aan discussies over wat de grote meester nu precies onder een bepaalde term verstond) en, wat belangrijker is, zelden wordt zo'n kader gebruikt zoals het behoort, als een generator van verklaring en predictie. In vooral het laatste decennium heeft in de psychologie een ingrijpende unificatie van de theoretische perspectieven plaatsgehad. De informatieverwerkingsbenadering van cognitieve processen functioneerde daarbij als de organiserende kern voor een veelomvattend kader, waarin zulke verschillende perspectieven als het associationistische gezichtspunt, het energetische, het sociale en het ontwikkelingsperspectief met elkaar in coördinatie zijn gebracht. Zeker moeten we oppassen het heil van deze ontwikkeling niet te overdrijven, maar er is voor onderwijskundigen in ieder geval geen reden meer hun kader te kiezen als een jas uit het rek. Het is voldoende kenbaar te maken dat zij een psychologische invalshoek hebben gekozen en niet, bijvoorbeeld, een sociologische. Het kader is dan duidelijk. De begrippen zijn de lezer duidelijk of kunnen in een willekeurig leerboek worden opgezocht en ieder begrip refereert aan een bepaalde onderzoeksliteratuur waarvoor één forum het werk van het onderscheiden van knollen en citroenen heeft uitgevoerd. In Van Streuns 'Heuristisch Wiskunde Onderwijs' komt de verankering van de constructie van een leergang Wiskunde (voor 4 VWO, wiskunde A) in de cognitieve onderwijspsychologie vooral aan de orde in het eerste hoofdstuk. Dit geeft een helder, uitstekend leesbaar, maar zeer selectief en beknopt overzicht van enerzijds de literatuur over wiskundendidactiek en anderzijds

datgene uit de cognitieve psychologie dat dienstig kan zijn voor het maken van didactische keuzes. Zoals gezegd is de behandeling van de aansluiting bij de algemene cognitieve theorie in dit hoofdstuk wel erg beknopt.

Dat maakt het voor de auteur nodig er steeds opnieuw op terug te komen in hoofdstuk II, bij de beschrijving van het ontwikkelingswerk en van de leerlingentekst zoals die uiteindelijk tot stand was gekomen, maar ook later verspreid in de tekst van de daaropvolgende hoofdstukken. Het was, ook wanneer het historisch gezien pas later een rol ging spelen, beter, want overtuigender en overzichtelijker, geweest, wanneer alle theoretische materiaal in het inleidende hoofdstuk was samengebracht; bijvoorbeeld in de vorm van beargumenteerde stellingen. Een belangrijk thema bij heuristisch d.w.z. op transfer gericht onderwijs is, bijvoorbeeld, het belang van het expliciet maken van oplosmethoden. De schrijver onderschrijft dit belang en komt er vaak op terug; steeds echter met slechts enkele regels, zodat het niet mogelijk is wat dieper te graven. Waarom is expliciet maken zo belangrijk? Geldt het alleen, of speciaal voor algemene heuristische methoden? En zo verder. Deze verspreide en steeds wat te beknopte behandeling is jammer om diverse redenen, maar vooral omdat als het om de beschrijving van het concrete lesmateriaal, in hoofdstuk II, gaat, het de lezer niet genoeg helder is welke stukken theorie hoe en in welke onderdelen van de leergang zijn verwerkt. Dat heeft tot effect dat zelfs deze recensent, toch redelijk op de hoogte en zeker bij voorbaat positief jegens heuristisch onderwijs gestemd, als hij moest vertellen hoe het programma is opgebouwd, gedwongen zou zijn steeds terug te vallen op de letterlijke test van Van Streun. Of hij zou de leerlingentekst erbij moeten nemen en zelf zijn abstracties tot stand moeten brengen en deze vervolgens moeten terugkoppelen aan Van Streuns verspreide theoretische presentaties. Dat is toch zeker geen optimale presentatie.

Het tweede hoofdstuk van 'Heuristisch Wiskunde Onderwijs' behandelt de opzet van een vergelijkend onderwijsexperiment dat werd uitgevoerd in het schooljaar '84-'85. Dat was het eerste jaar van de omschakeling van Wiskunde I en II naar Wiskunde A en B. Een gunstige omstandigheid was dat de intentie van de wetgever met Wiskunde A (waar de experimentele leergang op betrekking had) uitstekend aansloot bij de intentie van Van Streuns Heuristisch Wiskunde Onderwijs (HWO). Verspreid over zeven scholen, participeerden 452 leerlingen in 20 klassen 4VWO. Tien klassen kregen onderricht uit HWO. Vijf klassen kregen 'traditioneel' onderwijs uit Ruimte en Getal, of een variant daarvan, aangevuld met voor Wiskunde A typerende toepassingsopgaven. Deze onderwijsvariant wordt door Van Streun aangeduid met WEDT, hetgeen staat voor Wiskunde Eerst Dan Toepassingen. De resterende vijf klassen ontvingen onderwijs volgens de opvattingen over realistisch wiskunde onderwijs van Freudenthal en zijn groep (vakgroep O.W. en O.C.; voorheen I.O.W.O.). Het betreffende materiaal werd ontwikkeld door het HEWET-team dat de herverkaveling uitvoerde die leidde tot de nieuwe examenvakken Wiskunde A en B. Alle drie de varianten behandelen dezelfde stof; aard en aantal van de gebruikte opgaven zijn dezelfde. Het onderzoekdesign is uiteraard van het type 'niet-equivalente groepen'. Er zijn twee voortoetsen afgenomen (aan het einde van 3VWO) en twee qua conceptie corresponderende eindtoetsen na afloop, aan het einde van 4VWO. Daarnaast zijn, gedeeltelijk eveneens in een pre-post design, vragenlijsten afgenomen over meningen en houdingen m.b.t. de wiskunde en het genoten onderwijs.

Van Streuns beschrijving van de opzet en van de concurrerende onderwijsvarianten en de gebruikte instrumenten is gewetensvol, helder en uitvoerig. Bijzonder nuttig is de analyse per item van de wiskundige kenniscomponenten en de heuristieken die aan de orde zijn. Bij de beoordeling van de kwaliteit van zijn op cognitief onderwijspsychologische inzichten gestoelde HWO draait het natuurlijk vooral om de volgende punten. Waar het kennis en vaardigheden betreft, moet de meerwaarde van HWO het duidelijkst aan de dag

treden bij opgaven, wiskundig of toegepast, met een duidelijk probleemkarakter; dat wil zeggen opgaven waarbij bekende wiskundige begrippen en technieken pas na een transformatie, waarbij toepassing van een heuristische methode kan helpen, kunnen worden toegepast. In Van Streuns terminologie zijn dit de P-opgaven. Wat de opgaven betreft die direct geschikt zijn om met bekende begrippen en technieken aan te pakken (de BT-opgaven) is het gewenst dat HWO niet, althans niet te ver, achterblijft bij de andere varianten. Een derde punt betreft het feitelijk gebruik van hetzij heuristieken, hetzij begrippen en standaard algoritmen bij het aanpakken van P en BT-opgaven. Stel dat HWO inderdaad een voorsprong geeft in het oplossen van P-opgaven; voor de theoretische doorzichtigheid is het dan wenselijk dat dit is terug te vinden in de oplosprotocollen, kladblaadjes, werkbladen, bijvoorbeeld door blijken van een betere beheersing van de in HWO onderwezen heuristieken. Op grond van soortgelijke overwegingen kan men tot uitspraken komen over nagestreefde effecten m.b.t. de gemeten attitudes.

De resultaten worden in het derde hoofdstuk en de bijlagen daarvan tot in detail geanalyseerd. De wiskundedocent Van Streun treedt naar voren in de heldere en uitvoerige wijze waarop hij het geweld van de uitgevoerde multivariate analyses voor de lezer bezweert en doorzichtig maakt. Deze recensent zou overigens om leraars- en klasse-effecten te omzeilen (en als die er niet zijn komt het op hetzelfde neer) niet de individuele leerling maar de klas als eenheid van analyse hebben gekozen. Gelukkig zijn de resultaten van het vergelijkende experiment duidelijk genoeg om ook zonder het gebruik van welke turbotoets dan ook zichtbaar gemaakt te worden. Onderstaande tabel geeft de aantallen klassen boven en onder de mediaan van de twee voor- en de twee natoetsen, voor elk van de drie varianten.

	BT-3VWO			P-3VWO			BT-4VWO			P-4VWO		
	HWO	HEWET	WEDT	HWO	HEWET	WEDT	HWO	HEWET	WEDT	HWO	HEWET	WEDT
>med	4	2	4	3	4	3	6	1	3	7	2	1
<med	6	3	1	7	1	2	4	4	2	3	3	4

Het verwoorden van dit resultaat kan aan de lezer worden overgelaten. Minder duidelijk is het resultaat wat betreft het relatieve gebruik van heuristieken bij goed opgeloste P-opgaven. De project-hypothese is dat de leerlingen uit HWO gemiddeld de grootste verhoudingen oplossingen-met-heuristiek te zien zullen geven. De gevonden gemiddelden zijn echter: HWO: 0,84; HEWET: 0,89 en WEDT: 0,85. Univariaat getoetst blijkt de realistische aanpak, HEWET, de heuristische aanpak, HWO, te overtreffen. De overige verschillen zijn niet significant. Als verklaring post hoc beroept Van Streun zich op de verschuiving van de grens standaardopgave-probleem die bij goed onderwijs is te verwachten (zie de openingsregels van deze recensie). Hij doet dat duidelijk niet van harte. Het is mijns inziens ook overbodig. De P-opgaven zijn immers vooraf zo gekozen dat, gelet op het gegeven onderwijs, goede oplossingen-met-transformatie-en-heuristiek veel waarschijnlijker zijn dan kale algoritmische oplossingen. Gegeven dat uitgangspunt zou de hypothese eigenlijk behoren te zijn dat bij elke onderwijsvorm alle goede oplossingen van het heuristische type zullen zijn. De grotere beheersing van de heuristische methoden door HWO-leerlingen zou dan tot uitdrukking moeten komen in een groter aantal goede oplossingen. Het laatste doet zich voor, het eerste niet, het gemiddelde ligt hoog, maar is niet 1,0. Dat betekent dat er tegen de bedoeling in P-opgaven als standaard BT-opgaven moeten zijn opgelost. Uit het overzicht in bijlage 13 blijkt dit inderdaad zo te zijn, doch slechts bij één van de 17 aangewezen P-opgaven. Bij opgave 16a wordt tweederde van de goede oplossingen bereikt door een algoritmische berekening van kansen door het toepassen van de productregel en telregels. Behalve item 16 zijn alle opgaven door alle

leerlingen opgelost met toepassing van een heuristische methode. Op de gebeurtenissen bij één item bouwen we toch geen verstrekkende generalisatie? Gelet op het uitgangspunt was de hypothese voor het relatieve heuristiekengebruik dus wat te simpel gesteld. Welk alternatief hebben we?

De heuristieken waarvan steeds sprake is geweest (en die in de protocollen zijn gescoord) zijn die welke in HWO expliciet aan de orde komen: h1: het vertalen van de verbale P-opgave in een analytische weergave (e.g. een vergelijking); h2: werken met numerieke voorbeelden; h3: maken grafische voorstelling; h4: middelen-doelen analyse of 'working backwards'; h5: een bekende reeks oplosstappen wordt in de omgekeerde volgorde gebruikt; h6: inductie uit een reeks. Een mogelijkheid is sec te toetsen in hoeverre leerlingen uit de verschillende onderwijs-varianten verschillen in de mate waarin zij in staat zijn een van deze heuristieken toe te passen, indien de opdracht is dat te doen. Een andere gaat ervan uit dat niet het toepassen van een heuristiek het onderwijsdoel is, maar het spontaan (en goed) kiezen en toepassen van de in het gegeven geval beste heuristiek. Onder het laatste zou kunnen worden verstaan die methode die het makkelijkst leidt tot het voltooien van het werk met de toepassing van bekende begrippen en technieken. In de richting van dit alternatief gaat de analyse post hoc van Van Streun in het hoofdstuk Discussie. Hij draagt daar gegevens aan die erop wijzen dat de winst van 4VWO op 3VWO én van HWO op HEWET en WEDT vooral is gelegen in het vaker kunnen toepassen van h1, d.w.z. het vertalen van de verbale tekst in een wiskundige representatie. Sommigen zullen hun twijfel hebben of h1 überhaupt wel een heuristiek is; hoe dat zij, het is duidelijk dat h1 meer direct naar het toepassen van een standaardaanpak toewerkt dan de andere beschreven methoden, waarvan het succes vaak afhankelijk is van nog verdere transformaties en episodes van het type 'toen zag ik...'. Zou het kunnen zijn dat de meerwaarde van heuristisch wiskunde onderwijs vooral gelegen is in én het leren herkennen van de waarschijnlijk, na transformatie, toepasbare begrippen en technieken én in het leren uitvoeren van de methoden die de opgave-informatie geschikt maken als invoer voor een algoritme? Is het antwoord bevestigend, dan past tevredenheid: heuristisch onderwijs verricht geen wonderen.

Effecten in het motivationele vlak zijn van heuristisch onderwijs blijkens dit proefschrift niet zozeer te verwachten: wiskunde wordt blijkbaar niet leuker als het leuk wordt gebracht.

Proefschriften daarentegen worden véél leuker wanneer ze, zoals 'Heuristisch Wiskunde Onderwijs' door Van Streun, niet alleen met vakmanschap maar ook met inzet en 'mindful' zijn geschreven. Ik wens het boek veel lezers.

J. J. Elshout
Vakgroep Psychonomie
Weesperplein 8
1018 XA Amsterdam

B. De Vries: *Het leven en de leer. Een studie naar de verbinding van leren en werken in de stage*. ITS, Katholieke Universiteit Nijmegen. (dissertatie; verdedigd op 4 oktober 1988)

In deze dissertatie behandelt De Vries de stage als een onderwijskundig fenomeen. Sinds het midden van de jaren zeventig hebben opleidingen en onderwijskundigen zich moeite getroost om de stage te laten functioneren als een onderwijsmiddel teneinde curriculaire

doelstellingen in het beroepsonderwijs te realiseren. Het leerplaatsenproject, een van de studies waarop deze dissertatie is gebaseerd, was een van de eerste grote vernieuwingsprojecten gericht op deze doelstelling. De dissertatie heeft ten doel te beargumenteren waarom deze schoolcentrische benadering van de stage tot mislukken gedoemd is. Vanuit een meer arbeidsgerichte optiek worden voorstellen geformuleerd voor een andere structurele plaats voor stages in het beroepsopleidende traject voor jongeren.

In hoofdstuk 1 worden de twee onderwijskundige kernbegrippen uit de schoolcentrische benadering van de stage aangegeven: *sturing* en *integratie*. Sturing betreft het streven om de stage-situatie te beïnvloeden vanuit de opleiding; integratie betreft het streven om meer samenhang te brengen in het theoretisch en het praktisch deel van de opleiding.

Hoofdstuk 2 beschrijft op basis van een analyse van onderzoeks- en projectverslagen de ontwikkeling van de discussie rondom de praktijkcomponent van het beroepsonderwijs. Ten aanzien van de ontwikkelingen in de curriculumproblematiek moet De Vries constateren dat pogingen tot sturing en integratie weinig succesvol zijn geweest. De schoolcentrische benadering komt in de beschreven periode steeds meer onder kritiek te staan.

In hoofdstuk 3 wordt een evaluatie ondernomen van vier grote vernieuwingsprojecten rondom de stages. Het is spijtig dat het leerplaatsenproject zelf niet aan deze evaluatie is onderworpen, aangezien De Vries wel constateert dat de geëvalueerde projecten niet tot de gewenste sturing en integratie hebben geleid, maar hij geen inzicht verschaft in de remmende factoren die deze non-implementatie veroorzaken. Tussen de regels door geeft De Vries echter aan dat deze factoren vooral in de arbeidsorganisaties gezocht moeten worden: de organisatie van werk en productie laat zich in veel gevallen moeilijk sturen door op leren gerichte ingrepen (p. 74).

Het probleem is hiermee duidelijk gesteld: is er nog plaats voor een stage in een initiële beroepsopleiding? In het tweede deel van zijn dissertatie draagt De Vries bouwstenen aan voor de stelling dat die plaats er wel degelijk is en dat een beroepsopleiding zonder enige vorm van training on-the-job incompleet is. In hoofdstuk 4 wordt daartoe de opleidende stage gedefinieerd als een vorm van training on-the-job gericht op het aanleren van basis- en beroepskwalificaties die ook in andere functies op het betreffende terrein gebruikt kunnen worden.

Hoofdstuk 5 geeft een historische schets van het proces van de toenemende arbeidsdeling in de voorbereiding van jongeren op arbeid. Binnen het Nederlandse beroepsonderwijs neemt de dagschool een steeds belangrijkere plaats in. Stages hebben een correctieve werking ten aanzien van nadelen die structureel verbonden zijn aan een binnenschoolse theoretische opleiding. Deze nadelen betreffen zowel het belang van arbeidsorganisaties (minder functionele kwalificering en socialisering; ontbreken van selectiemogelijkheden en een reservoir aan goedkope arbeidskrachten) als dat van de jongeren (verlenging van onzelfstandigheid; derving inkomsten; verzwaring en vergroting van de nadruk op kennisinhouden) als ook dat van de samenleving. Pogingen om de stage te onderwerpen aan schoolse regels en didactiek ontnemen haar de mogelijkheid om correctief op te treden.

In hoofdstuk 6 worden de effecten van opleidingsinterventies onderzocht aan de hand van gegevens uit het leerplaatsenproject en een vervolgonderzoek van het ITS naar effecten van sturing van stages, uitgevoerd door collega's van De Vries. Als criteriumvariabelen worden in beide projecten gerapporteerde leerervaringen gehanteerd. Over het algemeen worden in beide studies redelijk veel leerervaringen gerapporteerd; stages hebben in de ogen van betrokkenen effect. Op de realisatie van deze effecten hebben opleidingsinterventies echter slechts marginaal invloed: kenmerken van de stagiair en kenmerken van de stageplaats zijn de belangrijkste voorspellers van leerervaringen. Van de opleidingskenmerken hebben alleen voorbereiding en begeleiding enige invloed. Beide

onderzoeken hebben echter plaats gevonden in 'natuurlijke' situaties, waarin geen sprake was van doelgerichte ontwikkelingsactiviteiten. De conclusie van De Vries dat een stage het meest effectief is indien deze is ingericht als training on-the-job waarbij de opleiding slechts een sturende taak heeft bij de voorbereiding en het stimuleren van reflectie, is ons inziens daarom overhaast.

In hoofdstuk 7 gaat De Vries op zoek naar verklaringen voor het geringe succes van opleidingsinterventies gericht op sturing en integratie met betrekking tot stages. Hij constateert een structurele cesuur tussen theorie en praktijk op een drietal terreinen. Vanuit het interorganisationale perspectief constateert De Vries dat voor het leren in de beroepspraktijk alleen de arbeidsorganisatie verantwoordelijk kan zijn. Vanuit het curriculaire perspectief wordt geconstateerd dat de beheersingsstructuur en de inhoudsstructuur van theoretisch en praktisch onderwijs principieel verschillen: theoretisch onderwijs wordt gestuurd vanuit de 'logica' van vakinhouden en didactische principes, terwijl training on-the-job wordt gestuurd door de 'logica' van het productieproces. Vanuit het socialisatieperspectief wordt een conflict geconstateerd tussen de leerlingrol en de beroepsrol.

In hoofdstuk 8 wordt nader ingegaan op de socialisatie van stagiairs gericht op hun toekomstige beroepsrol. Ook hiervoor wordt teruggevallen op materiaal uit het leerplaatsenproject. In een pretest-posttest design zijn veranderingen in kennis van en houding ten opzichte van aspecten van de arbeidsomgeving tijdens stages onderzocht. Dit is aangevuld met gegevens uit interviews met 90 stagiairs. Het kwantitatieve materiaal duidt niet op enige verandering in kennis of houdingsaspecten. Het kwalitatieve materiaal duidt erop dat stagiairs beschikken over strategieën waarmee zij discrepanties tussen het eigen normen- en waardensysteem en dat van de opleiding en de arbeidsorganisatie kunnen verwerken, zonder dat dat meteen tot aanpassingsverschijnselen leidt. Het interview als onderzoeksmethode blijkt de leerlingen meer bewust te maken van leerervaringen. Dit roept enige vraagtekens op omtrent de betrouwbaarheid van de onderzoeksgegevens, maar wijst tevens op de bruikbaarheid van reflectie-bevorderende gesprekken als aangrijpingspunt voor begeleiding en supervisie.

In het slothoofdstuk schetst De Vries de consequenties van zijn bevindingen voor de inrichting en positionering van stages in het beroepsopleidend traject voor jongeren. Hij kiest voor een aantal stages als autonome fase in het opleidingstraject, successief aan het theoretische gedeelte. Wel zou deze fase nog tot het initiële onderwijs gerekend moeten worden, om de mogelijkheden tot reflectie en begeleiding optimaal te benutten. Een viertal functies worden aan de stage toegeschreven: structureren (de stage als unieke brug tussen onderwijs en arbeid), kwalificeren, socialiseren en communicatie tussen opleiding en arbeidsveld.

De dissertatie van De Vries is een gedegen analyse van het fenomeen stage, verschenen op het moment dat ingrijpende veranderingen in de structuur van het beroepsonderwijs op handen zijn. In dat kader zal ook herbezinning op de stage aan de orde zijn. Dan zal ook duidelijk worden of de opvatting van De Vries om de stages als een eerste training on-the-job aan het einde van het initiële beroepsonderwijs te situeren door de diverse belanghebbenden wordt gedeeld.

Bij de redenering van De Vries zijn echter een aantal kanttekeningen te plaatsen. Ten aanzien van het onderwijssysteem geeft De Vries blijk van een erg pessimistische kijk: de term 'schools' heeft een sterk negatieve connotatie richting dociel, onzelfstandig en onwerelds. De Vries ziet weinig mogelijkheden om hier iets aan te veranderen; dit kan alleen worden gecompenseerd. Door de stage aan het einde van het opleidingstraject te plaatsen wordt dit schoolse karakter nog versterkt. De concentratie op het fenomeen stage

vermindert het zicht op de rol van praktijkleren gedurende het binnenschoolse traject. De voorstellen van De Vries moeten worden gezien in een breder perspectief, waarbij voor beroepsopleidingen een praktijkleerplan wordt ontworpen waarin ook zijn voorstellen een plaats kunnen krijgen.

Ten aanzien van het arbeidssysteem is De Vries juist erg optimistisch: de term 'curriculum on-the-job' veronderstelt een doelgericht scholingsbeleid op de arbeidsplaats, waarvoor weinig aanwijzingen bestaan. De Vries gaat uit van een hoge kwaliteit van de arbeid, een veronderstelling die hij zelf in zijn slothoofdstuk in twijfel trekt. Door de term 'curriculum on-the-job' verhult De Vries in feite dat een productieorganisatie niet automatisch ook een adequate leerorganisatie is. Een goede stage is afhankelijk van de bereidwilligheid van een werkorganisatie om een leeromgeving te creëren. Een minder optimistische kijk op stages leidt tot een enigszins andere structurele positie voor stages, nl. buiten het onderwijssysteem. Het is de vraag of een stage zoals opgevat door De Vries nog als onderwijsinstrument kan worden beschouwd, of dat het als een arbeidsmarktmechanisme moet worden gezien. In dit laatste geval hoort een stage dan ook binnen het arbeidssysteem te worden gedefinieerd met de daarbij horende rechtspositionele aspecten: een arbeidscontract en een passend loon.

De Vries beschouwt de stage als een compenserende reactie op de nadelen van het verzelfstandigd beroepsonderwijs. In zijn voorstellen blijven deze nadelen bestaan terwijl de stage teveel gericht wordt op functionele kwalificering, ondanks zijn pleidooi om de stage als een eerste training on-the-job binnen de structuur van het initiële beroepsonderwijs te houden. Hij beschouwt de stage primair als de overgang van school naar werk. Zijn pleidooi en argumentatie om de stage te ontdoen van de overheersende invloed van de schoolse didactiek wordt op hoofdlijnen door ons onderschreven. Het beroepsonderwijs aan jongeren verkeert momenteel in een situatie waarin gezocht wordt naar een nieuw machts evenwicht, waarin gezamenlijke verantwoordelijkheid het sleutelwoord is. Dat betekent, dat de heersende rol van de school betreffende de onderwijskundige inrichting van de stage afgezwakt dient te worden. Dat betekent echter tevens dat het arbeidsbestel serieus moet nadenken over het kwalificerend potentieel van arbeidssituaties. Zo niet, dan zijn zowel een afsluitende stage als ook vormen van begeleide inwerkperioden net zo gedoemd te mislukken als de door De Vries geanalyseerde vernieuwingsprojecten.

De verdienste van de studie van De Vries is, dat hij de kennis die beschikbaar is omtrent het fenomeen stage op een heldere en grondige wijze heeft geordend en gepresenteerd. De schoolcentrische benadering wordt op doeltreffende wijze ontleed en bekritiseerd. Dat daarbij de balans enigszins doorslaat is misschien onvermijdelijk en scherpt in elk geval de discussie. De opgave voor de jaren negentig is het juiste evenwicht te vinden tussen verschooling en vermaatschappelijking van het beroepsonderwijs.

A.F.M. Nieuwenhuis
RION, Instituut voor Onderwijsonderzoek
Postbus 1286
9701 BG Groningen

J.P.A. Snoek
APS, Algemeen Pedagogisch Studiecentrum
Postbus 7888
1008 AB Amsterdam



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

Greetje P.C. van der Werf, RION,
Postbus 1286, 9701 BG Groningen,
tel. 050-636657

Jrg. 14, nr. 1/2

April 1990

Inhoudsopgave

Convocatie ledenvergadering mei 1990	148
Beknopt verslag verenigingsjaar 1989	149
Ontvangen publikaties	151
Oproep Pedon	152

Convocatie ledenvergadering VOR

De eerstvolgende ledenvergadering van de VOR zal worden gehouden tijdens de Onderwijs Research Dagen op 10 en 11 mei 1990 te Nijmegen. Plaats en tijd worden nader bekend gemaakt.

Agenda

1. Opening
2. Mededelingen
3. Verslag vorige ledenvergadering (zie VOR-bulletin 13,5/6)
4. Verslag verenigingsjaar 1989 (zie dit VOR-bulletin)
5. Jaarrekening 1989 en verslag kascontrolecommissie (De jaarrekening kan worden opgevraagd bij de penningmeester van de VOR, dr. K. de Glopper, tel. 020-5550356)
6. VOR-gedragscode en reglement van Toezicht en Beroep (de gedragscode en het reglement kunnen worden opgevraagd bij de secretaris van de VOR, dr. M.P.C. van der Werf, tel. 050-636657)
7. Toekomstige structuur van de VOR
8. W.v.t.t.k.

Beknopt verslag verenigingsjaar 1989

1. Bestuurssamenstelling

Het bestuur was in 1989 als volgt samengesteld:

Prof. Dr. J. Claessen	: voorzitter
Dr. J. van Hout	: vice voorzitter
Dr. M.P.C. van der Werf	: secretaris
Dr. E. Roede	: penningmeester
Drs. K. Beintema	: lid
Dr. W. Gerritsen van der Hoop	: lid
Prof. Dr. G. Kanselaar	: lid
Prof. Dr. J. Scheerens	: lid

2. Onderwijs Research Dagen

De Onderwijs Research Dagen vonden plaats op 24 en 25 mei te Leiden. Prof. Dr. F.K. Kievit was voorzitter van de organisatiecommissie. Gastspreker was Prof. Dr. E. de Corte van de Katholieke Universiteit Leuven. Het aantal gepresenteerde papers bedroeg ruim 200.

3. Tijdschrift voor Onderwijs Research/Vor-bulletin

Het redactiesecretariaat van het Tijdschrift voor Onderwijs Research werd in 1989 voortgezet door Prof. Dr. B. Creemers. Het secretariaat van het VOR-bulletin werd gevoerd door de secretaris van het VOR-bestuur. Het bulletin vervulde voornamelijk de functie van mededelingenblad voor het bestuur. In totaal verschenen in 1989 vijf nummers van TOR/VOR-bulletin.

4. VOR aanmoedigingsprijs

De jury, onder leiding van Prof. Dr. W. van der Linden, kende de prijs van f 1000,- toe aan drs. H.J.A. Biemans. De titel van het winnende paper luidde: 'Effecten van een meta-cognitief trainingsprogramma'. Het paper verscheen als artikel in het Tijdschrift voor Onderwijs Research van december 1989.

5. Gedragscode voor onderwijsonderzoekers

De werkgroep gedragsregels bestaande uit Drs. S. Blaauwboer, Drs. K. Beintema, Dr. E. Roede en Drs. Ph. Groeneweg heeft in 1989 alle commentaren op de code, gegeven door het bestuur en de instituten en vakgroepen, verwerkt tot een afgeronde code die in november 1989 kon worden aangeboden aan de ledenvergadering. De ledenvergadering heeft de code met een aantal amendementen geaccepteerd en aangedrongen op een spoedige regeling voor een kamer van Beroep en een Commissie van Toezicht. Deze regeling wordt gepresenteerd op de ledenvergadering tijdens de ORD 1990.

6. *OTG/ANGO onderzoek*

Naar aanleiding van het besluit van SVO om de subsidiegelden voor OTG en ANGO gelden op een andere wijze te kanaliseren heeft het VOR-bestuur uitvoerig met de SVO overlegd om de gemeenschap van onderwijsonderzoekers te betrekken bij de nieuwe opzet. Dit heeft geresulteerd in een voorstel om te komen tot een of meer werkverbanden van onderzoekers binnen de VOR.

De toekomstige interne structuur van de VOR wordt op dit moment nader uitgewerkt. Hiertoe wordt overleg gevoerd met de OTG's en met werkgemeenschappen binnen Pedon. Op de ledenvergadering tijdens de ORD zullen nadere mededelingen worden gedaan.

7. *Subsidiëren congressen en symposia*

Het bestuur heeft in 1989 besloten een subsidie toe te kennen aan het Landelijk Overleg Studievaardigheden voor de organisatie van een symposium met als thema 'Economie van het Studeren'. Deze subsidie is betaald uit het jaarlijkse bedrag op de begroting ten behoeve van activiteiten van de leden.

8. *Ledenwerving*

Per 1 januari 1990 bedroeg het aantal leden van de VOR 445. Dit aantal komt vrijwel overeen met het aantal per 1 januari 1989.

9. *Financiën*

Zie kasverslag 1989 van de penningmeester van de VOR.

Ontvangen publikaties

- Beemster, J., Grift, W. van de, Jacobs, J., Koning, B. de & Schrijver-Witteveen, L. (Red.) (1989). In de klas. Praktijkboek voor effectief onderwijs. Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Dietvorst, C. & Mahieu, P. (Red.) (1989). Organisatiecultuur van scholen. Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Dochy, F.J.R.C., Segers, M.S.R. & Wijnen, W.H.F.W. (Eds.) (1990). Management information and performance indicators in higher education: an international issue. Assen: Van Gorcum.
- Foppen, J.W. (1989). Gistend beleid: veertig jaar universitaire onderwijspolitiek. 's-Gravenhage: VUGA.
- Franssen, H.A.M., Lagerweij, N.A.J. & Stokking, K.M. (1990). Leraren basisschool over veranderingen in het onderwijs. Utrecht: ISOR.
- Jong, O. de (1990). Rekenen aan reacties. Een didactische studie van chemische berekeningen in begin bovenbouw havo/vwo. Proefschrift RU Utrecht.
- Jong, T. de (Ed.) (1989). Computer simulations in an instructional context. First intermediate report on design components for simulations. DELTA project SAFE. Eindhoven: University of Technology, Department of Philosophy and Social Sciences.
- Leseman, P.P.M. (1989). Structurele en pedagogische determinanten van schoolloopbanen. Proefschrift EU Rotterdam. Rotterdam: Project Onderwijs en Sociaal milieu.
- Linden, W.J. van der & Kamp, L.J.Th. van der (Red.) (1989). Meetmethoden en data-analyse. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de onderwijsresearch no. 26).
- Nys, F. (1990). Probleemoplossend denken. Geel: Campinia.
- Putten, C.M. van & Beishuizen, M. (Red.) (1989). Informatietechnologie en media. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de onderwijsresearch no. 27).
- Reints, A., Driehuis, E. & Verberg, H. (1989). Het gebruik van het programma 'Allemaal Taal'. Utrecht: ISOR.
- Schoonman, W. (1989). An applied study on computerized adaptive testing. Proefschrift RU Groningen. Amsterdam/Lisse: Swets & Zeitlinger.
- Sterringa, B. & Petit, C. (1990). Seksespecifieke interactie in de onderwijsleersituatie. Eindverslag met bijlagen. Nijmegen: IMV-ISIO.
- Stichting Centrum voor Onderwijsonderzoek (1989). Jaarverslag 1988. Amsterdam: SCO.
- Tomic, W. & Sijde, P.C. van der (1989). Changing teaching for better learning. Amsterdam/Lisse: Swets & Zeitlinger.
- Versloot, A.M. (1990). Ouders en vrijheid van onderwijs. Schoolkeuze in de provincie Utrecht. Proefschrift RU Utrecht. De Lier: Academisch Boeken Centrum.

Oproep

Sinds twee jaar is er binnen NWO een eigen stichting voor pedagogisch onderzoek actief namelijk **PEDON**.

Binnen de Stichting PEDON is een **Werkgemeenschap Onderwijsonderzoek** opgenomen. Deze werkgemeenschap wil de forumfunctie tussen onderwijsonderzoekers behartigen naast het bevorderen van onderzoek ten aanzien van onderwijsvraagstukken (vanuit een pedagogische optiek).

De werkgemeenschapscommissie nodigt ieder uit die wetenschappelijk onderzoek verricht op het terrein van onderwijsvraagstukken om deel te nemen aan de werkgemeenschap.

De werkgemeenschap kent de volgende inhoudelijk georiënteerde onderzoeksgroepen

- **Opleiding en leerkrachtgedrag** (coördinatie: Th. Wubbels)
- **Curriculum en didaxologie** (coördinatie: W. Hoebe)
- **Schoolorganisatie** (coördinatie: N. Lagerweij)

Opgave is mogelijk bij de secretaris van de werkgemeenschap Dr. H. H. Tilema

RION Instituut voor Onderwijsonderzoek
Postbus 1286
9701 BG GRONINGEN

Het oog van de meester: opstelbeoordeling door schrijvers en leerkrachten

J.B. Hoeksma en K. de Glopper

Vrije Universiteit¹ en Universiteit van Amsterdam²

ABSTRACT

Research into the reliability and validity of essay ratings has given little attention to the existence of individual differences between raters. In this study we investigate the effect of between-rater differences in writing ability. Contrary to the hypothesis our results indicate that professional writers are no better judges than primary school teachers or well educated laymen.

INLEIDING

Aan het beoordelen van opstellen kleven problemen. Beoordelaars blijken maar in beperkte mate stabiel te zijn: bij beoordeling van dezelfde teksten op twee verschillende momenten blijkt de samenhang tussen de oordelen van het eerste en het tweede moment meestal verre van perfect. Ook de overeenstemming tussen verschillende beoordelaars laat vaak te wensen over. In veel gevallen voldoen de intra- en interbeoordelaarsovereenstemming niet aan de gangbare psychometrische eisen (Wesdorp, 1981).

De gebrekkige betrouwbaarheid kan het gevolg zijn van uiteenlopende effecten (De Groot, 1961; Wesdorp, 1981). Beoordelaars kunnen de beoordelingstaak verschillend opvatten. Er is dan sprake van het signifisch effect. Het halo-effect doet zich voor wanneer opvallende andere kwaliteiten dan de te beoordelen kwaliteit een storende uitstraling hebben. Bij het sequentie-effect wordt de beoordeling van een opstel door de nawerking van voorafgaande oordelen verstoord. Vrijheid in de beoordelingsschaal kan aanleiding geven tot normverschuiving (aanpassing van de score-verdeling aan de kwaliteit van het te beoordelen materiaal) of persoonlijke vergelijkingen (persoonlijke neiging tot het geven van scores met een bepaalde verdeling). Tenslotte is er sprake van het contaminatie-effect wanneer beoordelaars met de beoordeling niet alleen een objectief oordeel, maar ook andere doelen nastreven.

In de literatuur zijn diverse remedies voor de betrouwbaarheidsproblemen voorgesteld. Vrijwel alle maatregelen hebben betrekking op de taak van de beoordelaar. Met behulp van analytische schema's (Wolowitsj-Schelvis, 1975) wordt de beoordelingstaak gespecificeerd, ter bestrijding van het signifisch effect en het halo-effect. Met beoordelingsschalen, die bestaan uit voorbeeldopstellen (Hoeksma, 1985; Blok en Hoeksma, 1986), wordt het hoofd geboden aan normverschuiving en persoonlijke vergelijkingen. Door het inschakelen van onafhankelijke en ongebonden beoordelaars worden contaminatie-effecten vermeden. Sequentie-effecten worden voorkomen door variatie in de volgorde waarin de opstellen aan beoordelaars worden aangeboden.

In het streven naar verhoging van de betrouwbaarheid is naar verhouding weinig onderzoek verricht naar individuele verschillen tussen beoordelaars. Dat is opmerkelijk,

¹ Vakgroep Kinder- en Jeugdpsychologie, Van der Boechorststraat 1, 1082 BT Amsterdam

² Stichting Centrum voor Onderwijsresearch, Grote Bickersstraat 72, 1013 KS Amsterdam

want wellicht zijn sommige beoordelaars beter toegerust voor hun taak dan andere. Wel is meermalen getracht typen beoordelaars te onderscheiden (b.v. Cast, 1939, 1940; Morrison and Vernon, 1941; Diederich, French and Carlton, 1961; Britton, Martin and Rosen, 1966; Van Peet, 1970; Meuffels, 1985). In het algemeen wordt in de aangehaalde onderzoeken uitgegaan van de intercorrelatiematrix die resulteert uit de oordelen van een aantal beoordelaars. Op de correlatiematrix wordt een factoranalyse toegepast. Na rotatie worden de factorladingen geïnterpreteerd. Een tekortkoming van deze onderzoeksstrategie is dat beoordelaarstypen, zonder voorafgaande verwachtingen of theorie, post hoc, na het verzamelen van de gegevens, worden onderscheiden. De resultaten zeggen vermoedelijk dan ook meer over het interpretatievermogen van de onderzoekers dan over het bestaan van verschillende beoordelaarstypen.

Wanneer we wel proberen om vooraf beoordelaarstypen te onderscheiden, moeten we de volgende vraag beantwoorden. Waardoor zouden sommige beoordelaars beter toegerust zijn voor hun taak dan andere? Om wat meer licht op deze vraag te werpen, kunnen we te rade gaan bij de literatuur over schrijfprocessen (Hayes and Flower, 1980; Humes, 1983; Hillocks, 1986). Uit het verrichte onderzoek blijkt keer op keer dat goede schrijvers in vergelijking met slechte schrijvers hun eigen teksten beter reviseren. Goed reviseren vereist kennis over en toepassing van kwaliteitscriteria voor teksten. Met andere woorden: om goed te kunnen schrijven moet men zijn eigen teksten goed kunnen beoordelen. Het is een kleine stap om te veronderstellen dat goede schrijvers ook andersmans teksten goed kunnen beoordelen. Dat, met andere woorden, betere schrijvers betere beoordelaars zijn. Door hun kennis over kwaliteitscriteria voor teksten zullen goede schrijvers, zo veronderstellen wij, minder gevoelig zijn voor het signifiësch effect en het halo-effect.

LEERKRACHTEN, SCHRIJVERS EN LEKEN

In het onderwijs en in onderzoek worden opstellen doorgaans beoordeeld door leerkrachten. Wij veronderstellen dat zij in doorsnee over een gemiddelde stelvaardigheid beschikken. Schrijvers en dichters, zo nemen wij aan, beschikken over een grotere stelvaardigheid dan onderwijzers en docenten. Zeker wanneer het schrijvers van naam betreft.

In het onderzoek waarover wij hier rapporteren hebben wij de opsteloordeelen van schrijvers en leerkrachten onderling vergeleken. Op grond van de voorgestelde hypothese verwachtten wij dat schrijvers beter zijn toegerust voor hun taak dan onderwijzers. Wanneer de hypothese juist is, zullen de oordelen van de schrijvers onderling beter overeenstemmen en stabiel en meer valide zijn.

Wij hebben ons vooralsnog beperkt tot de voorspelling betreffende de overeenstemming tussen beoordelaars. De reden daarvoor is dat betrouwbaarheidsonderzoek logisch vooraf gaat aan validiteitsonderzoek. De beperking tot de interbeoordelaarsovereenstemming berust ook op een praktische overweging: wij verwachtten dat het verzamelen van herhaalde metingen bij de specifieke groep proefpersonen op bezwaren zou stuiten.

Aan onze toets van de hypothese dat een grotere schrijfvaardigheid leidt tot beter overeenstemmende opsteloordeelen, kleeft één belangrijk probleem. Schrijvers hebben minder ervaring dan docenten in het beoordelen van opstellen. Hun gebrek aan ervaring zou de 'betrouwbaarheidsvoorsprong' die wij hun op grond van hun schrijfvaardigheid toedichten, wel eens teniet kunnen doen. Om te controleren voor mogelijke effecten van ervaring hebben wij aan de vergelijking een derde groep toegevoegd: hoger opgeleide leken. De schrijfvaardigheid van deze groep wijkt naar verwachting niet veel af van die van de leerkrachten. Net als de schrijvers hebben de leken weinig of geen ervaring met het beoordelen van opstellen.

MATERIAAL EN METHODE

Zestig schrijvers en dichters van naam, een zelfde aantal onderwijzers uit het achtste leerjaar van de basisschool en dertig leken ontvingen per post vijftien opstellen met het verzoek deze te beoordelen. Deze steekproefomvang werd groot genoeg geacht om per groep in ieder geval tien bereidwillige beoordelaars te vinden.

Wij vroegen de beoordelaars om globale oordelen over de inhoud en over de stijl van de opstellen. De oordelen moesten uitgedrukt worden in schoolcijfers van 1 tot en met 10. De instructies luiden als volgt.

Wilt u de inhoud en stijl van de opstellen beoordelen? Bij inhoud kunt u bijvoorbeeld letten op: originaliteit, spanning, saaiheid, leuke wendingen en een aardig slot. Bij stijl kunt u letten op: variatie in zinsbouw, overgangen tussen zinnen, woordkeus en toon.

De instructie werd met opzet vrijblijvend gehouden. Wij veronderstelden dat zulke instructies 'natuurlijk' beoordelingsgedrag zouden ontlokken. Bovendien vreesden we dat sommige beoordelaars zich niet in gedetailleerde voorschriften zouden kunnen vinden. De te beoordelen opstellen droegen als titel "De Wandelende Boom" en waren afkomstig uit de in 1983-1985 uitgevoerde voorstudie "Periodieke Peiling van het Onderwijsniveau in Nederland" (Wesdorp e.a., 1986). Het betrof hier, zoals de titel al suggereert, fantasieverhalen. De vijftien opstellen waren zo geselecteerd dat zij een getrouwe afspiegeling vormden van de gehele prestatieverdeling van leerlingen in de achtste groep.

Ter voorkoming van sequentie-effecten werden de opstellen in twaalf verschillende aselechte volgordes aan de beoordelaars aangeboden. Om de beoordeling te vergemakkelijken werden de opstellen in getypte vorm aangeboden. Spel- en interpunctiefouten werden daarbij letterlijk overgenomen. De lay-out van de getypte opstellen weerspiegelde de indeling van de originelen.

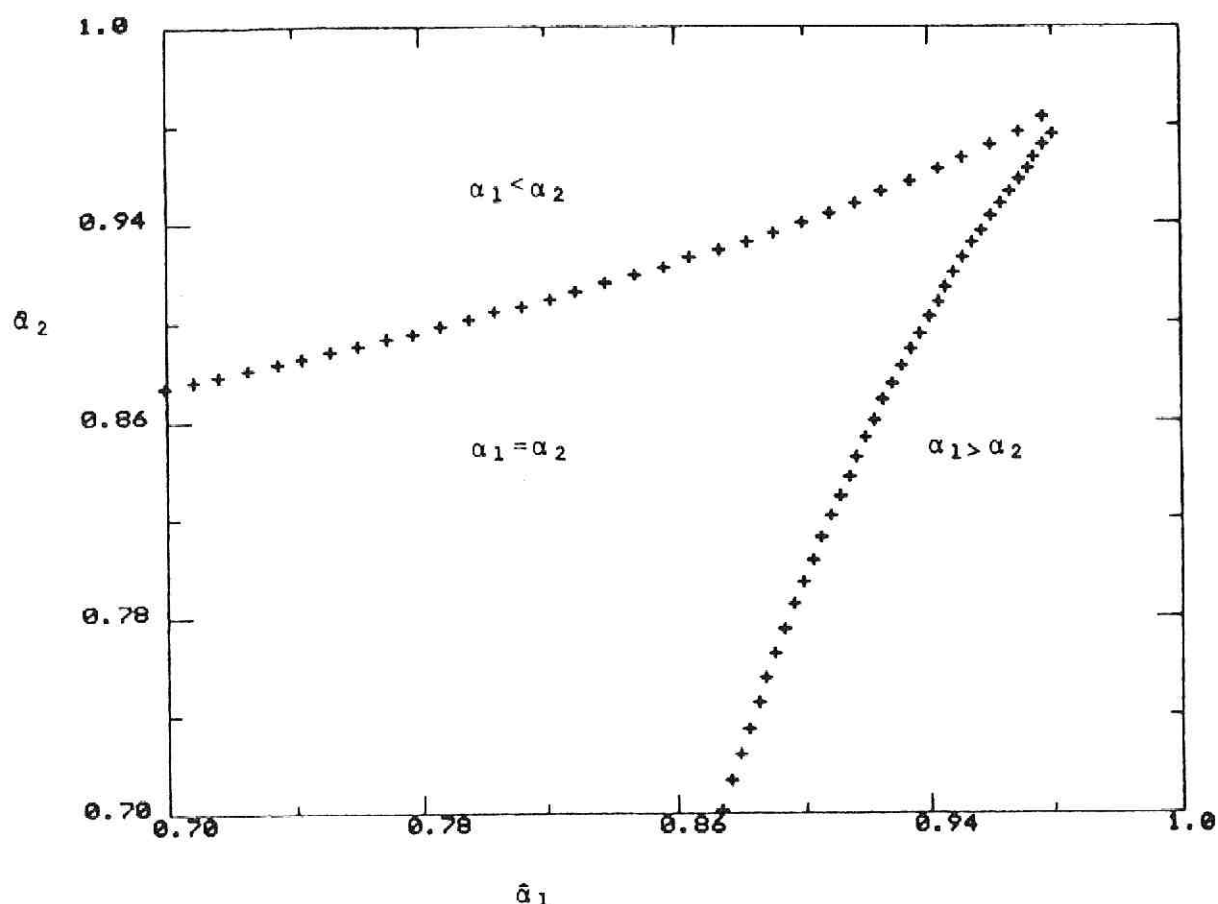
De overeenstemming binnen de verschillende groepen beoordelaars hebben wij vergeleken met een toets voor coëfficiënt alpha (Feldt e.a., 1987). Deze toets houdt rekening met onderlinge afhankelijkheid van de alpha-coëfficiënten; deze onderlinge afhankelijkheid ontstaat doordat de alpha's op dezelfde personen (i.c. opstellen) zijn gebaseerd. De toetsingsgrootte volgt de t-verdeling met $N-2$ vrijheidsgraden (N is het aantal items in de toets, c.q. het aantal beoordelaars) en wordt berekend als:

$$t = \frac{(\hat{\alpha}_1 - \hat{\alpha}_2)(N-2)^{1/2}}{[4(1-\hat{\alpha}_1)(1-\hat{\alpha}_2)(1-\hat{\rho}^2)]^{1/2}}$$

waarin $\hat{\rho}$ de correlatie is tussen de twee somscores van de twee toetsen (c.q. de jury-oordelen of gemiddelde oordelen van groepen beoordelaars).

Het bovengenoemde minimum van tien beoordelaars per groep achtten wij adequaat op grond van enige berekeningen met bovenstaande formule. De berekeningen bestonden uit het substitueren van waarden voor $\hat{\alpha}_1$, $\hat{\alpha}_2$ en N , met $\hat{\rho} = \sqrt{\min(\hat{\alpha}_1, \hat{\alpha}_2)}$. Figuur 1 toont de contouren van de verzamelingen uitkomsten die leiden tot de conclusies $\hat{\alpha}_1 < \hat{\alpha}_2$, $\hat{\alpha}_1 = \hat{\alpha}_2$ of $\hat{\alpha}_1 > \hat{\alpha}_2$, bij tweezijdige toetsing met een significantieniveau van .05 en bij tien beoordelaars per groep.

Uit de figuur blijkt dat bij een hoge waarde van de te vergelijken alpha's een klein verschil reeds tot verwerping van de nulhypothese leidt. Bij jury's van tien beoordelaars mogen alpha's van .80 en hoger verwacht worden.



Figuur 1. Waarden voor $\hat{\alpha}_1$ en $\hat{\alpha}_2$ die leiden tot verwerping van de nulhypothese $\hat{\alpha}_1 = \hat{\alpha}_2$ bij tweezijdige toetsing met $p = .05$ en tien beoordelaars per groep.

RESULTATEN

Elf van de zestig aangeschreven schrijvers voldeden aan ons verzoek de opstellen te beoordelen. Voor de leerkrachten was dit aantal vijftien en voor de leken drieëntwintig. Met toestemming van de beoordelaars zijn de namen van de schrijvers, leerkrachten en leken in een bijlage vermeld. Vrijwel alle leken bleken een opleiding op HBO- of universitair niveau afgerond te hebben.

Voor de door ons gekozen wijze van vergelijken van de onderlinge overeenstemming in de drie groepen zijn gelijke aantallen beoordelaars per groep gewenst. Daarom werden zowel uit de groep onderwijzers als uit de groep leken met behulp van een lijst aselechte getallen elf beoordelaars getrokken. Per groep is vervolgens de onderlinge overeenstemming tussen de beoordelaars bepaald met behulp van de homogeniteitscoëfficiënt alpha. De homogeniteit van de inhouds- en stijloordelen van de drie groepen staat vermeld in Tabel 1.

Tabel 1. Cronbachs alpha voor inhouds- en stijloordelen van schrijver, leerkrachten en leken

groep	inhoud	stijl
schrijvers	.72	.87
leerkrachten	.94	.95
leken	.92	.95

Tweezijdige toetsing bij een significantieniveau van .01, in plaats van .05 vanwege het aantal uitgevoerde vergelijkingen, had het volgende resultaat. De inhoudsoordelen van de schrijvers zijn minder homogeen dan die van de leerkrachten ($t=-5.63$, $df=9$; $p<.01$) en van de leken ($t=-5.33$, $df=9$; $p<.01$). De homogeniteit van de inhoudsoordelen van de leerkrachten en leken verschilt onderling niet ($t=1.31$, $df=9$; n.s.). De vergelijking van de homogeniteit van de gezamenlijke stijloordelen geeft overeenkomstige resultaten. De stijloordelen van de schrijvers zijn significant minder homogeen dan die van de leerkrachten ($t=-4.12$, $df=9$; $p<.01$) en de leken ($t=-4.25$, $df=9$; $p<.01$), terwijl die van de leerkrachten en leken niet verschillen ($t=.23$, $df=9$; n.s.).

Coëfficiënt alpha geeft nog geen inzicht in de structuur van de gezamenlijke oordelen in de drie onderscheiden groepen. De structuur van de oordelen en de betrouwbaarheid van individuele beoordelaars kunnen evenwel worden nagegaan aan de hand van meetmodellen, zoals beschreven door Jöreskog (1971). (Zie ook Lord and Novick, 1968). Deze modellen zijn gebaseerd op de klassieke decompositie van de geobserveerde variantie in ware en fouten-variantie.

Toegespitst op beoordelaars kunnen drie modellen worden onderscheiden. Op beoordelaars is het parallelle meetmodel van toepassing als zowel de ware score-varianties als de fouten-varianties van de afzonderlijke beoordelaars gelijk zijn. In dat geval zijn beoordelaars onderling volstrekt uitwisselbaar: zij gebruiken dezelfde schaal om hun oordeel uit te drukken en hun individuele betrouwbaarheden zijn gelijk. Er is sprake van tau-equivalente beoordelaars als de ware score-varianties gelijk zijn, terwijl de fouten-varianties verschillen. In dit geval gebruiken de beoordelaars dezelfde schaal, maar hun individuele betrouwbaarheden verschillen. Voor congenerieke beoordelaars geldt tenslotte dat zowel de ware score-varianties als de fouten-varianties tussen beoordelaars verschillen. In dit laatste geval gebruiken de beoordelaars verschillende schalen en verschillen hun individuele betrouwbaarheden. Alle drie meetmodellen veronderstellen één gemeten dimensie of factor.

De parameters in de genoemde modellen zijn geschat met Lisrel VI (Jöreskog and Sörbom, 1986). Daarbij werd gebruik gemaakt van de ongewogen kleinste kwadraten methode; bij gebruik van deze methode hoeft niet verondersteld te worden dat de gezamenlijke verdeling van de oordelen multivariaat normaal is. De passing van de modellen is beoordeeld aan de hand van de 'adjusted goodness of fit index' (agfi). De index is een maat voor de hoeveelheid door een model verklaarde variantie en covariantie, gecorrigeerd voor het aantal vrijheidsgraden van het model. Wij hanteren hier als vuistregel dat bij een agfi van .90 of groter van een goede modelpassing sprake is. Overigens moeten we opmerken dat het beperkte aantal waarnemingen (15 opstellen per beoordelaar) ons tot voorzichtigheid maant bij de interpretatie en generalisatie van de toetsingsresultaten.

Tabel 2 geeft een overzicht van de waarden van de agfi voor de drie modellen voor de inhoudsoordelen van de leerkrachten, schrijvers en leken. Uit de gegevens blijkt dat het parallelle model voor geen van de drie groepen beoordelaars past. Hetzelfde geldt met betrekking tot het tau-equivalente model. Het congenerieke model beschrijft de inhoudsoordelen van de leerkrachten en leken echter buitengewoon goed. Voor de schrijvers daarentegen

Tabel 2. 'Adjusted goodness of fit' indices voor parallelle, tau-equivalente en congenerieke meetmodellen voor inhoudsoordelen

groep	parallel	tau-eq.	congeneriek
schrijvers	.42	.55	.84
leerkrachten	.78	.83	.98
leken	.77	.81	.97

is ook de passing van het congenerieke model onder de maat.

De passing van de modellen voor de stijloordelen, weergegeven in Tabel 3, geeft een iets ander beeld. De parallelle en de tau-equivalente modellen geven voor geen van de drie groepen beoordelaars een acceptabele passing, terwijl het congenerieke model de oordelen van de leerkrachten, de leken en de schrijvers adequaat beschrijft. Voor de leerkrachten en de leken is de passing van het congenerieke model opnieuw uitzonderlijk goed.

Tabel 3. 'Adjusted goodness of fit' indices voor parallelle, tau-equivalente en congenerieke meetmodellen voor stijloordelen

groep	parallel	tau-eq.	congeneriek
schrijvers	.65	.74	.92
leerkrachten	.82	.86	.99
leken	.86	.88	.99

Ten slotte gaan wij na hoe de gemiddelde inhouds- en stijloordelen van de leerkrachten, de schrijvers en de leken onderling samenhangen. Tabel 4 geeft een overzicht van de correlaties tussen de gemiddelde oordelen voor inhoud en voor stijl. De correlaties zijn voor zowel inhoud als stijl zonder uitzondering hoog. Verder geldt dat de correlaties tussen de inhouds- en de stijloordelen binnen de groepen tevens hoog zijn: .92 (schrijvers), .97 (leerkrachten) en .92 (leken).

Tabel 4. Correlaties tussen de gemiddelde oordelen voor inhoud en stijl van schrijvers (Sch), leerkrachten (Lkr) en leken (Lek)

	Inhoud Sch	Lkr	Stijl Sch	Lkr
Lkr	.90		.94	
Lek	.93	.96	.94	.97

CONCLUSIE EN DISCUSSIE

De resultaten kunnen in vier punten worden samengevat:

1. De inhouds- en stijloordelen van de schrijvers zijn minder homogeen dan die van de leerkrachten en leken.
2. De inhoudsoordelen van de leerkrachten en de leken voldoen aan het congenerieke meetmodel, die van de schrijvers daarentegen niet. Dat betekent dat de leerkrachten zowel als de leken een gemeenschappelijk oordeel hebben over de inhoud van de opstellen. Schrijvers blijken geen gemeenschappelijk inhoudsoordeel te hebben.
3. De stijloordelen van zowel leerkrachten, schrijvers als leken voldoen aan het congenerieke model. Binnen ieder van de drie groepen bestaat een gedeeld oordeel over de stijl van de opstellen.
4. De gemiddelde inhoudsoordelen van de schrijvers, leerkrachten en leken hangen onderling sterk samen. Hetzelfde geldt voor de stijloordelen. De correlaties tussen de gemiddelde inhouds- en stijloordelen binnen de drie groepen zijn hoog.

De hypothese dat beoordelaars beter voor hun taak zijn toegerust naarmate zij over een grotere schrijfvaardigheid beschikken, wordt door de resultaten niet bevestigd. Het tegendeel

lijkt eerder waar. Leerkrachten en leken blijken beter voor de beoordelingstaak berekend dan schrijvers. Alvorens we enige verklaringen voor dit resultaat naar voren brengen willen we enkele kanttekeningen plaatsen bij de gevonden overeenstemming binnen de groepen en de samenhang tussen inhouds- en stijloordelen.

Leerkrachten en leken blijken met name de stijl van de opstellen redelijk betrouwbaar te kunnen beoordelen. Gebruikmakend van de Spearman-Brown formule kan uit Tabel 1 de betrouwbaarheid van een enkele beoordelaar worden berekend. Voor de leerkrachten en leken is de individuele betrouwbaarheid van het stijloordeel .68. De verklaring van deze naar verhouding hoge betrouwbaarheid moet worden gezocht in de spreiding in kwaliteit van de opstellen. De aangeboden opstellen waren immers zo geselecteerd dat zij een getrouwe afspiegeling vormden van de gehele range van schrijfsprestaties van leerlingen in de achtste groep.

De geconstateerde hoge correlatie tussen inhoud en stijl is niet ongewoon, gezien ander onderzoek (Blok en Hoeksma, 1986; De Glopper, 1988; Van den Bergh, 1988). De vraag in hoeverre de hoge overeenstemming tussen inhoud en stijl een beoordelaarsprobleem of een werkelijke empirische samenhang weerspiegelt, kunnen ook wij niet beantwoorden.

Voor de resultaten van de hypothesetoetsing kan een aantal verklaringen worden geopperd. Wij bespreken er hier enkele. Een eerste, mogelijke verklaring is hierboven reeds geïntroduceerd. Zij luidt dat de leerkrachten door hun ervaring met het beoordelen van opstellen hun veronderstelde achterstand in een voorsprong ombuigen. Deze verklaring lijkt ons niet adequaat, omdat de leken ondanks hun gebrek aan beoordelingservaring niet onderdoen voor de leerkrachten.

Een tweede verklaring luidt dat leerkrachten en leken bij het beoordelen van opstellen meer af gaan op oppervlakkige, gemakkelijk herkenbare kenmerken. Zo behalen zij een grotere overeenstemming dan schrijvers, die meer letten op creativiteit en subtiële stijlenmerken. Deze verklaring mag wellicht opgaan voor een enkele individuele beoordelaar, voor de groepen als geheel snijdt zij echter geen hout. De gemiddelde groepsoordelen stemmen bijna perfect overeen.

Het meest plausibel lijkt ons de verklaring dat de opvattingen van de schrijvers juist ten gevolge van hun vakmanschap uiteenlopen. Wij citeren uit een brief die de beoordelingen van een van de schrijvers vergezelde: "Schrijven is een vak dat moeizaam geleerd wordt en in de praktijk ontwikkeld moet worden.(..) Het is een ingewikkeld en langdurig proces tijdens welk de auteur gaandeweg criteria opstelt voor zichzelf en anderen. Met als gevolg: sterk afwijkende meningen." Zo bezien lijkt het significatief effect bij de schrijvers een belangrijke rol te spelen. De uiteenlopende opvattingen over de beoordelingstaak lijken juist deel uit te maken van het schrijverschap.

Vermoedelijk treedt het significatief effect vooral op bij verhalende teksten en globale oordelen die zonder dwingende instructie tot stand komen. In een situatie met minder pluriformiteit in kwaliteitscriteria, bij voorbeeld bij het beoordelen van zakelijke teksten, zou schrijfvaardigheid er mogelijk meer toe kunnen doen. In dat geval echter zou de keuze op andere goede schrijvers kunnen vallen: bij voorbeeld journalisten of schrijvers van beleidsteksten.

Ter illustratie van de eigenzinnige opvattingen van de schrijvers kan ten slotte het volgende citaat dienen: "U denkt misschien dat ik me er van af maak met al mijn tieners, maar dat is niet zo. Ik vind de verhalen buitengewoon vermakelijk en weinig conventioneel." Over de volgende zin, die in de ogen van leerkrachten geen genade zal vinden, is de betreffende schrijver zeer enthousiast: "Een lichte flits en daar loop opeens een boom met een wandelstokje. Mag ik die ergens in gebruiken?"

LITERATUUR

- Blok, H. & Hoeksma, J.B. (1986). *Schaalconstructie voor opstelbeoordeling*. Amsterdam: SCO.
- Bergh, H. van den (1988). *Examens geëxamineerd*. 's Gravenhage, SVO.
- Britton, J.N., Martin, N.C. & Rosen, H. (1966). *Multiple marking of English compositions: An account of an experiment*. Schools Council Examinations Bulletin 12. London: Her Majesty's Stationary Office.
- Cast, B.M.D. (1939). The efficiency of different methods of marking English compositions, I. *British Journal of Educational Psychology*, 9, 257-270.
- Cast, B.M.D. (1940). The efficiency of different methods of marking English compositions, II. *British Journal of Educational Psychology*, 10, 49-60.
- Diederich, P.B., French, J.W. & Carlton, S.T. (1961). Factors in the judgement of writing ability. *Research Bulletin R.B.* 61-65. Princeton N.J.: Educational Testing Service.
- Feldt L.S., Woodruff D.J. & Salih, F.A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement*, 11, 93-103.
- Gopper, K. de (1988). *Schrijven beschreven*. 's-Gravenhage, SVO.
- Hayes, J.R., & L.S. Flower (1980). Identifying the organization of writing processes. In: L.W. Gregg & E.R. Steinberg (eds.). *Cognitive processes in writing*. Hillsdale, N.J., Lawrence Erlbaum, 1980, 3-30.
- Hillocks, G. Jr. *Research on Written Composition. New Directions for Teaching*. Urbana, Illinois, NCRE/ERIC, 1986.
- Hoeksma, J.B. (1985). Opstelbeoordeling met lijnen en getallen. In: W.K.B. Koning (red.) *Taalbeheersing in theorie en praktijk* (362-369). Dordrecht: Foris.
- Humes, A. Research on the composing process. *Review of Educational Research*, 1983, 201-216.
- Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Jöreskog, K.G. & Sörbom, D. (1986). *LISREL: Analysis of linear structural relationships by the method of maximum likelihood* (version VI). Uppsala: University of Uppsala.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading Mass.: Addison-Wesley.
- Meuffels, B. (1985). Het beoordelen van opstellen: dimensioneel of typologisch? In: W.K.B. Koning (red.). *Taalbeheersing in theorie en praktijk* (386-392). Dordrecht: Foris.
- Morrison, R.L. & Vernon, P.E. (1941). A new method of marking English compositions. *British Journal of Educational Psychology*, 11, 109-119.
- Peet, A.A.J. van (1970). *Theorie en praktijk van opstelbeoordeling*. Amsterdam: Research Instituut voor de Toegepaste Psychologie.
- Wolowitsj-Schelvis, J.J. (1975). *Het ontwikkelen van een analytisch beoordelingsschema voor schrijfproducten van MAVO-leerlingen*. Amsterdam, R.I.T.P.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. 's-Gravenhage: Staatsuitgeverij.
- Wesdorp, H., Bergh, H. van den, Bos, D.J., Hoeksma, J.B., Oostdam, R., Scheerens J. & Triesscheijn, B. (1986). *De haalbaarheid van periodiek peilingsonderzoek*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 5-4-1989

Definitieve versie ontvangen 17-5-1990

BIJLAGE

SCHRIJVERS

J. Herzberg, T. Beckmann, J. Mendels, M. Bloem, A.H. Roemer, A. Schouten, M. Luccioni, M. van Keulen, W.F. Hermans, M. Höweler, B. den Uyl.

LEERKRACHTEN

J. Jong, J. Euwema, C. van Roessel, G. van der Veen, B.M. Rosa Biam, M. Pruis, W.G. Kuiper, M.M. Verwoerd, A. Heida, H. Hilberdink, S. Holwerda, G. Martin, H. Vergeer, H.A. Bijl. (Eén leerkracht wenste anoniem te blijven.)

LEKEN

J. van Dijk, L. Wesseling, E. Engels, B. Felix-Schollaart, M. Robroek, M. Jensma, S. Glaudé, E.T. Fühling, B. Schumacher, M. van den Berg, L. Bossers, P.M. de Glopper, P.J.C.M. Coumans, V. van Daal, F. Hecrema, M. de Jong, G. Kramer, A. Veldhuizen, M. Meerum Terwogt, F. Krooshof, G. Steenvoorden, R. de Sain, M. van der Beek.

An experimental study of the when and what of children's questions

Hans van der Meij*

Department for Educational Research, Rijksuniversiteit Leiden

ABSTRACT

Effects of children's prior knowledge on questioning behavior were examined. Pupils from fifth grade elementary school were presented with a set of words for which they had to indicate the meaning. Pupils could: (a) not ask a question, (b) ask a global question, or (c) ask a specific question. Pupils with little prior knowledge asked more questions, but their questioning behavior was more indiscriminate and less helpful. Pupils with much prior knowledge asked relatively more necessary questions and fewer unnecessary questions. They also tended to be more capable of using specific questions for finding the correct answer. In the discussion question asking is described as the outcome of a three-staged process, beginning with experiencing a problem, followed by determining the needed information and ending with the verbalization of the question. Possibilities for testing this model are discussed.

INTRODUCTION

Consider the following observations taken from an ordinary language lesson in elementary school.

Pupil 1: 'Can you help me?'

Teacher: 'Maybe. What is it that you want from me?'

Pupil 1: 'Eh, what does 'stout' mean?'

Teacher: 'What do you think?'

Pupil 1: '... I really don't know.'

Pupil 2: 'Can you say that someone who is stout weighs 100 pounds or so?'

Teacher: 'Perhaps, you'd better add another 100 pounds.'

Pupil 3: asks nothing at all.

These three pupils all faced the same problem, they did not know the meaning of the word 'stout' and it is interesting to see how they responded to this problem. The pupils 1 and 2 asked a question, pupil 3 did not. The question of pupil 1 is an unspecified request for help. It is a dependency-like reaction. As a result, the teacher dominates the interaction that follows. In contrast, there is the specific question of pupil 2. This pupil's question clearly shows that she knows what stout means and she controls the interaction. Pupil 3 asked nothing at all, although he clearly did not know the word. Why didn't he ask?

It takes various skills to ask a question. For example, to ask a question pupils must become aware of a problem. They must detect a knowledge gap or feel uncertain with what

* Present address: Department for Instructional Technology, Universiteit Twente, Postbus 217, 7500 AE Enschede.

they know. Asking a *specific* question goes beyond awareness. It requires skills in handling the problem or knowledge gap to determine what information is needed. Questioning also depends upon communicative skills. Pupils need to express their question so that the respondent understands its purpose.

There have been various attempts to train pupils to ask questions (e.g. Wong, 1985; Dillon, 1988) but there are not many extensive studies describing how questions arise and how they are being processed (Van der Meij, 1986). In the study that is reported here the central question is 'What are the effects of prior knowledge on question asking?'. More particularly, it is studied how prior knowledge affects the number and kinds of questions fifth-grade children ask, and how it affects the pragmatic properties of their questions.

Asking a question

Miyake & Norman (1979) were the first to study the relationship between prior knowledge and the number of questions asked. They suggested and found an inverted U-curve. People with less than average prior knowledge, the authors argued, are expected to ask fewer questions because they lack a proper framework or schema for asking questions. People with more than average prior knowledge ask fewer questions because they are likely to possess most of the relevant knowledge or are considered more capable of inferring or filling their knowledge gaps by guessing rather than by asking.

None of the studies that followed supported Miyake & Norman's hypothesis. Mostly, a linear and negative relationship was obtained. Less knowledgeable subjects always asked more questions than did more knowledgeable ones (e.g. Flammer, Kaiser, & Müller-Bouquet, 1981; Flammer, Grob, Leuthardt, & Lüthi, 1984; Smith, Tykodi, & Mynatt, 1988). It is suggested that there are two reasons for the discrepant findings: the participants' level of knowledge and the provision of immediate responses to questions (Flammer, 1986). Most of the experimental subjects were not absolute beginners. They often knew something of a topic, or they could dig up some relevant knowledge. In addition, unlike Miyake & Norman, the subjects received immediate responses to their questions which may have activated new schemas and thus bring along new questions. This is particularly so for subjects with little prior knowledge. Moreover, Fuhrer (1987, 1989) suggests that in naturalistic conditions the effects of prior knowledge are less clear and probably strongly affected by social considerations.

The subjects in all of these studies were adults. There are only two studies with pupils from elementary school. Nelson-LeGall & Glor-Scheib (1985) found a significant effect of prior knowledge on number of questions asked during math lessons but not for reading. Pupils with little mathematical knowledge asked more questions than pupils with an average or a high level. In an experimental study Van der Meij (in press) studied the same questioning behavior investigated by Nelson-LeGall & Glor-Scheib under experimental conditions. That is, pupils were presented with a set of words for which they could: (a) not ask a question, (b) ask process-oriented information (i.e. specific question), or (c) ask product-oriented information (i.e. an answer). A negative relationship was found between prior knowledge and number of questions asked.

Asking a particular type of question

Research on adult questioning focuses on the distinction between *closed* (e.g. yes-no) questions and *open* (e.g. who, what, where) questions. Flammer et al. (1981; 1984) hypothesized that more knowledgeable subjects prefer asking closed questions to open ones because they are likely to determine rather precisely the information that they need. But in the four experiments that were conducted no such effect was found. In contrast, Smith et al. (1988) did find that more knowledgeable subjects preferred asking closed

questions. They also discovered that when the novices became more expert, they too started to ask more closed questions. In short, the effect of prior knowledge on asking open or closed questions is an unsettled issue.

Research on pupils' questioning has focused on the distinction between process and product-oriented questions. *Process-oriented* questions supposedly reflect requests for information about the solution process. In contrast, *product-oriented* questions are requests for answers without interest in the 'how' of solution. Nelson-LeGall & Glor-Scheib (1985) found no effect of prior knowledge. Van der Meij (in press) found that pupils with little prior knowledge asked more product questions than did pupils with average or much prior knowledge. It is suggested that these pupils have more fragile knowledge and more fatalistic beliefs about the helpfulness of process information and hence elect to ask more product questions.

Pragmatic properties of questions

There are two issues on the pragmatic properties of questions: necessity and helpfulness. Necessity refers to the appropriateness or pertinence of the questions, helpfulness to their utility.

In many studies *necessity* is not an issue at stake. For question asking by pupils necessity is of obvious importance. Question asking when a child knows the solution or answer should be judged differently from question asking when a child does not know. In order to characterize questions as necessary or unnecessary one should therefore consider the pupil's capacity for coping. Necessity can be assessed only if there is (detailed) information about what a pupil knows or does not know. In the present study necessity is measured in an objective way by means of 'provisional answers'. That is, pupils give a temporary answer in advance to question asking (see Method). Van der Meij (in press) found that prior knowledge negatively effected necessity. This was ascribed to less spontaneous monitoring behavior of pupils with little prior knowledge.

Another important issue on the pragmatic properties of questions is their *helpfulness*. Questions can be effective means for closing a knowledge gap or increasing certainty of knowing. Research on peer interactions shows that helpfulness depends on factors such as the ability of the questioner, the kind of questions asked and the respondent's skills (Webb, 1982). Reported outcomes mostly indicate significant positive, but low, correlations with achievement ($r = 0.20$). Occasionally, negative effects are noted (Webb, 1983; Vedder, 1985). Systematic studies on the helpfulness of questions are wanting. Van der Meij (in press) found no effect of prior knowledge on helpfulness.

This study investigates the effects of prior knowledge on various aspects of questioning. Fifth graders are presented with a set of target words for which they have to indicate the meaning. The pupils are also shown five alternatives. One of these alternatives is the correct answer. For each word pupils can ask a global question, a specific question, or not ask a question. Global questions resemble open questions. They reflect a somewhat unspecified search for information about a target word. Specific questions are like closed questions. They pertain to fewer alternatives than global questions and are presumed to reflect someone's belief in knowing the target word by and large. Question asking is made easy for the pupils in that they select the question that they want to ask instead of producing their own questions. More knowledgeable children are expected to ask fewer questions than are less knowledgeable children. In addition, it is hypothesized that knowledge positively affects the necessity of the questions. For the other two aspects of questioning no hypotheses are advanced.

METHOD

Subjects

Participants in the experiment were 84 fifth-grade children (mean age 11 years exactly) from schools representing various socio-economic levels. Only subjects who ranked in the top 15% and bottom 15% on a vocabulary test were selected as participants. There were 51 pupils in the low knowledge group and 33 in the high knowledge group.

Materials

Prior knowledge is measured with the Stijnen Vocabulary Test (1975). The test contains 61 words, with each word presented in a sentence. The child must select the right answer from four alternatives. For example, item 25 is: 'They did everything *together*'; alternatives are (1) at the same time, (2) as it should, (3) in each other's company, and (4) completely on their own.' Internal consistency (KR 20) is .91.

Self-esteem may affect question-asking of pupils. For example, Van der Meij, Baarends & Leijh (1987) found an effect of self-esteem on repeated requests but not on single requests for assistance. In another study (Van der Meij, 1989) a significant effect was found on the pupils' confidence ratings per word, but no effect on questioning. In view of these outcomes self-esteem is treated as a covariate in the analyses. (Experienced) self-esteem is measured with a standardized test (Mooy, 1987). Test scores reflect the child's self-esteem in language. Exemplary items are: 'Do you have problems with language?' and 'Is language difficult for you?'. The five response alternatives range from never to always. Internal consistency (Cronbach's alpha) is .79. Preliminary analyses for homogeneity of regression revealed that, except for the number of questions asked, this assumption could be met at the .05 level. This indicates that self-esteem does not interact with knowledge in effects on the dependent variables.

Target words for the experimental task are drawn from the Primary Mental Abilities for fifth and sixth grade (Kema, 1978). In a pilot study the quality of the questions (i.e. helpfulness of the hints) was tested, and word difficulty assessed. The experimental task contains 20 target words. The (10) easy words had a mean of 90% correct. The (10) difficult words were answered correctly by 52% of the subjects. Two versions of the set of 20 words were used to control for any order effects.

Procedure

Under the direction of their teacher, all pupils completed the vocabulary and self-esteem test on two different days. A period of about three weeks expired during which outcomes were scored and participants for the experiment selected. Each pupil was escorted to a quiet place in the school. Once acquainted with the experimenter the following introduction was given: 'I want to know whether children of your age know the meaning of certain words. I will give you a word and five alternatives. One of the alternatives means exactly the same as the word that I give you. You must tell me which word that is.' This introduction is followed by 4 consecutive steps:

Step a: The target word is verbalized and a typed card showing the word and five alternatives presented. The pupil reads aloud the five alternatives. For example:

Exp.: Verbalizes the target word: To imitate.

Pupil: Reads aloud the alternatives

(a) To play

(b) To mimic

(c) To cry

- (d) To speak
- (e) To feign

Step b: The pupil must give an answer. The experimenter reacts by saying that that answer will be called the Provisional Answer. For example:

Pupil selects: 'To feign'

Exp.: 'We will call this your Provisional Answer.'

Step c: Pupils who do not ask a question give their Final Answer right away (i.e. proceed directly with step d). The pupils can ask for a global hint or a specific hint. Global hints typically exclude two alternatives. They always begin with: 'It has something to do with...'. Specific hints show the target word in a sentence. These hints exclude three alternatives. Both hints are typed on cards, to minimize memory load. There are blue cards for global hints, and yellow cards for specific hints. The pupil is told that he or she can ask for only one hint, not for both. Two examples of target words, alternatives and hints are given below.

Target word 1: Tardy. Alternatives: sleepy, slow, fast, crawling and snail. Global hint: 'It has something to do with speed'. Specific hint: 'A hare can be quick and a turtle can be *tardy*.'

Target word 2: Hesitantly. Alternatives: softly, nicely, fear, hoarse and reluctantly. Global hint: 'It has something to do with unsure'. Specific hint: 'Hans entered his new classroom *hesitantly*.'

In the introduction each pupil is shown two global hints and two specific hints and it is explained how these hints can be used in finding the correct answer. For example, after showing the global hint for the word hesitantly the following explanation is given: 'The answer can be softly, or fear, or reluctantly because these words have to do with hesitantly. It cannot be nicely and hoarse because they have nothing to do with unsure.' The specific hint is discussed as follows: 'If you say 'Hans ... hesitantly,' the answer can be softly, or reluctantly because you can put these words in the sentence to replace hesitantly. It cannot be fear, or nicely, or hoarse because you don't say 'Hans ... fear, or nicely, or hoarse.''. It is assured that the pupil understands how the hints assist in finding the right answer. The proceedings may be as follows:

Pupil: 'I will give my answer right away.'

Exp.: 'What is your Final Answer?'

Pupil: 'to feign'

or

Pupil: 'May I have a global hint?'

Exp. presents blue card: 'Can you read it out to me?'

Pupil reads: 'It has something to do with not being yourself.'

or

Pupil: 'May I have a specific hint?'

Exp. presents yellow card: 'Can you read it out to me?'

Pupil reads: 'In the play back show girls often *imitate* Madonna.'

Step d: The pupil gives a Final Answer. The experimenter gives no feedback about the Answer(s). In the introduction (s)he says: 'I will not tell you whether your answer is right or wrong'. For example:

Exp.: 'Now you must give your Final Answer.'

Pupil selects: 'To mimic'

The experiment begins as soon as the pupil is familiar with these steps. Instructions are repeated only when a pupil appears to have difficulties with remembering them. Occasionally, pupils asked whether or not they should ask a question. In such a case they were told that they had to figure that out for themselves. No external rewards or punishments were given in reaction to the question asking behavior of the pupils.

Scoring and data-analyses

There are two levels of prior knowledge. Dependent variables are: the number and type of questions asked, and pragmatic properties of questions. An explanation is in order to understand how the latter (i.e. necessity and helpfulness) were calculated.

Necessity is based on 'Provisional Answers' (PA's). Pupils give a PA before they decide about question asking. If the PA is wrong, and a question is asked, the question is considered necessary. If the PA is right, and a question is asked, it is called unnecessary. Pupils with much prior knowledge will know more target words. Necessity is therefore expressed as a percentage. That is, it shows the percentage of necessary (unnecessary) questions given an incorrect (correct) Provisional Answer. Thus, necessary question asking is the total number of necessary questions divided by the total number of incorrect PA's.

Helpfulness shows the utility of the questions. Its calculation also depends on PA's. Helpfulness is shown in: helpful or harmful question asking and in a gain score. Helpful question asking takes place when a pupil gives an incorrect PA, asks a question, and then gives a correct Final Answer (FA). Harmful question asking occurs when a correct PA, is followed by a question, and changed into an incorrect FA. Both measures are expressed as percentages. For example, helpful question asking shows the percentage of correct FA's following necessary questions. In formula: $P(\text{Correct FA} / \text{Question and Incorrect PA})$. The *gain score* shows how many more words (in an absolute sense) pupils answered correctly after questioning. Gain scores are expressed in standard deviations.

The computer program SPSSX is run for all (M)ANCOVA analyses. Preliminary analyses showed that there was no effect of sex. Task difficulty is treated as a within subjects factor and self esteem as a covariate in the analyses. The Pillais statistics is used to indicate the multivariate effects.

RESULTS AND DISCUSSION

Number and kinds of questions asked

A multivariate analysis of covariance (MANCOVA) of the no-questions, global questions and specific questions showed a significant effect of Prior Knowledge, $F(2,80) = 14.28$, $p < .001$. Subjects with little prior knowledge asked more global questions, $F(1,81) = 7.65$,

Table 1. Mean frequency of type of question asked as a function of prior knowledge and difficulty level

	Easy Words		Difficult Words	
	<i>M</i>	<i>S.D.</i>	<i>M</i>	<i>S.D.</i>
Little Prior Knowledge				
Global Questions	2.6	2.2	3.1	2.6
Specific Questions	2.7	2.6	4.9	2.8
Much Prior Knowledge				
Global Questions	0.8	1.2	2.6	2.0
Specific Questions	0.8	1.2	2.5	2.6

$p < .01$ and more specific questions, $F(1,81) = 9.00, p < .01$ (see Table 1). There was a main effect of Task Difficulty, $F(2,162) = 86.41, p < .001$ which was qualified by a Prior Knowledge by Task Difficulty interaction, $F(2,162) = 9.25, p < .001$. Whereas all pupils asked more questions for difficult items, pupils with little prior knowledge asked particularly many more specific questions for difficult items as compared with pupils with much prior knowledge (see Table 1)

A MANCOVA on the preference for a certain type of question revealed no effect of Prior Knowledge, $F(1,81) = 1.83, p > .10$. On average 52% of the questions were global, indicating that there was an even distribution of global and specific questions.

The negative effect of prior knowledge on number of questions asked appears to be rather robust, for it is upheld under the two quite different difficulty conditions. In general, there is no effect of prior knowledge on asking a particular type of question. However, there is an interesting interaction with level of difficulty. Pupils with little prior knowledge adapt their questioning well, perhaps even better than pupils with much prior knowledge, in dealing with difficult items. For such items they do not just ask more questions, they also seem to compensate for their ignorance by asking particularly many specific questions. (These questions are always more informative than global questions.)

Pragmatic properties of questions

A MANCOVA on *necessary and unnecessary* questions yielded a significant effect of Prior Knowledge, $F(2,80) = 34.69, p < .001$. Subjects with much Prior Knowledge asked more necessary questions, $F(1,81) = 5.16, p < .05$ ($M = 85\%$ versus $M = 74\%$) and fewer unnecessary questions, $F(1,81) = 29.72, p < .001$ ($M = 22\%$ versus $M = 52\%$). Task Difficulty affected necessity, $F(2,114) = 12.65, p < .001$. This was due mainly to a change in unnecessary questions, $F(1,116) = 25.32, p < .001$ (see Table 2). (Twenty-four subjects had missing data for the easy items -2 pupils with little prior knowledge and 22 pupils with much prior knowledge-. A Task Difficulty effect was therefore examined only for the remaining sixty subjects.)

Table 2. Mean percentage of necessary and unnecessary questions as a function of prior knowledge and difficulty level

	Easy Words		Difficult Words	
	<i>M</i>	<i>S.D.</i>	<i>M</i>	<i>S.D.</i>
Little Prior Knowledge				
Necessary Questions	77	33	74	28
Unnecessary Questions	43	30	67	33
Much Prior Knowledge	<i>M</i>	<i>S.D.</i>	<i>M</i>	<i>S.D.</i>
Necessary Questions	73	47	86	25
Unnecessary Questions	14	20	33	24

Necessary questions reflect justified doubts about one's knowledge. They are taken to signal adequate monitoring, an active control on what the pupil knows and what he or she does not know. Pupils with much prior knowledge ask more necessary questions than do pupils with little prior knowledge. Moreover, they also ask fewer unnecessary questions. That is, they hold off asking questions when there is no objective need for them. On both measures their questioning is thus more appropriate than that of pupils with little prior knowledge.

A MANCOVA on *helpful and harmful* questions yielded a marginal effect of Prior Knowledge, $F(2,75) = 2.50, p < .10$. Prior knowledge did not effect helpfulness; 64% of the questions were helpful. Pupils with much prior knowledge asked fewer harmful questions,

$F(1,76) = 5.21, p < .05$ ($M = 14\%$ versus $M = 30\%$). There was no effect of Task Difficulty.

Subjects with much prior knowledge profited more from *specific* questions than subjects with little prior knowledge. That is, their specific questions were less harmful, $F(1,57) = 9.22, p < .01$ (1% versus 20%) and also more helpful, $F(1,59) = 8.25, p < .01$ (95% versus 71%). Prior knowledge did not affect the utility of global questions. On average, 28% of the global questions were harmful, while 47% were helpful. In all, the pupils profited substantially from asking questions. On average the number of correct Final Answers was 1 standard deviation above Provisional Answers. Only two subjects actually gave fewer correct answers after questioning.

The ultimate goal of questioning is, of course, the reception of and processing of informative answers. Unfortunately, the study permits only for some tentative statements concerning the effects of prior knowledge on helpful and harmful questions. Subjects with much prior knowledge seem to be more capable of using specific questions to their advantage. Prior knowledge seems to have *no* effect on the helpfulness of global questions.

GENERAL DISCUSSION

Pupils with little prior knowledge asked more questions than did more knowledgeable ones. It seems reasonable to argue that this is caused by the fact that these pupils have more knowledge gaps and henceforth ask more. However, this would be an oversimplification of what it takes to ask a question. Question asking is likely to be the outcome of a three-staged process (see Dillon, 1988; Herrmann, 1985; Hoppe-Graff, Herrmann, Winterhoff-Spurk, & Mangold, 1985; Garrett in Smyth, Morris, Levy & Ellis, 1987). A pupil should first experience a problem. He or she must become conscious of something problematic. After that, the pupil must focus on the information that is needed. The pupil searches for information about a particular part of the problem. In this stage a pupil may solve his or her problem. For example, the memory search may be successful or the pupil may fill-in a default value for some missing information. In the third stage, the pupil must verbalize the question so that it leads to an optimal response. In other words, the problem must be translated into a question that an addressee will understand. In this stage questioners ought to become concerned about the knowledge that they share with the addressee. Asking whether you can say that 'John is stout' may be a wonderfully sharp typing of stoutness. However, without knowing John its communicative qualities are limited. By asking 'Can you say that John, who weighs 100 pounds, is stout?' such problems can be overcome.

The experimental design of this study minimizes the difficulties within the stages 2 and 3 of question asking by using a selection procedure. Pupils do not *produce* their own questions, they *select* one from a list provided by the experimenter. Obviously, stage 1 is the same in production as it is in selection. A major difference arises in stage 2. When pupils have to generate their own questions they must express themselves in such a way that what comes out as a question fits their (perception of the) problem. The selection procedure also creates a situation in which the pupils do not have to worry about communicating with the respondent (a problem of stage 3). The questions exist already, and the pupils can just say 'I want global (specific) information'. With such minimal constraints on asking questions pupils with little prior knowledge may have been set into a more favorable position for asking questions than they would normally be.

In the experiment there would appear to be only advantages to asking specific questions, no disadvantages. Surely then, pupils must have had other reasons than purely information gathering ones for asking global questions. What may these reasons be? First,

it is possible that the pupils ask global questions because these questions enable them to figure out the right answer themselves as much as possible. When they ask, it is merely to be helped a little, not all the way down to a correct answer. There is some empirical evidence for such an independence factor (Newman & Goldin, in press; Van der Meij, 1986). Therefore, if fifth-graders do solicit help, they may try to make it minimal. In such a case, global questions should come first. Second, pupils may ask global questions to make a favorable impression on the experimenter. Global questions *look* more strategic. They suggest a broad, surveying approach to the problem. In contrast, specific questions may be perceived as belonging to a hypothesis-testing approach. Thus, pupils may elect to ask global questions to make a good impression on the experimenter and in doing so ignore the fact that asking specific questions is more informative. In short, there may be social considerations that interact with or run counter to information-gathering purposes pupils have for choosing certain questions.

Evaluating questions is quite common. Teachers often value pupil questions, wondering whether the question is real or phony, whether it is a question for content or for certainty, before they respond. Unfortunately, evaluating the necessity of questions is uncommon in research on questioning of pupils. But clearly it should be, for it is a prominent factor that determines the respondents' reaction by and large. One way of assessing the background of the pupils questions is by asking for 'provisional answers' which makes it possible to measure necessity in a rather precise manner. It is debatable, however, whether question asking after a correct provisional answer should be considered unnecessary. Disregarding guesswork, it seems reasonable to argue that when pupils ask such a question they typically do so to increase confidence in knowing. They ask in order to confirm that they know the correct answer. By typing such questions as unnecessary, this basis is ignored.

The issue of qualifying questions is clearly a problematic one. In favor of objective criteria one could say that they are fairly easy to come by and that they provide for one common standard for comparing subjects. Subjective criteria, on the other hand, are important because they reflect someone's personal reasons for asking questions, or, conversely, they account for the responses of the addressee.

The present study offers a good starting point for further research on questioning because several important factors affecting questioning can be examined under carefully controlled conditions in various follow-up studies. First, one can make question asking more difficult in the experiment by having pupils make up their own questions. In producing their own questions pupils clearly have to face more problems in determining their need for help (stage 2 in questioning) and in verbalizing their questions (stage 3 in questioning). Second, one can study the effects of various kinds of feedback on question asking. For example, feedback can be given about the necessity for asking questions, or about the adequacy of the pupil's processing of responses. Third, the experiment has been set-up to confront pupils with many potential problems, and hence offers many opportunities for question raising and asking. However, the pupils can ask only one question per word. They do not have the possibility of asking a set of questions one after another in sequence, as happens in many realistic situations. As a first step in investigating a sequence of questions one could allow pupils to ask two or more questions about a word. For instance, one could let pupils ask a specific question after having asked a global one. Fourth, to assess effects of social constraints on question asking one could consider running the experiment on a computer (which can be accomplished rather easily). After novelty has worn off the computer group could then be compared with a group that interacts face-to face with an experimenter.

In 1969 Postman & Weingartner said that 'the art and science of asking questions is the most important ability man has yet developed'. I think they are right about the art of

asking questions. But are they right about the science of asking questions? Has it really developed?

REFERENCES

- Dillon, J.T. (1988). *Questioning and teaching: a manual of practice*. London: Croom.
- Flammer, A. (1986). *Asking a question: which one - if any?* Paper presented at the Memphis conference on Question Generation. Memphis, Tennessee.
- Flammer, A., Kaiser, H.R., & Müller-Bouquet, P. (1981). Predicting what questions people ask. *Psychological Research*, 43, 421-429.
- Flammer, A., Grob, A., Leuthardt, T., & Lüthi, R. (1984). Asking how to act. *Archives de Psychologie*, 52, 103-120.
- Fuhrer, U. (1987). Effects of social density and pre-knowledge on question asking in a novel setting. *Journal of Environmental Psychology*, 7, 159-168.
- Fuhrer, U. (1989). Effects of prior knowledge, crowding, and congruence of subjects' and others' goals on question asking in an unfamiliar setting. *Psychological Reports*, 64, 131-145.
- Good, Th.L., Slavings, R.L., Harel, K.H., & Emerson, H. (1987). Student passivity: a study of question-asking in K-12 classrooms. *Sociology of Education*, 60, 181-199.
- Herrmann, Th. (1985). *Allgemeine Sprachpsychologie*. München: Urban & Schwarzenberg.
- Hoppe-Graff, S., Herrmann, Th., Winterhoff-Spurk, P., & Mangold, R. (1985). Speech and situation: a general model for the process of speech production. In J.P. Forgas (Ed.), *Language and social situations* (pp. 81-95). New York: Springer.
- Kema, G.N. (1978). *Primary Mental Abilities. Nederlandse bewerking*. Amsterdam: Swets & Zeitlinger.
- Miyake, N., & Norman, D.A. (1979). To ask a question one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior*, 18, 357-364.
- Mooy, T. (1987). *Interactional multi-level investigation into pupil behavior, achievement, competence and orientation in educational settings*. Den Haag: SVO.
- Nelson-LeGall, S. & Glor-Scheib, S. (1985). Help-seeking in elementary classrooms: an observational study. *Contemporary Educational Psychology*, 10, 58-71.
- Newman, R.S., & L. Goldin (in press). Children's reluctance to seek help with schoolwork. *Journal of Educational Psychology*.
- Postman, N. & Weingartner, C. (1969). *Teaching as a subversive activity*. New York: DeLacorte Press.
- Smith, K.H., Tykodi, T.A., & Mynatt, B.T. (1988). Can we predict the form and function of spontaneous questions? *Questioning Exchange*, 2, 53-60.
- Smyth, M.M., Morris, P.E., Levy, Ph., & Ellis, A.W. (1987). *Cognition in action*. Hillsdale, N.J.: Erlbaum.
- Stijnen, P.J.J. (1975). *Woordenschat test [Vocabulary test]*. Nijmegen: Berkhout.
- Van der Meij, H. (1986). *Questioning. A study on the questioning behavior of elementary school children*. Den Haag: SVO.
- Van der Meij, H., Baarends, & Leijh (1987). *Effects of prior knowledge on questioning*. Internal Report 1, Vakgroep Onderwijsstudies, Rijksuniversiteit Leiden.
- Van der Meij, H. (1989a). *Self-esteem and questioning*. Unpublished manuscript. Vakgroep Onderwijsstudies, Rijksuniversiteit Leiden.
- Van der Meij, H. (in press). Effects of prior knowledge on question asking. *Zeitschrift für Pädagogische Psychologie*.
- Vedder, P. (1985). *Cooperative learning*. Den Haag: SVO.
- Webb, N.M. (1982). Student interaction and learning in small groups. *Review of Educational Research*, 52, 421-445.
- Webb, N.M. (1983). Predicting learning from student interaction: defining the interaction variables. *Educational Psychology*, 18, 33-41.
- Wong, B.Y.L. (1985). Self-questioning instructional research: a review. *Review of Educational Research*, 55, 227-268.

Manuscript ontvangen 2-10-1989

Definitieve versie ontvangen 2-3-1990

Het oplossen van vergelijkingsvraagstukken: een empirische toetsing van het model van Lewis en Mayer

L. Verschaffel, E. de Corte en A. Pauwels

Onderzoekscentrum voor Onderwijsleerprocessen, Katholieke Universiteit Leuven

ABSTRACT

In an attempt to explain pupils' and even adults' difficulties with particular types of compare problems, Lewis and Mayer (1987) have presented a model that simulates the comprehension processes when solving these problems. The basis of their model is the "consistency hypothesis", according to which students are more likely to make comprehension errors when the order of presentation of the terms in the relational statement of the problem is not consistent with the preferred order. To test this model, we carried out two eye-movement experiments in which either adults (Experiment 1) or third graders (Experiment 2) were administered a series of one-step compare problems. While the data of the adult subjects revealed no evidence in favor of the consistency hypothesis, the third graders' findings provided a good support of the Lewis and Mayer (1987) model.

INLEIDING

Onderzoek heeft aangetoond dat kinderen, maar ook nog vele volwassenen, moeite hebben met het representeren en oplossen van vergelijkingsopgaven zoals "Piet heeft 37 knikkers. Hij heeft 23 knikkers minder dan Tom. Hoeveel knikkers heeft Tom?".

Lewis en Mayer (1987) ontwikkelden een model van de processen die zich afspelen tijdens het begrijpen en oplossen van dergelijke opgaven. Steun voor dit model wordt volgens de auteurs geleverd door de resultaten van een eigen empirisch onderzoek waarin ze een aantal uit het model voortvloeiende voorspellingen betreffende de moeilijkheidsgraad van verschillende soorten van vergelijkingsvraagstukken vergeleken hebben met de feitelijke antwoordpatronen van studenten. Deze prestatiegegevens kunnen evenwel niet beschouwd worden als een overtuigende bevestiging van de cognitieve processen zoals beschreven in het model van Lewis en Mayer (1987). Er is dan ook behoefte aan bijkomende, meer procesgerichte empirische evidentie. In deze bijdrage brengen wij verslag uit van twee oogbewegingsexperimenten die werden uitgevoerd om het model van Lewis en Mayer (1987) verder te toetsen.

Eerst beschrijven we het model van Lewis en Mayer (1987) en de hypothesen die eruit voortvloeien. Daarna bespreken we de opzet en de resultaten van het eerste experiment, dat werd uitgevoerd bij universiteitsstudenten. Vervolgens komt het tweede experiment,

Lieven Verschaffel is bevoegdverklaard navorser van het Belgisch Nationaal Fonds voor Wetenschappelijk Onderzoek.

De auteurs danken I. Gielen en D. Christiaens voor hun hulp bij het verzamelen en analyseren van de oogbewegingsgegevens.

Correspondentie-adres: Lieven Verschaffel, Onderzoekscentrum voor Onderwijsleerprocessen, Katholieke Universiteit Leuven, Vesaliusstraat 2, B-3000 Leuven, België.

waaraan derdeklassers deelnamen, aan bod. Tenslotte staan we stil bij de implicaties van beide studies voor het model van Lewis en Mayer (1987).

HET MODEL VAN LEWIS EN MAYER

Van de drie basiscategorieën van aanvankelijke redactie-opgaven over-optellen en aftrekken (oorzaak-verandering, combinatie en vergelijking) zijn vergelijkingsvraagstukken ongetwijfeld het moeilijkst (De Corte, Verschaffel & Verschueren, 1982; Pauwels, 1987; Riley, Greeno & Heller, 1983). Bij dit soort van opgaven gaat het om twee hoeveelheden die vergeleken worden en om het verschil tussen beide. Binnen de categorie van de vergelijkingsopgaven kan men verder onderscheid maken naargelang van de aard van de onbekende hoeveelheid, m.n. één van de twee sets die vergeleken worden of de set die verwijst naar het verschil tussen beide. Het model van Lewis en Mayer (1987) heeft enkel betrekking op die vergelijkingsvraagstukken, waarbij één van de vergeleken sets én het verschil tussen beide gegeven is, terwijl de grootte van de tweede set dient bepaald te worden.

Deze auteurs maken daarbinnen onderscheid tussen consistente en inconsistente opgaven. Bij consistente opgaven komt de vergelijkende term overeen met de uit te voeren bewerking (er staat "meer" en er moet opgeteld worden of er staat "minder" en er moet afgetrokken worden) en is het grammaticale subject uit de eerste zin het object¹ uit de tweede zin (bijv. "Leen heeft 13 poppen. Lies heeft 3 poppen meer dan Leen. Hoeveel poppen heeft Lies?"). Bij inconsistente opgaven is er een conflict tussen de vergelijkende term en de uit te voeren bewerking (er staat "meer" maar er moet afgetrokken worden of er staat "minder" maar er moet opgeteld worden); bovendien is het grammaticale subject van de eerste zin tevens het subject van de tweede (bijv. "Leen heeft 13 poppen. Zij heeft 3 poppen meer dan Lies. Hoeveel poppen heeft Lies?").

Uit onderzoek is herhaaldelijk gebleken dat inconsistente vergelijkingsopgaven moeilijker zijn dan consistente (Briars & Larkin, 1984; Morales, Shute, & Pellegrino, 1985; Riley e.a., 1983). Ter verklaring van deze bevinding formuleerden Lewis en Mayer (1987) de *consistentiehypothese*. Zich baserend op het werk van Huttenlocher en Strauss (1968), stellen zij dat probleemoplossers een duidelijke voorkeur hebben voor één bepaalde volgorde waarin de verschillende informatie-elementen in vergelijkingsopgaven geïntroduceerd worden. Meer bepaald zouden zij de volgorde uit de consistente vraagstukken verkiezen, waarbij het subject van de eerste zin het object van de tweede zin is. Worden ze geconfronteerd met een inconsistente opgave, waarbij het subject van de eerste zin ook het subject is van de tweede, dan zullen zij die derhalve herstructureren tot een consistente. Dit impliceert dat zij (1) het subject en het object uit de tweede zin omwisselen, en (2) de rekenkundige operatie die met de vergelijkende term geassocieerd wordt, omkeren. Hierbij bestaat echter het risico dat zij nalaten om de tweede stap, nl. de omkering van de rekenkundige bewerking, uit te voeren, met als gevolg dat er een verkeerde-operatiefout wordt gemaakt (een aftrekking in plaats van een optelling of omgekeerd).

Lewis en Mayer (1987) ondersteunen hun hypothese met de resultaten van een empirisch onderzoek bij 96 universiteitsstudenten. Aan deze studenten werd een reeks van samengestelde redactie-opgaven gegeven waarvan één van de delen een vergelijkingsstructuur had. Bijvoorbeeld: "At ARCO gas sells for 1.13 Dollar per gallon. Gas at Chevron is 5 cents more per gallon than gas at ARCO. How much do 5 gallons of gas cost at Chevron?". De resultaten van deze studie lagen in de lijn van de verwachtingen: er werden meer verkeerde operaties uitgevoerd bij inconsistente dan bij consistente opgaven. Daarnaast stelden Lewis en Mayer (1987) een interactie vast tussen de grammaticale structuur van de

opgave en de vereiste operatie: binnen de inconsistente opgaven werden er meer verkeerde-operatiefouten gemaakt wanneer men moest optellen of vermenigvuldigen, dan wanneer men moest aftrekken of delen. Verwijzend naar Clark (1969), verklaren zij deze interactie als volgt. Een belangrijk verschil tussen inconsistente optel- en vermenigvuldigingsvraagstukken enerzijds en aftrek- en delingsopgaven anderzijds, is dat de eerste groep een *gemarkeerde* vergelijkende term² bevat, terwijl de tweede groep gekenmerkt wordt door een *ongemarkeerde* term. Volgens de auteurs zou men minder snel geneigd zijn om tijdens het herstructureren een gemarkeerde term om te keren, met als gevolg dat het risico op verkeerde operaties bij deze opgaven nog vergroot wordt.

Lewis en Mayer (1987) hebben hun hypothetische procesmatige verklaring voor foutieve antwoorden op bepaalde vergelijkingsopgaven vertaald in een formeel model van de onderliggende cognitieve schemata en processen. Dit model wordt weergegeven in Figuur 1. Ter verduidelijking geven we een beschrijving van de manier waarop de inconsistente opgave met gemarkeerde term "Jan heeft 8 knikkers. Hij heeft 5 knikkers minder dan Tom. Hoeveel knikkers heeft Tom?" begrepen en opgelost wordt.

In de stappen 1 tot en met 4 wordt de eerste zin van de opgave geëncodeerd. Eerst wordt de zin gelezen en het overeenkomstige schema wordt geactiveerd. Vervolgens identificeert de oplosser de naam ("Jan") en de hoeveelheid ("8") die in de eerste zin worden gegeven. Het geactiveerde schema wordt ingevuld als zijnde "Jan = 8".

De stappen 5 tot 11 hebben betrekking op het encoderen van de tweede zin. Deze zin wordt eveneens eerst gelezen. Vervolgens wordt het schema voor de vergelijkende zin geactiveerd. In stap 6 identificeert de oplosser de eerste naam ("hij" verwijzend naar "Jan"), de hoeveelheid ("5"), de vergelijkende term ("minder") en de tweede naam ("Tom"). Daarna wordt de vergelijkende term vertaald in een rekenkundig symbool ("minder" wordt vertaald in "-"). Vooraleer echter het schema aan te vullen met elementen uit de tekst, gaat de oplosser in stap 8 na of het subject van de tweede zin van de opgave overeenkomt met het subject van de eerste zin. Indien dit niet het geval is, heeft de opgave een consistente structuur. De oplosser gaat dan over naar stap 9-11 en vult het schema voor de vergelijkende zin aan.

Als de subjecten van beide zinnen wél identiek zijn, wordt er een herstructurering doorgevoerd. In stap R1 verwisselt de oplosser de twee namen uit de tweede zin, zodat de eigenaar van de onbekende set ("Tom") het subject van de zin wordt en deze van de gegeven set ("Jan") het object. Vervolgens moet ook de rekenkundige bewerking omgekeerd worden, doch de kans bestaat dat de oplosser deze omkering niet uitvoert (R2). In stap R3 gaat men na of de vergelijkende term gemarkeerd of ongemarkeerd is. Indien het - zoals in het voorbeeld - om een gemarkeerde term gaat, dan is de probabiliteit dat de rekenkundige bewerking niet omgekeerd wordt nog groter (R4). In stap R5 tenslotte is de reorganisatie beëindigd en wordt het schema aangevuld. Uiteindelijk creëert de oplosser in stap 12 een vergelijking die leidt tot de oplossing van het vraagstuk: ofwel de correcte vergelijking ("Tom = 8 + 5") ofwel de vergelijking die de verkeerde operatie bevat ("Tom = 8 - 5").

Volgens dit model ontstaan de verkeerde-operatiefouten op vergelijkingsvraagstukken dus vooral tijdens het herstructureren van de tweede zin, en met name in stap 2 daaruit. Dit impliceert dat deze fouten in principe enkel kunnen optreden bij inconsistente opgaven. Het model voorspelt verder dat de kans op een verkeerde-operatiefout nog vergroot wanneer de inconsistente opgave een gemarkeerde term bevat. In dit laatste geval kan er immers ook nog iets misgaan in stap 4 van de herstructureringsfase.

Zoals eerder vermeld kunnen de prestatiegegevens van de studenten die deelnamen aan het onderzoek van Lewis en Mayer (1987) bezwaarlijk beschouwd worden als een overtuigend bewijs voor de interne processen en structuren die in dit model vervat zitten.

SCHEMATA

1. Setschema (eerste zin)
(Set A) = (hoeveelheid a)
2. Schema voor vergelijkende zin (tweede zin)
(Set B) = (hoeveelheid b) relatie (set A)
3. Schema voor de algebraïsche vergelijking
(Set B) = (hoeveelheid a) rekenkundige operatie (hoeveelheid b)

PROCEDURES

Procedure voor het encoderen van de eerste zin

1. Schema voor hoeveelhedsbepalende zin activeren
2. Identificeren van (naam 1) en (aantal 1)
3. (naam 1) toewijzen aan (SET A)
4. (aantal 1) toewijzen aan (hoeveelheid a)

Procedure voor het encoderen van de tweede zin

5. Schema voor vergelijkende zin activeren
6. Identificeren van (naam 2), (aantal 2), (relatie) en (naam 3)
7. Vertalen van (relatie) in (rekenkundige operatie)
8. Indien (naam 2) = (naam 1) ga naar **herstructureringsfase**
9. (naam 2) toewijzen aan (SET B)
10. (aantal 2) toewijzen aan (hoeveelheid b)
11. (naam 3) toewijzen aan (SET A)

Procedure voor het creëren van de formele probleemrepresentatie

12. Vorm een algebraïsche vergelijking met de huidige parameters van SET, B, hoeveelheid a, rekenkundige operatie, en hoeveelheid b.

Procedure voor de herstructurering van de tweede zin

- R1 Omwisselen (naam 1) en (naam 2)
- R2 Omkeren van (rekenkundige operatie) met $\text{probabiliteit} = (1-p)$
- R3 Indien (relatie) gemarkeerd is
- R4 Vertalen van (relatie) in (rekenkundige operatie) met $\text{probabiliteit} = p$
- R5 Terugkeren naar Procedure voor het encoderen van de tweede zin

Fig. 1. Het oplossingsmodel van Lewis en Mayer (1987, p. 368)

Om het model op een meer rechtstreekse wijze te toetsen, verzamelden we oogbewegingsgegevens van zowel universiteitsstudenten als van leerlingen van de derde klas van de lagere school tijdens het lezen en oplossen van vergelijkingsvraagstukken.

Oogbewegingsregistratie bij het onderzoek naar cognitieve processen wordt verantwoord vanuit de veronderstelling dat er een sterk verband bestaat tussen de oogbewegingen van het subject enerzijds en datgene wat op dat moment cognitief verwerkt wordt

anderzijds. Een normaal oogbewegingspatroon bestaat uit een opeenvolging van fixaties - dit zijn perioden gedurende dewelke het oog relatief bewegingsloos is - en saccaden - dit zijn snelle sprongachtige bewegingen van de oogbol. Aangezien er gedurende de saccaden weinig of geen informatie wordt opgenomen, worden bij het onderzoek naar cognitieve verwerkingsprocessen de fixaties als de meest relevante variabelen beschouwd (Rayner, 1978; Just & Carpenter, 1987). Zo wordt aangenomen dat de fixatietijd op een bepaald tekstonderdeel een aanduiding geeft van de tijd die men nodig heeft om dat onderdeel te encoderen en te interpreteren. In het kader van onderhavige studie zijn we er dan ook van uitgegaan dat de fixatietijd van een leerling op een bepaalde zin uit de opgave een weerspiegeling is van de tijd die hij/zij nodig had om deze zin te verwerken.

HYPOTHESEN

Uit het model van Lewis en Mayer (1987) - verderop afgekort als het LM-model - kan een aantal hypothesen afgeleid worden die betrekking hebben op (1) de frequentie van verkeerde operaties, (2) de oplossingstijden en (3) de fixatietijden op de verschillende zinnen.

Wat de frequentie van *verkeerde operaties* (verderop afgekort als VO) betreft, voorspelt het LM-model een effect van de consistentie van de opgave: er zullen meer VO-antwoorden gegeven worden bij inconsistente dan bij consistente opgaven (verderop afgekort als I- en C-opgaven). Een tweede predictie heeft betrekking op een interactie tussen de consistentie van de opgave en de aard van de vergelijkende term: er zullen meer VO-antwoorden voorkomen bij I-opgaven met een gemarkeerde term ("minder") dan bij deze met een ongemarkeerde term ("meer").

In tegenstelling tot de studie van Lewis en Mayer (1987) is in onderhavige onderzoekingen - om redenen die we verderop nog zullen toelichten - gewerkt met enkelvoudige in plaats van samengestelde opgaven. Dit had tot gevolg dat we de twee bovenvermelde hypothesen in verband met de moeilijkheidsgraad van de opgaven enkel konden behouden in de studie met de derdeklassers; van universiteitsstudenten valt immers te verwachten dat ze alle enkelvoudige vergelijkingsopgaven - ook die met een inconsistente structuur en een gemarkeerde term - uiteindelijk tot een goed einde zullen brengen.

Inzake *oplossingstijden* kan uit het model de volgende hypothese afgeleid worden. Indien men een I-vraagstuk eerst dient te herstructureren tot een C-opgave, dan moet dit tot uiting komen in een langere oplossingstijd. Anderzijds laat het model niet toe een duidelijke hypothese te formuleren inzake het eventuele effect van de vergelijkende term op de oplossingstijd.

Zoals Figuur 1 aantoont, wordt het verschil in moeilijkheidsgraad en in oplossingstijd tussen I- en C-opgaven volledig veroorzaakt door een verschillende verwerking van de tweede opgavezin. Op grond daarvan kan men voorspellen dat de hogere oplossingstijd voor I-opgaven veroorzaakt zal worden door een langere fixatietijd op de tweede zin, terwijl de fixatietijden op de eerste en de derde zin niet systematisch zullen verschillen voor beide soorten van vraagstukken.

Tot slot impliceert het LM-model dat de transformatie van I- naar C-opgaven onmiddellijk na de eerste lezing van de tweede zin plaatsvindt, d.w.z. alvorens de vraagzin gelezen is (zie Figuur 1). Hieruit kan men afleiden dat de verschillen tussen C- en I-opgaven inzake oplossings- en fixatietijden reeds aangetroffen zullen worden *tijdens de eerste lezing van de eerste twee opgavezinnen*. Meer bepaald valt te verwachten dat deze fase langer zal duren bij I- dan bij C-opgaven en dat dit vooral te wijten zal zijn aan langere fixatietijden op de tweede zin van de opgave.

EXPERIMENT 1

Opzet

Subjecten

Negentien universiteitsstudenten namen op vrijwillige basis aan het onderzoek deel.

Opgaven

Elke student kreeg 30 enkelvoudige optel- en aftrekvraagstukken aangeboden. Zestien daarvan waren de eigenlijke testitems; de overige 14 waren bufferopgaven die tussen de testitems waren geschoven om stereotype antwoordtendenzen te vermijden. De ene helft van de testitems had een consistente structuur, de andere een inconsistente. Binnen deze twee categorieën waren er telkens vier opgaven met een ongemarkeerde vergelijkende term ("meer") en vier opgaven met een gemarkeerde term ("minder"). In Tabel 1 worden enkele voorbeelden van de vier gepresenteerde opgaventypes gegeven.

Tabel 1. Voorbeelden van testitems gepresenteerd tijdens Experiment 1

Term	Structuur	
	Consistent	Inconsistent
Ongemarkeerd	Bij BP kost benzine 26,3 fr per liter. Bij Q8 kost hij 0,9 fr meer dan bij BP. Hoeveel kost de benzine bij Q8?	Bij ESSO kost diesel 16,4 fr per liter. Dat is precies 0,7 fr meer dan bij ARAL. Hoeveel kost de diesel bij ARAL?
	Bert heeft een gewicht van 75 kilogram. Peter weegt 14 kilogram meer dan Bert. Hoeveel kilogram weegt Peter?	Anne heeft een gewicht van 57 kilogram. Zij weegt 12 kilogram meer dan Katrien. Hoeveel kilogram weegt Katrien?
Gemarkeerd	Bij Q8 kost de diesel 16,4 fr per liter. Bij BP kost hij 0,7 fr minder dan bij Q8. Hoeveel kost de diesel bij BP?	Bij ARAL kost benzine 26,3 fr per liter. Dat is wel 0,9 fr minder dan bij FINA. Hoeveel kost benzine bij FINA?
	Leen heeft een gewicht van 57 kilogram. Anja weegt 12 kilogram minder dan Leen. Hoeveel kilogram weegt Anja?	Dirk heeft een gewicht van 75 kilogram. Hij weegt 14 kilogram minder dan Johan. Hoeveel kilogram weegt Johan?

Deze vier types werden gecontroleerd voor het aantal zinnen, het aantal karakters per zin, de context waarrond de opgave was opgebouwd en de aard van de gegeven getallen. Zoals vermeld, werden er - in tegenstelling tot de studie van Lewis en Mayer (1987) - alleen enkelvoudige vraagstukken over optellen en aftrekken aangeboden. Deze keuze berustte op de volgende overwegingen. Ten eerste vereist het beschikbare apparaat voor oogbewegingsregistratie dat het hoofd van de proefpersoon tamelijk stabiel blijft. Daardoor kan de proefpersoon tijdens het oplossen geen beroep doen op externe hulpmiddelen (potlood en papier) waarvan men normaliter gebruik maakt bij het oplossen van samengestelde rekenopgaven. Verder opteerden we voor vraagstukken bestaande uit slechts drie regels tekst. Gezien het oogbewegingstoestel een onnauwkeurigheid van 0,5 visuele graad heeft, dienden de onderscheiden zinnen met voldoende tussenruimte gepresenteerd te worden.

Meer tekstregels zou deze tussenruimte verkleinen en dus de betrouwbaarheid van de metingen aantasten. Tenslotte zagen we geen enkele principiële reden om het toepassingsgebied van het LM-model te beperken tot samengestelde opgaven. Immers, de kern van dit model is dat een relationele zin met een inconsistente structuur moeilijk te interpreteren is en daarom eerst getransformeerd wordt in een zin met een consistente structuur vooraleer er verder gelezen wordt. Of deze relationele zin de tweede zin is van een enkelvoudig vraagstuk dan wel de tweede zin van een samengestelde opgave, verandert daar in principe niets aan.

Procedure

Alle studenten werden individueel onderzocht. Ze moesten de oplossing van het vraagstuk niet uitrekenen; de opdracht bestond er enkel in om telkens mondeling aan te geven welke operatie er met de twee gegeven getallen uit de opgave moest worden uitgevoerd om het juiste antwoord te bekomen. Door enkel naar de vereiste operatie te vragen werd voorkomen dat de invloed van de onafhankelijke variabelen die in deze studie centraal stonden, vertroebeld werd door factoren van rekentechnische aard.

De opgaven werden gepresenteerd op een tv-scherm. De proefpersoon nam op twee meter afstand daarvan plaats in een stoel, voorzien van hoofdsteunen. Terwijl hij de opgaven las en oploste, werden zijn oogbewegingen geregistreerd met behulp van een DEBIC 80 systeem, dat gebaseerd is op de pupilmidden-corneale reflectie methode (De Graef, Van Rensbergen & d'Ydewalle, 1985). De coördinaten van de opeenvolgende blikposities worden hierbij om de 20 milliseconden op twee verschillende wijzen geregistreerd: in de eerste plaats worden deze coördinaten vastgelegd op computerschijf; daarnaast wordt er een video-opname gemaakt van het scherm waarop de opgave gepresenteerd wordt, met daarbovenop de projectie van de blikpositie van het subject voorgesteld door het snijpunt van een verticale en een horizontale as (zie Figuur 2).

	Colruyt verkoopt elke dag 220 eieren.
	Aldi verkoopt er 75 meer dan Colruyt.
	Hoeveel eieren verkoopt Colruyt?

Fig. 2. Beeld van het scherm waarop de opgave gepresenteerd wordt, met daarover de projectie van de blikpositie van het subject voorgesteld door het snijpunt van een verticale en een horizontale as

De aldus verkregen ruwe oogbewegingsgegevens werden met behulp van een data-reductieprogramma herleid tot een opeenvolging van fixaties met een bepaalde positie en duur. Een fixatie werd geoperationaliseerd als een periode van minimum 100 milliseconden gedurende dewelke het oog haast onbeweeglijk op hetzelfde punt gericht blijft (Rayner, 1978). Deze fixaties vormden de basis voor het berekenen van de fixatietijden op de drie verschillende zinnen van de opgaven. Deze fixatietijden werden - evenals de totale oplossings tijden - onderworpen aan een variantie-analyse met de consistentie van de opgave (consistent of inconsistent) en de aard van de vergelijkende term (ongemarkeerd of gemarkeerd) als onafhankelijke variabelen (2×2 gerandomiseerde factoriële opzet). In dit verband dient opgemerkt dat de totale oplossings tijd niet gelijk is aan de som van de fixatietijden op de drie zinnen van de opgave. De oplossings tijd, die loopt vanaf het moment dat de opgave op het tv-scherm verschijnt tot op het moment dat de proefpersoon een antwoord geeft, bevat immers ook de tijdsintervallen tussen de opeenvolgende fixaties (de saccaden), evenals eventuele fixaties boven, onder of naast de opgavetekst.

Om de hypothesen betreffende de eerste fase van het oplossingsproces te toetsen werden de oogbewegingsprotocollen opgedeeld in twee fasen: een eerste fase waarin de proefpersoon de eerste en de tweede zin voor de eerste keer leest, en een tweede fase die aanvangt bij de eerste fixatie op de vraagzin en doorloopt tot het moment waarop het antwoord gegeven wordt. Alvorens de variantie-analyse uit te voeren, gingen we na of er tijdens alle oplossingsprocessen wel typisch leesgedrag voorkwam. Onder typisch leesgedrag verstaan we: meerdere opeenvolgende fixaties resp. op de eerste, de tweede zin en de derde zin uit de opgave. Slechts bij zes oogbewegingsprotocollen bleek dit niet het geval te zijn. Deze werden derhalve niet in de analyse van de eerste fase betrokken.

Resultaten

Zoals verwacht werden de meeste opgaven correct opgelost. In totaal werd 3.9% van de opgaven foutief beantwoord: 3.3% van de C-opgaven en 4.6% van de I-vraagstukken. We zullen derhalve niet verder stilstaan bij deze foutieve antwoorden, die allicht eerder te wijten waren aan een gebrek aan concentratie, dan aan moeilijkheden bij de probleemrepresentatie.

Bij de bespreking van de oplossings- en fixatietijden komen eerst de resultaten voor het totale oplossingsproces aan bod; daarna volgen de gegevens voor de eerste fase van het oplossingsproces.

Totale oplossingsproces

Duur. We verwachtten dat de studenten meer tijd zouden nodig hebben voor het oplossen van I- dan van C-opgaven. De variantie-analyse leverde echter niet het verwachte hoofdeffect van consistentie op. Er was zelfs een tegengestelde tendens: de studenten gaven iets sneller antwoord op de I- dan op de C-opgaven (8.78 sec. tegenover 9.39 sec.).

De analyse bracht wel een significant hoofdeffect van de vergelijkende term aan het licht ($F(1,18) = 12.58$, $MSe = 18.11$, $p < 0.01$). De subjecten hadden gemiddeld meer tijd nodig voor het oplossen van opgaven met een gemarkeerde term (9.95 sec.), dan voor opgaven met een ongemarkeerde term (8.22 sec.).

Er werd geen interactie-effect tussen beide onafhankelijke variabelen gevonden.

Fixatietijden op de verschillende zinnen. We voorspelden dat de totale fixatietijd op de tweede zin hoger zou zijn voor I- dan voor C-opgaven, terwijl de fixatietijd op de twee overige zinnen niet of nauwelijks zou verschillen voor beide opgavetypes. De gemiddelde fixatietijden op de drie zinnen worden weergegeven in Tabel 2.

Tabel 2. Gemiddelde fixatietijden (in milliseconden) op de drie zinnen bij Experiment 1

Term	Structuur		
	Consistent	Inconsistent	Totaal
Zin 1			
Ongemarkeerd	2000	2128	2064
Gemarkeerd	2766	2148	2457
Totaal	2383	2138	
Zin 2			
Ongemarkeerd	2280	2216	2248
Gemarkeerd	3012	2649	2831
Totaal	2646	2433	
Zin 3			
Ongemarkeerd	632	731	681
Gemarkeerd	821	852	836
Totaal	726	791	

Voor geen enkele van de drie zinnen vonden we een significant effect van de consistentie van de opgave. Zoals uit Tabel 2 blijkt, wijzen de gemiddelde fixatietijden op de tweede zin zelfs een tendens aan die tegengesteld is aan de voorspelling afgeleid uit het LM-model: de gemiddelde totale kijktijd op deze zin is immers lager voor I- dan voor C-opgaven (2433 msec. tegenover 2646 msec.).

Anderzijds bleek de fixatietijd op elk van de drie zinnen significant beïnvloed te zijn door de aard van de vergelijkende term ($F(1,18) = 9.67$, $MSe = 1213802$, $p < 0.01$ voor de eerste zin; $F(1,18) = 14.91$, $MSe = 1730030$, $p < 0.01$ voor de tweede zin; $F(1,18) = 5.45$, $MSe = 334564$, $p < 0.05$ voor de vraagzin). De zinnen uit opgaven met een gemarkeerde term werden alle langer gefixeerd dan deze uit opgaven met een ongemarkeerde term.

Voor de fixatietijd op de eerste zin vonden we tevens een significant interactie-effect tussen beide onafhankelijke variabelen ($F(1,18) = 7.47$, $MSe = 1413461$, $p < 0.05$). Het verschil tussen C- en I-opgaven kwam vooral tot uiting in het geval van een gemarkeerde vergelijkende term.

Eerste fase van het oplossingsproces

Volgens het LM-model vindt de herstructurering van de tweede zin uit een I-opgave reeds tijdens de eerste encoding van deze zin plaats; er wordt dus niet gewacht tot de opgavetekst volledig doorgenomen is. Vandaar onze hypothesen (1) dat het bij I-opgaven langer zal duren vooraleer met het lezen van de vraagzin begonnen wordt dan bij C-opgaven, en (2) dat dit vooral te wijten is aan de langere fixatietijd op de tweede zin gedurende deze eerste fase van het oplossingsproces.

Duur van de eerste fase. De analyse van de duur van deze eerste fase van het oplossingsproces bracht een significant effect van de consistentie van de opgave aan het licht ($F(1,18) = 7.40$, $MSe = 0.86$, $p < 0.05$). De gemiddelden lagen echter weerom niet in de verwachte richting: het duurde langer vooraleer de studenten de vraagzin uit een C-opgave begonnen te lezen (5.91 sec.), dan deze uit een I-opgave (5.32 sec.).

De aard van de vergelijkende term had eveneens een significant effect op deze afhankelijke variabele ($F(1,18) = 5.26$, $MSe = 142270$, $p < 0.05$): bij opgaven met een gemarkeerde term (5.85 sec.) duurde de eerste fase langer dan bij opgaven met een ongemarkeerde term (5.32 sec.).

We vonden geen significant interactie-effect tussen de consistentie van de opgave en de vergelijkende term.

Fixatietijden op de eerste twee zinnen. Net zoals bij het totale oplossingsproces, verwachtten we een significant effect van de consistentie van de opgave op de fixatietijd op de tweede zin van de opgave, maar niet op de fixatietijd op de eerste zin. De gemiddelde fixatietijden worden weergegeven in Tabel 3.

De variantie-analyse leverde inderdaad enkel een significant effect op van de consistentie van de opgave op de fixatietijd op de tweede zin ($F(1,18) = 5.33$, $MSe = 142270$, $p < 0.05$). Doch opnieuw lagen de gemiddelden niet in de voorspelde richting: bij de C-opgaven werd de tweede zin langer gefixeerd dan bij I-opgaven (1648 msec. tegenover 1448 msec.).

Tabel 3. Gemiddelde fixatietijden (in milliseconden) op de eerste twee zinnen tijdens de eerste fase van het oplossingsproces bij Experiment 1

Term	Structuur		
	Consistent	Inconsistent	Totaal
Zin 1			
Ongemarkeerd	1698	1824	1761
Gemarkeerd	2110	1702	1906
Totaal	1904	1763	
Zin 2			
Ongemarkeerd	1589	1371	1480
Gemarkeerd	1706	1525	1615
Totaal	1648	1448	

Het hoofdeffect van de vergelijkende term op de fixatietijden op de eerste en op de tweede zin tijdens de eerste fase was niet significant.

Net zoals voor het totale oplossingsproces stelden we ook hier een significant interactie-effect vast tussen beide onafhankelijke variabelen op de fixatietijd op de eerste zin ($F(1,18) = 10.67$, $MSe = 126996$, $p < 0.01$): bij de C-opgaven bleek er een duidelijk verschil te zijn tussen vraagstukken met een gemarkeerde en een ongemarkeerde term wat betreft de kijktijd op de eerste zin.

Samenvattend kunnen we stellen dat de resultaten van dit eerste experiment de consistentiehypothese van Lewis en Mayer (1987) geenszins ondersteunen. De oplossings- en fixatietijden van de studenten komen helemaal niet overeen met hetgeen op grond van het model kan worden verwacht: het duurde niet significant langer voordat een antwoord werd gegeven op de I- dan op de C-opgaven en er werd niet significant langer naar de tweede zin van de I- dan naar deze van de C-vraagstukken gekeken. Voor de eerste fase van het oplossingsproces werden er zelfs significante verschillen in de tegenovergestelde richting vastgesteld.

Ondanks het feit dat het LM-model niet toelaat om duidelijke hypothesen te formuleren betreffende de invloed van de vergelijkende term, stelden we vast dat deze factor in de meeste gevallen wel een significante invloed had op onze afhankelijke variabelen: de oplossings- en fixatietijden voor het totale proces evenals de duur van de eerste fase waren significant langer bij opgaven met een gemarkeerde term dan bij die met een ongemarkeerde term.

Een mogelijke verklaring voor de uitblijvende effecten van de consistentie van de opgave is dat de vraagstukken te gemakkelijk waren voor de studenten. Het kleine foutenpercentage en de snelle antwoordtijden suggereren dat deze studenten inderdaad als "experten" in het oplossen van de betreffende opgaven beschouwd konden worden. Bij subjecten voor wie dergelijke vraagstukken wél een uitdaging vormen, zou het LM-model dan wel opgaan. Om deze onderstelling te toetsen, besloten we om een tweede oogbewegingsstudie uit te voeren, ditmaal met derdeklassers als proefpersonen.

EXPERIMENT 2

Opzet

Subjecten

Aan het tweede experiment namen 15 leerlingen van het derde leerjaar van de lagere school (8 à 9 jaar) deel.

Opgaven en procedure

De leerlingen kregen 26 enkelvoudige vraagstukjes aangeboden: 16 vergelijkingsopgaven, aangevuld met 10 bufferitems. Net zoals in Experiment 1 bestonden de testitems voor de ene helft uit C- en voor de andere helft uit I-opgaven. Binnen elke categorie bevatten vier opgaven een ongemarkeerde term ("meer") en de andere vier een gemarkeerde term ("minder"). Voorbeelden van opgaven, die qua context enigszins verschilden van die uit

Experiment 1 en die weerom gecontroleerd werden wat betreft het aantal zinnen, het aantal karakters per zin, de getallen en de leestechische moeilijkheidsgraad, worden gegeven in Tabel 4.

Tabel 4. Voorbeelden van testitems gepresenteerd tijdens Experiment 2

Term	Structuur	
	Consistent	Inconsistent
Ongemarkeerd	Kris heeft 32 boeken. Raf heeft 13 boeken meer dan Kris. Hoeveel boeken heeft Raf?	Piet heeft 28 pennen. Hij heeft 17 pennen meer dan Ria. Hoeveel pennen heeft Ria?
	Lies kreeg 29 tulpen. Tim kreeg 14 tulpen meer dan Lies. Hoeveel tulpen kreeg Tim?	Bart kreeg 35 wafels. Hij kreeg 29 wafels meer dan Sam. Hoeveel wafels kreeg Sam?
Gemarkeerd	Ilse heeft 35 poppen. An heeft 29 poppen minder dan Ilse. Hoeveel poppen heeft An?	Miet heeft 38 ringen. Zij heeft 25 ringen minder dan Jo. Hoeveel ringen heeft Jo?
	Jan kreeg 28 auto's. Ivo kreeg 17 auto's minder dan Jan. Hoeveel auto's kreeg Ivo?	Inge kreeg 32 noten. Zij kreeg 13 noten minder dan Toon. Hoeveel noten kreeg Toon?

De onderzoeksprocedure en de gegevensverwerking verliepen op precies dezelfde wijze als in Experiment 1, met dit verschil dat hier ook het aantal VO-antwoorden aan een variantie-analyse werd onderworpen.

Resultaten

VO-antwoorden

Op basis van het LM-model verwachtten we bij deze derdeklassers meer VO-antwoorden bij I- dan bij C-opgaven en meer VO-antwoorden bij I-opgaven met een gemarkeerde dan met een ongemarkeerde term.

Een variantie-analyse met de consistentie van de opgave en de aard van de vergelijkende term als onafhankelijke variabelen, bracht een hoofdeffect van de consistentie van de opgave aan het licht ($F(1,14) = 19.64$, $MSe = 0.98$, $p < 0.01$). Het percentage VO-antwoorden bedroeg 3.3 voor C-opgaven tegenover 31.6 voor I-opgaven.

Er werden meer VO-fouten gemaakt bij de opgaven met een gemarkeerde dan bij die met een ongemarkeerde term (resp. 22.5 en 12.5%). Doch het hoofdeffect van de vergelijkende term was niet significant.

Terwijl we binnen de categorie van C-opgaven geen verschil in het aantal VO-antwoorden tussen beide soorten vergelijkende termen konden vaststellen (voor beide bedroeg dit 3.3%), bleken de derdeklassers dubbel zoveel fouten te maken op I-opgaven met een gemarkeerde term (41.6%) dan op die met een ongemarkeerde term (21.6%). Nochtans bleef het interactie-effect tussen consistentie en vergelijkende term eveneens boven het 5% significantieniveau.

Totale oplossingsproces

Duur. We verwachtten dat de leerlingen meer tijd zouden nodig hebben voor het oplossen van I- dan van C-opgaven. De resultaten van de variantie-analyse waren in overeenstemming met deze predictie. We vonden een significant hoofdeffect van de consistentie van de opgaven ($F(1,14) = 11.36$, $MSe = 94.71$, $p < 0.01$): bij I-opgaven gaven de leerlingen hun antwoord gemiddeld na 16.58 sec., bij C-opgaven reeds na 12.35 sec.

In tegenstelling tot de resultaten van Experiment 1 vonden we geen effect van de

Tabel 5. Gemiddelde fixatietijden (in milliseconden) op de drie zinnen bij Experiment 2

Term	Structuur		
	Consistent	Inconsistent	Totaal
Zin 1			
Ongemarkeerd	2906	3319	3113
Gemarkeerd	2764	4169	3466
Totaal	2835	3744	
Zin 2			
Ongemarkeerd	3759	4455	4107
Gemarkeerd	3443	5227	4335
Totaal	3601	4841	
Zin 3			
Ongemarkeerd	870	1175	1023
Gemarkeerd	748	1011	880
Totaal	809	1093	

vergelijkende term (14.14 msec. voor ongemarkeerde en 14.79 msec. voor gemarkeerde termen).

Tot slot werd er evenmin een interactie-effect vastgesteld.

Fixatietijden op de verschillende zinnen. Overeenkomstig het LM-model voorspelden we dat de langere oplossings tijden voor I-opgaven vooral tot uiting zouden komen in langere fixatietijden op de tweede zin. De gemiddelde fixatietijden op de drie zinnen worden weergegeven in Tabel 5.

De analyse op de fixatietijden op zin 2 leverde het verwachte effect op: de derdeklassers fixeerden de tweede zin uit een I-opgave gedurende 4841 msec. en de tweede zin uit een C-opgave slechts gedurende 3601 msec. ($F(1,14) = 8.27$, $MSe = 11154358$, $p < 0.05$). Maar ook de fixatietijd op de overige twee zinnen bleek significant beïnvloed door de consistentie van de opgave ($F(1,14) = 7.71$, $MSe = 6428539$, $p < 0.05$ voor zin 1; $F(1,14) = 4.82$, $MSe = 1005317$, $p < 0.05$ voor zin 3): de leerlingen keken ook naar deze zinnen langer bij I- dan bij C-vraagstukken.

De fixatietijden op de drie zinnen werden niet beïnvloed door de aard van de vergelijkende term.

Evenmin stelden we enig interactie-effect vast.

Eerste fase van het oplossingsproces

Aangezien alle leerlingen elke opgave systematisch lazen bij het begin van het oplossingsproces, werden alle oogbewegingsprotocollen bij deze analyses betrokken.

Duur van de eerste fase.

Zoals toegelicht bij de beschrijving van Experiment 1, verwijst de eerste fase van het oplossingsproces naar de eerste lezing van de eerste twee opgavezinnen.

Overeenkomstig de voorspelling vonden we een significant effect van de consistentie van de opgave op deze afhankelijke variabele ($F(1,14) = 12.91$, $MSe = 8.58$, $p < 0.01$): deze initiële fase duurde langer bij I-opgaven dan bij C-opgaven (zie Tabel 6).

Het hoofdeffect van de vergelijkende term bleef net boven het 5% significantieniveau.

De analyse leverde wel een significant interactie-effect op tussen consistentie en de vergelijkende term ($F(1,14) = 5.40$, $MSe = 1.07$, $p < 0.05$). Tabel 6 toont aan dat het gevonden verschil tussen C- en I-opgaven vrijwel helemaal op rekening te schrijven is van de opgaven met een gemarkeerde term. Dit is niet helemaal in overeenstemming met het LM-model. Volgens deze auteurs wordt de tweede zin uit een I-opgave altijd geherstructureerd, ongeacht de aard van de vergelijkende term. Dit impliceert dat ook bij vraagstukken met een ongemarkeerde term de duur van de eerste fase van het oplossingsproces beduidend langer zou moeten zijn bij een inconsistente dan bij een consistente structuur.

Tabel 6. Gemiddelde duur (in seconden) van de eerste fase van het oplossingsproces bij Experiment 2

Term	Structuur		
	Consistent	Inconsistent	Totaal
Ongemarkeerd	8.51	8.89	8.70
Gemarkeerd	7.96	10.30	9.13
Totaal	8.24	9.60	

Fixatietijden op de eerste twee zinnen. Overeenkomstig het LM-model, verwachtten we dat het effect van de consistentie van de opgaven op de fixatietijd zich vooral zou voordoen bij de tweede zin. De gemiddelde fixatietijden worden gegeven in Tabel 7.

Tabel 7. Gemiddelde fixatietijden (in milliseconden) op de eerste twee zinnen tijdens de eerste fase van het oplossingsproces bij Experiment 2

Term	Structuur		
	Consistent	Inconsistent	Totaal
Zin 1			
Ongemarkeerd	2190	2393	2292
Gemarkeerd	2011	2673	2342
Totaal	2100	2533	
Zin 2			
Ongemarkeerd	2494	2836	2665
Gemarkeerd	2398	3200	2799
Totaal	2446	3018	

De fixatietijd op de tweede zin bleek inderdaad significant beïnvloed te zijn door de consistentie van de opgave: tijdens de eerste fase van het oplossingsproces keken de leerlingen gemiddeld 3018 msec. naar de tweede zin in een I-opgave tegenover slechts 2446 msec. in een C-opgave ($F(1,14) = 25.51$, $MSc = 192259$, $p < 0.01$). Hoewel minder uitgesproken, vonden we echter ook een significant effect van de consistentie van de opgave op de absolute fixatietijd op de eerste zin ($F(1,14) = 7.26$, $MSc = 386912$, $p < 0.05$).

Beide afhankelijke variabelen werden niet systematisch beïnvloed door de aard van de vergelijkende term.

Evenmin werd er een interactie-effect vastgesteld.

DISCUSSIE

In onderhavig artikel brachten we verslag uit van twee oogbewegingsexperimenten die tot doel hadden de consistentiehypothese van Lewis en Mayer (1987) te toetsen. Volgens deze hypothese hebben probleemoplossers een duidelijke voorkeur voor vergelijkingsopgaven met een consistente structuur. Indien ze met een inconsistente opgave geconfronteerd worden, zullen zij deze eerst transformeren tot een consistente opgave.

De resultaten van een eerste experiment met universiteitsstudenten boden weinig of geen steun aan deze hypothese. Voor verscheidene afhankelijke variabelen was er een tendens in de tegenovergestelde richting; voor de eerste fase van het oplossingsproces bleken de verschillen zelfs statistisch significant te zijn. Bovendien was het effect van de aard van de vergelijkende term globaal genomen groter dan dat van de consistentie van de opgave.

De resultaten van een tweede experiment bij derdeklassers kwamen wel goed overeen met het model. Ten eerste maakten deze leerlingen significant meer VO-fouten op I- dan op C-opgaven. Ten tweede waren de totale oplossingstijden langer bij I- dan bij C-opgaven. De duur van de eerste fase van het oplossingsproces was eveneens langer voor I- dan voor C-opgaven, doch dit effect was het meest uitgesproken bij opgaven met een gemarkeerde relationele term. Wel keken de leerlingen niet alleen langer naar de tweede zin, maar naar alle zinnen van een inconsistente opgave. De aard van de vergelijkende term had in dit tweede experiment geen significante invloed op de afhankelijke variabelen.

Bij het overschouwen van de resultaten uit beide experimenten rijst in de eerste plaats de vraag hoe het komt dat we in de studie met de derdeklassers wél en in de studie met de

universiteitsstudenten geen empirische evidentie gevonden hebben voor de consistentiehypothese van Lewis en Mayer (1987). De meest plausibele verklaring is de geringe subjectieve moeilijkheidsgraad van de vraagstukken in de studie met de studenten. Duidelijke aanwijzingen hiervoor zijn het uiterst geringe foutenpercentage en de snelheid waarmee de antwoorden gegeven werden. De derdeklassers daarentegen maakten niet alleen veel meer fouten; bovendien hadden ze aanzienlijk meer tijd nodig. Klaarblijkelijk gaat de consistentiehypothese van Lewis en Mayer (1987) enkel op voor vergelijkingsvraagstukken met een behoorlijke moeilijkheidsgraad.

Verder is er de vraag hoe het komt dat de universiteitsstudenten in de eerste studie niet minder maar integendeel significant meer tijd nodig hadden voor het lezen van de eerste twee opgavezinnen van de consistente dan van de inconsistente vraagstukken. Mogelijk biedt de tekstverwerkingstheorie van Kintsch en Van Dijk (1978, 1983) hiervoor een verklaring. Men zou namelijk kunnen argumenteren dat volgens deze theorie de eerste twee zinnen uit een inconsistente opgave - vanuit een bepaald opzicht - meer beantwoorden aan de (impliciete) eisen van tekstcoherentie dan die uit een consistente opgave, en wel omdat de protagonist van de tweede zin van een inconsistente opgave dezelfde is als deze uit de eerste zin ("Piet heeft 3 appels. Hij ..."), hetgeen bij een consistente opgave niet het geval is ("Piet heeft 3 appels. An ..."). Wanneer we er van uitgaan dat de coherentie van een tekst een invloed heeft op het gemak waarmee de verschillende zinnen geïnterpreteerd en geïntegreerd worden, kunnen we verwachten dat de eerste zinnen van een inconsistente opgave sneller gelezen worden. Dat deze factor wel een rol heeft gespeeld in de studie met de universiteitsstudenten en niet in die met de jonge basisschoolleerlingen, zou hieraan kunnen toegeschreven worden dat laatstgenoemden deze impliciete regels in verband met tekstcoherentie nog niet verworven hebben.

Een laatste opmerkelijk resultaat is de onverwachts grote invloed van de aard van de vergelijkende term op de meeste afhankelijke variabelen in de eerste studie. Volgens het model van Lewis en Mayer (1987) heeft de aard van de vergelijkende term enkel een invloed op het aantal VO-fouten binnen de categorie van de moeilijke I-opgaven. In het eerste experiment werden er verscheidene hoofdeffecten gevonden van de aard van de vergelijkende term. Bovendien bleek de aanwezigheid van een gemarkeerde term vooral een negatieve invloed te hebben bij de C-opgaven. Een mogelijke verklaring voor het feit dat de relationele term wel een effect had in Experiment 1 en niet in Experiment 2, wordt geboden door Van Kuyk (1988). Volgens deze auteur is de grotere toegankelijkheid van ongemarkeerde termen te wijten aan hun hogere gebruiksfrequentie. Op basis hiervan zou men kunnen argumenteren dat het al dan niet gemarkeerd zijn van de relationele term wel invloed uitoefent op de denkprocessen van volwassen proefpersonen, maar nog niet op die van jonge kinderen met weinig ervaring in het gebruik van deze termen.

Tenslotte willen we er nog op wijzen dat er naast de in dit artikel beschreven consistentiehypothese in de literatuur ook nog een aantal andere ontstaanswijzen van VO-fouten op vergelijkingsopgaven vermeld worden. We denken daarbij o.m. aan de z.g. sleutelwoordstrategie (De Corte e.a., 1982; Nesher & Teubal, 1975; Sowder, 1988). Deze strategie komt kortweg hierop neer dat de leerling zich bij de keuze van de vereiste rekenoperatie helemaal niet laat leiden door een globale semantische representatie van de betekenis van de essentiële elementen en relaties uit het vraagstuk, maar louter op de aanwezigheid van één bepaald woord in de opgavetekst, bijv. het woord "meer" of "minder" die respectievelijk geassocieerd worden met optellen en aftrekken. Bij sommige opgaven, meer bepaald bij de vergelijkingsopgaven die door Lewis en Mayer (1987) consistent genoemd worden, leidt deze oppervlakkige sleutelwoordstrategie tot het correcte antwoord; bij andere opgaven daarentegen, waaronder de vraagstukken met een inconsistente structuur, resulteert diezelfde strategie in een VO-antwoord. Hoewel

verscheidene gegevens (o.m. de oplossingsstijden) de conclusie wettigen dat de sleutelwoordstrategie zeker niet naar voren geschoven kan worden als het belangrijkste verklaringselement voor de resultaten bekomen in Experiment 2, valt niet uit te sluiten dat de VO- antwoorden van sommige leerlingen via deze alternatieve weg tot stand gekomen zijn.

VOETNOTEN

- 1 "Object" verwijst in feite naar de "nabepaling bij de vergelijkende term". Omwille van de eenvoud van de uiteenzetting en naar analogie met de oorspronkelijke tekst van Lewis en Mayer (1987) hanteren we echter de term "object".
- 2 Bij deze verklaring beroepen Lewis en Mayer (1987) zich op de theorie van Clark (1969). Laatstgenoemde auteur stelt dat het onderscheid tussen *ongemarkeerde* en *gemarkeerde* termen berust op het feit dat de twee tegengestelde polen van een woordpaar (zoals "goed-slecht", "lang-kort", "groot-klein", en ook "meer-minder") niet zonder meer als symmetrisch te beschouwen zijn. De positieve of ongemarkeerde pool van het paar (bijv. "goed", "lang") kan namelijk niet alleen een contrasterende, maar ook een neutrale betekenis aannemen (bijv. "Hoe goed is hij?" of "Dit is maar 2 cm lang."). De ongemarkeerde pool heeft dus twee mogelijke betekenissen nl. een contrasterende en een neutrale, terwijl de gemarkeerde term er slechts één heeft, nl. een contrasterende. Omwille van hun geringere complexiteit zouden ongemarkeerde termen eerder verworven worden en gemakkelijker toegankelijk zijn. Volgens Van Kuyk (1988) heeft de snellere verwerving en de gemakkelijkere toegankelijkheid van ongemarkeerde positieve termen niet alleen te maken met een geringere complexiteit, maar ook met een hogere gebruiksfrequentie dan deze van negatieve gemarkeerde termen.

LITERATUUR

- Briars, D.J., & Larkin, J.H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction*, 1, 245-296.
- Clark, H.H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, 76, 387-404.
- De Corte, E., Verschaffel, L., & Verschuren, J. (1982). First graders' solution processes in elementary word problems. In A. Vermandel (Ed.), *Proceedings of the Sixth International Conference for the Psychology of Mathematical Education* (pp. 91-96). Antwerp: University of Antwerp.
- De Graef, P., Van Rensbergen, J., & d'Ydewalle, G. (1985). *User's manual for the Leuven eye-movement registration system* (Internal Report No. 52). Leuven: Laboratory for Experimental Psychology, University of Leuven.
- Huttenlocher, J., & Strauss, S. (1968). Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 7, 300-304.
- Just, M.A., & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Pittsburgh, P.A.: Carnegie-Mellon University.
- Kintsch, W., & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 5, 363-394.
- Lewis, A., & Mayer, R. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, 79, 363-371.
- Morales, R.V., Shute, V.J., & Pellegrino, J.W. (1985). Developmental differences in understanding and solving simple word problems. *Cognition and Instruction*, 2, 41-57.
- Nesher, P., & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics*, 6, 41-51.
- Pauwels, A. (1987). *Empirische toetsing van computermodellen over de ontwikkeling van de oplossingsvaardigheid bij eenvoudige redactie-opgaven over optellen en aftrekken*. (Niet-gepubliceerde licentiaatsverhandeling). Leuven: Onderzoekscentrum voor Onderwijsleerprocessen, Katholieke Universiteit Leuven.

- Rayner, K. (1978). Eye-movements in reading and information processing. *Psychological Bulletin*, 85, 618-680.
- Riley, M.S., Greeno, J.G., & 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.
- Sowder, L. (1988). Children's solutions of story problems. *Journal of Mathematical Behavior*, 7, 227-238.
- Van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York/London: Academic Press.
- Van Kuyk, J.J. (1988). Verwerven van grootte-begrippen. *Pedagogische Studiën*, 65, 281-290.

Manuscript ontvangen 2-1-1990

Definitieve versie ontvangen 27-3-1990

Notities en Commentaren

DE EERSTE ERVARINGEN MET EEN GEÏNDIVIDUALISEERDE TOETS BESTAANDE UIT 600 VRAGEN

H.J.M. van Berkel, E. de Graaff, J. Sprooten en H.G. Schmidt

Vakgroep Onderwijsontwikkeling & Onderwijsresearch, Rijksuniversiteit Limburg

De Faculteit der Gezondheidswetenschappen van de Rijksuniversiteit Limburg is een grote faculteit - 1850 studenten verspreid over vier jaargroepen - die bestaat uit zeven afstudeerrichtingen. Zoals aan iedere faculteit van de Rijksuniversiteit Limburg, worden de kennisdoelstellingen van de opleiding getoetst met behulp van de zogenaamde 'voortgangstoetsen'. De voortgangstoets is een toets die het kennisdomein van de hele studie omvat en waarvan de vragen een afspiegeling vormen van het eindniveau van de opleiding. De toets wordt driemaal per jaar afgenomen aan alle studenten van de faculteit. De studenten moeten dus gedurende hun vierjarige opleiding in totaal twaalf voortgangstoetsen afleggen. Omdat de toetsen zijn samengesteld uit vragen die uit hetzelfde domein afkomstig zijn (het einddoel van de opleiding), worden de twaalf toetsen geacht equivalent te zijn. Op die wijze ontstaat inzicht in de kenniscroei van de student gedurende de studie waarbij tevens kan worden gezien of de student het beoogde eindniveau zal halen (zie Wijnen en Van der Vleuten, 1985; Van Berkel, 1990 voor meer informatie over deze vorm van toetsing).

Het zal duidelijk zijn dat een groot aantal vragen is vereist in een kennistoets die de pretentie heeft de kennis te toetsen van het hele vierjarige curriculum. Er moet immers sprake zijn van een behoorlijke 'trefkans'. Dat wil zeggen, de toets moet de student voldoende kans geven om de geleerde kennis te uiten. Met andere woorden, het hele onderwijsprogramma, keuzevakken, bijvakken en het gemeenschappelijke programma, dient in de toets te zijn gerepresenteerd. In de medische faculteit zijn positieve ervaringen opgedaan met een toets van 250 vragen. Maar de medische faculteit kent een nauwelijks gedifferentieerd curriculum. In de Faculteit der Gezondheidswetenschappen is sprake van zeven afstudeerrichtingen. Bovendien bevat het curriculum veel keuze-onderdelen. Toch staat op de bul van de afgestudeerde dat hij/zij 'gezondheidswetenschapper' is. Dit betekent dat de faculteit moet toetsen of de afgestudeerde aan dit predikaat voldoet. Binnen de context van de toetsvorm kon de faculteit niet zeven afzonderlijke voortgangstoetsen maken. Enerzijds is het noodzakelijk één algemene toets te ontwikkelen die het kennisdomein van de gezondheidswetenschapper representeert, anderzijds is er de roep om het sterk gedifferentieerde onderwijs te toetsen. In het verleden heeft de faculteit dit dilemma opgelost door de toets uit twee gedeelten te laten bestaan: een algemeen gedeelte (negentig vragen) en een afstudeerspecifiek gedeelte (tachtig vragen). Dit bracht de nodige problemen met zich. Op de eerste plaats waren de algemene vragen niet algemeen genoeg, en wel om twee redenen: a. niemand in de faculteit had er enig idee van wat nu precies karakteristiek is voor een 'gezondheidswetenschapper' ('Wat is dat eigenlijk voor iemand?'); b. docenten in de faculteit voelen zich op de eerste plaats verantwoordelijk voor de afstudeerrichting waarmee zij bemoeienis hebben en dat maakt het wel erg verleidelijk om het accent niet te leggen op algemeenheden. Op de tweede plaats was het algemeen gevoel dat de toets van honderdzeventig vragen te klein in omvang was om een juist beeld te geven van het gehele kennisdomein. Op de derde plaats kreeg het keuze-onderwijs geen plaats in de toets.

De faculteit koos voor een, naar onze inzichten, volstrekt unieke en onorthodoxe

oplossing voor dit probleem. Zij ontwikkelde een toets bestaande uit zeshonderd vragen die wordt geacht het kennisdomein van de gezondheidswetenschappen en het totale onderwijsprogramma te representeren. Hoewel het aantal toetsvragen arbitraire aspecten heeft, is dit aantal niet helemaal toevallig. Uitgaande van de positieve ervaringen van de medische faculteit met een representatieve studietoets in een over het algemeen uniform curriculum, koos de faculteit in eerste instantie voor een toets bestaande uit zeven maal 250 vragen. Vervolgens werd de overlap van de verschillende afstudeerrichtingen onderzocht. Het is immers uiteraard niet zo dat de zeven afstudeerrichtingen totaal onafhankelijk zijn. Bijvoorbeeld, richtingen als Bewegingswetenschappen en Biologische Gezondheidskunde vertonen een ruime overlap die niet apart hoeft te worden getoetst. Er werd een blauwdruk gemaakt die 51 vakken omvatte en met behulp waarvan de zeven afstudeerrichtingen werden gekarakteriseerd. Wanneer de overlappingsen werden verwijderd uit de imaginaire toets van 1750 vragen, ontstond een toets van zeshonderd vragen waarin tevens de keuze- en bijvakken waren vertegenwoordigd.

De in twee parallelversies gesplitste toets wordt op één dag afgenomen in twee sessies van ieder drie uur.

Instructie

De onderstaande invulinstructie wordt aan de studenten gegeven.

welke vragen beantwoorden?

De zeshonderd vragen in de toets zijn gerubriceerd onder bepaalde vakken. Het is NIET de bedoeling dat je bij vraag 1 begint en vervolgens successievelijk alle andere beantwoordt. Zoek allereerst die vakken op waarvan je het idee hebt dat je de vragen kunt beantwoorden, bijvoorbeeld op basis van het gevolgd onderwijsprogramma, je vooropleiding en/of je persoonlijke interesse. Als je nog tijd over hebt, bekijk dan de vragen uit de vakken die in eerste instantie een lagere prioriteit hebben verkregen.

hoe antwoorden?

Iedere vraag kun je beantwoorden met 'juist', '?' of 'onjuist', maar je kunt ook helemaal geen antwoord geven. Het '?' vul je in als je de vraag hebt gelezen en hebt nagedacht over het mogelijke antwoord maar het goede antwoord niet weet. Als je de vraag slechts vluchtig hebt gelezen, of helemaal niet, vul dan niets in. Het maakt voor de scoreberekening niets uit of je de vraag overslaat of dat je de vraag met een '?' hebt beantwoord. In beide gevallen krijg je nul punten.

Analyses

Er heeft een uitgebreide rapportage plaatsgevonden van de eerste bevindingen. Hier gaan we alleen in op de belangrijkste resultaten. Het voor de faculteit aangepaste analyseprogramma gaat uit van de vier jaargroepen.

Tabel 1: Betrouwbaarheid (KR-20) en absolute scores - per jaargroep

Jaar	Aantal stud.	KR-20	G-F score	Goed score	Fout score	G+F score
1	367	0.96	20.4	42.0	21.7	63.7
2	508	0.96	44.9	86.2	41.3	127.5
3	388	0.94	49.0	87.1	38.1	125.2
4	290	0.96	62.3	110.4	48.1	158.5

Uit Tabel 1 kunnen de volgende conclusies worden getrokken:

- Over de vier jaargroepen vindt groei van kennis plaats.
- Deze groei betreft zowel het aantal goede antwoorden als de "goed min fout" score. Ook het aantal vragen dat foutief wordt beantwoord vertoont een stijging. Naarmate studenten meer weten, beantwoorden ze meer vragen en lopen dus vaker het risico met gedeeltelijke kennis vragen fout te beantwoorden. Deze bevinding is niet abnormaal. Ook de voortgangstoetsen in de medische en de juridische faculteit vertonen dit beeld.
- De groei tussen de cohorten twee en drie is relatief gering. Omdat het hier de eerste voortgangstoets van het cursusjaar betreft, gaat het in feite om het tweede cursusjaar. De groei in kennis in dat jaar, als er tenminste sprake is van groei, wordt door de voortgangstoets niet duidelijk zichtbaar gemaakt.
- De betrouwbaarheid van de toets is zeer hoog. Dit is niet verwonderlijk, omdat deze maat wordt beïnvloed door het aantal vragen in de toets. Hoe groter het aantal vragen, des te hoger de betrouwbaarheid, mits de vragen onderling gemiddeld bezien positief correleren.
- Een beginnend student aan de faculteit beantwoordt per saldo 20,4 vragen correct. Hieruit volgt dat nieuwkomers reeds een bepaalde kennis bezitten van gezondheidswetenschappen.
- De absolute G-F score loopt op tot 62,3 voor vierdejaars studenten. Of dit een hoge of lage score is, is niet eenvoudig te beargumenteren. Op de eerste plaats gaat het bij voortgangstoetsscores niet zozeer om de absolute hoogte, maar om groei in kennis aan te tonen (al is het natuurlijk bedenkelijk wanneer de scores zeer laag zijn). Op de tweede plaats is het niet duidelijk wat de maximaal te behalen score is omdat de voortgangstoets als het ware geïndividualiseerd is. De lengte van de toets is voor iedere student anders. Een indicatie voor de toetslengte is de laatste kolom in Tabel 1. Deze optelsom van goed en fout scores is een ondergrens voor de toetslengte omdat mag worden verondersteld dat de student dit aantal vragen serieus heeft bekeken. Zo beredeneerd beantwoorden de studenten gemiddeld ongeveer 70% van de serieus bekeken en beantwoorde vragen correct.

LITERATUUR

- Berkel, H.J.M. van, (1990). Assessment in a problem-based medical curriculum. *Higher Education*, 19, 123-146.
- Wijnen, W.H.F.W. & Vleuten, C.P.M. van der, (1985). Toetsing: hordenloop of voortgangscontrole? *Universiteit en Hogeschool*, 31, 270-279.

Ontvangen 5-3-1990

Boekbesprekingen

C.M. van Putten (1987), *Leerlingen van het individueel beroepsonderwijs nader beschouwd: Een vergelijkend onderzoek naar cognitieve, sociaal-emotionele en achtergrondkenmerken van leerlingen bij de aanvang van het IBO, LBO en VSO/MLK*. (Proefschrift R.U. Leiden, Promotores: Prof. dr. P. van den Broek, Prof. dr. J.P. van de Geer). XII + 257 pp.

In dit promotie-onderzoek op basis van SVO-project 0387, van 1977 tot 1983 in opdracht en met subsidie van het Ministerie van Onderwijs en Wetenschappen uitgevoerd, staat het Individueel Beroeps Onderwijs (IBO) centraal. Het IBO, uiteenvallend in het Individueel Landbouw Onderwijs (ILO), Individueel Huishoud- en Nijverheidsonderwijs (IHNO) en Individueel Technisch Onderwijs (ITO), maakte als gevolg van de verlenging van de leerplicht en de veranderingen in de doelstellingen en leerinhouden van het Lager Beroeps Onderwijs (LBO) een stormachtige ontwikkeling door: van een eerste ITO-opleiding in 1953 in Breda tot bijna 500 opleidingen met meer dan 60.000 leerlingen in 1984.

Hoewel in 1984 geen globale groei meer valt te constateren, staat tegenover een dan al jaren durende daling van ITO en IHNO, nog steeds een groei van het ILO. Evenals het LOM-onderwijs dat een vergelijkbare snelle groei doormaakte, moet volgens Van Putten het IBO worden gezien als een voorziening voor leerlingen waar het reguliere onderwijs (het LBO in dit geval) geen raad mee weet. Onder meer vanwege de zeer geringe doorstroming naar het LBO functioneert naar zijn mening het IBO in ieder geval niet als de beoogde didactische differentiatie met eenzelfde leerdoel als het LBO.

Achtergrond voor het onderzoek vormt de problematische positie die het IBO om verschillende redenen in het Nederlandse onderwijsbeleid inneemt. Allereerst is er de lang doorgezette groei zelf. Het gaat hier evenals bij het LOM om een grotendeels ongestuurde en onverwachte ontwikkeling waarop de overheid alleen al uit kostenoverwegingen greep moet zien te krijgen. In het kader van het streven naar integratie van het Voortgezet Onderwijs is het IBO eveneens een probleem. Vraag is, of en in hoeverre het IBO samen met het LBO in één geïntegreerde onderwijsvorm zal moeten opgaan. Zo niet dan wordt het probleem hoe het zich als een eventueel apart schooltype moet verhouden tot het Voortgezet Speciaal Onderwijs (VSO). Adviezen aan de regering zoals het advies van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) over de invoering van een algemene basisvorming in de eerste drie jaar van het Voortgezet Onderwijs, schenken aan het IBO nauwelijks of geen aandacht.

Naar de mening van Van Putten wordt in de discussie over de plaats van het IBO en de IBO-leerlingen in een toekomstig stelsel van geïntegreerd Voortgezet Onderwijs onvoldoende aandacht geschonken aan de kenmerkende verschillen tussen leerlingen. Om leerproblemen aan te pakken zoeken de regering en haar adviseurs een oplossing in het vergroten van de mogelijkheden van het reguliere onderwijs ('vergroting van de zorgbreedte'). Reacties uit het IBO zelf benadrukken eenzijdig de unieke problematiek van de leerlingen.

Doel van het onderzoek is informatie te verzamelen over kenmerken van IBO-leerlingen bij hun entree in het Voortgezet Onderwijs en deze te vergelijken met die van VSO- en LBO-leerlingen. Die informatie acht Van Putten van belang bij de besluitvorming over de structuur van het Voortgezet Onderwijs en bij de oplossing van vraagstukken in het bestaande onderwijs zoals met betrekking tot de doelstellingen en de differentiatie van het onderwijs en de toelating en de begeleiding van de leerlingen. Informatie over IBO-leerlingen is eerder door anderen verzameld, maar Van Putten acht dit materiaal

onvoldoende representatief omdat het ILO buiten beschouwing werd gelaten en naast de cognitieve vaardigheden en schoolvorderingen slechts weinig aandacht werd geschonken aan sociaal-emotionele gegevens en milieu-achtergronden.

In hoofdstuk 2 van het proefschrift wordt de onderzoeksopzet besproken. In de hoofdstukken 3, 4 en 5 komen achtereenvolgens de drie onderscheiden categorieën variabelen aan de orde: cognitieve leerlingkenmerken, sociaal-emotionele leerlingkenmerken, persoons- en achtergrondkenmerken. Ieder van deze hoofdstukken begint met een gedetailleerde beschrijving van afzonderlijke variabelen voor de onderzochte 436 VSO/MLK-leerlingen, 2091 IBO-leerlingen en 1699 LBO-leerlingen. Daarna volgt telkens een analyse van de samenhang tussen de variabelen, hetgeen in de hoofdstukken 3 en 4 'resulteert in de reductie van de data tot enkele factoren' (p. 9). Hoofdstuk 5 wijkt enigszins af doordat de analyse van de samenhang van de variabelen 'niet wordt afgesloten met een samenvatting van de veelheid aan informatie in enkele factoren' (p. 119). De auteur motiveert dit met 'het veelal nominale karakter van de achtergrondvariabelen en ook, zoals zal blijken, door de betrekkelijk geringe samenhang ervan' (p. 119).

In hoofdstuk 6 wordt nagegaan, hoe de verschillende schooltypen van IBO en LBO en het VSO/MLK zo goed mogelijk van elkaar kunnen worden onderscheiden. Het proefschrift sluit af met conclusies en aanbevelingen.

Ten aanzien van de *cognitieve leerlingkenmerken* komt Van Putten tot conclusies die opmerkelijk mogen heten: er zijn weliswaar verschillen tussen de IBO-gemiddelden en die van het VSO/MLK aan de ene en het LBO aan de andere kant, maar er is tussen IBO en LBO een grote overlap; een groot deel van de leerlingen heeft eenzelfde intelligentie- en schoolprestatieniveau. De verschillen tussen IBO en VSO/MLK zijn veel groter dan die tussen IBO en LBO en van overlap is daar nauwelijks sprake.

De *sociaal-emotionele variabelen* werden aan twee bronnen ontleend: een leerkrachtvragenlijst, die voor iedere leerling door twee leerkrachten werd ingevuld, en een leerlingvragenlijst. De leerkrachten bleken weinig verschil te maken tussen de 39 vragenlijst-items, terwijl de 3 geconstrueerde factoren slechts kleine verschillen lieten zien tussen de leerlinggroepen: LBO-leerlingen scoren in vergelijking met IBO- en in het bijzonder VSO/MLK-leerlingen iets positiever. Binnen het IBO wordt de IHNO-groep (meisjes) wat positiever beoordeeld qua inzet en gedrag dan de andere groepen. Teleurstellend is de lage interbeoordelaarsovereenstemming tussen de leerkrachten. Opvallend is dat leerkrachten met veel lesuren een negatiever oordeel geven, in het bijzonder met betrekking tot de beoordeling van storend gedrag. Ook met betrekking tot de 65 items van de leerlingvragenlijst was er sprake van antwoordtendenties en waren de verschillen tussen leerlinggroepen betrekkelijk klein in verhouding tot de verschillen binnen de groepen. De LBO-groep als geheel heeft een iets betere relatie met medeleerlingen dan de IBO-groep, maar de motivatie van IBO-leerlingen valt iets hoger uit dan die van de LBO-leerlingen.

Met betrekking tot de *persoons- en achtergrondkenmerken* constateert de auteur dat het IBO en het VSO/MLK in het eerste leerjaar een veel grotere groep oudere leerlingen van 14 à 15 jaar telt dan het LBO. Alle groepen laten over het algemeen een laag opleidings- en beroepsniveau van de ouders zien, terwijl ook het percentage arbeidsongeschikte vaders boven het landelijk gemiddelde ligt. Tussen de groepen bestaan op deze variabelen geen grote verschillen. Met betrekking tot de overige variabelen die naar de mening van de auteur van belang zijn voor de schoolloopbaan, blijkt een aantal omstandigheden (anderstaligheid, zeer grote gezinnen, het ontbreken van de vader thuis, en het niet lezen van dag- en weekbladen) in de steekproef te weinig voor te komen voor een zinvolle analyse. Een aantal andere omstandigheden (het spreken van een dialect, het aanwezig zijn van maar zeer weinig boeken, het van de ouders tot slechts 16 jaar of jonger naar school moeten) blijkt iets gunstiger uit te vallen voor de LBO-groep. Ook hier zijn de verschillen

tussen de groepen echter te gering.

Gezien de eerder vermelde resultaten kan het niet verbazen dat in hoofdstuk 6 de combinatie van schoolprestaties en intelligentiescores de drie groepen het best van elkaar blijkt te onderscheiden. De differentiatie tussen VSO/MLK en IBO lukt het best. IBO en LBO kennen een aanzienlijk overlappingsgebied, waarbinnen nog wel profielverschillen zijn aan te wijzen, maar deze zijn van veel geringere betekenis dan in voorgaand onderzoek is gevonden. Ook binnen het IBO en binnen het LBO treden slechts geringe verschillen tussen schooltypen op, die deels op verschillen tussen jongens en meisjes berusten.

Het proefschrift van Van Putten is een omvangrijke studie waarin een massa gegevens van uitzonderlijk grote groepen leerlingen uit IBO, VSO/MLK en LBO is opgenomen. Het proefschrift roept diverse vragen op met betrekking tot de motivatie voor het onderzoek, de toegepaste methodologie en de beleidsconsequenties die uit het onderzoek worden afgeleid. Als motivatie voor een onderzoek van een dergelijke omvang had men meer verwacht dan 'het actualiseren en het verbreden van de kennis van het IBO-populatie' (p. 8). Voor de keuze van de onderzochte variabelen zoekt men tevergeefs naar een overtuigende motivatie vanuit theoretische uitgangspunten of vanuit een analyse van de beleidsproblemen die de aanleiding vormden tot dit onderzoek. Weliswaar hebben eerdere onderzoekers minder aandacht besteed aan sociaal-emotionele gegevens en aan milieu-achtergronden, maar deze variabelen komen ook bij Van Putten slecht uit de verf. Niet alleen vertonen weinig van deze variabelen verschillen tussen de groepen maar de kwaliteit van de gekozen instrumenten laat nogal te wensen over. Dit kan mede de geringe verschillen tussen de groepen tot gevolg hebben gehad. Zowel bij de leerkracht- als bij de leerlingenvragenlijst was er sprake van antwoordtendenties, terwijl bij de leerkrachten de interbeoordelaarsovereenstemming gering was. Bij de persoons- en achtergrondkenmerken bleek een aanzienlijk aantal kenmerken zo weinig voor te komen dat geen conclusies getrokken konden worden voor verschillen tussen de groepen. Voor een onderzoek van een dergelijke omvang (4226 personen) had het voor de hand gelegen eerst via kleinere voorstudies te komen tot instrumenten van voldoende kwaliteit.

Ook de toegepaste methodologie geeft aanleiding tot vragen. We willen volstaan met twee punten. In de eerste plaats maakt het proefschrift een kwistig gebruik van multivariate technieken van meestal 'Leidse' snit: PCA, PRINCALS, HOMALS, CANALS, OVERALS. De argumentatie voor de toepassing van deze technieken is niet steeds duidelijk. Bijvoorbeeld het argument 'soms is het heel interessant om niet-lineaire relaties van data op intervalniveau te bestuderen' (p. 29) zal de doorsnee onderwijsonderzoeker niet overtuigen. Nadat de verschillen tussen de groepen univariaat al dan niet zijn aangetoond, worden ze via deze technieken nog eens multivariaat herhaald met in essentie dezelfde conclusies en in hoofdstuk 6 worden de verschillen andermaal geanalyseerd via CADA (Canonische Discriminant Analyse). In sommige gevallen kan het nuttig of zelfs noodzakelijk zijn om bij onoverzichtelijke grote databestanden verschillen op gecomprimeerde wijze multivariaat te beschrijven in plaats van univariaat. Het lijkt echter te veel van het goede om het beide te doen en dan ook nog eens via CADA. Een tweede punt betreft het feit dat nergens significanties voor de verschillen tussen de groepen worden vermeld. Dat is voor de doorsnee consument van onderwijsresearch zeer onbevredigend. De argumentatie is dat de steekproeven van die omvang zijn dat als het ware de populaties worden bestudeerd. Het bezwaar dat zelfs kleine verschillen significant worden bij grote steekproeven, kan echter gemakkelijk worden ondervangen door niet de significanties maar de overschrijdingskansen, een of andere systematische maat voor de sterkte van verschillen (bijvoorbeeld omega-kwadraat) of beide te vermelden.

Ten slotte zijn er vragen te stellen naar aanleiding van de beleidsconsequenties. In het

slothoofdstuk worden voor het huidige onderwijsstelsel en voor de (wellicht) komende herziening van het voortgezet onderwijs door invoering van de basisvorming, conclusies geformuleerd. Wat het huidige onderwijs betreft, gaat Van Putten vooral in op de pedagogische en didactische betekenis van de onderzoeksresultaten. Dat is nuttig en prijzenswaardig en hij zegt daar ongetwijfeld behartenswaardige dingen, waar veel docenten en opleiders hun voordeel mee kunnen doen. Voor het beleid relevante conclusies worden daar echter niet getrokken en minstens op één punt zou men die toch verwachten. Een zeer intrigerende vraag die het onderzoek opwerpt, is immers: hoe zit het eigenlijk met de toelatingsprocedures voor het IBO, als er, bij screening van LBO- en IBO-leerlingen met hetzelfde type instrumenten als bij het toelatingsonderzoek wordt gebruikt, zo'n grote overlap blijkt te zijn in leerlingkenmerken?

Wat beleidsconsequenties ten aanzien van de voorgenomen invoering van de basisvorming betreft is de auteur minder terughoudend. Zijn conclusies zijn echter nogal onverwacht. In het begin van de studie wordt gesteld, dat informatie over eventuele verschillen tussen IBO- en LBO-leerlingen van groot belang zijn bij de beleidsvorming rond de Basisvorming. Nu deze verschillen na afloop van het onderzoek betrekkelijk gering blijken te zijn, zou men als conclusie verwachten dat de Basisvorming zonder al te grote problemen ook voor de IBO-populatie kan worden doorgevoerd. Van Putten concludeert echter 'dat het aanvangsniveau van de IBO-leerlingen, uitzonderingen daargelaten, volstrekt ontoereikend is voor de aard, de omvang en het eindniveau van het curriculum van de basisvorming' (p. 220). Ook Van Putten laat er geen twijfel aan bestaan dat dit standpunt empirische ondersteuning behoeft door voor zowel IBO- als LBO-leerlingen naar het eindniveau te kijken, 'zodat een exacter beeld van de discrepantie tussen de huidige examenprestaties en de voorgestelde afsluitingseisen kan worden gegeven'. (p. 211). Een onderzoek waarin de huidige examenprestaties worden betrokken is door Van Putten echter niet uitgevoerd.

Wel haalt hij CBS-cijfers aan over de huidige deelname van LBO-leerlingen aan examens op de verschillende niveaus die nu onderscheiden worden. Daaraan valt echter geen steun te ontleen voor de stelling van de auteur dat de basisvorming voor IBO-leerlingen 'een drastische verhoging (betekent) van het minimaal vereiste niveau waarop de vakken moeten worden afgesloten' (p. 209). De deelnamecijfers zijn immers niet uitgesplitst naar IBO en LBO, zoals de auteur zelf opmerkt. Maar zelfs als dat in de toekomst wel zou gebeuren, worden we daar niet veel wijzer van. Er is immers weinig of niets bekend over wat de huidige examenniveaus precies inhouden: voor de examens op A- en B-niveau zijn er geen centraal vastgestelde en genormeerde examenopgaven, en het enige dat we met zekerheid over de A- en B-examens kunnen zeggen, is dat er zowel naar inhoud als naar niveau grote verschillen tussen scholen zijn. Het is, kortom, niet uitgesloten dat de auteur de haalbaarheid van de basisvorming voor IBO-leerlingen terecht in twijfel trekt, maar voor een adequate motivering van die stelling is een ander onderzoek nodig dan het hier gepresenteerde.

J.H.L. Oud
K.U.Nijmegen
Vakgroep Orthopedagogiek

M.M.B. Oud- de Glas
K.K. Nijmegen
Instituut voor Toegepaste
Sociale Wetenschappen

H. van den Bergh (1988), *Examens geëxamineerd*. (Proefschrift Universiteit van Amsterdam) Den Haag: SVO.

Het promotie-onderzoek van Van den Bergh is gestart naar aanleiding van een wijziging in de onderdelen tekstbegrip en schrijfvaardigheid van het examen voor LBO en MAVO. Uitgangspunt van zijn onderzoek zijn twee vragen:

- doen open en gesloten vragen tekstbegrip een beroep op dezelfde intellectuele vaardigheden en
- doen het traditionele opstel en functionele schrijfoopdrachten een beroep op dezelfde intellectuele vaardigheden?

Het meest pregnant worden deze vragen beantwoord in de stellingen bij het proefschrift. Op deze stellingen kom ik aan het eind van de bespreking van het proefschrift nog terug. Eerst zal ik de voornaamste uitkomsten van het onderzoek aan de orde stellen en daarbij vooral kijken naar de waarde van de uitkomsten.

Het proefschrift van Van den Bergh heeft de vorm van een serie tijdschriftartikelen. Dit heeft, zoals de auteur zelf in zijn voorwoord opmerkt, het voordeel van afzonderlijk leesbare hoofdstukken en een duidelijk doorzichtige structuur van elk hoofdstuk. De nadelen zijn de beknoptheid en de vrij grote overlap tussen de hoofdstukken. Naar mijn mening wegen in dit geval de voordelen tegen de nadelen op. Een belangrijke factor daarbij is de heldere, zakelijke schrijfrant van Van den Bergh.

Intellectuele vaardigheden vormen in beide vragen de onafhankelijke variabele. Als uitgangspunt om deze vaardigheden te beschrijven, kiest Van den Bergh het 'Structure of Intellect' model van Guilford. Bij de voor taalvaardigheid belangrijke cellen uit de Operaties *Modaliteit* Producten-matrix van dit model heeft hij in totaal 47 toetsen geconstrueerd. In hoofdstuk 6 toont Van den Bergh aan dat de correlatieve structuur tussen deze toetsen niet door eenvoudiger modellen, zoals het één-factor model van Spearman of het zeven-factoren model van Thurstone, verklaard kunnen worden en dat het gerechtvaardigd is de cellen uit het model van Guilford, zoals dat door hem is geoperationaliseerd, als onafhankelijke variabelen te hanteren.

In twee hoofdstukken gaat Van den Bergh in op de hoofdvragen: of open en gesloten vragen, respectievelijk opstel en functionele schrijftaken, eenzelfde beroep doen op de intellectuele vaardigheden zoals die met de Guilford-toetsen worden gemeten.

Hij komt tot de conclusie dat er geen substantieel verschil is tussen tekstbegriptoetsen met open en gesloten vragen. Zowel het design – random toewijzen van leerlingen aan de verschillende tekst- en vraagvormcondities – *als de analyse* – een meergroepenanalyse met LISREL – geven vertrouwen in de uitkomst. De hoge proporties verklaarde variantie in de tekstbegriptoetsen maken het bovendien onwaarschijnlijk dat andere, niet gemeten intellectuele vaardigheden een heel ander beeld te zien zouden geven. Aan de verklaarde variantie dragen de toetsen die de geheugenaspecten meten, overigens het meeste bij.

Voor opstel en functionele schrijftaken komt de auteur tot de conclusie dat beide vormen voor het meten van schrijfvaardigheid in gelijke mate een beroep doen op de onderscheiden intellectuele vaardigheden. Deze conclusie blijkt alleen te gelden als dezelfde beoordelings-procedure wordt gebruikt. Indien globaal wordt beoordeeld komen als intellectuele vaardigheden vooral coherentie, eenduidigheid en gevarieerdheid van woordkeuze aan bod. Scores op basis van een analytisch voorschrift worden meer bepaald door geheugenaspecten. Bij de beoordeling van schrijftaken, is de proportie variantie die door de Guilfordtoetsen verklaard wordt overigens beduidend kleiner dan bij de resultaten voor tekstverklaring. Hier is dus nog voldoende ruimte voor andere verklarende factoren.

Naast zijn hoofdonderzoek doet Van den Bergh verslag van drie andere deelonderzoeken: naar beoordelaarsbetrouwbaarheid, naar de correlatiestructuur van de oordelen over schrijftaken en naar verschillen tussen open en gesloten vragen.

Bij de beoordelaarsbetrouwbaarheid wordt een elegante procedure beschreven om in een opzet waarbij een aantal beoordelaars verschillende, overlappende porties teksten beoordelen, verschillende betrouwbaarheidsmodellen te toetsen. Het onderzoek naar de correlatiestructuur van de oordelen over de schrijftaken levert een duidelijke uitkomst op: er is een redelijk hoge samenhang tussen de beoordeelde aspecten binnen een taak, tussen taken is er nauwelijks correlatie.

Het onderzoek naar de open en gesloten vragen voor tekstbegrip levert een interactie op tussen vraagtype en schooltype (MAVO en LBO) en tussen vraagtype en itemmoeilijkheid. De interactie met schooltype geeft aan dat het verschil tussen MAVO en LBO op de meerkeuzevragen kleiner is dan op de open vragen.

Onderzoektechnisch valt op het onderzoek van Van den Bergh niet veel aan te merken. De designs zijn met zorg gekozen en de meest gebruikte analysemethode, LISREL, is weliswaar riskant, maar in het onderzoek wordt er zorgvuldig mee omgesprongen. De uitkomsten als zodanig zijn dan ook geloofwaardig. Wat moeilijker ligt het met de generaliseerbaarheid: onder welke omstandigheden mag aangenomen worden dat de conclusies geldig blijven. De voornaamste beperking daarbij vormt het beperkte instrumentarium. De toetsen bij het Guilfordmodel zijn alle nieuw gemaakt. Alleen de interne structuur is gevalideerd. Hoewel de relaties met de tekstbegrip-en schrijftaken indirecte evidentie leveren voor de validiteit, is het jammer dat de relatie met andere operationalisaties van de betrokken intellectuele vaardigheden niet is nagegaan. Ernstiger zijn de problemen bij de afhankelijke vaardigheden. Er zijn twee, nominaal parallelle tekstbegriptoetsen geconstrueerd, elk in twee vormen en er zijn vijf schrijftaken betrokken. Bij de tekstbegriptoetsen kan er tot vrij grote hoogte op vertrouwd worden dat de uitkomsten bij soortgelijke instrumenten repliceerbaar zijn. Voor de schrijftaken is dat in veel mindere mate het geval. Het betrokken onderzoek, maar bijvoorbeeld ook het onderzoek van De Glopper (1988) doet ernstig twijfelen aan de mogelijkheid naar andere schrijftaken te generaliseren. Dit laatste is een belangwekkende uitkomst van het onderzoek en kan de auteur dus niet aangerekend worden. Het beperkt echter wel de toepasbaarheid van de uitkomsten.

Tot slot drie meer subjectieve kanttekeningen.

Het proefschrift heeft mij twee zaken duidelijk(er) gemaakt. Op de eerste plaats dienen bij de beoordeling van schrijfvaardigheid niet alleen meerdere beoordelaars, maar ook verschillende beoordelingswijzen gebruikt te worden. Daarnaast dient ernstig getwijfeld te worden aan 'schrijfvaardigheid' als construct.

Het proefschrift 'Examens geëxamineerd' van Huub van den Bergh is er één uit een reeks van proefschriften van medewerkers van het SCO over taal, onder meer schrijven (Rijlaarsdam, 1986; De Glopper 1988; Blok, 1987). Uit deze reeks blijkt duidelijk hoe belangrijk het is dat een instituut en de onderzoekers de kans krijgen om expertise op te bouwen op een bepaald gebied.

In de stellingen bij het proefschrift geeft Van den Bergh op een wat andere manier antwoord op de onderzoeksvragen. Ik wil de lezer deze antwoorden niet onthouden.

Op de vraag of functionele schrijfopdrachten een beroep doen op dezelfde intellectuele vaardigheden als het traditionele opstel, antwoordt hij met de stelling dat functionele taalvaardigheden een begrip is dat beter zo snel mogelijk vergeten kan worden.

Op de vraag of open en gesloten vragen hetzelfde meten, slaat de stelling dat de jaarlijkse discussie rond de voor- en nadelen van meerkeuzevragen lijkt op een symbolische rondedans.

LITERATUUR

- Blok, H. (1987). *Taal voor alle dag: Feiten en meningen over het taalgebruik van LBO- en MAVO-leerlingen in alledaagse situaties*. Den Haag: SVO (Selectareeks).
- Glopper, K. de (1988). *Schrijven beschreven*. Den Haag: SVO (Selectareeks).
- Rijlaarsdam, G. (1986). *effecten van leerlingrespons op aspecten van stelvaardigheid*. Amsterdam: SCO.

Michel Zwarts
Cito
Postbus 1034
6801 MG Arnhem



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

Greetje P.C. van der Werf, RION,
Postbus 1286, 9701 BG Groningen,
tel. 050-636657

Jrg. 14, nr. 3

Juli 1990

Inhoudsopgave

Ontvangen publikaties

200

ONTVANGEN PUBLIKATIES

- Bos, K.T., Kesteren, B.J. van, Stoel, W.G.R. & Vermeulen, C.J.A.J. (1990). *Schoolverzuim: maatregelen die scholen kunnen nemen*. Amsterdam/Lisse: Swets & Zeitlinger (Forum 3).
- Bosman, R. & Louwes, W. (1989). *Gezinssituaties en onderwijskansen*. Een toetsing van verklaringen voor de schoolloopbaanverschillen tussen kinderen van eenoudergezinnen en twee-oudergezinnen. Nijmegen: ITS (OOMO-reeks).
- De Corte, E., Verschaffel, L., Schrooten, H., Indemans, R. & Hoedemaekers, E. (1990). *Logo als springplank voor het verwerven van denkvaardigheden bij zesdeklassers*. Leuven/Amersfoort: ACCO (Leuvense Pedagogische Bijdragen).
- Dekkers, H., Lieshout, F. van & Stortelder, M. (1989). *Onderwijskansen voor meisjes uit etnische minderheden*. Nijmegen: ITS.
- Franssen, H.A.M., Lagerwey, N.A.J. & Stokking, K.M. (1990). *Leraren basisschool over veranderingen in het onderwijs*. Utrecht: ISOR.
- Grift, W. van de & Kurek-Vriesema, C. (1990). *Schoolleidingen in de basisschool*. Amsterdam/Lisse: Swets & Zeitlinger (Forum 1).
- Heijden, M.K. van der & Rispens, J. (red.) (1990). *Zorgverbreding en afgrenzing*. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch nr. 29).
- Heijnen, G.W.H., Joostens, T.H. & Vroeijenstijn, A.I. (red.) (1990). *Kwaliteitszorg, waarborg voor kwaliteit in het hoger onderwijs*. Groningen: COWOG.
- Kessel, N. van & Kral, M. (1989). *Verlangd basisonderwijs in Veldhoven-Noord*. Nijmegen: ITS.
- Kloprogge, J. (1990). *Naar beter voorrangsonderwijs. Notitie over het onderwijsvoorrangsbeleid in 1989*. 's-Gravenhage: SVO.
- Langen, A. van & Jungbluth, P. (1990). *Onderwijskansen van migranten: de rol van sociaal-economische en culturele factoren*. Amsterdam/Lisse: Swets & Zeitlinger (Forum 6).
- Pelkmans, A. & Kuyk, J. van (1990). *De Stedelijke Scholengemeenschap Zutphen: opgave voor vandaag en morgen. Samenvatting rapportage marketingonderzoek*. Nijmegen: ITS.
- Petersen, B. & Prins, J. (1989). *Het functioneren van de propedeuse van PABO's*. Nijmegen: ITS/IOWO.
- Roede, E., Kok, F. & Vermeulen, A. (1989). *Zelfstandigheid motiveert leerlingen*. Amsterdam: SCO (SCO-SCO-rapport nr. 205).
- Ruijsenaars, A.J.J.M. & Hamers, J.H.M. (red.) (1989). *Dyslexie: ernstige lees- en spellingsproblemen*. Leuven/Amersfoort: ACCO. f 38,75.
- Ruys, A. & Mientjes, A. (1989). *Werving en selectie van schoolmanagers. Een onderzoek naar de sollicitatieprocedures voor schoolmanagers in het basis- en voortgezet onderwijs*. Nijmegen: ITS.
- Tesser, P., Werf, M.P.C. van der, Mulder, L. & Weide, M.G. (1989). *De beginfase van het voorrangsonderwijs*. Nijmegen/Groningen: ITS/RION.
- Verhaak, C. & Tesser, P. (1989). *Limburgse spijbelopvangprojecten en hun relatie tot het onderwijsvoorrangsbeleid*. Nijmegen: ITS.
- Vries, A. Ph. de (1990). *Afkijken geboden: ervaringen van consulenten, praktijkopleiders, directeuren en vakexaminatoren met de praktijkopleiding in het agrarisch leerlingwezen*. Wageningen: Vakgroep Agrarische Onderwijskunde Landbouwuniversiteit (Agrarisch Onderwijskundige Reeks).
- Vries, B. de & Boer, P. den (1989). *Omwille van de kwaliteit. Onderzoek naar voorwaarden en mogelijkheden van herstructurering van studie- en beroepskeuzevoorlichting en afstemming onderwijs-arbeid op regionaal niveau*. Nijmegen: ITS.

Effecten van een training in basale gespreksvaardigheden

S.M. Bögels¹ en R. Kreutzkamp²

¹Vakgroep Onderwijsontwikkeling & Onderwijsresearch, Rijksuniversiteit Limburg

²Vakgroep Geestelijke Gezondheidskunde, Rijksuniversiteit Limburg

ABSTRACT

During their study students in the Mental Health Sciences are systematically trained in interviewing skills, diagnostic skills, basic therapeutic skills and reporting skills. In addition, they are required to apply these skills in interviews with simulated patients. The purposes of the present study were to evaluate effects on students of a short training in interviewing skills and to compare the level of proficiency in applying these skills as acquired by students with the level of proficiency displayed by professional helpers who were not trained. Results on a behavioral and cognitive measure indicated that the training was effective in learning interviewing skills to students. Furthermore, students attained levels comparable to those of professional helpers on the behavioral measure and patient-satisfaction. However, professionals judged themselves as more interpersonal effective than students.

INLEIDING

Een centrale doelstelling in het onderwijs van gezondheidszorgopleidingen (Geneeskunde, Psychologie, Verpleegkunde, Gezondheidswetenschappen) is het bewerkstelligen van een effectieve communicatie tussen hulpverlener en patiënt. Daarvoor is het nodig dat de hulpverlener zich een aantal communicatieve vaardigheden eigen maakt. Aan de Rijksuniversiteit Limburg wordt bij de vakgroep Onderwijsontwikkeling en Onderwijsresearch vaardigheidsonderwijs voor studenten van de faculteit der Gezondheidswetenschappen ontwikkeld, verzorgd en geëvalueerd.

Eén van de zeven afstudeerrichtingen van Gezondheidswetenschappen heet Geestelijke Gezondheidskunde. In de opleiding worden theorie en praktijk van de geestelijke gezondheidszorg zoveel mogelijk geïntegreerd, door parallel aan het theoretisch onderwijs steeds vaardigheidsonderwijs aan te bieden. Observatievaardigheden, gespreksvaardigheden, diagnostische vaardigheden, therapeutische vaardigheden en rapportage vaardigheden worden in de verschillende trainingen behandeld. De trainingen zijn gebaseerd op het microtraining model van Ivey (1971). Microtraining is een instructiemodel dat complex menselijk gedrag verdeelt in gedragseenheden, en vervolgens deze gedragseenheden traint door instructie, rollenspel, observatie en onmiddellijke concrete feedback.

Een nieuw element bij deze vorm van training is het gebruik van zogeheten simulatiecliënten. Dit zijn daartoe ingehuurd personen die veelal in de geestelijke gezondheidszorg werkzaam zijn en ten behoeve van het onderwijs een cliënt uit de praktijk naspelen. Ze worden voorbereid op het spelen van een rol en het geven van feedback. Iedere training die studenten volgen, wordt afgesloten met een 'simulatiecliëntcontact', waarbij de student de tijdens de training verworven vaardigheden zo goed mogelijk dient toe te passen. De student heeft dan de mogelijkheid om zijn vaardigheden te toetsen aan een (gesimuleerde) praktijksituatie.

¹Postbus 616, 6200 MD Maastricht

Empirische evaluatie van een dergelijk vaardigheidsprogramma is om verschillende redenen van groot belang. Ten eerste dient de vraag te worden beantwoord of studenten datgene leren dat bij de opzet van het programma beoogd werd. Aantoonbare effecten vormen een voorwaarde voor continuering van programma's als deze. Ten tweede kan door middel van meting van vaardigheden systematische feedback aan studenten worden gegeven over de mate waarin de vaardigheden beheerst worden. Zulke feedback is uit onderwijskundig oogpunt essentieel voor het verwerven van die vaardigheden (Riccardi & Kurtz, 1983); men leert niet van ervaring als deze niet begeleid wordt door informatie over de juistheid van de prestatie.

Sinds het werk van Ivey (1971) over microtraining, is er een groot aantal studies gepubliceerd op het gebied van evaluatie van effecten van interpersoonlijke vaardigheidstrainingen voor studenten en hulpverleners in de gezondheidszorg (zie onder meer de overzichtsartikelen van Carrol en Monroe, 1980, Ford, 1979, Kurtz, Marshall en Banspach, 1985, Marshall, Sharping en Bell, 1979). Marshall et al. (1979) vonden na een analyse van 122 studies over de effecten van vaardigheidstrainingen aan maatschappelijk werkers, psychologen, psychiaters en counselors, dat slechts bij de helft van de studies significante resultaten werden gerapporteerd. Carrol en Monroe (1980) kwamen in hun analyse van 33 studies over de effecten van vaardigheidstrainingen aan medische studenten, artsen en verpleegsters tot een veel hoger percentage effectieve trainingen, namelijk 80%. Bij de meeste van de door Carrol en Monroe geanalyseerde studies werd echter gebruik gemaakt van een pretest-posttest design zonder controlegroep, een pré-experimenteel design dat gevoelig is voor vertekende factoren zoals bijvoorbeeld het effect van testen. Bij de overzichtsstudie van Marshall et al. daarentegen ging het voornamelijk om werkelijk experimentele designs. De overige overzichtsstudies rapporteerden geen gegevens over de effectiviteit van trainingen. Bij de meeste effectiviteitsstudies werd als maat voor effectiviteit van de training uitsluitend gekeken naar statistisch significante verschillen. Statistische significantie zegt echter weinig over de sterkte en relevantie van de gevonden verschillen voor de onderwijspraktijk. Een manier om inzicht te krijgen in de betekenis van het vaardigheidsniveau dat door training bereikt wordt, is de prestaties van getrainde proefpersonen te vergelijken met die van passende referentiegroepen, namelijk hun latere beroepsgroep. Stone (1982) noemde dit de sociale vergelijkingsmethode. Op basis van deze vergelijking kan ook de vraag, aan welk eindniveau de trainee dient te voldoen, worden beantwoord.

Hoewel er veel onderzoek werd verricht naar de effecten van trainingen en trainingsprocedures, bleef de validiteitsvraag, namelijk zijn de getrainde vaardigheden relevant voor het bewerkstelligen van een effectieve communicatie, veelal onbeantwoord. Zo is er bijvoorbeeld weinig bekend over de invloed van vaardigheden van de hulpverlener op cliëntvariabelen, zoals de tevredenheid van de cliënt over het contact met de hulpverlener (Ford, 1979, Stone, 1982, Kurtz et al., 1985). Training kan vele potentiële veranderingen tot gevolg hebben; veranderingen in het gedrag van de trainee, veranderingen in de wijze waarop de trainee denkt, veranderingen in de beleving van de persoon op wie de vaardigheden zijn toegepast. De prestaties van de trainee in termen van observeerbare gedragsverandering stonden bij evaluatieonderzoek echter meestal centraal. Wanneer evaluatieonderzoek zich uitsluitend concentreert op observeerbare gedragsverandering bestaat het risico dat de trainee op een mechanische wijze bepaald gedrag vertoont waarbij de functie van dat gedrag uit het oog wordt verloren. Het is daarom van belang trainingen vanuit meerdere invalshoeken te evalueren; het overte gedrag van de trainee, de cognitieve activiteit van de trainee, en de ervaring van de cliënt.

Op basis van bovenstaande overwegingen werd een onderzoek uitgevoerd bij studenten Geestelijke Gezondheidskunde die een training basale gespreksvaardigheden volgden. Het

onderzoek had tot doel:

1. De effecten van een training in basale gespreksvaardigheden uit het vaardigheidsprogramma voor studenten Geestelijke Gezondheidskunde op verschillende variabelen te onderzoeken; het gedrag en de cognities van de student en de satisfactie van de simulatiecliënt.
2. Inzicht te krijgen in de validiteit van de getrainde vaardigheden.
3. De mate van beheersing van basale gespreksvaardigheden van studenten te vergelijken met de vaardigheden van hun latere beroepsgroep.

METHODE

Proefpersonen

Zestien van de zesendertig tweedejaarsstudenten Geestelijke Gezondheidskunde namen vrijwillig deel aan het onderzoek. De reden dat veel studenten niet wilden deelnemen aan het onderzoek was dat zij er tegen opzagen om een gesprek met een simulatiecliënt te voeren. De onderzoeksgroep was derhalve mogelijk niet representatief voor de studentenpopulatie. In de discussie wordt hierop teruggekomen. De onderzoeksgroep bestond uit vijf mannen en elf vrouwen, de gemiddelde leeftijd was 22.4 jaar (range 19 tot 45 jaar). Eén student had eerdere ervaring in de geestelijke gezondheidszorg. Twaalf hulpverleners, zeven mannen en vijf vrouwen, die werkzaam waren in verschillende instituten voor geestelijke gezondheidszorg, namen als referentiegroep deel aan de studie. Vier van hen waren artsen in opleiding tot psychiater, zes waren (sociaal)psychiatrisch verpleegkundigen en twee waren psychologen. De gemiddelde leeftijd was 29.5 jaar (range 23 tot 37 jaar), de gemiddelde duur van hun ervaring in de geestelijke gezondheidszorg was 4.5 jaar. Alleen relatief jong ogende hulpverleners werden geselecteerd, om te voorkomen dat de beoordelaars op grond van leeftijdsverschillen studenten zouden kunnen onderscheiden van professionele hulpverleners. De werkzaamheden van de hulpverleners bestonden onder meer uit het voeren van intake-, begeleidings- en therapiegesprekken met cliënten. Probleemverhelderende gespreksvoering behoorde derhalve tot hun taken.

Onderzoeksopzet

Er werd gebruik gemaakt van een pré-pré-post, of 'time-series', design om verbetering in gespreksvaardigheden bij de studenten te onderzoeken. Drie weken voorafgaand aan de training doorliepen de studenten een gedragstest en vulden enkele vragenlijsten in (zie Metingen). Vlak voor de start van de training werden deze metingen herhaald. Na afloop van de drie weken periode waarin studenten aan de training deelnamen, werden de metingen opnieuw afgenomen. Het theoretisch onderwijs aan de Rijksuniversiteit Limburg is georganiseerd in zes-weekse blokken, waarin een bepaald thema wordt behandeld. De wachtperiode en de trainingsperiode vonden plaats in hetzelfde blok. Het theoretisch onderwijs in de wachtperiode verschilde daarom niet wezenlijk van het theoretisch onderwijs tijdens de training.

Na de laatste meting van de studenten ondergingen de hulpverleners dezelfde gedragstest éénmaal en vulden eveneens de vragenlijsten in.

Training

De training 'Probleemverhelderen' bestond uit 5 bijeenkomsten van 3 uur en werd in groepen van 9 à 10 studenten gevolgd. Studenten dienden vooraf relevante literatuur met betrekking tot de vaardigheden (Lang en Van der Molen, 1984) te bestuderen.

Doelstelling van de training was het adequaat kunnen voeren van een probleemverhel-

derend gesprek. Deze samengestelde vaardigheid werd opgesplitst in een aantal deelvaardigheden, namelijk: kleine aanmoedigingen, open vragen, parafraseren van inhoud, reflecteren van gevoel, samenvatten, concretiseren, situatie verduidelijken. De training startte met een videovoorbeld van het geïntegreerd gebruik van alle deelvaardigheden, zodat studenten een indruk kregen van wat zij in de training zouden gaan leren. Daarna werd tijdens iedere bijeenkomst één of twee deelvaardigheden geïntroduceerd. De nieuwe deelvaardigheid werd eerst verduidelijkt door middel van een voorbereidende oefening. Vervolgens werd de deelvaardigheid in een rollenspel getraind, waarbij de trainee de nieuwe vaardigheid naast de reeds geleerde vaardigheden diende toe te passen. Tot slot werd het geïntegreerd en functioneel gebruik van alle vaardigheden goefend. Door middel van video-opnames en observaties van de trainer en een student-observator kregen de studenten onmiddellijke en concrete feedback op hun gedrag als hulpverlener.

Metingen

Gedragsmeting

Voor de beoordeling van de gespreksvaardigheden werd een scoringslijst ontwikkeld naar voorbeeld van de Maastrichtse Anamnese en Advies Scoringslijst (MAAS). De MAAS is een uitvoerig onderzocht instrument, ontwikkeld voor de beoordeling van medische interviewvaardigheden (Kraan en Crijnen, 1987). De voor dit onderzoek ontwikkelde lijst bestond uit 10 items, waarvan 7 betrekking hadden op gespreksvaardigheden die van belang zijn in een probleemverhelderend gesprek. Het betrof de volgende vaardigheden: aandachtgevend gedrag, vragen stellen, parafraseren van inhoud, reflecteren van gevoel, samenvatten, concretiseren en regulerende vaardigheden (voor een omschrijving van de vaardigheden zie Lang en Van der Molen, 1984). De overige 3 items hadden betrekking op meer algemene aspecten van een gesprek: houding, gesprekstempo en mate waarin het probleem werd verhelderd. Het adequaat en functioneel gebruik van elke vaardigheid werd beoordeeld op een vijfpuntsschaal van 1 (zeer slecht) tot 5 (uitstekend). Tot slot werd de beoordelaars gevraagd de proefpersoon een schoolcijfer te geven met betrekking tot zijn gedrag als hulpverlener; het betrof hier een globaal eindoordeel.

Simulatiecliëntsatisfactie

De simulatiecliënten werd verzocht direct na afloop van de gedragsmeting een simulatiecliëntsatisfactielijst in te vullen. De lijst bestond uit 7 beweringen, waarvan 6 betrekking hadden op het interviewgedrag, het empathisch vermogen en begrip van de hulpverlener, zoals ervaren door de cliënt. De zevende bewering betrof een gedragsintentievraag. Op een zespuntsschaal vulde de cliënt in in hoeverre zij het met de bewering eens was, waarbij de score 1 stond voor geheel oneens en de score 6 voor geheel eens.

De beweringen waren:

De interviewer toonde begrip voor mij en mijn probleem

De interviewer gaf mij de ruimte om te vertellen wat me bezig hield

De interviewer ging in op hetgeen ik tijdens het gesprek naar voren bracht

De interviewer had aandacht voor de verschillende aspecten van mijn probleem

De interviewer drong door tot de kern van mijn probleem

Ik voelde me op mijn gemak gesteld door de interviewer

Bij deze interviewer zou ik terugkomen

Self-efficacy

De cognitieve meting die in dit onderzoek werd gebruikt, had betrekking op de overtuiging van de hulpverlener van de effectiviteit van zijn gedrag in het gesprek met de cliënt ("self-efficacy"). Bandura (1977) veronderstelde dat de mate waarin mensen overtuigd zijn van

de effectiviteit van hun gedrag een belangrijke determinant is van hun gedrag, omdat effectiviteitsverwachting de intensiteit en volharding van het gedrag zou vergroten. Omgedraaid zou het ontwikkelen van vaardigheden bijdragen aan iemands overtuiging van zijn effectiviteit. Self-efficacy werd gemeten door de proefpersonen voorafgaand aan de gedragsmeting te vragen op een zespuntsschaal aan te geven hoe effectief zij dachten verschillende gespreksvaardigheden te kunnen toepassen (1=helemaal niet effectief, 6=uiterst effectief). De belangrijkste vaardigheden uit het probleemverhelderend gesprek: aandachtige houding, luisteren, doorvragen, omgaan met gevoelens, samenvatten en regulerende vaardigheden, werden beschreven. Tevens werd de proefpersonen gevraagd op dezelfde schaal aan te geven hoe effectief zij een probleemverhelderend gesprek dachten te kunnen voeren.

Spanning

Te grote spanning kan voor de studenten een belemmerende factor zijn bij een simulatiecliëntcontact. Temeer daar het voor de studenten die in het onderzoek participeerden de eerste keer was dat zij een gesprek met een simulatiecliënt voerden, kon verwacht worden dat de spanning hen parten zou spelen. Om hiervoor te controleren werd na de gedragsmeting aan de proefpersonen gevraagd hoe gespannen zij tijdens het gesprek geweest waren (1 = helemaal niet gespannen, 6 = uiterst gespannen).

Procedure

Na binnenkomst van de proefpersonen in de onderzoeksruijnte, werd hen gevraagd de self-efficacy lijst in te vullen. Vervolgens lazen zij een brief van de huisarts, die aangaf dat haar patiënt sinds enige tijd last had van hoofdpijn, waarvoor geen somatische oorzaak te vinden was en medicatie slechts tijdelijk verlichting bracht. Zij kregen van de proefleider de instructie gedurende 20 minuten een gesprek met deze cliënt te voeren, met de bedoeling het probleem van de cliënt te verhelderen. De proefpersonen haalden de cliënt zelf uit de wachtkamer en namen haar mee naar een video-opnameruimte. De ruimte was ingericht met twee gemakkelijke stoelen en een lage tafel en er waren twee camera's opgesteld. Eén camera bracht de hulpverlener en cliënt in beeld, de andere camera maakte iedere 5 minuten gedurende 15 seconden een close-up van de hulpverlener. Na afloop van het gesprek werd de proefpersoon gevraagd aan te geven hoe gespannen hij was geweest. Van de video-opnames werden de 5e tot de 15e minuut geselecteerd voor verdere analyse. De reden voor deze selectie was de grote variëteit in wijze van openen van het gesprek, sommige proefpersonen begonnen met een inleiding van 5 minuten terwijl anderen onmiddellijk met de eerste vraag begonnen. Hetzelfde gold voor de afsluiting. De 5e tot de 15e minuut gaf beter vergelijkbaar materiaal.

Vier vrouwelijke simulatiecliënten, die regelmatig als cliënt bij het vaardigheids-onderwijs optraden, werden geïnstrueerd om één van de rollen te vertolken. Er waren drie voorgeschreven rollen, rol A, B, en C. De rollen kwamen overeen wat betreft de aanmeldingsklacht (langdurige en frequente hoofdpijn); de specifieke onderliggende problematiek over het werk, de relatie, of de gezinssituatie verschilde per rol. Eén simulatiecliënt vertolkte rol A, een andere rol B, en twee rol C. De helft van de studenten voerde bij de eerste voormeting een gesprek met cliënt A, en de helft met cliënt B. Bij de tweede voormeting kregen de studenten die bij de eerste voormeting cliënt A hadden gezien nu cliënt B, en vice-versa. Bij de nameting voerden alle studenten een gesprek met cliënt C. De hulpverleners voerden daarna ook een gesprek met cliënt C. De werkwijze van de hulpverleners in dit gesprek was vergelijkbaar met die van de getrainde studenten; eerst werd de aanmeldingsklacht geëxploreerd en daarna werd verder ingegaan op factoren die met de klacht in verband zouden kunnen staan, zoals problemen op het werk of in het gezin.

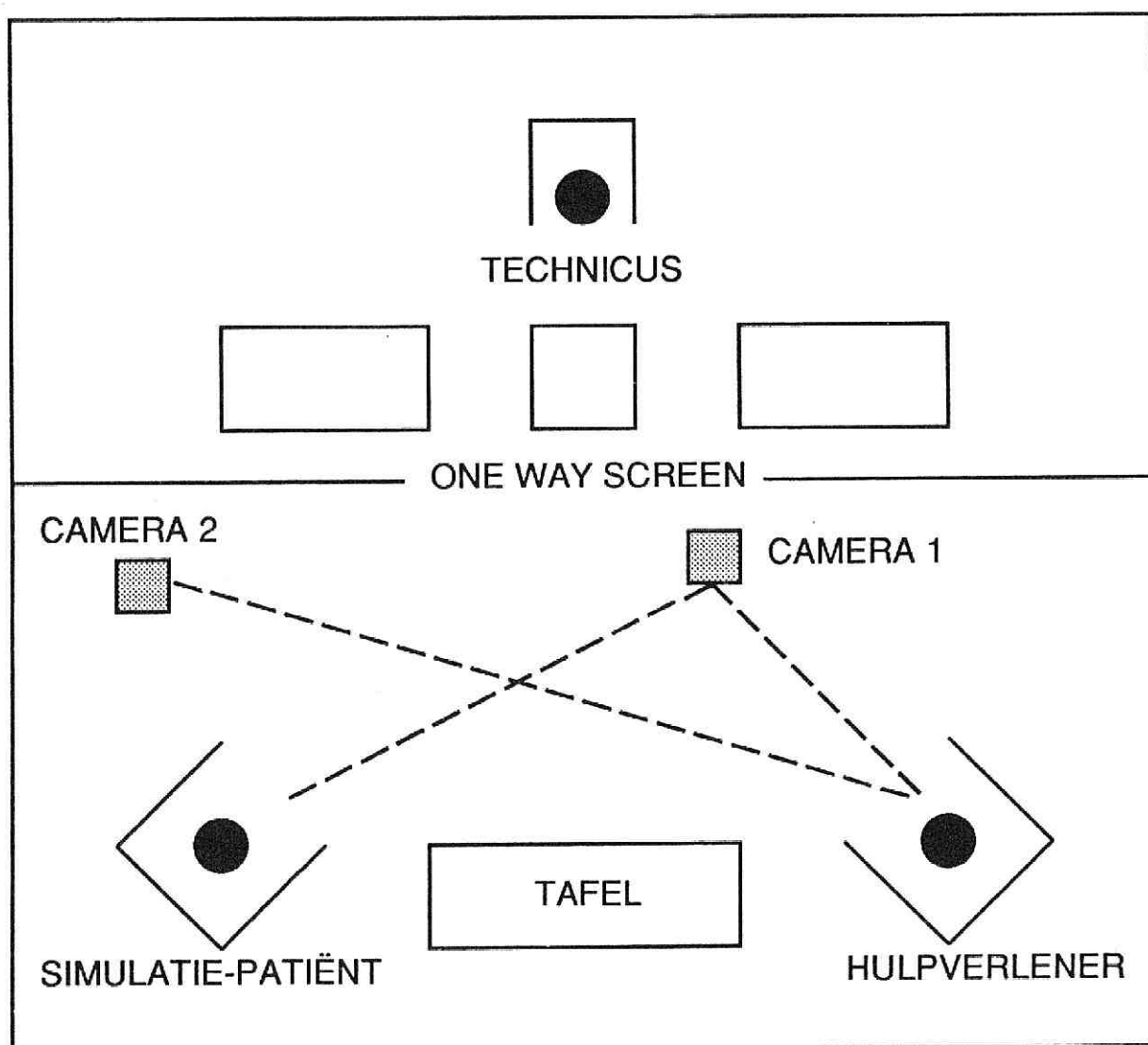


Fig. 1. Een plattegrond van de testkamer

Voor het beoordelen van de gesprekken werd gekozen voor deskundige beoordelaars, dat wil zeggen mensen met kennis over gespreksvaardigheden en bekendheid met het werkveld van de geestelijke gezondheidszorg. Vier psychologen, twee mannen en twee vrouwen, werden gevraagd de videofragmenten te beoordelen. Zij hadden ervaring in het begeleiden van vaardigheidstrainingen en waren werkzaam (geweest) in de geestelijke gezondheidszorg. De beoordelaars werd uitgelegd dat het een onderzoek betrof naar effecten van training in gespreksvaardigheden en dat aan het onderzoek zowel studenten als hulpverleners uit de praktijk deelnamen. Zij werden vooraf uitvoerig geïnstrueerd in het hanteren van het beoordelingsinstrument. De in totaal 60 videofragmenten (3 per student en 1 per hulpverlener) werden aselekt aan de beoordelaars aangeboden. Elk fragment werd door elke beoordelaar gescoord. De beoordeling geschiedde in plenaire sessies van vier uur. Het materiaal werd 'blind' beoordeeld, dat wil zeggen de beoordelaars waren niet geïnformeerd of de opname een student danwel hulpverlener betrof en evenmin of het een voormeting danwel een nameting betrof.

RESULTATEN

Interbeoordelaarsbetrouwbaarheid

Voor de berekening van de interbeoordelaarsbetrouwbaarheid is gekozen voor een gewogen Cohen's kappa. De gewogen kappa is een overeenstemmingsmaat voor ordinale en intervalschalen, waarbij gecorrigeerd wordt voor overeenstemming ten gevolge van toeval. Als ondergrens voor de kappa werd een waarde van .40 vastgesteld, welke geïnterpreteerd moet worden als 'middelmatig' (Pritchard, 1981).

Tabel 1. Gewogen cohen's kappa per categorie over de 4 beoordelaars

	VM I N=16	VM II N=16	NM st. N=16	M hulpv. N=12
1. Aandachtgevend gedrag	.49	.45	.44	.59
2. Vragen	.49	.21	.35	.31
3. Parafraseren	.42	.46	.43	.61
4. Reflecteren	.60	.52	.67	.56
5. Samenvatten	.45	.32	.51	.65
6. Concretiseren	.51	.33	.53	.67
7. Reguleren	.41	.55	.53	.50
8. Gesprekstempo	.42	.25	.61	.59
9. Houding	.40	.51	.44	.65
10. Probleemverhelderen	.59	.31	.45	.66

De meetbetrouwbaarheid per categorie was in het algemeen middelmatig, bij de nameting van studenten en met name bij de meting van hulpverleners oversteeg de betrouwbaarheid vaker de waarde .60 (hoog/substantieel). De interbeoordelaarsovereenstemming over het totaal van de scores op de 10 categorieën en over het globale eindoordeel werd berekend met de Pearson product moment correlatie-coëfficiënt. Op totaalniveau was de betrouwbaarheid bevredigend; de gemiddelde correlatie tussen alle beoordelaars was bij de eerste voormeting .71, bij de tweede voormeting .58, bij de nameting van studenten .75 en bij de meting van de hulpverleners .74. De interbeoordelaarsbetrouwbaarheid met betrekking tot het globale eindoordeel was eveneens bevredigend; bij de eerste voormeting .80, bij de tweede voormeting .65, bij de nameting van studenten .75 en bij de meting van hulpverleners .78. Het totaal van de scores op de 10 categorieën en het globale eindoordeel vertoonden een dermate sterke samenhang (Pearson correlaties variërend van .94 tot .98) dat besloten werd deze variabelen bij elkaar op te tellen. De score op de gedragsmeting wordt in het hierna volgende gevormd door het gemiddelde van de vier beoordelaars op de 10 categorieën plus het globale eindoordeel.

De interne consistentie van de schaal voor simulatiecliëntsatisfactie was zeer hoog (Cronbach's $\alpha = .95$).

Effecten van de training

De effecten van de training worden gepresenteerd in Tabel 2 en in Figuur 3, 4, en 5. Toetsing vond plaats met behulp van multivariate en univariate variantie-analyses over de verschillen van de tweede voormeting minus de eerste voormeting en de nameting minus de tweede voormeting.

Er werd geen significant 'overall' verschil gevonden tussen de eerste en tweede voormeting ($F_{Hot.(3,13)}=2,49$, $p=.106$). Daarom mag worden verondersteld dat

verschillen tussen de tweede voormeting en de nameting kunnen worden toegeschreven aan het effect van de training en niet aan het effect van de gedragsmeting, de tijd, of het theoretisch onderwijs, dat immers in de wachtperiode niet wezenlijk verschilde van dat in de trainingsperiode (zie onderzoeksopzet).

Tussen de tweede voormeting en de nameting werd een significant 'overall' verschil gevonden ($F_{\text{Hot.}(3,13)}=24.49$, $p=.000$) ten gunste van de nameting. De univariate F-toetsen laten zien dat dit verschil vooral wordt veroorzaakt door een significant verschil op de gedragsmeting ($F(1,13)=27.29$, $p=.000$) en in iets mindere mate door een significant verschil op de self-efficacy meting ($F(1,13)=10.81$, $p=.005$). Geen verschil werd gevonden op de simulatiecliëntsatisfactiemeting ($F(1,13)=.46$, $p=.51$). De verbetering van studenten op de gedragsmeting bedroeg bijna 2 standaarddeviaties.

Studenten gaven aan zich bij alle drie de metingen ongeveer even gespannen te voelen, geen significante verschillen werden gevonden tussen de eerste en tweede voormeting ($F(1,15)=.24$, $p=.63$) en tussen de tweede voormeting en nameting ($F(1,15)=.81$, $p=.38$).

Tabel 2. Gemiddelden en standaarddeviaties van de afhankelijke variabelen van studenten op 3 meetmomenten en hulpverleners.

			M	s.d.
gedragsmeting	stud.	VM I	27.8	6.7
		VM II	30.5	6.0
		NM	41.2	7.0
	hulpv.		39.2	10.3
simulatiecliëntsatisfactie	stud.	VM I	29.4	9.4
		VM II	27.9	7.3
		NM	29.3	5.2
	hulpv.		32.8	6.8
self-efficacy	stud.	VM I	24.7	3.0
		VM II	24.8	4.2
		NM	27.2	2.4
	hulpv.		31.1	1.5
spanning	stud.	VM I	2.9	1.0
		VM II	2.7	0.7
		NM	2.6	0.6
	hulpv.		2.5	0.8

Verbanden van vaardigheden met overige variabelen

In Tabel 3 worden de correlaties van de gedragsmeting met de simulatiecliëntsatisfactie en de cognitieve meting gepresenteerd, uitgesplitst naar meetmoment en soort hulpverlener. Er werd een substantiële samenhang gevonden tussen de gespreksvaardigheden, zoals beoordeeld door de beoordelaars, en de satisfactie van de simulatiecliënten met het gesprek. Self-efficacy vertoonde bij studenten weinig samenhang met de gespreksvaardigheden, bij de hulpverleners werd er zelfs een negatieve samenhang tussen self-efficacy en vaardigheden gevonden.

Tabel 3. Correlaties van de gedragsmeting met simulatiepatientsatisfactie en self-efficacy, uitgesplitst naar de metingen van studenten en hulpverleners

			simulatiecliënt-satisfactie	self-efficacy
gedragsmeting	stud.	VMI	.48*	.19
		VMII	.54*	.19
		NM	.81***	-.07
	hulpv.		.54*	-.54*

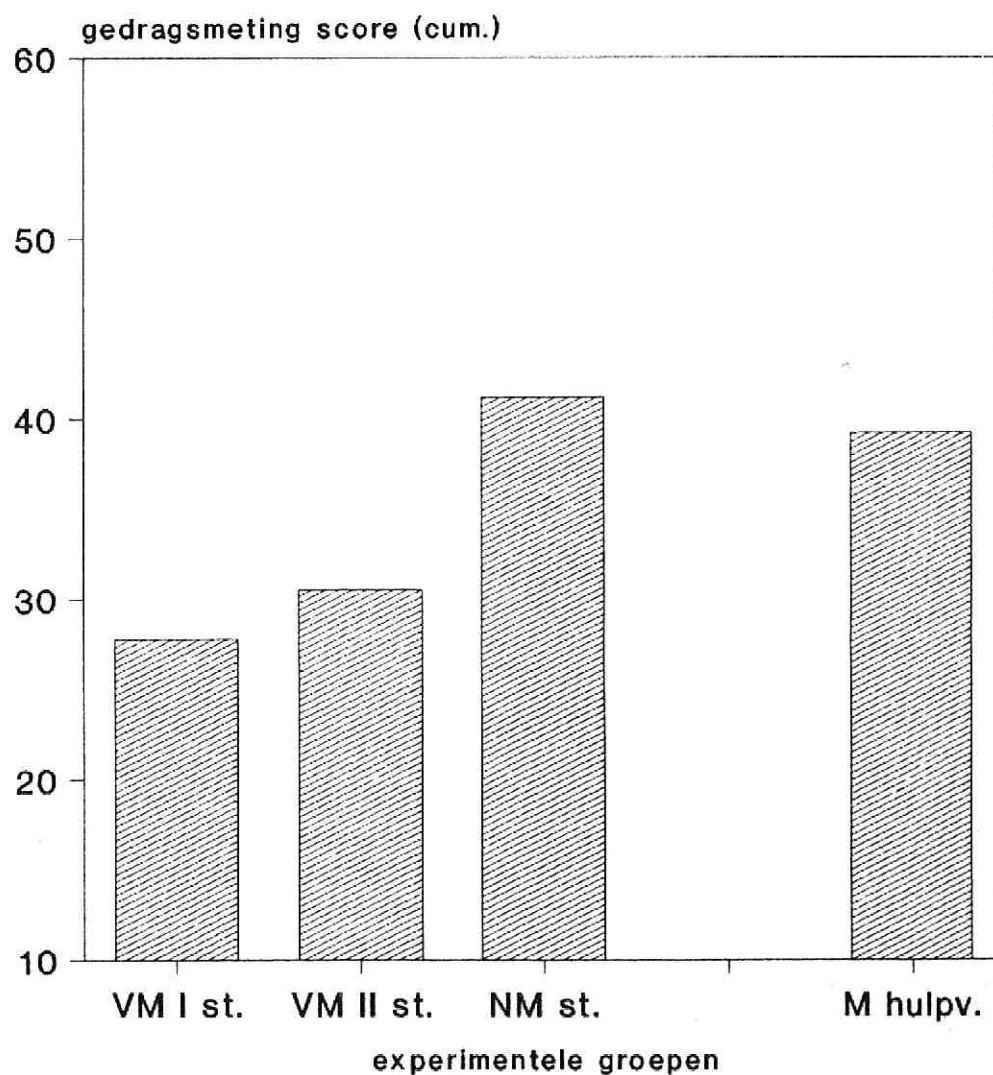
* = $p < 0.05$ *** = $p < 0.001$ 

Fig. 2. Gedragsmeting van studenten (3 meetmomenten) en hulpverleners

Vergelijking van getrainde studenten met een referentiegroep

De studenten werden bij de nameting vergeleken met een groep hulpverleners (zie Tabel 2 en Figuren 2, 3, en 4). Multivariate variantie-analyse leverde een significant 'overall' verschil op ($F_{Hot. (3,24)}=8.40, p=.001$) ten gunste van de hulpverleners. De univariate variantie-analyses laten zien dat dit verschil wordt veroorzaakt door de self-efficacy meting ($F(1,26)=25.21, p=.000$) en niet door de gedragsmeting ($F(1,26)=.36, p=.55$) en de simulatiecliëntsatisfactie ($F(1,26)=2.4, p=.13$). Hulpverleners onderscheidden zich dus van studenten in de verwachtingen die zij hadden van de effectiviteit van hun gedrag; zij hadden hogere verwachtingen van hun prestaties dan studenten.

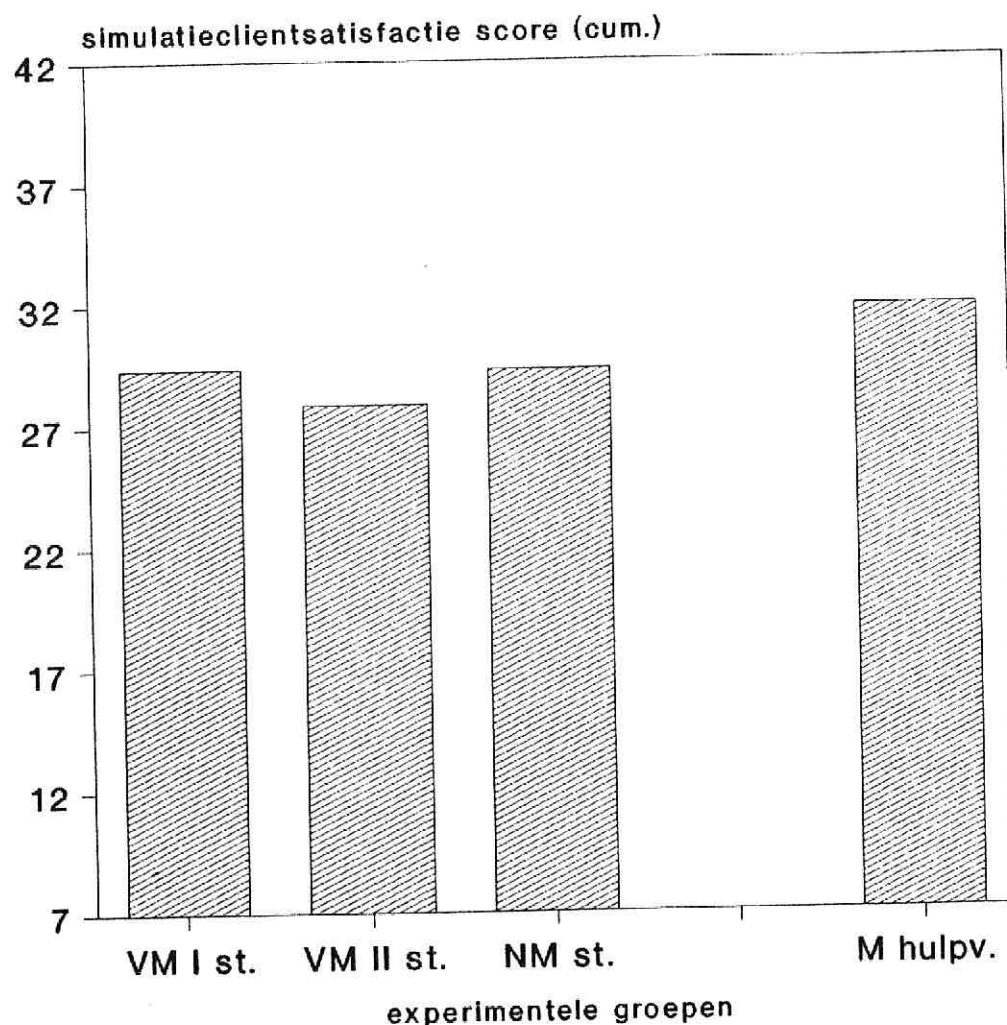


Fig. 3. Simulatiecliëntsatisfactie van studenten (3 meetmomenten) en hulpverleners

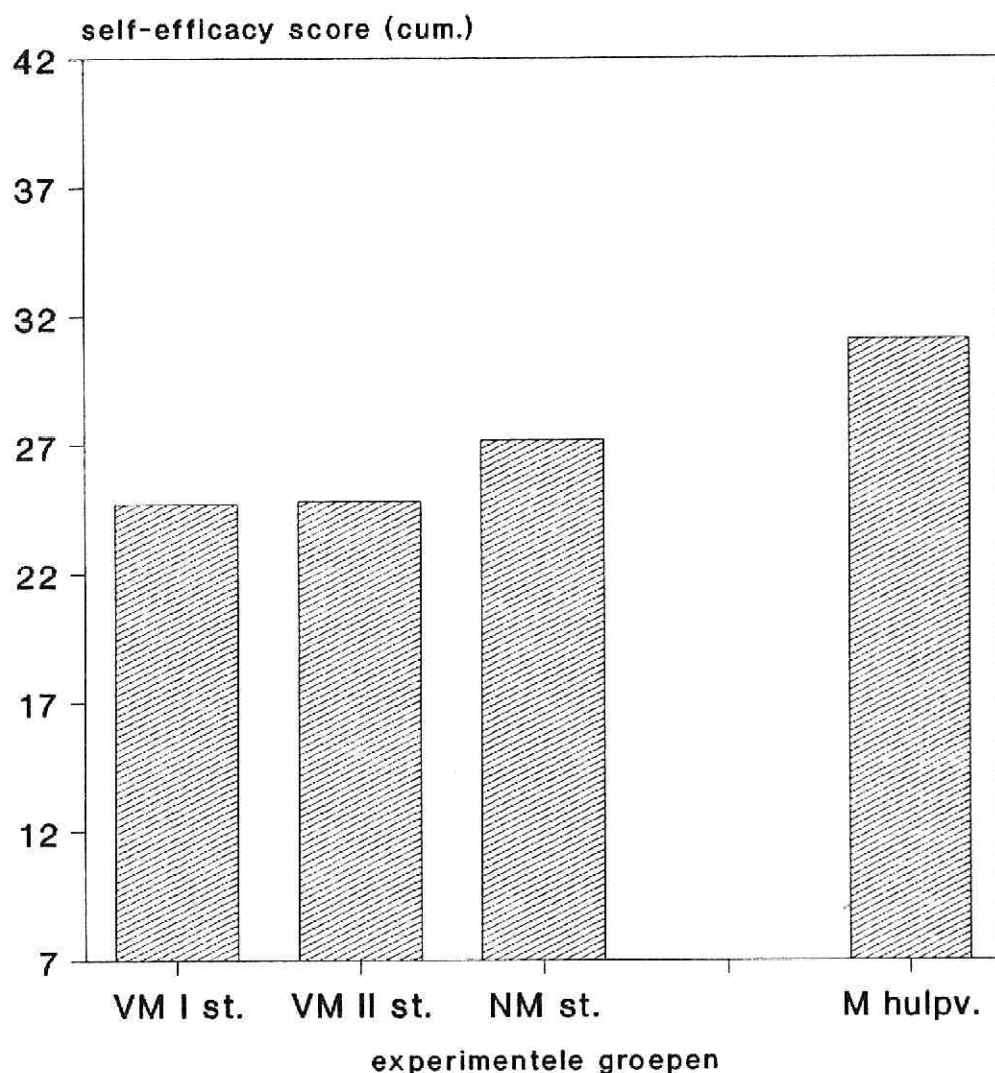


Fig. 4. Self-efficacy van studenten (3 meetmomenten) en hulpverleners

DISCUSSIE

Deze studie beoogde ten eerste antwoord te geven op de vraag wat de effecten van een kortdurende vaardigheidstraining zijn. De training bleek, althans op korte termijn, effectief in het bewerkstelligen van gedragsverandering in de gewenste richting. Er werd tevens een cognitieve verandering gevonden: na training hadden studenten hogere verwachtingen van hun interpersoonlijke effectiviteit dan voor de training. De satisfactie van de simulatiecliënten veranderde niet door training. Een mogelijke verklaring voor deze bevinding wordt gevonden in een methodologische tekortkoming van deze studie. Bij de nameting van studenten traden andere simulatiecliënten op dan bij de beide voormetingen, waardoor de simulatiecliënten die optraden bij de nameting een andere baseline hadden dan de simulatiecliënten die optraden bij de voormetingen (namelijk vaardiger studenten). Het is mogelijk dat de simulatiecliënten van de nameting daardoor strenger waren in hun oordeel. Verder onderzoek naar het effect van training op cliëntvariabelen is dan ook gewenst.

De studenten die participeerden in het onderzoek deden dit op basis van vrijwilligheid. Wat betreft de studenten die deelname weigerden bestond de indruk dat een deel erg opzag tegen een gesprek met een simulatiecliënt. De studenten die aan de studie deelnamen waren derhalve mogelijk niet representatief; zij zouden minder angstig, minder vermijdend en wellicht ook vaardiger kunnen zijn dan de studenten die niet deelnamen aan het onderzoek. Van een training volgens het microteaching-model waarbij beoogd wordt vaardigheden te verbeteren en waarbij door de gedragsmatige aanpak angst en vermijding weinig kans krijgen zouden ook de meer angstige, vermijdende en onvaardige studenten veel kunnen profiteren. Verwacht kan daarom worden dat wanneer alle studenten aan het onderzoek hadden deelgenomen, vergelijkbare trainingseffecten zouden zijn gevonden.

De tweede vraag van deze studie betrof de validiteit van de getrainde vaardigheden. Er werd een substantiële samenhang gevonden tussen het vaardigheidsniveau, zoals beoordeeld door de beoordelaars, en de tevredenheid van de simulatiecliënt met het contact. De gevonden samenhang tussen het vaardigheidsniveau en simulatiecliëntsatisfactie ondersteunt de validiteit van de getrainde vaardigheden; het goed en functioneel toepassen van deze vaardigheden ging samen met een grotere tevredenheid van de simulatiecliënt. Dit gold zowel bij student-hulpverleners als bij professionele hulpverleners.

Er werd bij studenten geen verband gevonden tussen het vaardigheidsniveau en verwachting van de interpersoonlijke effectiviteit. Bij de hulpverleners werd zelfs een negatief verband gevonden tussen vaardigheidsniveau en self-efficacy. Hoewel het kleine aantal proefpersonen in de groep hulpverleners voorzichtigheid gebiedt ten aanzien van conclusies die uit de gevonden negatieve samenhang kunnen worden getrokken, kan het gebrek aan samenhang, dat ook bij de studenten steeds werd gevonden, als volgt worden verklaard. Mogelijk is het zo dat een bepaalde mate van onzekerheid (dat wil zeggen onderschatting van eigen effectiviteit) samenhangt met het vermogen tot empathisch begrip, dat een grote rol speelt bij het adequaat voeren van een probleemverhelderend gesprek. Of anders gezegd, mogelijk dat een te hoge self-efficacy resulteert in een onaantastbare houding van de hulpverlener die interfereert met het vermogen met de cliënt mee te voelen. Voor de door Bandura (1977) geclaimde relatie tussen geloof in eigen effectiviteit en effectief gedrag werd geen steun gevonden. Self-efficacy lijkt derhalve geen relevante cognitieve variabele te zijn bij het effectief gebruik van probleemverhelderende gespreksvaardigheden. In toekomstig onderzoek naar effecten van training in gespreksvaardigheden zouden andere cognitieve variabelen betrokken kunnen worden, zoals self-statements (Merluzzi en Glass, 1981). In verschillende onderzoeken werd een relatie gevonden tussen datgene wat men tegen zichzelf zegt (self-statements) en hoe men zich gedraagt.

De derde vraag van deze studie betrof de validatie van de resultaten van de training: Hoe verhouden de prestaties van getrainde studenten zich tot die van hulpverleners? Getrainde studenten bleken niet te onderscheiden van hulpverleners wat betreft hun vaardigheden en (niet significant) wat betreft de tevredenheid van de cliënt. Hulpverleners geloofden alleen meer in hun eigen effectiviteit. Deze resultaten zijn op zijn minst opmerkelijk te noemen. Op grond van de voorsprong in opleiding en ervaring die de hulpverleners hadden ten opzichte van de studenten zou men verwachten dat de hulpverleners beter zouden presteren dan studenten.

Bij deze bevinding kunnen echter verschillende kanttekeningen worden geplaatst. Ten eerste werden in dit onderzoek alleen basale gespreksvaardigheden gemeten. Bij observatie van de gesprekken viel het bijvoorbeeld op dat verschillende hulpverleners hun cliënten tijdens het gesprek confronteerden met tegenstrijdigheden in hun verhaal, terwijl getrainde studenten dit weinig deden. Verwacht kan worden dat hulpverleners beter zullen presteren dan tweedejaarsstudenten wat betreft gevorderde gespreksvaardigheden, waarbij een

therapeutisch doel wordt nagestreefd. Ten tweede werden in dit onderzoek alleen kortetermijn effecten gemeten. Het is onduidelijk in hoeverre de gevonden effecten beklijven. Mogelijk zullen studenten op de lange termijn 'terugvallen' ten opzichte van hulpverleners, die hun vaardigheden immers dagelijks in de praktijk bijhouden. In het huidige curriculum Geestelijke Gezondheidskunde is het echter onmogelijk lange-termijn effecten van één training zuiver te meten, omdat studenten gedurende het gehele curriculum vaardigheidstrainingen volgen, waarbij de vaardigheden uit eerdere trainingen in latere trainingen opnieuw aan de orde komen. Ten derde werd de gedragsmeting ontwikkeld op basis van einddoelstellingen van de training. Studenten wisten derhalve mogelijk beter wat er van hen verwacht werd dan hulpverleners.

Gezien de recente bezuinigingsrondes bij de universiteiten, die vooral de kwaliteit van het onderwijs bedreigen, is het van belang dat de waarde van vaardigheidsonderwijs met empirische gegevens wordt ondersteund en dat deze vorm van onderwijs een vastere plaats in het curriculum van universiteiten in gaat nemen.

Toekomstig onderzoek naar vaardigheidstrainingen zou zich onder meer moeten richten op twee validiteitsvragen. De eerst vraag betreft de validatie van de geselecteerde vaardigheden: zijn deze vaardigheden relevant voor het doel dat met de training beoogd wordt? Naast metingen van de ervaring van de cliënt tijdens het contact met de hulpverlener is het van belang cliënt-outcome metingen te gebruiken, dat wil zeggen metingen van de vooruitgang van de cliënt op variabelen die in het hulpverleningscontact worden beoogd. De tweede vraag betreft de validatie van de effecten van vaardigheidstrainingen. Door trainingseffecten te kwantificeren en door het door training bereikte vaardigheidsniveau te vergelijken met het niveau van de latere beroepsgroep, wordt inzicht verkregen in de betekenis van het trainingseffect. Om tenslotte inzicht te krijgen in de optimale hoeveelheid trainingen en trainingsuren in een universitair curriculum zou het niveau van alle vaardigheden (in termen van de einddoelstellingen van de studie) van studenten in verschillende fasen van hun studie vergeleken moeten worden met dat van de latere beroepsgroep.

DANKWOORD

De volgende mensen worden bedankt voor hun bijdrage aan het onderzoek: de studenten, simulatiecliënten en hulpverleners die als proefpersoon hebben deelgenomen, Els Graafsma, Rense Hoekstra, Jeanette Hommes en Rutger Lulofs voor het beoordelen van de videobanden, Gerrie van Wunnik voor de technische assistentie, en Henk Schmidt en Rudolf Ponds voor hun commentaar op een eerdere versie van het manuscript.

LITERATUUR

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Carrol, J.G. & Monroe, J. (1980). Teaching clinical interviewing in the health professions. *Evaluation of the Health Professions*, 3, 21-45.
- Ford, J.D. (1979). Research on training counselors and clinicians. *Review of Educational Research*, 49, 87-130.
- Glass, G., McGaw, B. and Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills CA: Sage.
- Ivey, A. (1971). *Microcounseling*. Springfield, Il: Charles C. Thomas.
- Kraan, H.F. and Crijnen, A.A.M. (1987). *The Maastricht History-taking and Advice Checklist: Studies of instrumental utility*. Proefschrift, Amsterdam: Velden van den Hezelaar.

- Kurtz, P.D., Marshall, E.K. and Banspach, S.W. (1985). Interpersonal skill training research; a 12 year review and analysis. *Counselor Education and Supervision*, 249-263.
- Lang, G. en van der Molen, H. (1984). *Psychologische gespreksvoering*. Baarn: Nelissen B.V.
- Marshall, E.K., Charping, I.W. and Bell, W.J (1979). Interpersonal Skills training: A review of the research. *Social Work Research & Abstracts*, 15, 10-16.
- Merluzzi, T.V. and Glass, C.R. (1981). *Cognitive assessment*. New York: Guilford Press.
- Pritchard, M.J. (1981). Temporal reliability of a questionnaire measuring psychological responses to illness. *Journal of Psychosomatic Research*, 25, 63-66.
- Riccardo, V.M. and Kurtz, S.M. (1983). Methods for teaching and assessing communication in health care. *Communication in health care (appendix B)*. Springfield, Ill.: Charles C. Thomas.
- Stone, G.L. (1982). Evaluating the effectiveness of skills training programs. In Marshall, E.K. and Kurtz, P.D. and associates. *Interpersonal Helping skills*. San Francisco: Jossey-Bass publishers, 447-467.

Manuscript ontvangen 1-6-1989

Definitieve versie ontvangen 7-6-1990

Gebruik van verzorgingsactiviteiten in het basisonderwijs *

D. Koster, Th. van Batenburg en H. H. Tillema
RION, Instituut voor Onderwijsonderzoek

ABSTRACT

In evaluating the education support system in (primary) education, one has to deal with a range of problems. The most difficult ones are establishing criteria, agreement on standards and the definition of inputs. At this moment evaluation of the support system is premature: instead explorative impact assessment is proposed in order to find input factors of external support programs which relate to the utilization of support programs. In this article, results of this type of explorative studies are provided.

EVALUEREN VAN ONDERWIJSVERZORGING

In dit artikel wordt een onderzoek besproken dat door het RION in 1987 en 1988 is uitgevoerd. Het onderzoek valt binnen de context van evaluatie van onderwijsverzorging. Het standpunt wordt verdedigd dat, voordat van evaluatie sprake is, eerst nagegaan dient te worden of er sprake is van doorwerking van het aanbod van onderwijsverzorgingsinstellingen op leerkrachten.

Evaluatie-onderzoek kan gegevens verschaffen over de effectiviteit van de onderwijsverzorging, maar is om een aantal redenen problematisch:

- een specificatie van *doelen* van de onderwijsverzorging is niet voorhanden, het is niet op voorhand duidelijk welke concrete eisen aan de dienstverlening van de verzorgingsinstellingen gesteld mogen worden (doelprobleem);
- er is geen overeenstemming over de effectcriteria en effectstandaarden (criteriumprobleem);
- de aard van de dienstverlening is dermate divers dat het niet eenvoudig is het *object van de evaluatie* te benoemen (objectprobleem).

Ondanks deze problemen wordt het belang van evaluatie van onderwijsverzorging erkend. Herhaaldelijk is geprobeerd aan te geven langs welke lijnen evaluatie-onderzoek zich dient te bewegen.

In de Heroverwegingsnota (1982) werd nagegaan of effecten van onderwijsverzorging af te meten zijn aan de kwaliteit van het onderwijs. Geconcludeerd werd dat:

- de kwaliteit van onderwijs niet eenduidig en kwantificeerbaar valt vast te stellen;
- de onderwijsverzorging slechts één van de kwaliteitsbevorderende instrumenten is;
- de onderwijsverzorging niet in een rechtlijnige relatie tot de kwaliteit van het onderwijs staat.

Uit de laatste conclusie blijkt dat evaluatie van onderwijsverzorging niet zonder meer

* Onder Projectnr. 6004 gesubsidieerd door SVO - Instituut voor Onderzoek van het Onderwijs te Den Haag

gekoppeld kan worden aan de kwaliteit van het onderwijs (bijvoorbeeld in termen van leerprestaties). Evaluatie van onderwijsverzorging dient zich te richten op de effecten van de dienstverlening en niet op de effecten van onderwijs. Nog steeds (zie Hoeben, 1987) is over dat laatste standpunt een discussie gaande onder evaluatie-onderzoekers, waarbij de 'harde' lijn ('kwaliteit moet blijken', zie bijvoorbeeld de Groot, 1986) tegenover de 'zachtere' lijn ('onderwijsverzorger zijn geen leraren') staat. Onderwijsverzorger stellen doorgaans dat ze niet verantwoordelijk zijn voor bijvoorbeeld leerlingprestaties (zie bijvoorbeeld Dingemans & Paans, 1986).

Evaluatie van onderwijsverzorging komt ook aan de orde in de programmeringsstudies van SVO (Stokking, 1984; Hoeben & Stokking, 1986). Met name de onderwijsverzorgingsinstellingen zelf blijken behoefte te hebben aan een overall-studie naar de activiteiten, die de onderwijsverzorging aanbiedt en naar de effecten van de verzorging.

Vanuit het beleid is inmiddels de programmering van evaluatieonderzoek, met betrekking tot onderwijsverzorging, gestart. Begin 1987 werd door de Minister een commissie 'Programmavoorstel evaluatie onderwijsverzorging' (Carpay, Brenninkmeijer & Hofstee, 1987) ingesteld. De effectiviteit van onderwijsverzorging wordt opgevat als een functie van de kwaliteit van het aanbod en de acceptatie/het gebruik ervan door scholen.

Aansluitend is begin 1989 een taakgroep evaluatie onderwijs (TEVON) ingesteld die uitvoering gaat geven aan het evaluatieplan (zie Wiegersma, Giesbers & Nauta, 1989). Gelet op het verbrokkeld karakter van het verrichte onderzoek naar onderwijsverzorging en de nog geringe beschikbare kennis, is de vraag gerechtvaardigd of het in dit stadium mogelijk is evaluatie-onderzoek naar onderwijsverzorging te verrichten. De voorstellen van TEVON gaan in feite niet veel verder dan inventarisatie en beschrijving van het verzorgingswerk. Ook richt men zich op evaluatie van taakstellingen. Naar onze mening is dat in de praktijk een verkeerd uitgangspunt. Van evaluatie-onderzoek is volgens Hofstee (1986) pas sprake wanneer evaluatiecriteria van maatregelen of activiteiten die worden ondernomen om de gestelde doelen te bereiken, vastliggen. In de Wet op de Onderwijsverzorging (WOV) worden de volgende weinig concrete doelen genoemd: versterking van het probleemoplossend vermogen en handhaving/verbetering van de kwaliteit van het onderwijs.

Over de invulling van criteria, de interventies en de doelen bestaat (nog?) geen consensus; over het opzetten en uitvoeren van evaluatie-onderzoek evenmin. Een groot struikelblok is de operationalisering van de relatie tussen doelen, genoemd in de WOV en de middelen (verzorgingsactiviteiten). In de doelen van de WOV wordt namelijk niet aangegeven hoe het probleemoplossend vermogen van scholen en de kwaliteit van het onderwijs gemeten moeten worden.

Een (voorlopige) oplossing is af te zien van effectmeting op het niveau van de leerling en te zoeken naar indicatoren bij het te evalueren object (de onderwijsverzorging zelf). Tegen deze achtergrond wordt in dit artikel een studie gepresenteerd, waarin is onderzocht welke factoren het gebruik van onderwijsverzorgingsaanbod op basisscholen beïnvloeden en waarin is nagegaan welke factoren door dit gebruik beïnvloed worden.

THEORETISCH MODEL EN BESCHRIJVING VAN VARIABLEN

Om de probleemstelling te beantwoorden is een model opgesteld dat getoetst is met LISREL. De gegevens werden via vragenlijsten van leerkrachten op basisscholen verkregen. Dit model wordt in Figuur 1 gepresenteerd. Er wordt invloed van de gebruikerscontext op het gebruik verwacht, evenals van de kwaliteit, bruikbaarheid en verspreiding. Daarnaast wordt een indirect effect van kwaliteit en bruikbaarheid via

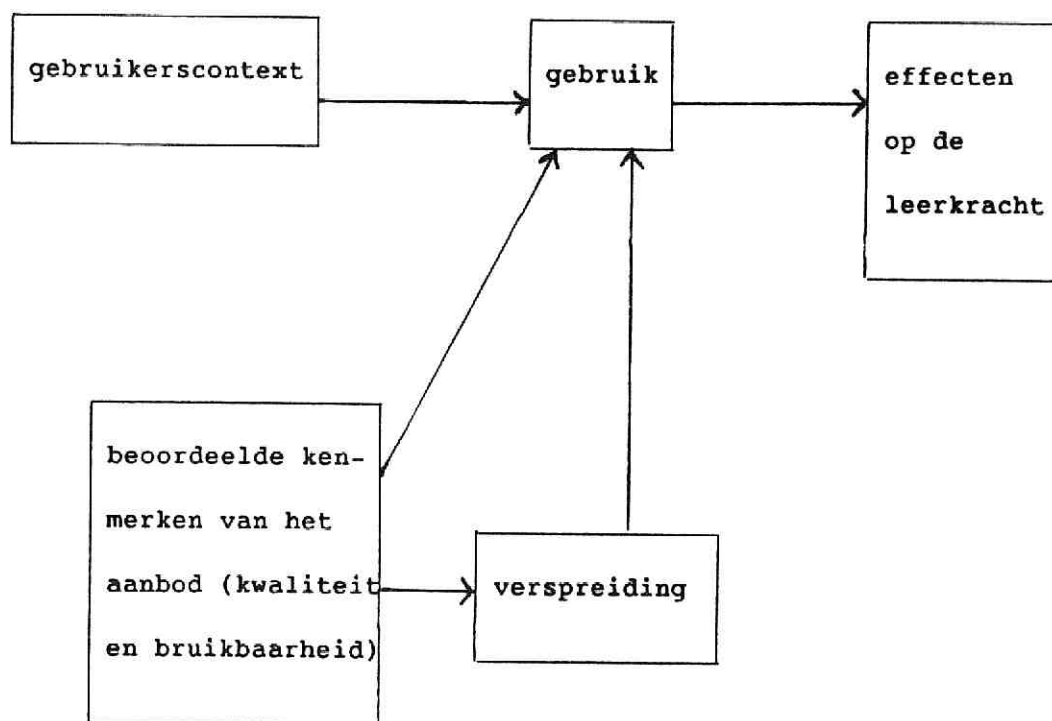


Fig. 1.

verspreiding verwacht. Tenslotte hopen we de invloed van het gebruik op de leerkracht vast te stellen. De belangrijkste variabelen uit dit model worden eerst besproken, waarna het model uitgebreid besproken wordt. Daarna wordt op de representativiteit van het aanbod ingegaan.

Voor een meer uitgebreide beschrijving van de variabelen en voor gegevens over betrouwbaarheid van scholen e.d. verwijzen we naar Tillema, Koster en Van Batenburg (1989) en naar Koster, Tillema en Van Batenburg (1989).

Modelvariabelen

Het onderzoek is gericht op de exploratie van relaties tussen aanbod, gebruik en effecten van verzorgingsactiviteiten in het basisonderwijs. In dit onderzoek is onderscheid gemaakt tussen dimensies van gebruik (hoe vaak wordt iets gebruikt, hoeveel onderdelen e.d.), conceptueel gebruik (de invloed van het gebruik op ideeën en opvattingen van de gebruiker) en niveau van gebruik (aanvullend, getrouw en adaptief gebruik). We bespreken elk kort:

1. Een met het programma HOMALS berekende gewogen objectscore op grond van vier dimensies: omvang, frequentie, intensiteit en duur van gebruik. De vraag was of er sprake zou zijn van een achterliggende latente variabele (van weinig frequent, geringe omvang, duur, intensiteit naar hoog frequent etc.). Er was inderdaad sprake van één achterliggende latente variabele.
2. Een somscore als gevolg van 10 items met betrekking tot 'conceptueel gebruik'. Hier was sprake van een goede schaal.
3. Een score voor het niveau van gebruik op grond van scoringsregels (zie Tillema, Koster & van Batenburg, 1989).

Gebruik werd in verband gebracht met:

- *Gebruikerscontext*. Binnen de gebruikscontext zijn de volgende variabelen onderscheiden: 'kindgerichtheid', 'prestatiegerichtheid', 'organisatiebeleving', 'besluitvorming', 'persoonlijke doelmatigheidsbeleving' en 'resultaatverwachting'. Beide laatstgenoemde variabelen verwijzen naar het construct 'efficacy' (zie o.a. Imants & Vriens, 1987). Alle zes variabelen vormden een goede schaal (zie Tillema, Koster & van Batenburg, 1989). De variabele 'behoefte aan het produkt' is hierbij eveneens onderscheiden. Dit is de enige variabele in deze rij die specifiek betrekking heeft op de betreffende activiteit.
 - *Beoordeelde kenmerken van het aanbod*. De volgende variabelen zijn onderscheiden: *kwaliteit*. Om de kwaliteitskenmerken van produkten en diensten te bepalen, zijn 15 items geformuleerd die met meer objectieve kenmerken te maken hebben (in het vervolg '*objectieve kwaliteit*' genoemd) en vier items die met meer subjectieve kenmerken te maken hebben (in het vervolg '*subjectieve kwaliteit*' genoemd). Voorbeelden van de eerste zijn 'Ik kan er goed zelfstandig mee werken' en 'De activiteit biedt genoeg aanwijzingen om het op school te gebruiken'; Voorbeelden van de tweede zijn 'Ik vind de activiteit praktisch' en 'Ik vind de activiteit goed gestructureerd'.
 - *Bruikbaarheid*. Voor negen aspecten (ten aanzien van vooral het didactisch handelen) is gevraagd aan te geven in hoeverre de betreffende activiteit leerkrachten in het algemeen zou kunnen helpen. Een voorbeeld van zo'n aspect is 'het ontwikkelen en verbeteren van didactische vaardigheden.'
- Ook deze variabelen vormden een goede schaal.
- *Verspreiding*. Deze term verwijst naar de manier waarop een activiteit op school werd geïntroduceerd. Allereerst is nagegaan of er met betrekking tot de activiteit persoonlijk contact is geweest met een onderwijsverzorgingsinstelling. Als dit het geval was, werden zes vragen over de directheid van dit contact ingevuld; een voorbeeldvraag is: 'De activiteit werd afgestemd op onze schoolsituatie.' Dit werd '*directheid van verspreiding*' genoemd. Daarnaast zijn vragen gesteld over de '*waardering van de verspreiding*.' Voorbeelden van items zijn 'Ik heb informatie gekregen over de activiteit', 'Ik heb inzicht gekregen in de gebruiksmogelijkheden' en 'Er is aangegeven hoe ik er in de dagelijkse situatie mee moet omgaan.' Ook deze twee variabelen vormden een goede schaal.
 - *Effecten*. Als effecten worden drie variabelen onderscheiden: 'gedragseffecten', 'behoefte aan aanvullende ondersteuning' en 'nut':
 - Gedragseffecten*. De onderwijsverzorging is erop gericht het gedrag van leerkrachten te beïnvloeden. Er werd dan ook aan leerkrachten gevraagd inschattingen te maken over hun gedragsveranderingen als gevolg van het gebruik van de activiteit. Bij negen aspecten is gevraagd aan te geven in hoeverre er door het gebruik een bijdrage werd geleverd. Het betrof dezelfde aspecten als bij het construct 'bruikbaarheid', maar nu is gevraagd of de beoogde gedragsverandering in feite heeft plaatsgevonden.
 - Behoefte aan aanvullende ondersteuning*. Bij dezelfde negen aspecten is gevraagd of de leerkracht door/na het gebruik behoefte had aan verdere ondersteuning. Dit effect is onderscheiden van het hierboven besproken effect daar uit veel onderzoek (zie voor een overzicht Hoeben & Stokking, 1986) blijkt dat als gevolg van gebruik van onderwijsverzorging, leerkrachten vaak meer behoefte eraan krijgen.
 - Nut*. Er werden acht items voorgelegd die vooral te maken hadden met tijdsbesparing die men bij het gebruik ervaart.
- De items bij deze drie variabelen zijn eveneens gesommeerd daar betrouwbaarheidsanalyses een homogene schaal opleverden.
- Daar een belangrijke voorwaarde voor LISREL het normaal verdeeld zijn van de variabelen (Boomsma, 1983) is, is nagegaan of aan deze voorwaarde voldaan is: er bleken geen schendingen van normaliteit te zijn.

Representativiteit van het aanbod en steekproef

In het onderzoek is ernaar gestreefd een representatieve set van verzorgingsactiviteiten samen te stellen (het woord 'activiteiten' omvat produkten en diensten). Uit jaarverslagen, werkplannen, catalogi en diverse brochures van de verzorgingsinstellingen (CITO, LPC, SLO, SVO en de OBD) werden 54 activiteiten geselecteerd, waarbij in een vooronderzoek bij basisscholen is bepaald wat de mate van bekendheid, gebruik, belang en behoefte is. Op grond van deze scores werden 'hoog' en 'laag' scorende activiteiten geselecteerd. De lijst van 54 werd gereduceerd tot 30. Daarnaast hebben zes experts uit de onderwijsverzorging in twee ronden de representativiteit van de activiteiten beoordeeld (zie Tillema, Koster & van Batenburg, 1989).

Uiteindelijk zijn 23 activiteiten geselecteerd die als redelijk representatief voor het aanbod van onderwijsverzorgingsinstellingen beschouwd kunnen worden.

Per activiteit (23) werden ± 20 scholen gezocht waar de betreffende activiteit in gebruik was (zgn. 'gebruikers') en vijf scholen waar de betreffende activiteit niet gebruikt werd, doch waar men er wel bekend mee was (zgn. 'niet-gebruikers').

In totaal zijn 794 scholen telefonisch benaderd. Hiervan gaven 536 scholen aan mee te willen werken. Ze ontvingen de vragenlijst. Deze werd door 339 scholen ingevuld en teruggestuurd. Het maximum aantal vragenlijsten voor één school was twee. Op deze manier resteerden 540 bruikbare vragenlijsten (411 'gebruikers' en 129 'niet-gebruikers').

RESULTATEN

Allereerst worden de correlaties besproken, daarna komt de modeltoetsing aan bod.

Correlaties

Gebruikscontext/schoolsituatie en gebruik

In Tabel 1 worden de correlaties tussen de gebruikerscontext en gebruik gepresenteerd.

Tabel 1. Correlaties: contextvariabelen en gebruik

	1	2	3	4	5	6	7	8	9
1 Kindgerichtheid	—								
2 Prestatiegerichtheid	.24	—							
3 Organisatiebeleving	.22	-.05	—						
4 Besluitvorming	.23	-.02	.38	—					
5 Pers. doelmatigheidsbeleving	.22	.09	.30	.24	—				
6 Resultaatverwachting	.28	.07	.48	.37	.59	—			
7 Niveaus van gebruik	.03	.03	-.04	.03	.07	.04	—		
8 Dimensies van gebruik	.02	.03	.08	-.01	.07	.11	.34	—	
9 Conceptueel gebruik	.23	.12	.09	.07	.17	.20	.31	.48	—

De hoogste correlatie in de tabel is die tussen 'persoonlijke doelmatigheidsbeleving' en 'resultaatverwachting' (.59). De verwachting dat 'persoonlijke doelmatigheidsbeleving' en 'resultaatverwachting' samenhangen met het gebruik, komt niet uit. Dit geldt eveneens voor de overige contextvariabelen. De hoogste correlaties tussen contextvariabelen en gebruik treden op bij conceptueel gebruik, maar geen enkele correlatie is substantieel; de hoogste is .23.

Kwaliteit, bruikbaarheid, verspreiding en gebruik

In Tabel 2 worden de onderlinge relaties en de relaties met gebruik gepresenteerd.

Tabel 2. Correlaties: aanbod, verspreiding en gebruik

	1	2	3	4	5	6	7	8
1 Kwaliteit (obj.)	—							
2 Kwaliteit (subj.)	.69	—						
3 Bruikbaarheid	.25	.24	—					
4 Directheid verspreiding	.09	.08	.39	—				
5 Waardering verspreiding	.49	.41	.48	.43	—			
6 Niveaus van gebruik	.22	.19	.21	.16	.31	—		
7 Dimensies van gebruik	.26	.27	.37	.41	.49	.34	—	
8 Conceptueel gebruik	.31	.33	.74	.44	.56	.31	.48	—

De correlaties tussen de eerste vijf variabelen in de tabel (2x kwaliteit, bruikbaarheid en 2x verspreiding) zijn redelijk hoog. Alleen de variabele 'directheid van verspreiding' correleert niet met de kwaliteit van activiteiten. De correlaties van deze vijf variabelen met gebruik zijn zwak tot hoog positief (.71). De hoogste correlatie is die tussen conceptueel gebruik en de bruikbaarheid. De correlaties met conceptueel gebruik zijn ook hier hoger dan de correlaties met niveaus en dimensies van gebruik.

Gebruik en effecten

In Tabel 3 worden de correlaties met gebruik gepresenteerd.

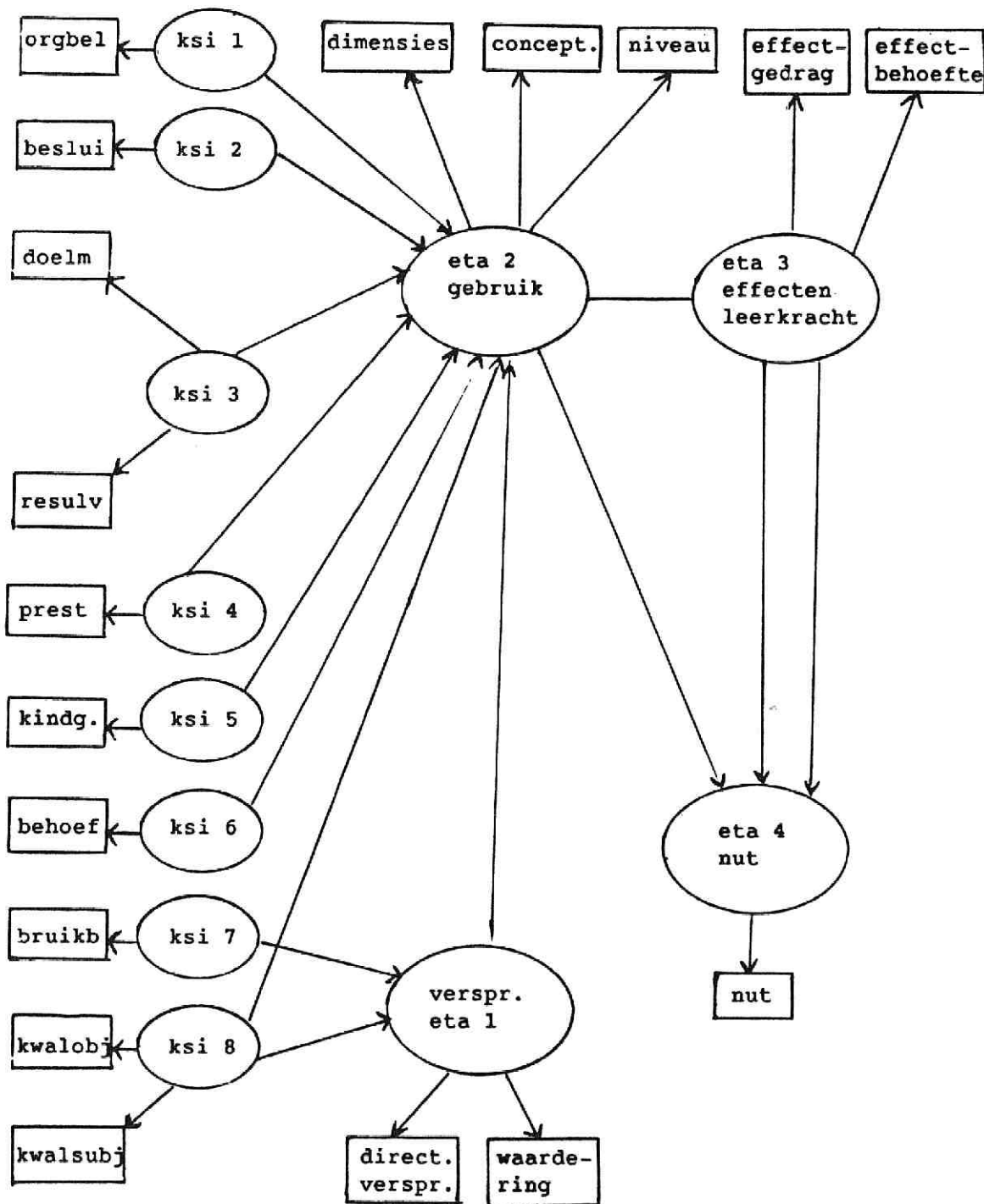
Tabel 3. Correlaties: gebruik en effecten

	1	2	3	4	5	6
1 Niveaus van gebruik	—					
2 Dimensies van gebruik	.34	—				
3 Conceptueel gebruik	.31	.48	—			
4 Effecten (gedrag)	.22	.42	.66	—		
5 Effecten (behoefte)	.14	.23	.44	.60	—	
6 Nut	.18	.42	.54	.52	.31	—

De effectmaten ('effecten(gedrag)', 'effecten(behoefte)' en 'nut') correleren onderling redelijk hoog. De correlaties met de drie gebruiksvariabelen zijn zwak tot hoog positief. De hoogste is die tussen conceptueel gebruik en effecten(gedrag).

Toetsing van het model

In Figuur 2 staat het aanvangsmodel met de theoretisch verwachte relaties afgebeeld. Dit model is getoetst met LISREL VI. Er is gebruik gemaakt van Maximum-Likelihood-schattingen, waarbij de covariantiematrix geanalyseerd is.



* De manifeste variabelen zijn met een □ aangegeven, de latente met een ○

Fig. 2. Het aanvangsmodel

Om de stabiliteit van het model te onderzoeken, is het model ontwikkeld op een halve dataset. De totale dataset is ad random in tweeën gesplitst. Deze eerste helft wordt in het vervolg met S1 aangeduid, de andere helft met S2. Het aantal analyse-eenheden in S1 was 192, dat in S2 197. Om de fit te verbeteren moesten er twee paden bijgetrokken dienden te worden: één van 'bruikbaarheid' naar eta 3 ('effecten voor de leerkracht') en één van 'kwaliteit' naar eta 4 ('nut'). Hierna kon het model niet verbeterd worden. Dit model werd 'gefit' op S2.

Op grond van verschillende criteria kan geconcludeerd worden dat de fit redelijk is:

- Slechts een paar 'normalized residuals' zijn iets groter dan 2 (een door Jöreskog & Sörbom (1985) genoemd criterium).
- De Goodness of Fit Index is .90 (bij S1: .89) en de Root Mean Square Residual is .03 (idem als bij S1).
- Er zijn geen negatieve 'coefficients of determination'. Er zijn ook geen grote standaardfouten.
- De punten in de q-plot (plot van de 'normalized residuals') vallen zo goed als samen op de 45° lijn, dat wil zeggen dat er geen afwijking van lineairiteit is.
- χ^2 is 194.1 (df=98). Op grond van dit criterium mag niet geconcludeerd worden dat het model verworpen dient te worden. Bekend is dat het aantal cases ongeveer lineair gerelateerd is aan de hoogte van χ^2 (zie bijvoorbeeld Boomsma, 1983). Aangezien de power van de χ^2 -toets bij veel vrijheidsgraden relatief groot is, wordt ook wel een beslissingsregel gehanteerd die niet op de overschrijdingskans gebaseerd is; als χ^2 kleiner is dan twee keer het aantal vrijheidsgraden, is er sprake van een passend model (Boland, 1988).

Omdat de χ^2 van 194.1 minder dan twee keer zo groot is als het aantal vrijheidsgraden (df=98), kan geconcludeerd worden dat het model volgens dit criterium niet slecht fit. Bij S1 is χ^2 iets groter dan twee keer het aantal vrijheidsgraden, namelijk 208.9.

Een vergelijking van belangrijke criteria (en parameters) leverde geen grote verschillen op tussen het model met betrekking tot S1 en S2. Omdat S1 gebruikt is om het model te ontwikkelen, is het correct S2 als basis voor het model te gebruiken. Het definitieve model met daarin de directe effecten ziet er als volgt uit (om de figuur overzichtelijk te houden, staan twee directe effecten onderin vermeld):

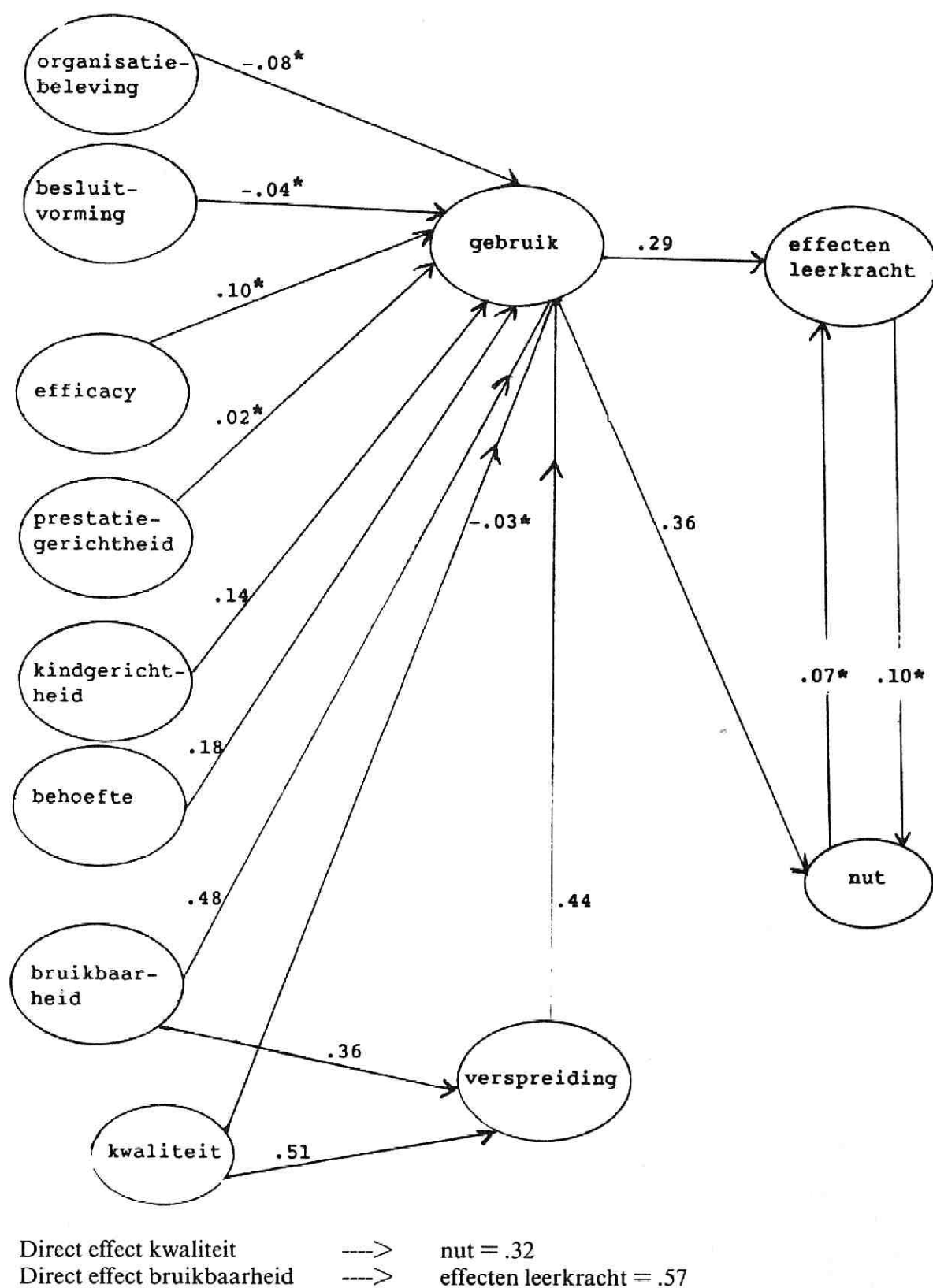


Fig. 3. Het eindmodel (* = niet significant): de diverse effecten

De invloed van schoolcontextvariabelen op gebruik is zeer gering. Andere factoren zijn belangrijker, vooral de 'bruikbaarheid' heeft een groot direct effect op 'gebruik' en op 'effecten voor de leerkracht'. De kwaliteit doet er minder toe, wel is er een indirect effect van 'kwaliteit' via 'verspreiding' op 'gebruik' (effectgrootte = $.51 \times .44 = .22$); er is ook een indirect effect van bruikbaarheid via verspreiding op gebruik, dat iets kleiner is (.16). Het gebruik heeft een redelijk sterk direct effect op 'nut' en 'effecten voor de leerkracht'.

Na deze toetsing op grond van alle 23 activiteiten, is nagegaan of er verschil is tussen bepaalde *aanbodkenmerken*. Alle activiteiten zijn ingedeeld in diverse algemene kenmerken als 'initiatief' (extern-intern), 'aard' (vakgericht-onderwijsgericht) en soort instelling (bijv. OBD-niet OBD). Vooral tussen OBD-en niet OBD-activiteiten blijken er grote verschillen te zijn. In Tabel 4 worden deze verschillen gepresenteerd.

Tabel 4. Instelling en modelvariabelen

	OBD	niet-OBD	P-waarde
1. Conceptueel gebruik	3.0	2.6	.00*
2. Niveau van gebruik	2.6	2.3	.00*
3. Dimensies van gebruik	0.3	0.3	.00*
4. Kindgerichtheid	3.6	3.6	.70
5. Prestatiegerichtheid	2.8	2.8	.33
6. Organisatiebeleving	3.9	3.9	.53
7. Besluitvorming	4.3	4.3	.66
8. Pers. doelmatigheidsbeleving	4.1	4.2	.49
9. Resultaatverwachting	3.9	4.0	.26
10. Behoeft	3.4	3.1	.00*
11. Kwaliteit subjectief	3.3	3.4	.23
12. Kwaliteit objectief	3.1	3.2	.04
13. Bruikbaarheid	2.9	2.6	.00*
14. Directheid verspreiding	0.7	0.2	.00*
15. Waardering verspreiding	3.3	2.9	.00*
16. Effecten (bijdrage)	2.7	2.4	.00*
17. Effecten (behoefte)	2.5	2.2	.00*
18. Nut	2.9	2.7	.00*

* $p < .01$

OBD-activiteiten worden bijna altijd significant hoger beoordeeld dan de activiteiten van de overige instellingen.

Gezien dit verschil tussen OBD ($n=195$) en niet-OBD-activiteiten ($n=191$), is nagegaan in hoeverre het ontwikkelde model paste op beide groepen. De fit met betrekking tot de OBD-activiteiten was redelijk ($\chi^2=159.2$; $df=98$; $GFI=.92$). Dit model zou verbeterd kunnen worden door de meetfouten van 'conceptueel gebruik' en 'dimensies van gebruik' te laten correleren. Hiervoor zagen we echter geen aanleiding. Er zijn geen 'fitted residuals' $> .10$. Slechts drie 'normalized residuals' zijn iets groter dan 2 (deze horen alle drie bij 'behoefte aan verdere ondersteuning'). De punten in de q-plot van deze 'normalized residuals' vallen zo goed als samen op de 45° lijn. Dit alles bij elkaar genomen duidt op een 'moderate fit'. Van de variantie in gebruik is 85% verklaard.

Vervolgens is bij niet-OBd-activiteiten nagegaan in hoeverre het ontwikkelde model paste. De variabele 'directheid van verspreiding' is weggelaten vanwege een scheve verdeling (dit hangt samen met het type verzorging). Het model blijkt niet goed te passen ($n=191$; $\chi^2=189,1$; $df=83$; $GFI=.90$); zes fitted residuals zijn groter dan .10, vijf normalized residuals zijn groter dan 2 (één is bijna 4). Zinnvolle verbeteringen zijn nauwelijks mogelijk.

Een mogelijke verklaring voor deze resultaten is de grotere diversiteit aan activiteiten binnen de niet-OBd-activiteiten (tijdschriften, schoolleiderstraining, toetsen, en dergelijke). Binnen deze elf niet-OBd-activiteiten is echter een groep te onderscheiden, waarbij de activiteiten wel min of meer gelijksoortig van aard zijn. Dit zijn de volgende zeven activiteiten met een sterk produktkarakter: de drie CITO-activiteiten, de drie SLO-activiteiten en de activiteit 'individualisering en differentiatie van leesonderwijs'. Het aantal analyse-eenheden is 104. Het model bleek voor deze activiteiten te passen ($\chi^2=154,1$; $df=83$; $GFI=.87$).

CONCLUSIES EN DISCUSSIE

Een probleem in onderzoek naar verzorgingsaanbod is de mate van representativiteit van het geselecteerde aanbod. Ook ontbreekt een compleet overzicht van het aanbod.

Na een vooronderzoek bij leerkrachten en experts is een set van 23 verzorgingsactiviteiten opgenomen. Deze set kan op grond van uitspraken door experts en leerkrachten als redelijk representatief voor het totale aanbod beschouwd worden.

De modeltoetsing is eerst uitgevoerd voor de set activiteiten en alle instellingen gezamenlijk. De schoolinterne factoren bleken praktisch geen invloed te hebben op het gebruik van verzorgingsactiviteiten. Wel is er een zwak verband tussen deze perceptievariabelen bij de leerkracht (bijv. 'persoonlijke doelmatigheidsbeleving' en 'kindgerichtheid') en het conceptueel gebruik. De relatie tussen behoefte aan en gebruik van activiteiten is tamelijk gering.

De verzorgingsfactoren (aanbod en verspreiding) hebben grote invloed op gebruik. De belangrijkste variabele hierbij is de bruikbaarheid van het aanbod. Deze variabele is in het onderzoek uitgewerkt als een inschatting van effecten die de leerkracht verwacht van een activiteit op (verschillende aspecten van) het leerkrachthandelen. Leerkrachten gebruiken de verzorgingsactiviteiten meer, naarmate ze inzien wat de gebruikswaarde en mogelijkheden zijn van een activiteit op concrete aspecten van hun handelen. Deze invloed is er zowel bij OBd als bij niet-OBd-activiteiten. Bruikbaarheid blijkt eveneens sterk samen te hangen met verspreiding. De verspreiding heeft een positieve invloed op gebruik. Naarmate er meer direct contact tussen de aanbieder en de afnemer is, neemt het gebruik bij de laatste toe. De invloed van kwaliteit op gebruik verloopt via verspreiding; er is geen direct effect op gebruik. Naarmate leerkrachten de kwaliteit beter beoordelen, vinden ze dat het 'nut' voor de praktijk groter is. Naast gebruik zijn andere effectmaten onderscheiden. Verondersteld is dat verzorgingsactiviteiten via het gebruik doorwerken op verbetering van het leerkrachthandelen en nut (baten voor het onderwijs). Gebruik heeft een substantiële invloed op deze effectmaten. Gebruik (van het aanbod) blijkt vooral toegeschreven te kunnen worden aan factoren die de onderwijsverzorging zelf in handen heeft, namelijk bruikbaarheid, kwaliteit en verspreiding. Deze bevindingen betreffen meestal de onderwijsverzorgingsactiviteiten in het algemeen. De resultaten gelden niet direct voor afzonderlijke activiteiten. Zo blijkt het model te passen op de set OBd-activiteiten, maar slechts op een subset binnen de niet-OBd-activiteiten.

Implicaties voor evaluatieonderzoek naar onderwijsverzorging

Ten behoeve van de voorgenomen evaluatie van de onderwijsverzorging (Commissie Carpay, 1987) kunnen vanuit dit onderzoek de volgende twee punten onder de aandacht gebracht worden:

1. Het aanbod dat naar scholen wordt gedaan is als uitgangspunt genomen en niet bijvoorbeeld de taakstellingen van instellingen. Dit heeft voordelen, omdat is aangegeven wat het onderzoeksdomein is. Taakstellingen van de instellingen zijn daarentegen vaak divers en globaal en verwijzen niet of zeer moeilijk, naar te meten effecten of criteria.
2. De meting van effecten is en blijft een moeilijk probleem in evaluatie-onderzoek van onderwijsverzorging. Van (door het onderwijs) gebruikt aanbod valt na te gaan wat de effecten zijn in het onderwijs. Dit leidt ertoe dat men in evaluatie-onderzoek enerzijds verzorgingswerk kan relateren aan doelmatigheidsindicatoren (hoeveelheid inzet, output) ter verbetering van te realiseren gebruik en anderzijds het gerealiseerde onderwijs kan relateren aan outputindicatoren.

Evaluatie-onderzoek naar onderwijsverzorging is, ondanks onderkende problemen in de uitvoering ervan, van belang omdat het kan bijdragen aan een meer professionele ondersteuning van het onderwijs.

LITERATUUR

- Becker, H.A. & Porter, A.L. (1986). *Impact-assessment today*. Utrecht: Van Arkel.
- Bedrijvige school* (1989). Amsterdam: ABOP.
- Boland, T. (1988). De ontwikkeling van de leesattitude in het basisonderwijs: een longitudinale benadering. *Tijdschrift voor Onderwijsresearch*, 13, 1, 3-15.
- Boomsma, A. (1983). *On the robustness of LISREL against small sample size and non-normality*. Proefschrift RU Groningen.
- Carpay, J.A.M., Brenninkmeyer, G. & Hofstee, W.K.B. (1987). *Evaluatie van de onderwijsverzorging*. Den Haag: Staatsuitgeverij.
- Dingemans, M. & Paans, P. (1986). *Evaluatie van onderwijsbegeleiding*. Bergen op Zoom.
- Groot, A.D. de (1986). *Begrip van evalueren*. Den Haag: VUGA Uitgeverij.
- Herooverweging collectieve uitgaven Rijksbegroting (1982). Deelrapport 36. *De verzorgingsstructuur onderwijs*. Tweede Kamer stuk 16625., nr. 44, 114-120.
- Hoak Nota (1986). Hoger onderwijs: Autonomie en kwaliteit, Ministerie O & W.
- Hoeben, W.Th.J.G. (1987). *Evaluatie van de onderwijsverzorging, een samenhangend programma van evaluatie-onderzoek*. Groningen: RION.
- Hoeben, W.Th.J.G. & Stokking, K.M. (1986). *Probleemverkenning onderwijsverzorging. Stand van zaken 1985*. Groningen/Utrecht.
- Imants, J. & Vriens, M. (1987). *Het meten van doelmatigheidsbeleving bij leerkrachten basisschool*. Leiden: Psychological reports educational psychology EDU-87-01.
- Jöreskog, K.G. & Sörbom, D. (1985). LISREL VI. *Users guide*. University of Uppsala, 2nd printing.
- Koster, D., Tillema, H.H. & Batenburg, T. van (1989). The impact of the educational support system in the Netherlands. *Impact Assessment Bulletin*, 7, 83-97.
- Saris, W.E. & Stronkhorst, L.H. (1984). *Causal modelling in nonexperimental research: an introduction to the LISREL approach*. Amsterdam: Sociometric Research Foundation.
- Stokking, K.M. (1984). *Probleemverkenning onderwijsverzorging*. Utrecht.
- Tillema, H.H., Batenburg, Th. van & Koster, D. (1989). Het meten van gebruik. In R. Vandenberghe en R. van der Vegt (Red.). *Onderwijsvernieuwing*. ORD-bundel nr. 23. Lisse: Swets & Zeitlinger B.V.
- Tillema, H.H., Koster, A.J. & Batenburg, T.A. van (1989). *Aanbod en gebruik van onderwijsverzorging*. Groningen: RION.

Wiegersma, S., Giesbers, J.H.G.I. & Nauta, A.P.N. (1989). *Evaluatie onderwijsverzorging*. Eerste advies van de taakgroep evaluatie onderwijsverzorging.

Manuscript ontvangen 22-1-1990

Definitieve versie ontvangen 21-6-1990

De betrokkenheid van schoolleiders in het V.O. bij vernieuwingen in het onderwijs*

P.C. van der Sijde en E. Pruis

Universiteit Twente, Onderzoekcentrum Toegepaste Onderwijskunde

ABSTRACT

In this study the involvement of schoolleaders with two different educational innovations are being compared. These innovations are "the examination social and political studies" and "computers in education". It appears from the results of this comparison, that the differences can be ascribed to the different nature of the innovations: the involvement of the schoolleader with "computers in education" is higher than with "the examination social and political studies".

INLEIDING

In het voortgezet onderwijs hebben zich in de afgelopen periode veel vernieuwingen voorgedaan (bijv. Mavo-project en middenschool) en veel dienen zich aan (bijv. basisvorming). Een onderwijsvernieuwing kan veranderingen c.q. vernieuwingen op drie niveaus inhouden. Door Fullan (1982) worden de volgende onderscheiden: (1) het mogelijke gebruik van nieuwe of gereviseerde materialen (curriculum materialen of technologieën), (2) het mogelijke gebruik van nieuwe doceer-benaderingen (strategieën of activiteiten), en (3) de (mogelijke) veranderingen van visies en opvattingen (onderwijskundige aannames en aan materialen ten grondslag liggende ideeën). Het proces van de invoering van een vernieuwing doorloopt drie met elkaar verweven fasen (Van de Akker, 1989): initiatie - implementatie - institutionalisatie. In deze studie staan met name de implementatie en de institutionalisering centraal.

Indicaties voor het feit dat een vernieuwing als geïnstitutionaliseerd beschouwd kan worden zijn (1) de acceptatie van de vernieuwing door de actoren (docenten, leerlingen, schoolleiding), (2) de verwachting dat de vernieuwing geaccepteerd en gecontinueerd wordt, (3) dat de vernieuwing niet meer als zodanig wordt gezien, (4) de vernieuwing is een deel van de school geworden en (5) dat er tijd en geld is om de vernieuwing door te voeren (Crandall, 1989). Institutionalisering hangt af van de relatieve winst van de vernieuwing voor de school, de complexiteit en de uniciteit van de vernieuwing (Paul, 1977, geciteerd in: Clark, Lotto & Astuto, 1989).

In deze bijdrage vormt de betrokkenheid van de schoolleider (rector, directeur, conrector, adjunct directeur) bij een onderwijskundige vernieuwing het object van studie. De betrokkenheid wordt in het kader van deze bijdrage geoperationaliseerd als de aard van het begeleidingsgedrag van de schoolleider. De rol van de schoolleider ligt vooral in de voorwaardelijke sfeer als 'resource provider' in de breedste zin van het begrip (Bamburg & Andrews, 1989). Met name kan worden verwacht dat hij/zij regelmatig toetst of de visies van de bij de verandering betrokken actoren (voornamelijk docenten) stroken met de vernieuwing (zie bijv. Van der Vegt & Bruining, 1984).

* Deze studie is verricht in het kader van het onderzoeksproject 'Evaluatie van de experimentele invoering van maatschappijleer als facultatief examenvak' (SVO-project 8031)

In deze bijdrage willen wij de betrokkenheid van schoolleiders bij onderwijsvernieuwingen bestuderen door twee vernieuwingen met elkaar te vergelijken. Deze vernieuwingen verschillen van elkaar in de mate van relatief belang voor de school, complexiteit en uniciteit. In het vervolg van deze bijdrage zullen we eerst ieder van de twee vernieuwingen schetsen en vervolgens het onderzoek beschrijven.

DE TWEE VERNIEUWINGEN

De betrokkenheid van schoolleiders wordt bij een tweetal vernieuwingen met elkaar vergeleken, dit zijn 'de invoering van het eindexamen maatschappijleer' en 'invoering van computers in het onderwijs'.

Sinds het cursusjaar 1986/1987 op het Lbo en Mavo, en sinds 1987/1988 op het Havo en Vwo is het voor leerlingen mogelijk op een aantal scholen ($n=40$) maatschappijleer als examenvak te kiezen. Tot die tijd was maatschappijleer een verplicht vak op school, maar zonder examen. De invoering van het examen (wat door ons wordt opgevat als een onderwijsvernieuwing, zie ook Kempa, 1983) heeft bij docenten en schoolleiders geleid tot veranderingen in visie op het vak (Van der Sijde & Pruis, 1990) en daarnaast o.a. tot veranderingen in lesuitvoering (Van der Sijde & Pruis, 1990). Het belang van een examenvak maatschappijleer voor de school bestaat uit het kunnen aanbieden van een grotere keuzemogelijkheid aan leerlingen. De complexiteit van de vernieuwing bij maatschappijleer is, vanuit de schoolleider gezien, relatief gering, terwijl er sprake is van uniciteit.

Veel scholen voor voortgezet onderwijs beschikken over computers (Inspectie, 1986) als gevolg van landelijke stimuleringsprojecten (NIVO) en andere (lokale) projecten. De vernieuwing 'computers in het onderwijs' is in tegenstelling tot het 'examen maatschappijleer' geen enkelvoudige vernieuwing, maar een bundeling van vernieuwingen waarbij verschillende personen en componenten betrokken zijn (Plomp & Van de Akker, 1988), met andere woorden een complexe vernieuwing, waarvan het (relatieve) belang voor de school groot is. De mate van uniciteit is gering, immers de invoering van de computer moet op een gelijksoortige, maar vakeigen wijze plaatsvinden.

METHODE

Proefpersonen

Aan het onderzoek werd deelgenomen op vrijwillige basis door 50 schoolleiders (rectoren, directeuren, conrectoren, adjunct-directeuren) in het voortgezet onderwijs. 28 Schoolleiders waren betrokken bij het vernieuwingsexperiment 'eindexamen maatschappijleer' en 22 schoolleiders bij de invoering van 'computers in het onderwijs'.

Vragenlijst

De vragenlijst die aan de schoolleiders is voorgelegd bestaat uit een groot aantal items, waarvan we ten behoeve van deze bijdrage alleen rapporteren over de betrokkenheid bij de vernieuwing 'examen maatschappijleer' en 'computers in het onderwijs'. De betrokkenheid van schoolleiders wordt geoperationaliseerd met behulp van de Schoolleiderinterventieschaal (SIS), die door Van de Grift (1987) is ontwikkeld en aan een psychometrische (Rasch-)analyse is onderworpen. De SIS meet het begeleidingsgedrag van een schoolleider; de aard van het begeleidingsgedrag van de schoolleider geeft aan hoe betrokken hij/zij is bij de vernieuwing. In Tabel 1 wordt een overzicht van de items en hun moeilijkheidsgraad gegeven.

Tabel 1. Items en moeilijkheidsgraad van Schoolleiderinterventie schaal van Van de Grift (1987)

1	Het bestuur betrekken bij de vernieuwing	-2.75
2	Met mensen van buiten de school over de vernieuwing spreken	-1.75
3	Met ouders over de vernieuwing spreken	-1.61
4	Tijdens vergaderingen op school terugkomen op eerder besproken problemen in het gebruik van de vernieuwing	-1.61
5	Met onderwijsgeevenden individueel over de problemen met de vernieuwing in hun klas spreken	-1.47
6	Met de onderwijsgeevenden individueel over de vorderingen van de leerlingen met de vernieuwing spreken	-1.47
7	Tijdens vergaderingen op school de leerlingvorderingen bij de vernieuwing bespreken	-.85
8	Tijdens de vergaderingen op school aan de onderwijsgeevenden gedetailleerd verslag van de voortgang van de vernieuwingsactiviteiten in hun klas vragen	-.27
9	De vorderingen van de leerlingen met de vernieuwing bijhouden	-.27
10	Schriftelijk de vorderingen van de leerlingen met de vernieuwing bijhouden	-.17
11	Nabesprekingen met de onderwijsgeevenden houden over wat de schoolleider met betrekking tot de vernieuwing in hun klas heeft gezien	.98
12	Observeren in de klassen wat er met de vernieuwing gebeurt	1.18
13	Duidelijk aan een onderwijsgevende laten weten wanneer het vernieuwingsproces in zijn/haar klas onvoldoende vordert	1.22
14	Schriftelijk bijhouden wat de onderwijsgeevenden in hun klas met de vernieuwing doen	1.91
15	Naar aanleiding van klasseobservaties werksuggesties geven aan de onderwijsgeevenden voor het gebruik van de vernieuwing	1.97
16	In de klassen observeren hoe de door de schoolleider gegeven werksuggesties bij de vernieuwing door de onderwijsgeevenden in de praktijk gebracht zijn	2.85

Ieder item werd door de schoolleiders op een vijf-puntsschaal beoordeeld en voor de analyse werden de antwoorden gedichotomiseerd (1 $\rightarrow$ 0 en 2,3,4 en 5 $\rightarrow$ 1). De totaalscore werd met behulp van omreken Tabellen omgezet in een score. Het scoringsbereik van de vragenlijst ligt in de praktijk tussen -4 en +4; een hoge score betekent dat de schoolleider sterk betrokken is bij de vernieuwing en daarbij een groot scala interventiemethoden hanteert, een lage score betekent het omgekeerde. Voor details over deze vragenlijst verwijzen we naar Van de Grift (1987).

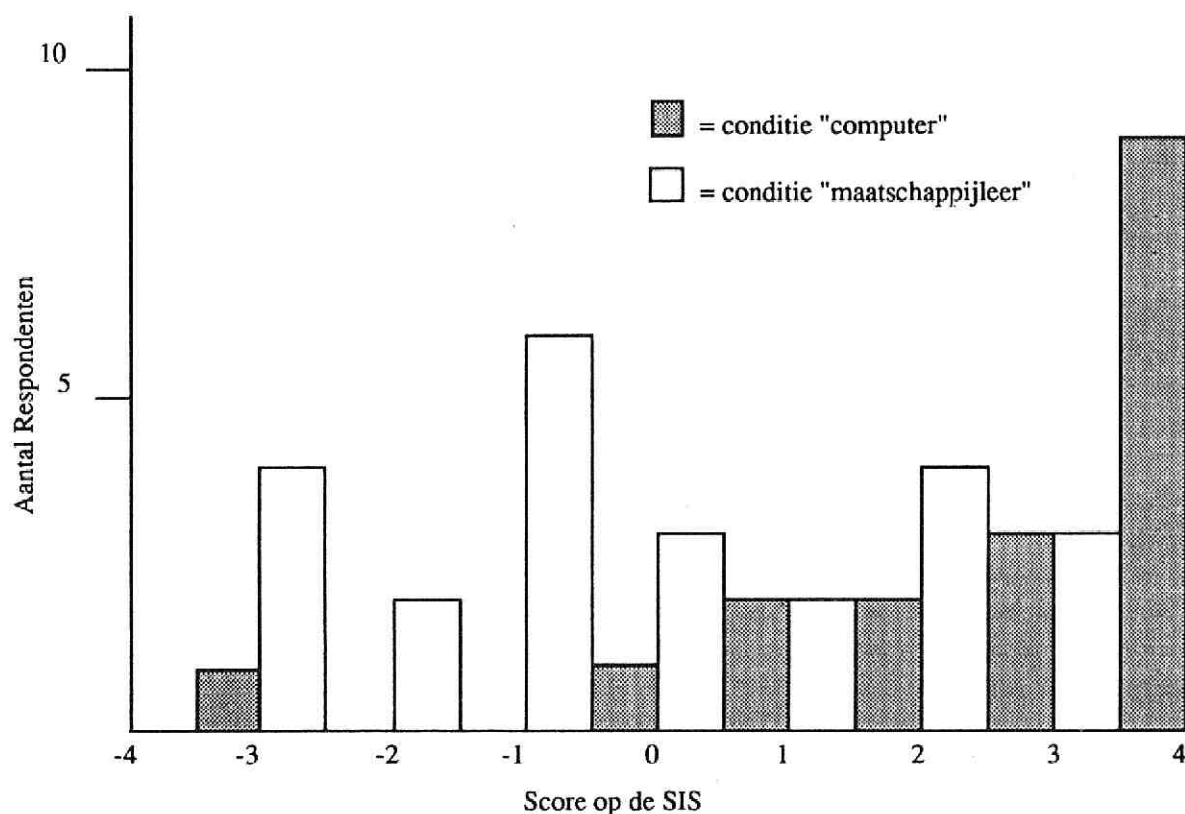
Naast de SIS wordt in deze bijdrage tevens gerapporteerd over de perceptie van de eigen betrokkenheid bij de vernieuwing in vergelijking met de betrokkenheid tot (andere) schoolvakken. Hierover is één item in de vragenlijst opgenomen.

Procedure

De vragenlijst werd aan de schoolleiders per post toegestuurd, zodat iedereen deze op een voor hem/haar gelegen tijdstip kon invullen.

RESULTATEN

Van de 28 schoolleiders die de vragenlijst invulden voor de vernieuwing 'examen maatschappijleer', hebben 25 van hen de SIS ingevuld. De gemiddelde score van deze groep schoolleiders bedraagt 0.434 (standaard deviatie: 2.02). Van de 22 schoolleiders die de vragenlijst invulden voor de vernieuwing 'computers in het onderwijs', hebben 19 van hen de SIS ingevuld. De gemiddelde score van deze groep schoolleiders bedraagt 2.198 (standaard deviatie: 2.12). Figuur 1 geeft een overzicht van de scores van beide groepen schoolleiders.



Wanneer er aan de schoolleiders gevraagd wordt of in hun perceptie hun betrokkenheid bij (andere) schoolvakken anders is, blijkt dat het merendeel van de schoolleiders vindt dat dit niet het geval is (zie Tabel 2): $t_{48}=1.16$, p = niet significant. (NB. Alle schoolleiders hebben deze vraag ingevuld).

Tabel 2. Afwijking van de betrokkenheid bij vernieuwing in vergelijking met andere vakken

conditie	nee (1)	weinig (2)	ja (3)	gemiddelde	s.d.
"Maatschappijleer"	20	5	3	1.39	.69
"Computers"	10	6	6	1.82	.85

DISCUSSIE

Uit de resultaten blijkt, dat volgens de schoolleiders zelf, zij niet méér betrokken zijn bij de vernieuwing in kwestie, dan zij betrokken zijn bij ieder ander vak op school. Dit resultaat geeft de algemene zelfperceptie van de schoolleider aan. Wanneer echter een aantal specifieke vragen wordt voorgelegd aangaande hun gedrag, dat wil zeggen activiteiten die zij uitvoeren, dan blijkt er een groot verschil tussen de schoolleiders in beide condities.

De betrokkenheid van de schoolleiders in de conditie 'computers' is groter dan die van de schoolleiders in de condities 'maatschappijleer'. Dit laatste zou het gevolg kunnen zijn van het verschil tussen de bundeling van vernieuwingen (zoals 'computers') en de enkelvoudige vernieuwing (zoals 'maatschappijleer'). Daarnaast is, volgens Crandall (1989), een vernieuwing geïnstitutionaliseerd, wanneer het door iedereen als vanzelfsprekend wordt beschouwd. Die vanzelfsprekendheid geldt in ieder geval voor de bestaande

vakken. De betrokkenheid van de schoolleider is in geval van een geïnstitutionaliseerd vak (maatschappijleer) minder dan bij een nog niet geïnstitutionaliseerde technologische vernieuwing (computers). Dit verklaart echter niet, waarom de schoolleiders op een rechtstreekse vraag naar hun betrokkenheid antwoorden, dat deze niet verschilt van hun 'normale' betrokkenheid. Een verklaring voor deze, schijnbare, tegenstelling zou de intentiegeladenheid van de algemene zelfperceptie van de schoolleider kunnen zijn; de schoolleider wil in eerste instantie alle vakken gelijk behandelen.

De betrokkenheid van de schoolleider bij het 'eindexamen maatschappijleer' is in vergelijking met de referentiegroep van Van de Grift als gemiddeld te noemen en de betrokkenheid van de schoolleiders bij de 'invoering van de computer in het onderwijs' als boven het gemiddelde. Met betrekking tot de invoering van de computer is een relativisering op zijn plaats. Plomp, Pelgrum en Steerneman (1990) concluderen uit hun onderzoek dat de integratie van de computers noch op school- noch op vakniveau op een aantal, volgens de onderzoekers, leidende scholen plaatsgevonden had. Deze ogenschijnlijke discrepantie (hoge betrokkenheid, lage implementatie) kan verklaard worden uit het feit dat er sprake is van twee verschillende zaken: de integratie van de computer op de school en in de vakken, en de betrokkenheid van de schoolleider. De schoolleider laat zijn of haar betrokkenheid bij onderwijsvernieuwingen mede afhangen van het feit of de vernieuwing alleen van schoolintern of ook van schoolextern belang is (Paul, 1977). De betrokkenheid van een schoolleider zal daarom voor een vernieuwing die voor de (status van de) school belangrijk is (computers), hoger zijn dan voor een vernieuwing die in feite als vanzelfsprekend gepercipieerd wordt (het examen maatschappijleer).

LITERATUUR

- Bamburg, J.D. & Andrews, R.L. (1989). Putting effective school research to work: the process of change and the role of the principal. In B. Creemers, T. Peters, & D. Reynolds (eds.), *School effectiveness and school improvement*. Lisse: Swets & Zeitlinger.
- Clark, D.L., Lotto, L.S. & Astuto, T.A. (1989). Effective schools and school improvement: a comparative analysis of two lines of inquiry. In J.L. Burdin (ed.), *School leadership, a contemporary reader*. Newbury Park: Sage Publications.
- Crandall, D.P. (1989). Conditions for successful policies and practices in the field of new technologies: some lessons from schoolimprovement research. In U. Hameyer & S. Loucks-Horsley (eds.), *New technologies and school improvement: support policies and practices*. Leuven: Acco.
- Fullan, M. (1982). *The meaning of educational change*. New York: Teacher College Press.
- Inspectie voor Speciale Diensten (1986). *Onderwijs en informatietechnologie: een beschrijving van de stand van zaken in het schooljaar 1985/1986*. 's-Hertogenbosch: Inspectie voor Speciale Diensten (Inspectierapport 7).
- Kempa, R. (1983). Bestimmungsfaktoren curricularer Lehrereignisse: Pruefungsbestimmungen und Stundentafeln. In U. Hameyer, K. Frey & H. Haft, *Handbuch der Curriculum-Forschung*. Basel: Beltz Verlag.
- Plomp, Tj. & Van de Akker, J.J.H. (1988). *Computer integration in the curriculum: promises and problems*. Paper presented at AERA, New Orleans.
- Plomp, Tj., Pelgrum, W.J.U. Steerneman, A.H.M. (1990). Influence of computer use on schools' curriculum: limited integration. *Computers & Education*, 14, 159-171.
- Van de Akker, J.J.H. (1989). Implementatie: wat is de beste methode? In *Onderwijskundig ontwerpen*. Heerlen: Ou.
- Van de Grift, W. (1987). *Implementatie van vernieuwingen: de rol van de schoolleider*. Leiden: dissertatie R.U.L.
- Van der Sijde, P.C. & Pruis, E. (1990). *Adoptie en implementatie van het eindexamen maatschappijleer als vernieuwing*. Enschede: Toegepaste Onderwijskunde.

- Van der Sijde, P.C. & Pruis, E. (1990). De status van en opvattingen over het vak maatschappijleer: resultaten van een onderzoek onder docenten en schoolleiders. In H. Dieteren & P. van der Sijde (red), *Maatschappijleer gewaardeerd*. Leiden: Stenfert Kroese.
- Van der Vegt, R. & Bruining, G.R.P. (1984). Implementation of innovations and school leadership. In R. van der Vegt, G.R.P. Bruining, & H. Knip (eds), *School leadership and implementation: on implementing educational change in schools*. Nijmegen: Vakgroep Interdisciplinaire Onderwijskunde K.U.N. (Mededelingen nr. 14).

Manuscript ontvangen 23-4-1990

Definitieve versie ontvangen 4-7-1990

Notities en Commentaren

ONDERWIJSONDERZOEK TE KIJK GEZET

COMMENTAAR OP HET RAPPORT 'SCHAAL EN KWALITEIT IN HET BASIS-ONDERWIJS'

Bert P.M. Creemers en Wijnand Th.J.G. Hoeben

RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen

Het Nederlandse onderwijs in de jaren zeventig werd gekenmerkt door grootscheepse vernieuwingsprojecten, die door onderwijskundigen werden geïnitieerd, geadviseerd dan wel ondersteund. Onderwijspolitiek en onderwijskunde sloten destijds naadloos aan. Sinds de resultaten van die vernieuwingsprojecten in de loop van de jaren tachtig voor en na zichtbaar werden, is gebleken dat de onderwijskundige ondersteuning niet zozeer was gebaseerd op kennis van en inzicht in de resultaten van de geadviseerde activiteiten. De gedrevenheid om de onderwijspraktijk te verbeteren is gepaard gegaan met een hoge dosis 'wishful thinking' (Scheerens, 1983). We hebben de indruk dat onderwijsonderzoekers lering uit deze ervaringen hebben getrokken door bij de voor advisering noodzakelijke mix tussen empirie en speculatie meer aandacht te geven aan de kwaliteit van hun empirische argumentatie. Dit had uiteraard tot gevolg dat onderwijskundige adviezen wat voorzichtiger werden.

Het rapport 'Schaal en kwaliteit in het basisonderwijs' van de Projectgroep Schaalvergroting Basisonderwijs (PSB, 1990) blijkt een breuk met deze trend. Het rapport adviseert een drastische schaalvergroting van de Nederlandse basisscholen en leunt daarbij heel sterk op een onderwijskundige argumentatie voor een optimale schoolgrootte van ongeveer 500 leerlingen en een onderwijskundig absolute ondergrens van 250 leerlingen. We zullen hieronder betogen dat de argumentatie een onaanvaardbaar laag empirisch gehalte en een onaanvaardbaar hoog speculatief gehalte heeft. De onderwijskundige argumentatie heeft bovendien weinig oog voor empirische tegenargumenten. Het geheel blijft daarmee onder de maat van de eisen die je mag stellen aan een goed doorwrocht advies.

Aangezien het onderwijskundig advies wordt gebruikt om een diep ingrijpende schaalvergroting in het Nederlandse basisonderwijs voor te stellen, waarbij ongeveer de helft van de Nederlandse basisscholen zou moeten worden opgeheven, ruim eenderde van de basisschool-leerlingen zou moeten worden verplaatst en een passant ongeveer f 700 miljoen zou worden bespaard, hadden de opstellers van het advies er rekening mee moeten houden dat hun argumenten op een goudschaaltje zouden worden gewogen. Op zijn minst hadden ze alternatieve bestedingsmogelijkheden met bijbehorende kwaliteitseffecten in hun analyses moeten betrekken. Het gestelde in Bijlage 3 draagt op een ongunstige manier bij aan de beeldvorming over onderwijskunde en onderwijsonderzoek: de auteurs zetten het onderwijsonderzoek te kijk.

KWALITEITSVERBETERING DOOR SCHAAALVERGROTING?

Adviezen voor beleid en praktijk van onderwijs berusten altijd op een mix van empirische en theoretisch-speculatieve argumenten. Het zal niet altijd eenvoudig zijn om op een goede verhouding uit te komen. Ter vermijding van een naadloos aansluiten tussen onderwijskunde en onderwijsbeleid, waarvan sprake was in de eerste helft van de jaren zeventig, en met name ter vermijding van de onderwijskundige overmoed die aanleiding gaf tot deze naadloze aansluiting, hebben sommigen de neiging een tamelijk vergaande distantie van onderzoek ten opzichte van het onderwijsbeleid te bepleiten. Dit komt men zowel binnen de beleidscontext als binnen de wetenschappelijke context tegen (Van Deursen, 1990; Koster, 1990).

Onderwijskunde en onderwijsonderzoek blijven zich echter afspelen binnen de spanningsrelatie tussen onderwijsverbetering en wetenschapsontwikkeling. Bij wetenschapsontwikkeling kunnen theoretische speculaties niet worden gemist: mits dergelijke speculaties tot een empirische beslissing worden gebracht, dat wil zeggen mits zij aanleiding geven tot voorspellingen die worden getoetst, houden zij de empirische discussie en daarmee de wetenschapsontwikkeling gaande. Bij adviezen voor onderwijsverbetering kunnen (aan theorieën ontleende) argumenten van speculatieve aard evenmin worden gemist: theorieën gaan altijd verder dan de waargenomen empirische werkelijkheid en vormen een (speculatief) interpretatiekader voor deze werkelijkheid. Ook hierbij geldt dat de empirische werkelijkheid niet uit het oog mag worden verloren en de toetssteen blijft van op verbetering gerichte adviezen (Creemers & Hoebe, 1984).

Wetenschappelijke adviezen zullen beter zijn naarmate hun empirisch gehalte hoger is en hun speculatieve gehalte lager. Wetenschappelijke adviseurs dienen daarbij zeer kritisch te zijn op de speculatieve gedeelten van hun argumentatie. Een absolute benedengrens voor het kritische gehalte van een argumentatie is ons inziens het recht doen aan empirische tegenargumenten in het advies.

Het advies dat schaalvergroting van de Nederlandse basisscholen de kwaliteit van hun onderwijs zal verbeteren, berust op twee soorten van argumenten. Het betreft empirische argumenten met betrekking tot de relatie tussen schoolgrootte en gemiddelde toetsscores voor de vakken rekenen en taal. Anderzijds betreft het argumenten van onderwijskundige aard, waarbij processen binnen de school worden gepresenteerd als een speculatief interpretatiekader voor de verklaring van een (empirisch?) verband tussen schoolgrootte en toetsscores.

De empirische argumentatie

De empirische argumenten houden te weinig rekening met de tegenstrijdige effecten van schoolgrootte op verschillende kwaliteitsindicatoren en met de beperkingen van het secundair geanalyseerde materiaal. Behalve toetsscores voor de vakken taal en rekenen werden in de analyses ook verwijzingspercentages naar het speciaal onderwijs als kwaliteitsindicator genomen. Na correctie voor de samenstelling van de leerlingpopulatie en voor de denominatie van de school bleek uit deze analyses een ongunstige invloed van de schoolgrootte op de verwijzingspercentages. De gunstige invloed van schoolgrootte op toetsscores is niet veel groter dan de ongunstige invloed ervan op verwijzingspercentages (beta's van .11, resp. .08). Het eerste cijfer wordt echter geïnterpreteerd als opmerkelijk en aanzienlijk, als een weliswaar bescheiden maar onmiskenbare invloed; het tweede cijfer wordt geïnterpreteerd als een zeer zwakke relatie waarvoor het moeilijk is een plausibele verklaring te vinden (PSB, 1990, p. 20-23). De relatie tussen schoolgrootte en het verwijzingspercentage naar het speciaal onderwijs wordt in de rest van het advies gemakshalve vergeten, terwijl de onderwijskundige argumentatie van het omgekeerde

verband uitgaat: schoolvergroting wordt als middel gezien om de trend van stijgende verwijzingspercentages te keren (PSB, 1990, p. 101-102). In beide gevallen worden resultaten van eerder onderzoek genegeerd, waarbij invloed van schoolgrootte op verwijzingsresultaten niet kon worden aangetoond (Wijnstra, 1987; Meijer, 1988).

Gegeven de verhoudingsgewijs geringe bijdrage van schoolgrootte in de verklaring van verschillen in gemiddelde CITO-scores en van verschillen in gemiddelde verwijzingspercentages, zou het gepast zijn geweest als de empirische analyses wat meer voorbehouden hadden gemaakt ten gevolge van een aantal onvolkomenheden. We hebben er begrip voor dat bij secundaire analyses wordt geroeid met de beschikbare riemen, maar bij het trekken van conclusies en met name bij het geven van beleidsadviezen dient rekening te worden gehouden met de beperkingen van het beschikbare materiaal.

Op de eerste plaats zou een voorbehoud gerechtvaardigd zijn geweest in verband met de gebruikte indicatoren voor de samenstelling van de leerlingenpopulatie. De leerlinggewichten die in de formatieregeling in het kader van het OVB (Onderwijsvoorrrangsbeleid) worden gehanteerd, zijn niet op basis van wetenschappelijke overwegingen, maar vooral op basis van beleidsmatige (en daarbij vooral budgettaire) overwegingen vastgesteld (Verhaak & Jungbluth, 1990). Andere, meer gebruikelijke, indicatoren voor het milieu van herkomst van leerlingen, zoals opleiding en beroep van ouders, verklaren in het algemeen meer variantie in leerprestaties. De door schoolgrootte te verklaren variantie wordt dus minder. In verschillende door het RION uitgevoerde onderzoeken verdwijnt een aanvankelijk bescheiden schoolgrootte-effect na correctie voor dergelijke variabelen. De effecten zijn bovendien zo gering dat het vaak niet de moeite waard was erover te rapporteren of dit slechts in het voorbijgaan te doen. De empirische gegevens werden daarom niet prominent gepresenteerd. Als bovendien wordt gecontroleerd voor intelligentiescores van leerlingen, dan is de omvang van een school een volstrekt te verwaarlozen factor in de verklaring van verschillen tussen prestatie-indicatoren (zie bijv.: Van Batenburg, 1988; Harskamp, 1988; Van der Werf, 1988; Brandsma & Knuver, 1989). Andere onderzoeksinstituten hebben waarschijnlijk vergelijkbare ervaringen.

Een tweede beperking is gelegen in de selectie-effecten die aanwezig zijn in de databestanden van het CITO. De eindtoets basisonderwijs wordt afgelegd ten gevolge van een beleidsmatige beslissing door de school. In sommige gevallen is ook de selectie van leerlingen die de eindtoets afleggen een gevolg van een beleidsbeslissing door de school. Alleen indien deze beleidsbeslissingen als random beslissingen mogen worden geïnterpreteerd, zou een analyse van het databestand van het CITO naar representativiteit een basis geven voor statistische generalisatie. Secundaire analyses van het PPON-materiaal zouden betere uitgangspunten voor een statistische generalisatie hebben gegeven. In de PPON-rapporten worden overigens geen schoolgrootte-effecten gerapporteerd (Wijnstra, 1988; Zwarts, 1990).

Een laatste probleem betreft de stabiliteit van gemiddelde toetsscores per school over een langere tijdsperiode. Bosker en Scheerens (1989) rapporteren correlaties van taal-, resp. rekenscores tussen opeenvolgende jaren, die variëren van .35 tot .65. Dergelijke correlaties geven aanleiding tot de voorspelling dat er maar weinig scholen zullen zijn die over een reeks van jaren vergelijkbare gemiddelde toetsscores laten zien. Een poging om ten behoeve van een outlier-studie een steekproef van scholen samen te stellen die, naast een bepaalde leerlingsamenstelling, in drie opeenvolgende jaren CITO-scores binnen de hoogste 20 en binnen de laagste 20 percentielen hadden, mislukte dan ook vanwege het geringe aantal scholen dat aan deze stabiliteitscriteria voldeed (Van den Berg, 1988). Deze stabiliteitsgegevens noodzakten tot grote voorzichtigheid bij het trekken van conclusies en het geven van beleidsaanbevelingen over schoolgrootte op basis van analyses van CITO-gegevens die op slechts één schooljaar betrekking hebben. Het stabiliteitsprobleem geldt a

fortiori voor kleine scholen waar de gemiddelde toetsscore is gebaseerd op de resultaten van drie tot zes leerlingen.

We concluderen dan ook dat de projectgroep bij haar empirische argumentatie onvoldoende rekening houdt met de beperkingen van het geanalyseerde materiaal, en dat zij, waar empirische gegevens worden gepresenteerd, op een ongerechtvaardigde manier toedeneert naar een vooropgezette conclusie.

De 'onderwijskundige' argumentatie

Voor een verklaring van de verschillen in schoolprestaties tussen scholen van verschillende omvang wordt in het rapport verwezen naar de onderwijskundige argumentatie van Claessen en Lagerweij (PSB, 1990, p. 87-106). Daarbij worden processen van schoolorganisatie, van schoolbeleid, van schooleffectiviteit en van pedagogisch-didactische kwaliteit enerzijds geïnterpreteerd als belangrijk voor een hoge kwaliteit, en anderzijds geïnterpreteerd als gevolgen van de grootte van een school.

Met betrekking tot schoolorganisatie worden versterking van management (een volledig vrijgestelde directeur), een competent bestuur, flexibele inzetbaarheid van leraren en een grotere onafhankelijkheid van de onderwijsverzorging genoemd als kwaliteitsverhogende factoren, die op hun beurt in gunstige zin worden beïnvloed door de school te vergroten. Het empirische gehalte van deze argumentatie is uitzonderlijk laag. De kwaliteitsbevorderende invloed van schoolleiders is empirisch omstreden, zeker voor wat betreft schoolleiders in het Nederlandse basisonderwijs (Hoeben, 1989; Van de Grift, 1990). Voorzover onderwijsonderzoek tot een gunstige invloed van schoolleiders concludeert, is het echter een invloed van onderwijskundig leiderschap en niet van managementvaardigheden. De kwaliteitsbevorderende invloed van de overige aspecten van de organisatie van een grote school heeft voorsnog geen empirische ondersteuning (Scheerens, 1989).

Aan voorwaarden voor een eigen schoolbeleid kan volgens het rapport beter worden voldaan bij grote scholen. Deze voorwaarden hebben vooral betrekking op vergroting van de bestuurlijke vermogens van scholen, op vergroting van hun interne capaciteiten, op vergroting van hun diagnostiserend vermogen en op verhoging van hun omgevingsresponsiviteit. De effecten van deze voorwaarden op kwaliteitsverbetering worden in de argumentatie eerder gepostuleerd, dan verklaard. Empirische ondersteuning van de argumentatie kan niet worden gevonden. Het speculatieve gehalte is derhalve uitzonderlijk hoog.

Factoren van schooleffectiviteit worden onderscheiden in factoren van organisatorische aard en in processen, die moeten leiden tot de beoogde verbeteringen. De factoren van organisatorische aard komen voort uit de bekende lijstjes van Amerikaans effectiviteitsonderzoek. Het rapport onderkent dat niet alles wat voor scholen buiten Nederland geldt, ook hier van toepassing is. Hieruit worden echter geen consequenties getrokken. Om aan het empirisch gehalte van hun argumentaties recht te doen, hadden de auteurs in ieder geval rekening moeten houden met het gegeven dat in de Nederlandse situatie de invloed van de volgende factoren empirisch omstreden is: het onderwijskundig leiderschap; soort en beschikbaarheid van externe ondersteuning; inservice training van leraren; ouderparticipatie (Scheerens, 1989). Het effect van schoolgrootte op innovatievorderingen is afwezig of niet significant (Hofman et al., 1984; Harskamp & Rijken, 1988; Van der Werf, 1988). Het kwaliteitsbevorderende effect van processen die leiden tot onderwijsvernieuwing is empirisch gezien omstreden (Hoeben & Creemers, 1988; Hoeben, 1989). De relatie van de processen en factoren van schooleffectiviteit met de omvang van een schoolorganisatie wordt in het advies gepostuleerd, maar niet beargumenteerd. Het speculatieve gehalte van de argumentatie is derhalve hoog en de empirische argumentatie houdt te weinig rekening met empirische tegenargumenten.

De processen van pedagogisch-didactische kwaliteit worden onderscheiden in het vermogen tot individualisering, specialisatie en loopbaanperspectief van leraren, en in zorgverbreding. Van deze drie aspecten worden positieve kwaliteitseffecten verwacht, die gunstiger zullen zijn bij grote scholen: deze aspecten worden geïnterpreteerd als vragend om schoolorganisatorische oplossingen, waarvoor een school met een voldoende aantal leerlingen en leraren gewenst is. Daarbij wordt geen rekening gehouden met de elders in het rapport gepresenteerde empirische gegevens dat grotere scholen verhoudingsgewijs meer leerlingen doorverwijzen naar het speciaal onderwijs, hetgeen iets zou moeten zeggen over hun capaciteiten voor zorgverbreding. Evenmin wordt veel aandacht besteed aan empirische gegevens over interne differentiatie, zowel in de relatie tot kwaliteit van het onderwijs als in relatie tot schoolgrootte (Reezigt & Weide, 1989). Zonder empirische argumentatie, maar in overeenstemming met de ideologie van 'groot is mooi' kiest het rapport voor externe differentiatie tussen klassen van eenzelfde jaargroep. Op effecten van deze voorkeur wordt niet ingegaan. Gezwegen wordt over de waarschijnlijke selectiviteit die gaat ontstaan in Nederlandse basisscholen met een structureel onderscheid tussen klassen voor meer en klassen voor minder getalenteerde leerlingen. Daarom hoefden de auteurs zich ook niet bezig te houden met onderzoek dat uitwijst dat selectieve systemen niet beter zijn voor goede leerlingen, maar wel slechter zijn voor minder goede leerlingen (Kifer, 1989).

Het valt op dat argumenten ontleend aan empirisch onderzoek over de kwaliteit van instructie en over de daarmee samenhangende professionele kwaliteit van leraren in de onderwijskundige argumentatie voor een grote school ontbreken. Aangezien dit kwaliteits- en effectiviteitsfactoren betreft op het niveau van de klas en van de afzonderlijke leraar (gunstige gelegenheden tot leren, actieve betrokkenheid van leerlingen, regelmatige succeservaringen, structuur brengen in het onderwijs, vaart houden in het onderwijs, en adequate tempo-aanpassingen; zie Berliner, 1987), en aangezien zij nauwelijks in verband zijn te brengen met schoolgrootte, hoeft dit niet helemaal te verbazen. Kwaliteits- en effectiviteitsfactoren op het niveau van de klas en van de afzonderlijke leraar kunnen daarentegen wel in verband worden gebracht met waar het in het onderwijs eigenlijk om gaat, namelijk de kwaliteit van het dagelijkse werk met leerlingen. In een wetenschappelijk onderbouwde argumentatie van kwaliteitsverbetering door schaalvergroting zou daarom op zijn minst mogen worden verwacht dat een poging werd gedaan om kwaliteiten van de instructie door de leraar in verband te brengen met de omvang van de school. Een dergelijke poging zal naar onze mening waarschijnlijk niet kunnen slagen. Schaal en kwaliteit hebben daarvoor te weinig met elkaar te maken. De argumentatie op het niveau van de school is zwak, zoals we hiervoor hebben aangetoond. Een argumentatie met behulp van kwaliteiten van leraren kan wel leiden tot verbeteringsvoorstellen, maar dergelijke voorstellen kunnen nauwelijks meer worden geïnterpreteerd als gevolgen van schaalvergroting. Dat Claessen en Lagerweij dit ook niet eens hebben geprobeerd, suggereert echter des te meer dat zij naar hun advies van schaalvergroting hebben toegeredeneerd. Het versterkt de algehele indruk dat de onderwijskundige onderbouwing van het rapport een bijzonder laag kritisch gehalte heeft.

De argumentatie van de optimale schoolgrootte tenslotte, komt helemaal uit de lucht vallen. In de statistische analyses van schaal en kwaliteit hanteert het rapport een school van 300 leerlingen en meer als grootste schaalcategorie. In de financiële productiviteitsanalyses wordt berekend dat de kosteneffectiviteit optimaal is bij een omvang van 400-450 leerlingen. De onderwijskundige argumentatie gaat voorbij aan het statistische optimum van 300 leerlingen en meer en aan het financiële optimum van 400-450 leerlingen en komt uit op een onderwijskundig optimum van 500 leerlingen. Het doorslaggevende onderwijskundige argument lijkt te bestaan uit de voorkeur voor een school met een vrijgestelde

schoolleider en met 16 klassen. Deze voorkeur blijkt op zichzelf reeds op drijfzand te berusten, zoals boven is aangegeven. Maar deze voorkeur gaat tevens voorbij aan de statistische en de financiële argumenten elders in het rapport.

Concluderend menen we te mogen stellen dat de onderwijskundige argumentatie van de stelling dat schaalvergroting van de basisschool tot een optimum van 500 leerlingen een verbetering van de kwaliteit tot gevolg heeft, een uitzonderlijk hoog speculatief gehalte heeft, zich bijzonder weinig gelegen laat liggen aan empirische argumenten, en bijzonder ongevoelig is voor empirische tegenargumenten. De onderwijskundige argumentatie van het rapport is daarmee wetenschappelijk gezien onder de maat.

DE AFWEGING VAN SCHAAL EN KWALITEIT

Om schaalvergroting van het Nederlandse basisonderwijs ter discussie te stellen was geen onderwijskundige argumentatie van een mogelijk verband tussen schaal en kwaliteit nodig geweest. Het financiële argument dat ten gevolge van schaalvergroting *f* 700 miljoen op een andere manier in het basisonderwijs besteed kunnen worden, zou een voldoende aanleiding zijn geweest om een diepgaande wetenschappelijke en politieke discussie over de voor- en nadelen van schaalvergroting voor de kwaliteit van het basisonderwijs te doen starten. Dat hebben de auteurs niet gedaan. Het rapport houdt zich zelfs niet bezig met alternatieve bestedingen van het genoemde bedrag, noch met de mogelijke kwaliteitseffecten van dergelijke alternatieve bestedingen. Met name vanuit het oogpunt van kwaliteitsverbetering zou een discussie over alternatieve in kwaliteit te investeren bestedingen van een bedrag van deze omvang de discussie hebben geplaatst op het niveau waar ze thuishoort: de afweging van de kosten van het verhoudingsgewijze dure vigerende beleid en de kwaliteitseffecten van alternatieve bestedingen van hetzelfde bedrag. Nu lijkt het er te zeer op dat de projectgroep een argumentatie heeft gezocht voor een aanzienlijke bezuiniging. Openbare mededelingen van het tegendeel overtuigen niet. Een verantwoord beleidsadvies over schaal en kwaliteit in het basisonderwijs, dat budgettaire neutraliteit als uitgangspunt neemt, behoort zich bezig te houden met kwaliteitseffecten van alternatieve bestedingen van het te bezuinigen bedrag. In tegenstelling tot het Ministerie van Financiën (1984) weet het Ministerie van Onderwijs en Wetenschappen, althans zijn ambtenaren en onderwijskundige adviseurs, kennelijk niet hoe het hoort.

Naar onze mening is een dergelijke analytische evaluatie van het vigerende beleid in vergelijking met beleidsalternatieven waarbij rechtstreeks in kwaliteitsverbetering wordt geïnvesteerd uit de met schaalvergroting bespaarde middelen, de enige manier om schaal en kwaliteit zinvol met elkaar in verband te brengen. De poging van Claessen en Lagerweij om de op zijn best zwakke verbanden tussen schaal en kwaliteit oorzakelijk te interpreteren was bij voorbaat gedoemd te mislukken. Het beschikbare onderwijsonderzoek laat een dergelijke oorzakelijke interpretatie nu eenmaal niet toe. Dat hadden zij moeten weten. Schaalvergroting kan alleen maar met behulp van flankerend beleid tot kwaliteitsverbetering leiden. Schaalvergroting is daarbij niet eens noodzakelijk. Het is alleen maar nuttig om geld vrij te maken voor alternatief, op kwaliteitsverbetering gericht beleid.

Beleidsalternatieven zijn er te over. In een korte opsomming beperken we ons tot alternatieven in de sfeer van incentives, dat wil zeggen beleidsmaatregelen die stimuleren dat scholen en leraren de kwaliteit van hun werk verbeteren. Bij het aantreden van de huidige bewindslieden was er sprake van vertrouwenwekkende maatregelen, die het klimaat in het Nederlandse onderwijs zouden verbeteren; daarbij werd in eerste instantie gesproken over salarisverbeteringen. Daarnaast kan worden gedacht aan het afbouwen van de verstikkende bureaucratische regelgeving. Op zichzelf zou dit reeds een stimulerend

effect hebben; het zou bovendien besparingen opleveren die voor flankerend beleid kunnen worden bestemd. Afbouwing van regelgeving wijst in de richting van vergroting van de autonomie van scholen. Op zichzelf heeft autonomie waarschijnlijk weinig kwaliteitseffecten. Gecombineerd met beleid en regelgeving waarbij feitelijke kwaliteit wordt beloond en waarbij scholen middelen voor eigen kwaliteitsverbetering worden aangereikt, heeft de autonome school waarschijnlijk wel kwaliteitseffecten. Beloning van feitelijke kwaliteit door boven een redelijk niveau van de gangbare inputfinanciering elementen van outputfinanciering te introduceren, zou naar ons idee in een analytische evaluatie wel eens kunnen blijken de grootste kwaliteitseffecten te hebben.

Om schaal en kwaliteit tegen elkaar af te wegen, is dus een ander rapport met een andere aanpak nodig. Het vigerende beleid moet in vergelijking met enkele mogelijke beleidsalternatieven analytisch worden geëvalueerd. Een van de beleidsalternatieven zou in ieder geval moeten bestaan uit schaalvergroting die gepaard gaat met de genoemde stimulerende maatregelen. Bij een analytische evaluatie zou wel eens kunnen blijken dat stimulerende maatregelen meer effect hebben bij grotere scholen. Iets vergelijkbaars zou kunnen blijken bij een ander beleidsalternatief dat bestaat uit maatregelen om de eerder genoemde kwaliteiten van leraren in de klas te verbeteren. Een onderdeel van dit alternatief zou er uit kunnen bestaan om schoolleiders meer bevoegdheden te geven het functioneren van leraren op hun school te bewaken, te evalueren en zo nodig te verbeteren. Deze maatregel zou wel eens beter uitvoerbaar kunnen zijn op grotere scholen met vrijgestelde directeuren. Dit zijn twee voorbeelden van argumenten die schaal en kwaliteit indirect met elkaar verbinden doordat ze elkaar met behulp van flankerend beleid kunnen versterken. Het rapport dat nodig is, dient langs deze lijn alternatieven voor flankerend beleid uit te werken en in hun effecten op kwaliteit te vergelijken met het huidige beleid. Alleen dan spreken we niet meer over bezuiniging, maar over kwaliteitsverbetering. Alleen dan zijn goede afwegingen mogelijk.

Het bekritisieren van de kwaliteit van het rapport is in het belang van het onderwijs, in het belang van de kwaliteitsverbetering in het onderwijs, maar ook in het belang van het Nederlandse onderwijsonderzoek. Betere adviezen zijn mogelijk. Daarbij is het uiteraard de vraag of deze een zelfde inhoud zullen hebben. Een volgens bovenstaande lijnen opgebouwd advies zou in ieder geval een serieuze discussie en een politieke afweging uitlokken. Het gegeven advies heeft echter noch een discussie over schaalvergroting noch over kwaliteitsverbetering in het Nederlandse basisonderwijs tot gevolg, laat staan een politieke afweging. We hebben de indruk dat de eerste massale en afwijzende reacties het onderwerp in feite reeds van de agenda hebben afgevoerd. Dat vinden we jammer, met name omdat daarmee ook het traject van een discussie over alternatieve investeringen in kwaliteitsverbetering wordt afgesneden.

LITERATUUR

- Batenburg, Th.A. van (1988). *Een evaluatie van taalmethoden*. Groningen: RION, dissertatie.
- Berg, G. van den (1988). *Effectieve instructie en klascontext*. Een longitudinaal onderzoek naar de effecten van instructie van variabelen op het leren van instrumentele kennis en vaardigheden. Groningen: RION (subsidie-aanvraag).
- Berliner, D.C. (1987). Simple views of effective teaching and a simple theory of classroom instruction. In D.C. Berliner & B.V. Rosenshine (eds). *Talks to Teachers*. New York: Random House.
- Bosker, R. & Scheerens, J. (1989). Criterion definition, effect size and stability, Three fundamental questions in school effectiveness research. In B.P.M. Creemers, T. Peters & D. Reynolds (eds.). *School Effectiveness and School Improvement*. Proceedings of the Second International Congress (p. 241-252). Amsterdam: Swets & Zeitlinger.

- Brandsma, H.P. & Knuver, J.W.M. (1989). *Basisschoolkenmerken als determinanten van het functioneren van leerlingen*. Deelrapport II: Resultaten. Groningen: RION.
- Creemers, B.P.M. & Hoebe, W.Th.J.G. (1984). *Onderzoek tussen onderwijsontwikkeling en wetenschapontwikkeling*. Harlingen: Flevodruk.
- Deursen, C.I. van (1990). Het nut van onderwijskunde. *Tijdschrift voor Onderwijswetenschappen*, 20, 222-229.
- Grift, W. van de (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement*, 1, 26-40.
- Harskamp, E.G. (1988). *Rekenmethoden op de proef gesteld*. Groningen: RION, dissertatie.
- Harskamp, E.G. & Rijken, V.M. (1988). *Periodieke peiling stand van zaken basisscholen*. Eindrapport. Groningen: RION.
- Hoebe, W.Th.J.G. (1989). Educational innovation or school effectiveness: a dilemma? In B. Creemers, T. Peters & D. Reynolds (eds.). *School Effectiveness and School Improvement*. Proceedings of the Second International Congress (p. 157-167). Amsterdam: Swets & Zeitlinger.
- Hoebe, W.Th.J.G. & Creemers, B.P.M. (1988). Political backgrounds and innovation strategies as influences on middle school implementation. *International Journal of Educational Research*, 12, 533-546.
- Hofman, R., Alderen, M. van & Span, B. (1984). *Achterblijvers in het innovatieproces basisschool*. Omvang, kenmerken en motieven. Den Haag: SVO.
- Kifer, E. (1989). What IEA studies say about curriculum and school organisation. In A.C. Purves (ed.). *International comparison and educational reform*. Alexandria: Association for Supervision and Curriculum Development.
- Koster, K.B. (1990). Onderwijskunde: van idee tot professie? *Tijdschrift voor Onderwijswetenschappen*, 20, 230-239.
- Meijer, C.J.W. (1988). *Verwijzing gewogen*. Een studie naar de determinanten van verwijzing naar speciaal onderwijs. Groningen: RION, dissertatie.
- Ministerie van Financiën, Afdeling Beleidsanalyse (1984). *Evaluatiemethoden, een introductie*. Den Haag: Staatsuitgeverij.
- Projectgroep Schaalvergroting Basisonderwijs (1990). *Schaal en kwaliteit in het basisonderwijs*. Zoetermeer: Ministerie van Onderwijs & Wetenschappen.
- Reezigt, G.J. & Weide, M.G. (1989). *Effecten van differentiatie. Resultaten survey-onderzoek*. Deelrapport IV, Groningen: RION.
- Scheerens, J. (1983). *Evaluatie-onderzoek en beleid: methodologische en organisatorische aspecten*. Den Haag: SVO, dissertatie.
- Scheerens, J. (1989). *Wat maakt scholen effectief? samenvattingen en analyses van onderzoeksresultaten*. Den Haag: SVO.
- Verhaak, Chr. & Jungbluth, P. (1990). Extra leerkrachten voor arbeiderskinderen en migranten; de formatieregeling basisonderwijs als beleidsinstrument voor de onderwijsvoorrang. In C.A.C. Klaassen & P.L.M. Jungbluth (red.). *Onderwijs en samenleving* (p. 337-350). Onderwijs Research Dagen 1990. Nijmegen: ITS.
- Werf, M.P.C. van der (1988). *Het schoolwerkplan in het basisonderwijs*. Ontwikkeling, implementatie en opbrengst. Lisse: Swets & Zeitlinger.
- Wijnstra, J.M. (1987). *Op de grens van basis- en speciaal onderwijs*. Lisse: Swets & Zeitlinger.
- Wijnstra, J.M. (red.) (1988). *Balans van het rekenonderwijs in de basisschool*. Uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs. Arnhem: CITO.
- Zwarts, M. (red.) (1990). *Balans van het taalonderwijs aan het einde van de basisschool*. Uitkomsten van de eerste taalpeiling einde basisonderwijs. Arnhem: CITO.

Boekbesprekingen

W. van Esch, *Schoolbesturen en de vrijheid van onderwijs. Bestuurlijke en organisatorische factoren van invloed op de benutting van beleidsruimte door bijzondere schoolbesturen*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen, 1988 (Proefschrift Universiteit van Amsterdam).

Juridisch gezien vormen schoolbesturen het bevoegde gezag van de scholen en zij kunnen als zodanig verantwoordelijk worden gesteld voor de kwaliteit van het onderwijs in de onder het bestuur ressorterende scholen. Reeds in 1964 stelde Idenburg dat er 'merkwaardig weinig gegevens over schoolbesturen bekend zijn' en 'het zou intussen de moeite waard zijn voor het gehele terrein van het bijzonder onderwijs te onderzoeken hoe schoolbesturen zijn samengesteld en welke rol zij daadwerkelijk in de leiding der scholen spelen.'

Pas recentelijk worden resultaten gepresenteerd van empirische studies inzake het functioneren van schoolbesturen. Wil van Esch is de eerste (in een reeks) die zijn onderzoek met een dissertatie afrondt. In dit proefschrift is een schat van informatie bijeengebracht over samenstelling en functioneren van schoolbesturen in het bijzonder onderwijs.

In het inleidende hoofdstuk schetst Van Esch het kader waarbinnen het onderzoek moet worden geplaatst, namelijk binnen de SVO-programmering inzake bestuurlijk-juridische theorievorming op het terrein van onderwijs. Hij noemt zijn onderzoek zelf echter praktijkgericht, omdat '... het praktische probleem is geformuleerd door vertegenwoordigers van (een deel van) het onderwijsveld' (p. 5). Wat dat praktische probleem inhoudt wordt echter niet verder uitgelegd.

De kern van het onderzoek van Van Esch betreft de vraag naar factoren die van invloed zijn op verschillen in benutting van beleidsruimte inzake de vrijheid van inrichting door schoolbesturen. De onderzoeksvraagstelling wordt gepreciseerd vanuit twee invalshoeken: een historisch-juridisch-sociologische (H2) en een bestuurlijk-organisatorische (H3). Opvallend is dat de specificatie van het centrale begrip onder studie 'beleidsvrijheid' pas plaatsvindt in paragraaf 3.3.2 (p. 79). Beter zou zijn geweest deze paragraaf voorafgaand aan H2 en H3 te presenteren. Daarmee zou een regelmatige terugkoppeling van het begrip 'beleidsvrijheid' (onderverdeeld in formele, feitelijke en gepercipieerde) mogelijk zijn gemaakt en dit zou de leesbaarheid van beide hoofdstukken zeer ten goede zijn gekomen. De historisch-juridisch-sociologische invalshoek (H2) gebruikt Van Esch om de achtergronden van de strijd voor vrijheid van onderwijs te begrijpen en de gevolgen daarvan voor ons onderwijssysteem en voor de schoolbesturen. Hij maakt daarbij gebruik van secundaire bronnen. Echter voor een empirische studie waarin de vrijheid van onderwijs leidraad is en wordt gespecificeerd in het begrip 'beleidsvrijheid' lijkt gebruik van primaire - meer op de vraagstelling toegespitste - bronnen noodzakelijk.

Hoofdstuk 3 betreft het bestuurlijk-organisatorisch perspectief, waarbij inzichten uit organisatietheorieën, zoals de contingentietheorie, configuratietheorie, interorganisatiele en netwerktheorie worden gehanteerd om het functioneren van schoolbesturen in een kader te plaatsen. Met name de organisatietheorie van Mintzberg wordt als referentiekader gebruikt. Van Esch onderscheidt uiteindelijk de volgende clusters van mogelijke invloedsfactoren: (1) achtergrondkenmerken van schoolbesturen, (2) bestuursinterne factoren, (3) factoren in relatie met andere interne organisatie-eenheden en (4) factoren in de omgeving van schoolbesturen. Het is jammer dat hij deze hoofdstukken niet afsluit met een synthese van hoofdstuk 2 en 3, waarbij hij tracht de inzichten uit theorieën

en clusters van invloedsfactoren tot een consistent geheel samen te brengen. In feite is dit een van de grootste zwakheden van de dissertatie: er wordt niet getracht te komen tot een samenhangend model van factoren, waaraan de resultaten kunnen worden 'getoetst'.

Hoofdstuk 4 betreft de onderzoeksofzet. Het onderzoek bestaat uit gevalsstudies (10x) en een survey. Juist omdat nog zo weinig bekend is over schoolbesturen, lijkt dit een logische werkwijze. Het is dan ook beslist jammer dat niet meer gebruik is gemaakt van gegevens uit de gevalsstudies, voorafgaand aan de survey. Met name het feitelijk gebruik van 'beleidsvrijheid' door schoolbesturen zou op die wijze in al z'n omvang kunnen worden geëxploreerd om vervolgens generaliseerbare kennis daarover vast te stellen in het survey. Het responspercentage van het survey is 42.6%. Daarvan is 18% alleen bestuurder voor het voortgezet onderwijs (p. 170). Het is mij bepaald niet duidelijk waarom deze groep in de steekproef is opgenomen (waarover ook in de paragraaf voor respons niets wordt vermeld). Temeer daar in het historisch-juridisch overzicht (H2) ook alleen maar wordt ingegaan op het basisonderwijs en op p. 80 van dat hoofdstuk juist wordt gewezen op het verschil: 'De beleidsruimte inzake de vrijheid van oprichting, van inrichting en van richting is in het basisonderwijs anders en groter dan in het voortgezet onderwijs.' Een aanvullend probleem inzake de responsgroep betreft het gegeven dat ongeveer 20% (van die 42.6%) van de enquête niet is ingevuld door voorzitter of secretaris van het schoolbestuur, maar door schoolleiding of administratiekantoor. Het lijkt mij niet onwaarschijnlijk (en onderzoek door mij uitgevoerd bevestigt dat) dat deze groeperingen anders antwoorden. Wanneer we deze groeperingen zouden uitsluiten blijft een respons van $\pm 26.4\%$ over. Het is dan ook volstrekt onduidelijk waarom een non-respons-onderzoek (hetgeen eenvoudig met de telefonische derde rappel gecombineerd had kunnen worden) achterwege is gelaten.

Hoofdstuk 5 gaat in op de resultaten van de gevalsstudies. Over de theoretische achtergrond en de selectie van specifieke variabelen is Van Esch wel erg kort, hij verwijst slechts naar H2 en H3. Via de gevalsstudies wordt nagegaan in hoeverre de beleidsruimte anders is voor confessionele (4x) versus niet-confessionele (6x) schoolbesturen. Van Esch concludeert dat scholen tot op grote hoogte ruimte krijgen om een eigen godsdienstige of levensbeschouwelijke weg te gaan. Tevens moet hij vaststellen 'wat de kern van het onderwijs betreft en de rol van de grondslag daarbij kunnen we concluderen, dat de meeste schoolbesturen zich hoogstens controlerend gedragen' (p. 168).

In hoofdstuk 6 worden de resultaten van een survey onder 390 schoolbesturen besproken. In paragraaf 6.1 wordt nagegaan in hoeverre de verschillende richtingen (A.B., R.K., P.C.) op de onafhankelijke en afhankelijke variabelen verschillen. Via factoranalyses wordt in paragraaf 6.2 getracht de grote hoeveelheid variabelen te reduceren. Dit hoofdstuk zou een stuk leesbaarder zijn geworden wanneer deze paragraaf als bijlage was opgenomen. Informatie over de mate van betrouwbaarheid van de ontwikkelde schalen wordt node gemist. Bij navraag werd alsnog een lijst toegezonden waaruit blijkt dat 40% van de schalen (17:43) een interne consistentie (alpha-coëfficiënt) bezitten van onder de .70 en zelfs 74% van onder de .80. Hierbij is 18 maal sprake van 2-item "schalen". Waarom heeft Van Esch er eigenlijk niet voor gekozen om een hoofdstuk instrumentontwikkeling op te nemen. De 'operationalisaties' van variabelen staan nu verspreid over verschillende hoofdstukken, hetgeen veel bladerwerk vergt van een ook wetenschappelijk geïnteresseerde lezer.

Hoofdstuk 6 wordt afgesloten met regressie-analyses, waarbij als afhankelijke variabelen optreden: (1) inschatting van beleidsruimte, (2) benutting ruimte voor grondslag, (3) gebruik externe documenten en (4) gebruik wettelijke mogelijkheden. Van de eerste twee afhankelijke variabelen wordt respectievelijk 24 en 22% variantie verklaard. Van de overige twee slechts 4 en 6%. Van Esch concludeert 'dat schoolbesturen met een sturende

taakopvatting, met veel cohesie, die berekend zijn voor hun taak en waar de inspectie een stimulerende rol vervult de beleidsruimte groter inschatten' (p. 244) en dat 'grote confessionele schoolbesturen met veel (basis)scholen onder zich, die er een sturende taakopvatting op na houden en veel cohesie vertonen en voor hun taak zijn berekend meer van de beleidsruimte inzake de vormgeving van de grondslag gebruik maken' (p. 246).

Het afsluitende hoofdstuk betreft een systematische, heldere en lezenswaardige samenvatting met conclusies. Vooral een belangrijk hoofdstuk dus voor hen die geïnteresseerd zijn in het reilen en zeilen van schoolbesturen maar liever geen 300 bladzijden willen doorworstelen.

Al met al een dissertatie waarin een schat aan informatie is bijeengebracht, waarbij echter een uitvoerige nabeschuiving wordt gemist. Juist omdat op dit moment schoolbesturen op allerlei fronten, door deregulering en autonomievergroting, veranderende bekostiging, de roep om kwaliteit en dergelijke, onder druk staan. Niet in het minst door de recent weer sterk ter discussie staande begrippen 'richting' en 'beheersvorm'. Een dissertatie met als titel 'schoolbesturen en de vrijheid van onderwijs' zou dan ook zeker niet voorbij mogen gaan aan een nabeschuiving over de implicaties van de onderzoeksresultaten voor het toekomstige beleid van schoolbesturen en overheid.

R. H. Hofman,
RION, Instituut voor Onderwijsonderzoek,
Rijksuniversiteit Groningen,
Postbus 1286, 9701 BG Groningen

J.M.H.W. Willems, *Studietaken als instructiemiddel: een onderzoek naar de sturende werking van opdrachten bij studieteksten vanuit een informatietheoretisch kader*. Proefschrift K.U. Nijmegen, Nijmegen: Instituut voor Onderzoek van het Wetenschappelijk Onderwijs. ISBN 90 901 794 1

In zijn proefschrift, getiteld 'Studietaken als Instructiemiddel', beschrijft J. Willems een onderzoek naar de sturende werking van opdrachten bij studie-teksten.

Willems deed onderzoek bij studenten uit het Wetenschappelijk Onderwijs. Het idee om studie-taken binnen dit Wetenschappelijk Onderwijs te introduceren sproot voort uit het gegeven dat deze 'studenten vaak voldoende kennis opdoen, maar deze kennis niet weten te gebruiken bij het aanpakken van wetenschappelijke of maatschappelijke problemen'. Een bekend verschijnsel dat niet alleen beperkt blijft tot het Wetenschappelijk Onderwijs.

Binnen dit W.O. tracht de docent het leren van de studenten enigszins te sturen, zodat niet volstaan wordt met (eenzijdige) informatie-overdracht in de vorm van hoor-colleges en met te bestuderen schriftelijk materiaal. De docent geeft studenten opdrachten die niet altijd even gespecialiseerd zijn. Willems bedoelt hiermee dat deze opdrachten nauwelijks enige indicatie geven voor de activiteiten die van de studenten worden verwacht. In het gunstigste geval is wel te verwachten dat studenten tot activiteiten worden aangezet, maar welke activiteiten dat zullen zijn, en tot welke leer-processen en leerprestaties die zullen leiden is onduidelijk. Hij pleit er dan ook voor om deze opdrachten nader te specificeren. Dit zijn de door hem genoemde 'studietaken': opdrachten die aan bepaalde eisen voldoen.

Hiermee geeft Willems een duidelijk onderscheid aan tussen vragen en doelstellingen als sturende gedeelten bij teksten enerzijds en opdrachten anderzijds. Doelstellingen geven aan wat studenten na het leren moeten kunnen of weten; vragen verlangen een activiteit. Studietaken, indien duidelijk gespecificeerd, daarentegen geven zowel aan *wat* er na het

leren moet worden beheerst en welke activiteiten van de studenten worden verwacht, maar ook *hoe* deze activiteiten moeten worden uitgevoerd.

Willems stelt zich middels zijn proefschrift tot doel 'na te gaan of het mogelijk is door middel van opdrachten bij studieteksten dié leeractiviteiten en leerprocessen in gang te zetten, die er uiteindelijk toe leiden dat de studenten niet alleen kennis opdoen, maar ook met deze kennis kunnen werken'.

Zijn proefschrift is als volgt opgebouwd:

Na een korte inleiding, waarin hij een model specificceert, dat de relaties tussen onderwijsmaatregelen en leeractiviteiten, -processen, -resultaten en -prestaties aangeeft en dat de rode draad door zijn proefschrift vormt, komt hij tot een nadere beschrijving van een informatieverwerkingsmodel. Dit model beschrijft de informatieverwerkingsprocessen die tijdens het leren een rol spelen en de effecten die er het gevolg van zijn. Het vormt het kader van waaruit zijn onderzoek wordt gezien. In dit hoofdstuk volgt een gestructureerde en heldere uiteenzetting van de informatieverwerkingsprocessen en -structuren en van de effecten daarvan. De beschrijving van deze processen en effecten loopt parallel aan de door hem in de inleiding onderscheiden leerprocessen, leerresultaten en -prestaties. Jammer vind ik dat de 'leeractiviteiten' (althoewel geen processen), die juist door studietaken nader gespecificeerd zouden moeten worden, bij deze beschrijving niet apart worden belicht, maar in de paragraaf 'leerprocessen' worden ondergeschoven.

In zijn derde hoofdstuk geeft hij een overzicht van onderzoeksresultaten met doelstellingen, vragen en studietaken als hulpmiddelen bij het bestuderen van teksten. Daarbij maakt hij een onderscheid in vier verschillende typen van studie-taken, die verschillen in de soort van activiteit die voor een studietaak moet worden ondernomen. Hij besluit, op basis van het grootst verwachte effect van elk van deze studietaken op verschillende typen leerprestaties en op basis van de doelstelling van het W.O., allereerst de effecten van far-transfertaken bij studietaken te onderzoeken. Dit zijn taken waarbij de student 'op basis van gegeven informatie regels moet ontwerpen om in een ander werkelijkheidsdomein te kunnen handelen'. Deze taken zouden naar verwachting een positief effect hebben op het produktieve denken (inzicht en probleem-oplossen).

In het vierde hoofdstuk beschrijft hij het onderzoek, waarin hij de effecten van studietaken bij far-transfertaken nader onderzoekt. De resultaten laten zien dat er geen verschil is in de aard van de leeractiviteiten, in de gemeten leerprocessen en in de leerresultaten tussen de groepen met en de groepen zonder studietaken. Wel blijken er verschillen in leerprestaties ten gunste van de controle groep te zijn. Dit effect verklaart Willems door de moeilijkheidsgraad van de taken, de normale studiestrategie van de student en de perceptie van de taak. Vanuit dit onderzoek formuleert hij een aantal randvoorwaarden waaraan moet worden voldaan om de kans op positieve effecten van studietaken te optimaliseren. Vervolgens gaat Willems nader in op het ontwerpen van een module en op de plaats van studietaken daarbinnen.

In het zesde hoofdstuk worden vervolgens een drietal vooronderzoeken beschreven, waarin de module op de kwaliteiten wordt geëvalueerd en aangepast. Uiteindelijk leidt dit tot een module en tot verschillende materialen, bestaande uit: een inleiding, waarin met name de doelstelling wordt omschreven; een oplosmethode voor inzichtsvragen, bestaande uit een vijftal stappen van een heuristisch karakter; studieteksten; studietaken en toetsen. Bij deze studietaken wordt verlangd, dat de student een redeneeractiviteit uitvoert. Bij zo'n redeneer-taak moet de student 'op basis van de gegeven informatie nieuwe informatie produceren die tot hetzelfde domein behoort als de gegeven informatie'.

Al deze materialen worden in een experimentele setting op hun effecten getoetst.

Vanwege de mogelijkheid dat leerstijlen van studenten kunnen interfereren met het

effect van de studie-taken, besluit Willems in zijn experiment voor deze leerstijlen te controleren. In het volgende hoofdstuk wordt daarom nader op deze leerstijlen ingegaan en wordt een vragenlijst ontworpen, die bestaat uit een vijftal open vragen over inzicht, toepassing, begrijpen en goed onderwijs. Deze vragenlijst is vervolgens onderzocht op betrouwbaarheid en validiteit.

In het laatste onderzoek worden vervolgens de verschillen in effecten tussen drie condities onderzocht. Geconcludeerd wordt dat het mogelijk is om met behulp van studietaken het leren van de studenten zo te sturen dat er dat geleerd wordt wat is beoogd. Deze studie-taken moeten volgens Willems wel aan een aantal voorwaarden voldoen; het aantal studietaken mag niet al te groot zijn, er moet een systematische opbouw in studietaken worden aangebracht, de studietaken moeten worden ingepast in het geheel van onderwijsmaatregelen, er moet sprake zijn van een adequate terugkoppeling en studietaken moeten de student de mogelijkheid bieden om het repertoire van metacognities uit te breiden.

De beschreven studietaken, aan de hand waarvan studenten studieteksten bestuderen, zijn te beschouwen als procedures waarmee de cursist wordt ondersteund bij het zelfstandig verwerken van de leerstof. Deze procedures sluiten sterk aan bij ondermeer de handelingstheoretische procedure voor het oplossen van problemen van Mettes en Pilot.

Het werken aan de hand van concrete opdrachten heeft ook veel gelijkenissen met het probleem-gestuurd onderwijs, zoals dat aan de RijksUniversiteit van Maastricht en in de studierichting Personeelswetenschappen van de Katholieke Universiteit Brabant wordt gehanteerd. Gelijkenissen bestaan in het bestuderen van de relevante lesstof aan de hand van opdrachten of probleemsituaties. Verschillen bestaan in de aard van deze opdrachten of problemen en in de aard van de voorgeschreven oplossingsmethodieken.

De oplosmethode die de cursist tijdens zo'n studietaak voorgelegd krijgt is van een algemene aard en bevat heuristische aanwijzingen over de activiteiten die de cursist zou kunnen uitvoeren. Tijdens de verschillende studie-taken wordt de cursist ook om deze activiteiten gevraagd. Daarmee is binnen de methode duidelijk aangegeven *hoe* de student bij de verschillende stappen te werk moet gaan en welke activiteiten van hem worden verwacht.

Binnen het probleem-gestuurd onderwijs daarentegen zijn naar mijn idee deze activiteiten veel minder duidelijk gespecificeerd. Dit betekent dat er in de methodiek van Willems een duidelijkere sturing vanuit de onderwijsmaatregel aanwezig is, die 'als het ware de metacognities die de studenten zelf niet bezitten vervangen'. Door deze compensatie zouden de doelstellingen van het W.O., die te maken hebben met het oplossen van wetenschappelijke en maatschappelijke problemen, ook voor de groep van lerenden, die deze metacognities niet bezitten, kunnen worden gerealiseerd. Juist in een maatschappelijke context, waarin inzichten en probleem-oplossende vaardigheden belangrijk zijn, kunnen studietaken daarbij een wezenlijke bijdrage leveren. Tevens verwacht ik, dat het toepassingsgebied niet alleen tot het W.O. beperkt hoeft te blijven, maar dat het ook voor andere opleidingsgebieden, waaronder de bedrijfsopleidingen, relevant zou kunnen zijn.

Maar dat is nog niet alles; de methodiek van Willems bevat naar mijn idee nog meer mogelijkheden.

De externe sturing die middels de studiemethodiek van Willems wordt aangereikt, zou, ook naar zijn zeggen, mogelijk kunnen interfereren met de geprefereerde studie-strategie (leerstijl) van de cursisten. Dit betekent dat de noodzaak tot differentiatie naar leerstijlen van de cursisten van zeer groot belang is. In de toekomst zou kunnen worden nagegaan op welke manier dat binnen de 'studietaken-methodiek' mogelijk is.

Ten tweede is het naar mijn idee, zeker voor de doelgroep van Willems, van belang dat

studenten in de loop van de gehele universitaire opleiding, steeds meer worden gestimuleerd het eigen leerproces te sturen. Dit betekent dat het vragen naar steeds complexere leeractiviteiten, zoals bij de studietaken gebeurt, niet voldoende is. Het louter aanreiken van 'metacognities' zal niet bij iedere student betekenen, dat hij de aangereikte externe sturing automatisch gaat overnemen. Geïntegreerd studievaardigheidsonderwijs vereist niet alleen het activeren van leeractiviteiten en -processen, maar ook een soort bewustwordingsproces van de relevantie ervan. Daarnaast moet de student in de gelegenheid worden gesteld om zelf te beslissen welke leeractiviteiten en -processen hij in bepaalde situaties prefereert. Dit onderstreept wederom de noodzaak tot differentiëren. Bezien zou moeten worden op welke wijze deze aspecten het best binnen de methodiek gerealiseerd zouden kunnen worden.

Beide hiervoor genoemde punten zou ik louter willen typeren als ideeën, die binnen de studie-taken van Willems zeer goed mogelijk zijn. Op die manier ontstaat een differentiërende onderwijs- of opleidingsmethodiek, die in staat is tekorten in (meta-)cognitieve vaardigheden van cursisten op te vangen en zo mogelijk te stimuleren. Zo'n opleidingsmethodiek verdient in een context waarin zelfstandig probleem oplossen een belangrijke vereiste vormt, een (nog grotere) prioriteit.

Dit alles betekent tot slot, dat het proefschrift van Willems voor velen interessant zou kunnen zijn.

Voor onderzoekers en wetenschappers is het proefschrift voornamelijk interessant doordat het ideeën aanreikt over de wijze waarop zelfstudie kan worden bevorderd. Tevens geeft het aan de hand van het informatieverwerkingsmodel een heldere theoretische visie van de relaties tussen onderwijsmaatregelen, studentkenmerken, leeractiviteiten, -processen, -resultaten en -prestaties.

Toekomstig onderzoek zou kunnen nagaan in hoeverre differentiatie naar leerstijlen en studievaardigheden binnen de methode mogelijk is en in hoeverre het zelfstandig leren met de methode daadwerkelijk kan worden gestimuleerd. Deze bevindingen zijn ook voor de praktijksituatie van alle dag belangrijk.

Ik verwacht dat voor onderwijsontwikkelaars, onderwijsdeskundigen en onderwijsbegeleiders in de opleidingspraktijk vooral de hoofdstukken drie en vijf en het laatste hoofdstuk van het proefschrift interessant zullen zijn. Zij vinden er een beschrijving van studiepakketten, waarin zelfstandige leeractiviteiten een centrale plaats innemen. Zulke pakketten verdienen in een opleidingssituatie, waarin zelfstudie een belangrijke rol speelt, prioriteit. Het gebruik van deze studiepakketten hoeft daarbij niet noodzakelijk tot het Wetenschappelijk Onderwijs beperkt te blijven: te denken valt zeker ook aan bijvoorbeeld de opleidingspraktijk in het bedrijfsleven, de Open Universiteit en de Volwassenen Educatie.

Christa C.J. Teurlings
Sectie Onderwijs- en Opleidingspsychologie
Katholieke Universiteit Brabant.
Postbus 90153
5000 LE Tilburg

M. Sprangers, *Response shift and the retrospective pretest: on the usefulness of retrospective pretest - posttest designs in detecting training related response shifts*. (Proefschrift Universiteit van Amsterdam). 's-Gravenhage: SVO, 1989 (Selecta reeks).

Zoals de titel aangeeft, behandelt het proefschrift van Miriam Sprangers het probleem van de zgn. "response shift" en een mogelijke remedie, de retrospectieve voormeting. Het probleem doet zich voor in pretest-posttest (control group) designs, waarbij de afhankelijke variabele een "self-report" van de proefpersoon is, eventueel naast een objectievere maat.

Het inleidend hoofdstuk geeft een korte en heldere schets van het probleem van de "response shift": pretest-posttest (control group) designs voor de evaluatie van een training of behandeling, in combinatie met een "self-report" score als afhankelijke variabele, gaan ervan uit dat de maatstaf waarmee een subject zichzelf meet, niet verandert tijdens de behandeling. Echter, juist indien een behandeling effectief is, valt te verwachten dat deze aanname ongeldig is. Zo kan een goede training in studievoordigheden de cursisten (a) een beter begrip geven van wat hen nu eigenlijk gevraagd wordt zelf te beoordelen, en (b) hen er van bewust maken wat er tekort schiet aan hun vaardigheden. Doet een dergelijke "response shift" zich voor, dan kan dat het effect van de behandeling maskeren¹: Een vergelijking van pre- en posttest scores in de experimentele groep gaat mank doordat de subjecten zichzelf tijdens de pretest soepeler beoordelen dan tijdens de posttest. Een vergelijking tussen de posttest scores van experimentele en controle groep is eveneens misleidend, want de "shift" is het gevolg van de behandeling en treedt dus niet op in de controle groep. Deze zal zichzelf dan ook op de posttest soepeler beoordelen dan de experimentele groep.

Een in 1979 voorgestelde en onderzochte methode om deze "response shift" te detecteren en onder controle te houden, is de zgn. "retrospectieve voormeting", door Sprangers aangeduid als "thentest" of "retrotest": tegelijk met de nameting wordt aan de proefpersonen gevraagd, aan te geven hoe ze vóór de behandeling waren. Vergelijking van retrotest met pretest scores geeft een schatting van de "response shift" en vergelijking van retrotest met posttest scores een schatting van het effect van de behandeling, tenminste: onder de aanname dat de retrotest scores op hun beurt niet vertekend zijn. Een van de controles hierop is het gelijktijdige gebruik van een tweede, objectievere, afhankelijke variabele. In het geval van een studievoordigheidstraining kan dat een studievoordigheidstoets zijn. Indien het verschil tussen posttest en retrotest self-report meer overeenstemt met het objectieve effect - verschil tussen voormeting en nameting op de toets - dan het verschil tussen posttest en pretest self-report, steunt dit de retrotest scores tegenover de pretest. Enkele oudere onderzoeken suggereren dat "response shift" optreedt en dat een retrospectieve voormeting een bruikbare oplossing hiervoor is.

Het werk van Sprangers had als doel het optreden van "response shifts" te repliceren, en nader te onderzoeken onder welke omstandigheden pretest en retrotest scores divergeren en of de laatste een remedie is tegen deze "shift". Daarbij komen ook drie alternatieve verklaringen voor verschillen tussen pretest en retrotest scores aan bod, die erop neer komen dat de retrotest self-reports vertekend zijn door "demand characteristics", sociale wenselijkheids- en geheugeneffecten. Door handige experimentatie, o.m. placebo's, worden deze alternatieven getoetst ten opzichte van de "response shift" verklaring. De

¹ Onder "response shift" worden niet gerekend de welbekende hertest- en rijpingseffecten die onafhankelijk van de behandeling in zowel experimentele als controle groep op kunnen treden en juist met behulp van een controle groep onder controle worden gehouden.

experimenten betreffen de evaluatie van het effect van studie-, communicatie- en andere vaardigheidstrainingen.

De eerste experimenten (hfd. 2,3) geven evidentie voor een "response shift" en voor de retrotest als remedie: objectievere scores tonen een trainingseffect in overeenstemming met de posttest minus retrotest self-reports, en in tegenstelling tot de posttest minus pretest self-reports. Uit verdere experimentatie (hfd. 4) blijkt o.m., dat indien een objectievere vaardigheidstest vooraf gaat aan de "self-report" pretest, geen "response shift" optreedt. Dit bevestigt het vermoeden dat de shift het gevolg is van zelfoverschatting van de proefpersonen op de pretest, doordat zij dan nog geen extern criterium hebben om zichzelf aan af te meten.

Na een onderbreking van drie hoofdstukken, waarover zometeen meer, gaan de hoofdstukken 8 en 9 hierop door. Gedemonstreerd wordt, dat te hoge pretest scores en de "response shift" ook kunnen worden voorkomen door een "bogus pipeline" techniek, i.e. door de proefpersonen voorafgaand aan de pretest te suggereren dat de validiteit van hun self-reports kan worden getoetst aan objectieve maten. Dit alles suggereert meteen een potentieel alternatief voor de retrotest: voorkom "bias" in de pretest self-reports door een objectieve pretest en/of de "bogus pipeline".

In de hoofdstukken 5 en 6 wordt onderzocht in hoeverre de discrepantie tussen pretest en retrotest self-report scores te wijten is aan vertekening van de retrotest scores door "demand characteristics" en "effort justification", in plaats van aan "bias" in de pretest. Bij toevoeging van een placebo-konditie aan het design (het geven van een irrelevante training), blijkt dat niet alleen in de experimentele groep, maar ook in de placebo-groep de post- minus retrotest self-report scores een verbetering van de vaardigheid suggereren, terwijl alleen in de experimentele groep een objectieve verbetering optreedt. Nadere inspectie leert dat dit placebo-effekt in één experiment (hfd. 5) zit in te lage retrotest scores, en in twee experimenten (hfd. 6) in te hoge posttest en "direct improvement" self-reports. Iets soortgelijks trad op in hoofdstuk 3 waar de controlegroep objectief geen, maar subjectief een even groot, post- minus pretest score verschil vertoonde als de experimentele groep, ofschoon daar geen sprake was van een placebo. Uit een literatuurstudie (hfd. 7) konkludeert de auteur uiteindelijk dat zowel "response shift" als "subject bias" optreden.

De auteur lijkt hier over het hoofd te zien dat twee van haar drie experimenten (nl. hfd. 6) een placebo-effekt tonen in de posttest in plaats van in de retrotest scores. Deze laatste blijken niet lager te zijn dan de pretest scores in de placebo-groepen, maar alleen in de experimentele groepen. Dit ondermijnt (a) de "subject bias" verklaring als concurrent van de "response shift" verklaring voor het in hfd. 1-4 gevonden verschil tussen pretest en retrotest scores, en (b) het gebruik van de posttest self-reports, of die nu wordt vergeleken met pretest of retrotest scores.

Wat verder opvalt bij vergelijking van de tabellen in hfd. 6 is, dat een objectieve posttest van de criterium-vaardigheid voorafgaand aan de posttest self-reports en "direct improvement" self-reports, de "bias" in deze laatste twee wel iets lijkt te reduceren, maar niet elimineert. Immers, eerder bleek een objectieve pretest de "bias" in de pretest self-reports te kunnen voorkomen (hfd. 4).

Het slothoofdstuk zet alles nog eens op een rijtje en gaat beknopt in op voor- en nadelen van self-reports ten opzichte van objectievere maten en van de retrotest ten opzichte van de gewone pretest self-reports. De auteur neemt een tamelijk genuanceerd middenstandpunt in en beveelt de retrotest aan als aanvulling op, niet als substituut voor, hetzij de pretest self-reports, hetzij objectievere maten.

Over het algemeen was de dissertatie van Sprangers een aangename en nuttige ontmoeting. De inhoud overtuigt sceptici van het nut van een retrotest naast een pretest, en suggereert leuke toepassingen. De schrijfstijl was zonder meer prettig. Naast deze lof uiteraard ook een paar kritische nootjes.

Allereerst zouden een paar bijlagen met de letterlijke inhoud van de gebruikte self-report meetinstrumenten zeer welkom zijn. Het ontbreken van gegevens over de inhouds- en inhoudelijke validiteit hiervan steekt wat af bij de gedetailleerde behandeling van de met die instrumenten gemeten experimentele effecten.

Verder zijn er enkele data die de bruikbaarheid van post minus retro verschilcores relativeren: hoewel deze scores net als de objectieve maten een experimenteel effect in termen van groepsgemiddelden laten zien, zijn (a) de trainingseffecten in termen van de self-report scores meestal (hfd. 4-8) veel kleiner dan de objectieve effecten, en (b) de korrelaties tussen self-reported en objectieve verschilcores zeer laag (hfd. 3-4).

Het eerste is wellicht te wijten aan te ongevoelige self-report instrumenten. De oorzaak van het tweede ligt vermoedelijk in een lage betrouwbaarheid van de self-report en/of objectieve verschilcores. Op basis van de gerapporteerde betrouwbaarheden van de self-report scores en korrelaties tussen post- en retrotest scores, valt te ramen dat de betrouwbaarheid van de verschilcores tussen .15 en .30 zal liggen. Een lage korrelatie tussen self-reported en objectieve verschilcores mag dan ook niet verwonderen, zeker niet als deze laatste ook behept zijn met 'error'.

Uiteindelijk blijft ondergetekende een beetje zitten met de vraag, in hoeverre "self-reports" bruikbaar zijn als instrument om het effect van een training of behandeling te meten. De traditionele pre-post self-report vergelijking en de vergelijking tussen experimentele en controle groep wat betreft posttest self-report, gaan mank aan het "response shift" fenomeen. De retrospectieve voormeting biedt hulp, maar lijdt zelf ook aan vertekening (hfd. 5).

Beide methoden, de pre-post en de retro-post vergelijking worden dan ook steeds gevalideerd aan objectievere maten voor verandering. Hoewel de auteur in het slothoofdstuk opmerkt dat ook objectieve maten zo hun nadelen hebben, wordt hierop te weinig ingegaan om de scepticus te overtuigen. Juist in de onderwijswereld waar de experimenten zich op richten, lijkt me dat redelijk betrouwbare en valide objectievere maten te konstrueren zijn, vooral als het criterium-gedrag van sterk cognitieve aard is.

Maar deze dissertatie richt zich dan ook niet op de vergelijking tussen self-reports en objectievere maten, maar op die tussen pre-test en retro-test self-report. En in dat opzicht is het onderzoek aardig geslaagd. Minstens één scepticus is bekeerd tot een positievere instelling ten aanzien van retrospectieve tests.

Verder lijken "self-reports" en retrotests zeker in de wereld van onderwijs en vaardigheidstrainingen wel zinvol, ook als er een objectievere maat voorhanden is. Het is voor de cursusleiding immers prettig om te weten, dat niet alleen objectief, maar ook in de waarneming van de cursisten, de training een positief effect heeft gehad. Want daar hangt toch van af of er een markt voor de cursus zal zijn.

Gerard J. P. van Breukelen
K. U. Nijmegen
Vakgroep Mathematische Psychologie
Postbus 9104
6500 HE Nijmegen



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

Greetje P.C. van der Werf, RION,
Postbus 1286, 9701 BG Groningen,
tel. 050-636657

Jrg. 14, nr. 4

Oktober 1990

Inhoudsopgave

Verslag VOR-ledenvergadering d.d. 10 mei 1990	252
Congressen, Symposia, Cursussen, Trainingen	254
Ontvangen publikaties	255

Verslag VOR-ledenvergadering d.d. 10 mei 1990

Aanwezig: Van Batenburg, Claessen, Dronkers, Hoebe, Van Hout, Leune, F. van der Linden, W. van der Linden, Mooij, Post, Roede, Rijkers, Simons, Span, Stijnen, Van der Sijde, Taconis, De Tempe, Tillema, Van der Werf

1. Mededelingen

Na opening van de vergadering door de voorzitter wordt een wijziging in de agenda aangebracht: agendapunt 5 (jaarrekening 1989) wordt verplaatst naar de vergadering van november; toegevoegd wordt het agendapunt 'nieuwe bestuursleden'.

2. Ingekomen stukken

Geen.

3. Verslag vorige vergadering

Geen opmerkingen.

4. Verslag verenigingsjaar 1989

Geen opmerkingen.

5. Jaarrekening 1989

Vervalt

6. Gedragscode en reglement van toezicht en beroep

In de ledenvergadering van december 1989 is de gedragscode door de leden vastgesteld. Er is toen aangedrongen op het instellen van een Kamer van Toezicht en een College van Beroep. Het reglement van deze commissies wordt nu ter goedkeuring aangeboden aan de leden. Wanneer dit reglement wordt geaccepteerd, zal in de ledenvergadering van november een voordracht worden gedaan voor de samenstelling van de commissies. Het reglement wordt bij acclamatie door de ledenvergadering goedgekeurd. Dronkers merkt nog op dat het bij de behandeling van zaken door de Kamer van Toezicht niet altijd strikt om een klacht hoeft te gaan. De Kamer kan zich ook buigen over vragen die meer betrekking hebben op een advies. Het bestuur beaamt dit.

7. Toekomstige structuur VOR

Het VOR-bestuur is voornemens om de VOR intern te structureren op AERA-achtige wijze. Daartoe denkt het bestuur aan het instellen van werkgroepen/divisies. In samenwerking met de OTG-besturen wil het nagaan hoe deze werkgroepen vorm kunnen krijgen. Eventueel zal de organisatie van de ORD inhoudelijk daarbij worden aangesloten.

De ledenvergadering wijst erop dat niet alleen aansluiting gezocht moet worden bij de OTG's. Er zijn ook andere bestaande groepen. Aansluiting bij een SVO- of NWO-structuur ligt niet voor de hand; dit zijn subsidie-structuren. De VOR moet daar los van blijven staan. Besloten wordt dat het VOR-bestuur eerst de functie en taken van de werkgroepen/divisies zal uitwerken. Mogelijkheden zijn: netwerkfunctie, publikaties, stimuleren van SIG's e.d. Voor de vorming van de werkgroepen zal vooral worden

uitgegaan van de AERA en OTG-structuur. Een voorstel daartoe zal worden uitgewerkt voor de volgende ledenvergadering. Daarin worden ook de taken en functies van de werkgroepen uiteengezet.

Naar aanleiding van dit punt wordt gevraagd naar de rol van de VOR bij de reviewing van de door SVO aanbestede PSO-programma's. Het bestuur heeft hierover aan SVO meegedeeld dat hieraan dit jaar niet wordt meegewerkt, omdat de structuur hiervoor vooralsnog ontbreekt.

8. Voordracht nieuwe leden VOR-bestuur

Gerritsen v.d. Hoop heeft zich teruggetrokken uit het bestuur. Volgens het huishoudelijk reglement mag het bestuur bij tussentijdse vacatures deze zelf opvullen en achteraf aan de leden voorleggen. Het bestuur heeft besloten voor de vacature Simons te benoemen in het bestuur. Daarnaast is besloten het bestuur uit te breiden. Hiervoor is Hoebe benaderd. Beide voordrachten worden bij acclamatie door de leden goedgekeurd.

9. Rondvraag

Gevraagd wordt of het klopt dat de VOR haar doelgroep wil uitbreiden naar de beroepsgroep onderwijskundigen. Hierop wordt geantwoord dat dit klopt. De VNSU-kamer onderwijskunde heeft dit verzocht, omdat veel onderwijskundigen ongeorganiseerd zijn, maar wel behoefte hebben aan een organisatie waarbij zij zich thuis voelen. De VOR zal binnenkort bekijken wat zij deze groep en eventuele andere groepen kan bieden.

Congressen, Symposia, Cursussen, Trainingen

CONFERENTIE "VAN ANDERS NAAR BETER"

Actief inspelen op veranderingen in secundair en tertiair onderwijs

Het Instituut voor Lerarenopleiding, Onderwijsontwikkeling en Studievaardigheden (IVLOS) van de Rijksuniversiteit Utrecht organiseert op 25 januari 1991 een conferentie over onderwijsverbetering.

De doelgroepen zijn docenten, directie- en stafleden en begeleiders op scholen voor voortgezet onderwijs en instellingen voor hoger onderwijs in de regio Utrecht.

Overal in het onderwijs worden pogingen gedaan om, temidden van de vele veranderingen die vanuit het ministerie op het veld afkomen, tot verbeteringen te komen. De conferentie is bedoeld om het onderwijsveld in de gelegenheid te stellen kennis te nemen van de nieuwste inzichten over onderwijsverbetering. Daarbij staan drie invalshoeken centraal: de organisatie van onderwijs, docentgedrag en leerprocessen.

De conferentie zal worden geopend door staatssecretaris drs. J. Wallage. Daarna belichten drie inleiders de genoemde invalshoeken: prof. dr. J.H.G.I. Giesbers, dr. Th. Wubbels en prof. dr. C.F. van Parreren.

In de middag kunnen de congresdeelnemers kiezen uit twee series van parallelle workshops, waarin een aantal thema's concreet uitgewerkt wordt.

De conferentie wordt gehouden in het universiteitscentrum De Uithof.

De kosten zullen, inclusief lunch en congresbundel, f 125,-- bedragen.

Er is slechts een beperkt aantal (100) plaatsen beschikbaar.

Een brochure met inschrijfformulier kan worden aangevraagd bij het conferentiesecretariaat: Mevr. R. Gompelman, IVLOS lerarenopleiding aan de Rijksuniversiteit Utrecht, Postbus 80120, 3508 TC Utrecht, telefoon 030-532194.

Ontvangen publikaties

- Boekkooi-Timminga, E. (1989). *Models for computerized test construction*. Proefschrift Universiteit Twente. De Lier: Academisch Boeken centrum.
- Boshuizen, H.P.A. & Smidt, H.G. (1990). *The role of biomedical knowledge in clinical reasoning by experts, intermediates and novices*. Maastricht: Rijksuniversiteit Limburg (Onderzoek van Onderwijs no. 41).
- CITO (1990). Jaarverslag van het Instituut voor Toetsontwikkeling over 1989. Arnhem: CITO.
- Dijkstra, S., Hout Wolters, B.H.A.M. van & Sijde, P.C. van der, (1990). *Research on instruction: design and effects*. New Jersey, Englewood Cliffs: Educational Technology Publications.
- Eck, E. van & Volman, M. (red.) (1990). *Determinanten van seksespecifieke keuzen en prestaties*. Amsterdam: SCO.
- Hulshof, M.J.F. (1990). *Welke kwalificaties vraagt het wetenschappelijk onderwijs. Literatuurstudie voor de ARVO naar gevraagde kennis en vaardigheden in de propedeuse*. Nijmegen: IOWO.
- ITS. (1989). Jaarverslag.
- Meerum Terwogt-Kouwenhoven, K. (1990). *Niet gewogen, toch te licht bevonden. Analyse van de rendementsproblematiek aan de universiteit*. Proefschrift Universiteit van Amsterdam. Kampen: Mondiss.
- Santema, M., Maandag, D. W. & Wichers, J.E. (red.) (1989). *Onderwijs en bedrijfsleven in regionaal perspectief. Kansen en uitdagingen voor de negentiger jaren*. Groningen: Instituut voor Onderwijskunde, RUG.
- Schuurs, U. (1990). *Leren schrijven voor lezers. Het effect van drie vormen van probleemgericht schrijfonderwijs op de zinsbouwvaardigheid*. Proefschrift Universiteit Twente. Enschede: Faculteit Wijsbegeerte en Maatschappijwetenschappen (WMW-publikatie 3).
- Zwarts, M. (red.) (1990). *Balans van het taalonderwijs aan het einde van de basisschool: uitkomsten van de eerste taalpeiling einde basisonderwijs*. Arnhem: CITO (PPON-reeks nr. 2).

deel 7

Kwalificatie en curriculum

**T.F. Brandsma,
W.J. Nijhof,
J.C. Kamphorst**

Deze literatuurstudie is uitgevoerd in het kader van het onderzoek naar de achtergronden, toepassingen en effecten van de zogenoemde 'naieve' benadering van aansluitingsproblemen. De overheid streeft steeds meer naar institutionalisering van beroepsprofiel en opleidingsprofielontwikkeling, als basis voor de leerplanontwikkeling voor het voortgezet beroepsonderwijs. Van een aantal Duitse en Amerikaanse modellen, gericht op de aansluiting onderwijs - arbeidsmarkt/opleiding - beroepspraktijk, wordt in deze studie inzicht gegeven in de historische, methodologische, inhoudelijke en bruikbaarheidsaspecten. Het boek biedt daarmee een kritische reflectie op de (grootschalige) toepassing van methoden voor beroepsprofielconstructie binnen de context van het Nederlandse beroepsonderwijs.

ISBN 90 265 1093 4
f 27,50, 136 blz.

deel 8

Tekstopbouw en stelvaardigheid

**T. Jansen,
M. Overmaat**

Tekstopbouw is een belangrijk aspect van schrijfprodukten. Goede instructie over het organiseren en structureren van ideeën voor een te maken tekst is dus noodzakelijk. De auteurs onderzochten een procesmethode die uitgaat van mentale processen tijdens het schrijven en vergeleken die met een modelmethode die werkt met het lezen en bestuderen van modelteksten. Gekeken werd naar effecten op prestatie en attituden van leerlingen, op cognitieve processen en op strategieën die leerlingen beheersen. Deze stelmethoden bereiken zeer gunstige resultaten in een relatief korte periode. Schrijven kun je leren! Een belangrijk onderzoek voor de theorieontwikkeling over het stelonderwijs en een verrijking voor de didactische praktijk.

ISBN 90 265 1100 0
f 29,50, 156 blz.

In de boekhandel

S&Z

SWETS & ZEITLINGER B.V.

Heereweg 347, 2161 CA Lisse, Tel. 02521-35111

Multi-niveau onderzoek: uitgangspunten en toepassingen

Voorwoord

P. van den Eeden en G.W. Meijnen

Vakgroep Methoden en Technieken van Sociaal-Wetenschappelijk Onderzoek, Vrije Universiteit en Vakgroep Onderwijskunde, Universiteit van Amsterdam

Het onderzoek, waarin behalve leerlingen óók klassen en scholen voorkomen, is zo oud als het onderwijsonderzoek. Toch heeft het lange tijd geduurd voordat de onderzoekers vragen gingen stellen over hoe een dergelijk onderzoek dient te worden opgezet en welke valkuilen zich daarbij kunnen voordoen. Deze ontwikkeling heeft geleid tot het uitkristalliseren van een aparte onderzoeksopzet voor het gecombineerd bestuderen van leerlingen, klassen en scholen. Deze opzet is bekend geworden onder de naam "multi-niveau onderzoek" of ook wel "multilevel onderzoek".

Ofschoon het multi-niveau onderzoek van toepassing is op alle gebieden waar meerdere sociale niveaus van belang zijn, heeft deze onderzoeksopzet zich juist in de bestudering van het onderwijs kunnen ontwikkelen. De redenen daarvan liggen voor de hand: in het onderwijs is er een duidelijke afbakening tussen de verschillende klassen en scholen aanwezig en de klassen en scholen zijn in redelijk grote aantallen voorhanden. Een praktische bijkomstigheid is dat in het onderwijs veel onderzoek wordt verricht dat speciaal gericht is op de vraag hoe klassen en scholen zodanig kunnen worden gemanipuleerd dat de resultaten van de leerlingen worden verbeterd. De multi-niveau onderzoeksopzet blijkt zich voorts goed te lenen voor evaluatieonderzoek. De nadruk kan hierbij liggen op zowel de bevordering van gelijke onderwijskansen alsmede op de verhoging van de effectiviteit van het onderwijs als zodanig.

Nu er gedurende de laatste tien jaar belangrijke vorderingen zijn geboekt in het multi-niveau onderzoek, is het een goede gedachte om een apart themanummer samen te stellen dat gericht is op deze thematiek. We hebben gekozen voor een formule waarbij vooral aan de hand van een aantal voorbeelden wordt geïllustreerd hoe multi-niveau onderzoek in het onderwijs wordt aangepakt. Wellicht herkennen andere onderwijsonderzoekers de aangesneden problemen en kunnen zij zich richten naar geboden oplossingen.

Het themanummer opent met een overzichtsartikel over het multi-niveau onderzoek in het onderwijs. Daarin wordt aangegeven hoe het type probleemstelling eruit ziet en welke consequenties dat heeft in de diverse fasen van het onderzoek. Een kritische bespreking van enkele recent in Nederland verrichte onderzoeken wordt daaraan toegevoegd.

Vervolgens wordt in een bijdrage van Terwel en Van den Eeden een analyse gepresenteerd die gericht is op de vraag of de effectiviteit van een bepaalde wiskundendidaktiek door kenmerken van de classesituatie wordt beïnvloed en of deze invloed zich in de verschillende niveau-groepen gelijkelijk voordoet. Hier wordt de multilevel benadering toegepast op de problematiek van de "differentiële effecten".

Mooij behandelt de effectiviteit van het computergebruik in het leesonderwijs op basisscholen. Om de effecten voor de leerlingen van de aanwezigheid van een computer in de klas nauwkeurig vast te stellen, controleert hij daarbij voor mogelijk verstorende variabelen zowel op het niveau van de leerling als op dat van de klas.

De bijdrage van Roeleveld, De Jong en Koopman gaat over de invloed van scholen. Nagegaan wordt hoe effectief een school is, eveneens door controle van mogelijk relevante variabelen op het individuele en het hogere niveau (de school). Tevens proberen zij vast te stellen of de effectiviteit van een school een constant gegeven is.

Dit themanummer bevat als laatste een artikel van Bosker en Snijders over statistische aspecten van het multilevel onderzoek. Op basis van een onderzoek naar de power van de gebruikte toetsen kunnen zij enige suggesties geven over de gewenste aantallen leerlingen, klassen of scholen.

Daarnaast is een overzicht van relevante literatuur en van beschikbare computerprogramma's opgenomen.

Ontvangen 11-9-1990

Onderwijsonderzoek vanuit multi-niveau perspectief

P. van den Eeden en G.W. Meijnen

Vrije Universiteit en Universiteit van Amsterdam***

ABSTRACT

This paper aims to describe multilevel research by means of the central terms 'approach', 'theory', 'model' and 'procedure'. The diverse accents of this type of research design are stressed. Thereafter three recent Dutch multilevel studies are commented. Some remarks of expected developments of multilevel research are added.

INLEIDING

De kern van het onderwijsonderzoek is gedragswetenschappelijk van aard. Het verklaren en voorspellen van de ontwikkeling van leervorderingen van individuen, kinderen dan wel volwassenen, maken het hart uit van de onderwijskunde als wetenschappelijke discipline.

In alle geïndustrialiseerde landen is de overdracht van maatschappelijk relevante kennis en vaardigheden op een volgende generatie 'gecollectiviseerd' en geïnstitutionaliseerd in de vorm van afzonderlijke organisatorische kaders. Het leren van kinderen is daarmee voor een belangrijk deel leren in groepsverband geworden. De klas of jaargroep is daarvan de vanzelfsprekende uitdrukkingsvorm. Organisatorisch maken klassen weer deel uit van scholen of soortgelijke instellingen (universiteiten, opleidingen).

Organisaties kennen behalve een formele en informele structuur ook een eigen cultuur (opvattingen docentencorps, omgangsregels, etc.). Tezamen reguleren deze waarschijnlijk mede de gedragsalternatieven van zowel docenten als leerlingen. De leervorderingen van leerlingen zijn derhalve wellicht mede afhankelijk van de modaliteit van de context. In termen van verklaringen van waargenomen gedrag (leervorderingen) zouden deze dus niet alleen gezocht moeten worden op het basisniveau van de persoonskenmerken van de individuele leerling maar ook op een tweede, hoger, niveau van de kenmerken van de klas als leeromgeving of zelfs op dat van een derde niveau, namelijk de school. Niet uitgesloten is dat ook nog een niveau hoger, namelijk dat van de lokale gemeenschap (wijk; buurt; regio), direct of indirect, het leergedrag van een individuele leerling zou kunnen beïnvloeden.

De voorwaardelijkheid van de gekozen formuleringen omtrent de invloed van de 'hogere' niveaus wordt ingegeven door de overweging dat de fundamentele discussie over de relevantie van deze 'hogere' niveaus nog steeds niet is afgesloten. Een discussie die van oudsher wordt aangevoerd door (wetenschaps)theoretici afkomstig uit resp. de psychologie en de sociologie. Kernvraag daarbij is steeds in hoeverre de oorzaken van individueel gedrag zijn terug te voeren op een lager verklaringsniveau (het reductionisme-debat). Kan aan het 'sociale' (vgl. Durkheim) een eigenstandige, autonome verklaringsbron worden toegeschreven of is alles te herleiden tot psychologische dan wel zelfs fysieke eigenschappen van het individu?

Deze wetenschapstheoretische discussie is de laatste jaren echter meer en meer op de achtergrond geraakt en vervangen door een wellicht vruchtbaarder inzet: het ontwerpen van theoreti-

* Vakgroep Methoden en Technieken van Sociaal-Wetenschappelijk Onderzoek, Vrije Universiteit, Koningslaan 22-24, 1075 AD Amsterdam.

** Vakgroep Onderwijskunde, Universiteit van Amsterdam, Prinsengracht 227, 1015 DT Amsterdam.

sche noties over de invloed van sommige contextaspecten op individueel gedrag en het beproeven van de daaruit afgeleide hypothesen.

Het onderwijs is door zijn vormgeving een institutie bij uitstek waarbinnen deze problematiek bestudeerd kan worden. Bovendien is de maatschappelijke interesse voor de toetsing van dit type hypothesen bij voorbaat gegeven. Een antwoord op de vraag welke verschillen tussen scholen relevant zijn (en in welke mate) voor de leervorderingen van kinderen kan op veel aandacht rekenen, zowel van de kant van ouders, leerkrachten en beleidsmakers als van wetenschappers.

De toegenomen belangstelling voor multi-niveau (MN)-onderzoek is het resultaat van een proces dat in Nederland reeds ruim tien jaar gaande is. Kenmerkend voor de voortgang van MN-onderzoek is echter dat met betrekking tot de methodisch-technische problemen die ermee gepaard gaan meer vooruitgang wordt geboekt dan op het terrein van theorievorming. Dat heeft ertoe geleid dat met steeds grotere precisie de variantiebronnen worden vastgesteld (individu, groep, interactie individu/groep) van het vertoonde gedrag, maar dat we over de processen die daaraan ten grondslag liggen nog grotendeels in het ongewisse zijn gebleven. Nederlandse onderzoekers nemen, ook internationaal, een vooraanstaande plaats in wat het MN-onderzoek betreft. De reden daarvan kan voor een gedeelte worden gezocht in de activiteiten van een speciale daarop gerichte onderzoeksgroep van de SoMO: de 'multilevel-groep' (Van den Eeden en Hüttner, 1984; Oosthoek en Van den Eeden, 1986; Van den Eeden en Hauer, 1986). Als voorbeelden van MN-onderzoeken in het onderwijs kunnen gelden Mensen en Guldemond (1987), Oosthoek (1985), Hüttner (1987), Koehler en Van den Eeden (1987), Bosker en Van der Velden (1988), Mooij (1988), Kreft (1988), Kreft en De Leeuw (1987) en Roeleveld, Koopman, De Jong en Van den Eeden (1989).

Hoewel de MN-benadering dus geenszins nieuw is, is een aantal ideeën daaromtrent nu dusdanig uitgekristalliseerd dat met behulp van enkele termen de uitgangspunten van MN-onderzoek redelijk zijn te structureren. In aanmerking komt een viertal termen, die min of meer in elkaars verlengde liggen. Het gaat hierbij om de termen 'benadering' (verwijzend naar de grondnotie van een onderzoek, ook 'paradigma' of 'visie' genoemd), 'theorie' (verwijzend naar een stelsel van uitspraken met de pretentie voor gevonden gegevens een verklaring te bieden, dat voorafgaand aan een onderzoek reeds aanwezig is of tijdens het onderzoek moet worden ontwikkeld), 'model' (een formele uitdrukking van de theorie) en 'procedure' (de technische uitwerking van het model ter analyse van een of meer datasets).

VIER CENTRALE TERMEN

In het navolgende wordt kort ingegaan op de invulling van deze vier termen bij het MN-onderzoek, en wordt de relevantie van de invulling aangegeven.

Benadering

De eerste term is 'benadering'. Het gaat hier om een duidelijk omliggende notie die ten grondslag ligt aan een onderzoek en die in het geval van het MN-onderzoek kenmerken van een anti-reductionistisch onderzoeksprogramma heeft gekregen. Een benadering weerspiegelt dikwijls een bepaalde maatschappij- en mensvisie of een bepaalde verklaringsstructuur. Voor de MN-benadering zijn de volgende karakteristieken van belang.

1. De samenleving is opgebouwd uit sociale categorieën die hiërarchisch ten opzichte van elkaar zijn gestructureerd. De grenzen tussen de niveaus variëren van scherp tot vloeiend. In de meeste scholen behoort een leerling eenduidig tot een klas en die klas eenduidig tot een school. Soms is afbakening problematischer, zoals bij een leerling die tot meerdere groepen behoort of bij leerlingen uit diverse klassen die éénzelfde keuzevak volgen. In weer andere gevallen is er sprake van overlapping, zoals bij een indeling van leerlingen naar school en naar wijk van herkomst.

2. Elke hiërarchische laag heeft een relatieve autonomie van functioneren. In zoverre er sprake is van autonomie van processen per niveau gelden daarmee autonome verklaringen. Zo worden op het individuele niveau, nl. dat van de leerlingen, leervorderingen in verband gebracht met motivatie en intelligentie. Op klassenniveau daarentegen worden beoordelings- en instructieprocessen door de docent in verband gebracht met de sociale samenstelling en het gemiddelde prestatieniveau van de groep leerlingen. Bij scholen wordt het beheer en het sociale klimaat in samenhang gezien met de omvang en de samenstelling van het docentencorps. De processen van een der niveaus kunnen niet herleid worden tot een ander. Een dergelijke autonomie blijkt uit de noodzaak om de variabelen van onderzoek niveau-specifiek te definiëren. Zo is 'intelligentie' van een leerling iets geheel anders dan 'intelligentie' van een klas. Wanneer een dergelijke niveau-specifieke definitie afwezig is, spreekt men van de 'semantische fout'. Meestal staat een onderzoeker bij deze definitiekwesties niet stil en worden de betekenissen van variabelen van verschillende niveaus door elkaar gebruikt. Soms is deze verwarring het gevolg van aggregatie: men berekent dan bijvoorbeeld gemiddelden van individuele kenmerken per klas en men behoudt dan de betekenis van de individuele kenmerken (De Glopper, 1988; Van der Wolf, 1984; Van Marwijk Kooij, 1984). Om dergelijke verwarring te voorkomen is het raadzaam dat een onderzoeker zijn/haar gegevens verzamelt op het niveau waarop de variabelen zijn gedefinieerd.

Het uitgangspunt van de relatieve autonomie van niveaus houdt verder in dat de verklaringen ook dienen plaats te vinden op de niveaus waar de processen worden geconstateerd. Bij een individualistische verklaring reduceert men bijvoorbeeld klasprocessen tot processen die zich op het niveau van een leerling afspelen. Omgekeerd tracht men wel individuele processen op het niveau van de klas te verklaren. Indien processen van meerdere niveaus in een onderzoek worden betrokken dienen de verschillende niveaus in de structuur van de verklaring terug te komen. Een dergelijk anti-reductionistisch standpunt verwerpt één-niveau verklaringen van MN-processen.

3. Complementair aan het uitgangspunt van de autonomie van niveaus is het uitgangspunt van de relatieve afhankelijkheid. Het is gebruikelijk bij verklaringen tevens aandacht te besteden aan de lagere niveaus waarvan de effecten op de hogere niveaus doorwerken (zoals het gedrag van de leerlingen van een klas doorwerkt in het functioneren van die klas). Het is ook gebruikelijk bij verklaringen tevens aandacht te besteden aan de hogere niveaus die als randvoorwaarden dienen voor processen op lagere niveaus (zo is de samenstelling van een klas een randvoorwaarde voor het leerproces van de individuele leerling). Voor het laatste geval heeft Lindenberg (1976) de aanpak systematisch uitgewerkt. Deze bestaat erin dat eerst op het individuele niveau naar verklaringen wordt gezocht. Vervolgens worden de individuele verklaringen getransformeerd naar het groepsniveau onder vermelding van de collectieve aanvangsvoorwaarden.

De afhankelijkheid tussen de verschillende niveaus kan in een research-probleemstelling worden gespecificeerd. Die afhankelijkheid is er al bij het definiëren van de eenheden en van de variabelen (vgl. verhouding leerling-klas). Van meer belang is de afhankelijkheid in termen van een veronderstelde causale relatie tussen de variabelen van de verschillende niveaus. Het is niet voldoende alleen maar in het algemeen te veronderstellen dat een hoger niveau op een lager niveau effect uitoefent. Noodzakelijk is ten minste de specificatie van een variabele van het lagere niveau. Deze noodzakelijke specificatie van de lager-niveau variabele bestaat in het aangeven dat die verschillen betrekking hebben op bijvoorbeeld de studieprestatie van de leerlingen.

Een betere specificatie wordt verkregen door òók de relevante hoger-niveau variabele aan te geven. In ons voorbeeld zou men kunnen aannemen dat de grootte van een klas invloed uitoefent op de studieprestaties van de leerlingen. In dit voorbeeld is 'de klas' het hogere niveau, 'de leerling' het lagere niveau, 'grootte' een variabele van het hogere niveau en 'studieprestatie' een variabele van het lagere niveau. Door aldus de niveaus en hun variabelen te specificeren wordt de structuur van een MN-probleemstelling aangegeven. In Nederland zijn er tal van

studies te vinden waarvan de probleemstelling een dergelijke structuur vertoont. Enkele voorbeelden zijn: Meijnen (1984), Van Marwijk Kooij (1984), Mooij (1987) en De Glopper (1988).

Het gegeven voorbeeld is uitgegaan van de veel voorkomende aanname dat een variabele van het hogere niveau invloed uitoefent op een variabele van het lagere niveau. Deze aanname is tamelijk willekeurig; ook het omgekeerde is mogelijk. Het onderscheid in 'contextuele' en 'emergente' verklaringswijzen komt daarmee overeen. Een contextuele verklaringswijze legt de onafhankelijke variabelen op het hogere niveau en de afhankelijke variabele op het lagere. Bij de emergente verklaringswijze gebeurt dat juist omgekeerd. Het is in zekere zin een traditie, en niet alleen in Nederland, vrijwel uitsluitend de contextuele verklaringswijze te hanteren. In wezen gaat het hier om een theoretische traditie.

De MN-benadering kent derhalve de volgende stappen. Bij de formulering van een probleemstelling op het onderhavige gebied dienen de verschillende niveaus door de onderzoeker als het ware worden "losgepeld". Vervolgens is het van belang op elk niveau de relevante variabelen te onderscheiden. Deze variabelen behoren exclusief bij de verschillende niveaus. Zo hoort 'studieprestatie' bij een leerling en 'instructiestijl' bij een klas. Verder moet de causale ordening van de niveaus worden aangegeven; een dergelijke keuze hangt af van de gehanteerde theorie.

Theorie

De tweede term waarmee het MN-onderzoek kan worden gekarakteriseerd is 'theorie'. In dit verband verstaan we daar onder een stelsel van uitspraken ter verklaring van empirisch vastgestelde feiten. De elementen van een theorie zijn begrippen, die verwijzen naar niveau-specifieke variabelen. Begrippen zijn weer in relatie met andere begrippen gedefinieerd en als zodanig vormen de uitspraken over die relaties evenzeer de elementen van een theorie. Tesaamen vormen die uitspraken een theoriernetwerk. Een belangrijke vraag is hoe in het MN-onderzoek een theoriernetwerk eruit ziet.

In een theoriernetwerk komen de diverse niveaus en hun variabelen voor. De elementaire relaties zijn daar complex van aard: ze vormen geen binaire relaties, maar relaties waarbij drie variabelen betrokken zijn: twee variabelen van het lagere niveau, waartussen een proces wordt verondersteld, en een derde variabele bevindt zich op het hogere niveau die een conditie vormt voor het op het lagere niveau veronderstelde proces. Een voorbeeld van een MN-theorie is de uitspraak "Het instructietype toegepast in een klas (de conditie-variabele van het hogere niveau) bevordert de invloed van de motivatie van een leerling (de onafhankelijke variabele van het lagere niveau) op zijn of haar studieprestatie (de afhankelijke variabele van het lagere niveau).

In het onderzoeksonderzoek zijn uitspraken met een dergelijke structuur niet nieuw: de bekende onderzoeken naar aptitude-treatment-interactie (ATI) zijn daarop gebaseerd. Het lijkt echter dat men er tot heden niet in is geslaagd om behalve de formele notie van conditionering een min of meer algemene inhoudelijke theorie te formuleren. Veel MN-onderzoek lijkt beschrijvend van aard te zijn, zonder dat de mechanismen die tussen de niveaus werkzaam zijn worden benoemd. Toch is er een aantal gunstige uitzonderingen te noemen. Eén zo'n uitzondering is de moderne versie van de theorie van de leeromgeving (Dar en Resh, 1986). Een andere poging van het aangeven van tussenniveau mechanismen is de oorspronkelijke formulering van de normatieve en comparatieve referentiegroep (Merton, 1957). Recentelijk heeft De Vos (1986) een variant van de nutstheorie ontwikkeld die in dit geval eveneens vruchtbaar lijkt. Daarnaast is er de endogene feedback-benadering (Mensen en Guldemon, 1987), die enige specificaties oplevert over de mechanismen van sociale beïnvloeding in een klas. Hoe nuttig dergelijke aanzetten ook zijn, de precieze mechanismen blijven nog betrekkelijk onderbelicht.

Kortom, het MN-onderzoek kan worden gekarakteriseerd aan de hand van de structuur van de gebruikte theorie. Kenmerkend is het voorkomen van uitspraken met een afhankelijke variabele van het lagere niveau en twee onafhankelijke variabelen: een van het hogere niveau en een van het lagere niveau. Een dergelijk gestructureerde theorie blijft vrijblijvend wanneer die niet aan data wordt getoetst. Voorafgaand aan een dergelijke toetsing zal er een geschikt model van analyse moeten worden gekozen.

Model

Daarmee is de derde term genoemd die hier wordt gehanteerd ter karakterisering van MN-onderzoek: de term "model". Een model is een geformaliseerde uitdrukking van de begrippen en hun onderlinge relaties van een theorie (Bertels en Nauta, 1974). Een dergelijke uitdrukking neemt doorgaans de vorm aan van een stelsel van vergelijkingen. Bij het type theorie gehanteerd bij het beantwoorden van MN-probleemstellingen hoort ook een specifiek type model.

Het model van de 'twee-stappen procedure' voor MN-analyse (Van den Eeden en Saris, 1984) is daar een voorbeeld van. In dat model wordt eerst het proces op het individuele niveau gemodelleerd. Daarna wordt getracht de individuele groepspecifieke parameters door groepsvariabelen te verklaren. De recent ontwikkelde "random coëfficiënt-modellen" voor de analyse van hiërarchisch gestructureerde data hebben deze notie verder verfijnd. Bij de random coëfficiënt-modellen is het mogelijk de binnengroep-regressie als een variabele op te nemen in de tussen-groep-regressie en de onverklaarde variantie te splitsen in een individuele component en een groepscomponent.

De vergelijkingen zien er dan in hun elementaire vorm binnen het regressiemodel als volgt uit. De eerste stap betreft het proces op individueel niveau.

$$Y_{ij} = a_j + b_j X_{ij} + z_{ij} \quad (1)$$

met i = de individuele leerling, en j = de klas; z_{ij} is de storingsterm met variantie s^2 .

In de tweede stap is er sprake van een tussengroepsregressie van het intercept a_j en de helling b_j op de klasvariabele U_j . Dit levert op:

$$a_j = A_0 + A_1 U_j + d_j \quad (2a)$$

$$b_j = B_0 + B_1 U_j + e_j \quad (2b)$$

met d_j en e_j de storingstermen met variantie g^2 en l^2 . Het model dat wordt gepresenteerd door de vergelijkingen 1 en 2 staat onder diverse namen bekend, zoals random coëfficiënt-model (De Leeuw en Kreft, 1986), variance decomposition model (Aitkin en Longford, 1986, Winer, 1971) en hiërarchisch lineair model (Raudenbush en Bryk, 1986). Het vertoont overeenkomst met de "separate equation approach" van Boyd en Iversen (1979) en het model van de twee-stapsprocedure, omdat bij beide slechts de A_0 , A_1 , B_0 , en B_1 -coëfficiënten kunnen worden geschat. A_0 staat voor een algemene constante, A_1 voor het direct effect van U , B_0 voor het direct individueel effect en B_1 voor het effect van U op de invloed van X op Y . Met behulp van d_j en e_j kan worden vastgesteld in hoeverre de verschillen in a_j en b_j meer dan toevallig zijn. Bovendien kan de correlatie tussen de storingstermen per groep worden berekend (intra-class correlatie), die duidt op gemeenschappelijkheid van ervaringen van leerlingen van eenzelfde klas.

Het model kan worden gesplitst in een vast deel en een random deel:

$$Y_{ij} = A_0 + A_1 U_j + B_0 X_{ij} + B_1 U_j X_{ij} + \quad (\text{vast deel})$$

$$d_j + e_j X_{ij} + z_{ij} \quad (\text{random deel})$$

Deelmodellen van het random coëfficiënt-model kunnen worden gevonden door onderdelen uit deze vergelijking weg te laten. Het vast deel met een ongesplitste storingsterm is het gebruikelijke model van regressierekening, en in deze vorm wordt het ook wel het model voor "contextuele analyse" genoemd. Het vaste deel zonder de U_j - en de $U_j X_{ij}$ -variabelen en de $e_j X_{ij}$ -term en het random deel met $e_j X_{ij} = 0$ vormt het traditionele model van covariantieanalyse. Het vaste deel zonder U_j - en de d_j -term staat bekend als het model voor het vaststellen van de "slopes-as-outcomes".

Wanneer de term $e_j X_{ij}$ gelijk aan 0 is, is het mogelijk het random deel eenvoudig te splitsen in een individueel deel en een groepsdeel. De verhouding daartussen geeft aan op welk niveau er het meest zinvol variabelen dienen te worden toegevoegd om de passing van het model bij de

data te verhogen. Hoe deze toevoeging van variabelen in concreto gebeurt betreft niet zozeer het model, maar eerder de procedure van analyse.

Procedure

De vierde term ter karakterisering van het MN-onderzoek is 'procedure'. Een procedure geeft aan hoe een onderzoeker daadwerkelijk zijn analyse uitvoert, uitgaande van het model. Voor MN-analyse is met name de analysestrategie en de schattingsprocedure van belang.

In de strategie van analyse worden er keuzes gemaakt, uitgaande van het model en een aantal theoretische overwegingen. In het MN-onderzoek kan een dergelijke keuze inhouden welke delen van het model belangrijker worden geacht dan andere. Zo kan een onderzoeker op grond van zijn theoretische inzichten besluiten dat de coëfficiënten voor de individuele effecten (B_0) het meest belangrijk zijn, daarna de coëfficiënten voor de verklaring van hellingverschillen door groepsvariabelen (B_1), gevolgd door de verklaring van de interceptverschillen door groepsvariabelen (A_1), en dat het algemeen intercept (A_0) onbelangrijk is. Hij zou ook meer aandacht aan de tussen-helling-variantie (I^2) kunnen besteden dan aan de tussen-intercept-variantie (g^2).

Ook is de keuze voor het feit of een onderzoeker start van een volledig gespecificeerd model en achtereenvolgens de niet significant van nul verschillende coëfficiënten aan 0 gelijk stelt totdat het model maximaal vereenvoudigd is, interpreteerbaar is en een voldoende passing bij de data heeft, een theoriegeleide. De onderzoeker kan ook, omgekeerd, starten met een zo leeg mogelijk model en in achtereenvolgende runs de onverklaarde variantie trachten te binden door introductie van telkens nieuwe individuele en groepsvariabelen.

In het algemeen verschillen de problemen bij deze strategie niet van die bij de strategie van andere modellen. Omdat er ook in het MN-onderzoek meerdere modellen acceptabel zijn, zal hier de onderzoeker zich voor ogen dienen te houden waar hij zijn theoretische prioriteiten legt. Wel is het een specifiek probleem bij de MN-analyse dat er zowel in het vast deel als in het random deel van het model individuele en groepsvariabelen aan de orde zijn.

Bij zijn keuze voor een uiteindelijk model laat de onderzoeker zich leiden door twee criteria: een voldoende passing van het model bij de data en van significante coëfficiënten. De bekende ordinary least squares (OLS)-schattingsprocedure is bij de random coëfficiënten-modellen onbruikbaar. Het gebruik van deze procedure levert inefficiënte en zelfs verkeerde schattingen op (De Leeuw en Kreft, 1986). Vandaar dat bij de random coëfficiënt-modellen gebruik wordt gemaakt van andere, meer geschikte schattingsprocedures (maximum likelihood, Bayesiaanse procedures, generalized least squares en weighted least squares). Op het ogenblik zijn er in Nederland drie geavanceerde speciale computer-programma's voor de analyse met behulp van random coëfficiënt-modellen operationeel: VARCL van Longford, HLM van Raudenbush, Bryk en Congdon, en ML-3 van Goldstein en Prosser (zie de bijlage van dit themanummer). Met deze programma's kan een betrekkelijk groot aantal variabelen van zowel individueel niveau als groepsniveau worden behandeld. Bovendien hoeft het aantal niveaus niet tot twee beperkt te worden zoals in het bovenstaande is gebeurd. Hoe bruikbaar deze programma's ook zijn, het zou wenselijk zijn indien ze werden gecombineerd met modellen voor padanalyse en met modellen met ongemeten variabelen. De combinatie met de laatstgenoemde modellen is van belang omdat causaliteits- en operationaliseringsvragen dan beter te beantwoorden zijn.

Samenvattend: het MN-onderzoek kent in de praktijk tal van toepassingen. Deze zijn in sterke mate afhankelijk van andere eerder genoemde karakteristieke kenmerken van het gehanteerde model, de gebruikte theorie en de probleemstelling van een MN-onderzoek. Deze kenmerken hangen onderling nauw samen omdat daarin meerdere niveaus voorkomen. Uiteraard dient het databestand eveneens een hiërarchische structuur te bezitten (Koopman, Van den Eeden en Jungbluth, 1989).

TOEPASSINGEN IN NEDERLAND

Het bovenstaande is nog vrij algemeen. Het is wenselijk indien het werd geïllustreerd met een aantal voorbeelden van concreet uitgevoerd MN-onderzoek. Dan zal blijken dat de beschrijvingen in de concrete gevallen nuanceringen behoeven. Bovendien kunnen de voorbeelden worden gehanteerd om de karakterisering van het MN-onderzoek met behulp van het viertal gebruikte termen te verduidelijken. Zoals gezegd, zijn er in Nederland inmiddels reeds meerdere studies met een MN-probleemstelling verricht. Een enkele daarvan worden hieronder besproken, waarbij de keuze voor de studies vrij willekeurig is geweest. Bij de besprekingen wordt ingegaan op een aantal mogelijke valkuilen.

Van der Wolf (1984)

De studie van Van der Wolf heeft een probleemstelling waarin zowel de school als de klas voorkomen, waardoor de studie past binnen de MN-benadering. Deze probleemstelling bevat drie onderzoeksvragen. De eerste luidt als volgt: "Wat is de bijdrage van de school aan schooluitval van de kinderen?" De tweede onderzoeksvraag luidt: "Zijn er verschillen tussen scholen wat betreft de bijdragen aan de schooluitval van kinderen?". De laatste vraag is van meer praktische aard: "Zijn er voorstellen te ontwikkelen om scholen een bijdrage te laten leveren aan het terugdringen van schooluitval in het regulier onderwijs?", en daarom zal hier verder geen aandacht aan worden besteed. In de twee andere onderzoeksvragen komt de structuur van de MN-benadering duidelijk naar voren: het gaat om twee niveaus: de school en de leerling. Alleen de afhankelijke variabele (schooluitval) die zich op het niveau van de leerling bevindt is aangegeven. In de formulering van de probleemstelling zijn de onafhankelijke variabelen op school- en leerlingniveau niet gespecificeerd, waardoor het lijkt dat het onderzoek slechts een vergelijkend karakter heeft. Dat is evenwel niet het geval, hetgeen samenhangt met de exploratieve opzet van Van der Wolf's onderzoek.

Op het niveau van de leerling zijn enkele belangrijke onafhankelijke variabelen: agressiviteit, eenzaamheid en negatief beeld van de gang van zaken in de school. Op groepsniveau wordt de belangrijkste onafhankelijke variabele gevormd door 'type school' met de categorieën: prestatiegerichte school, cognitief gerichte welzijnsschool en affectief gerichte welzijnsschool. Daarnaast betrok Van der Wolf een aantal variabelen van klasse- dan wel leerkrachtniveau in zijn onderzoek. Het onderzoek is uitgevoerd onder 1411 leerlingen bij dertien basisscholen in Amsterdam.

Eerst besteedt Van der Wolf uitvoerig aandacht aan de constructie van de variabelen, bestaande uit diverse schalen, clusters en typen. Dan wordt nagegaan in hoeverre de typen scholen samenhangen met de schoolvariabelen, ontleend aan het klassenniveau. Daartoe worden de gemiddelden van deze variabelen tussen de typen scholen vergeleken. Dit leidt tot onder meer de volgende conclusie: "De 'prestatiegerichte' scholen (...) hebben minder steunzoekende leerkrachten. Er is minder geneigdheid tot samenwerking bij de onderwijsgevers. Ook is er minder behoefte aan interne en externe hulp bij het werk. Er bestaat tevredenheid over het werk en de samenwerking met collega's. De cognitieve welzijnsscholen hebben coöperatieve en steunzoekende leerkrachten. De 'controleerende' component speelt ook een duidelijke rol." (p.96).

Vervolgens worden de gemiddelden van de individuele onafhankelijke variabelen naar schooltype vergeleken. We noemen hier ter illustratie de conclusie van Van der Wolf voor het prestatiegerichte type school: "Op de prestatiegerichte scholen worden door de kinderen meer eenzame kinderen genoemd in de sociometrische bepaling.(...) Daar staat tegenover dat de kinderen een positief beeld hebben van eigen functioneren, minder schoolangstig zijn en minder hoog scoren op 'wachten en vervelen'. De agressieve reactie komt het minste voor binnen dit scholencluster." (p.97).

Tenslotte worden per schooltype de gemiddelden van de door Van der Wolf onderscheiden indicatoren voor schooluitval (zittenblijven, doorstroom en verwijzing buo) vergeleken. Opnieuw noemen we ter illustratie zijn conclusie voor de prestatiegerichte scholen: "Op de cluster

1 scholen is sprake van weinig doorverwijzing naar het buitengewoon onderwijs en meer doorstroom naar de hogere vormen van voortgezet onderwijs. De zittenblijf percentages zijn het hoogst. Duidelijk is dat van dit scholencluster op twee van de drie schooluitvalindicatoren een gunstige score haalt. Daarom kan worden gesteld dat dit scholencluster relatief weinig bijdraagt aan de schooluitval van kinderen." (p.100).

Het lijkt hier niet zinvol uitvoerig te blijven stilstaan bij de wijze waarop de verschillende typen van onafhankelijke variabelen zijn gehanteerd door Van der Wolf. Belangrijker is om na te gaan hoe Van der Wolf's studie vanuit de MN-benadering zou dienen te worden opgezet. Omdat in de probleemstelling meerdere niveaus zijn aangegeven mag men aannemen dat er een theorie wordt geformuleerd over de te verwachten samenhang tussen de genoemde variabelen van het school- en leerlingniveau. Dat is niet het geval. Van der Wolf geeft echter alleen een notie over de algemene invloed van schoolklimaat op het schoolgebeuren. Er wordt geen theorie geformuleerd op het niveau van de leerling. In de notie van de klimaatinvloeden op het schoolniveau is het onderscheid in de twee aangebrachte niveaus verdwenen. Daarmee wordt er een reductieve verklaringswijze op op groepsniveau gehanteerd van MN-processen.

In de onderzoeksliteratuur komt een dergelijke groepsreductie bij herhaling voor, met name bij evaluatieonderzoek. Naast de reductieve verklaringswijze op het groepsniveau treft men bij veel onderzoeken een ander type reductie aan: die tot het individuele niveau. De theorie wordt dan uitsluitend geformuleerd op het niveau van het individu, bijvoorbeeld de leerling, en de groepskenmerken worden daarbij impliciet gezien als een individuele stimulus of perceptie die voor alle leden van eenzelfde groep gelijk is. Hoewel een onderzoeker het recht heeft een dergelijke assumptie te hanteren, blijft het een ernstige vraag of de assumptie van gelijkheid van stimulus of perceptie per groep wel realistisch is. Indien een onderzoeker zich in zijn/haar verklaring uitsluitend richt op hetzij het groepsniveau danwel op het individuele niveau, is er geen sprake meer van een MN-theorie.

In het model van analyse dat Van der Wolf hanteert heeft er eveneens een reductie plaatsgevonden. De analyse vindt op de verschillende niveaus apart plaats. Op het individuele niveau wordt de correlatie nagegaan tussen de genoemde onafhankelijke variabelen en de uitval van de leerling. Daarnaast wordt nagegaan of de gemiddelden van de individuele schooluitval tussen de verschillende schooltypen verschilt. Daarmee is er geen sprake meer van een MN-model. Er wordt voor elk niveau een apart analysemodel gehanteerd. Het analysemodel komt daarmee niet overeen met het MN-karakter van de probleemstelling, en daarmee heeft Van der Wolf het zich onmogelijk gemaakt om zijn probleemstelling te beantwoorden.

Het splitsen van modellen over de verschillende niveaus, kan men wel vaker aantreffen (bijvoorbeeld De Jong, 1987). Op elk niveau apart wordt dan een analyse uitgevoerd, waarbij de afhankelijke variabele op dezelfde individuele waarnemingen zijn terug te voeren. Op het groepsniveau wordt dan het gemiddelde daarvan genomen. Een dergelijke analyse-opzet geeft aan dat een onderzoeker heeft moeten veronderstellen dat de processen op de verschillende niveaus in principe onafhankelijk van elkaar verlopen. Het is zeer de vraag hoe plausibel een dergelijke veronderstelling is. Bovendien veronderstelt de onderzoeker dat de afhankelijke variabelen op het individuele en het groepsniveau identiek zijn, met andere woorden, dat de berekening van gemiddelden geen verschuiving teweeg brengt in de betekenis en de verdeling van de variabele. Deze veronderstelling is niet realistisch.

Het splitsen van modellen mag niet worden verward met het geval dat er een model wordt opgesteld waarin de onafhankelijke variabelen zich op het individuele en het groepsniveau bevinden en de afhankelijke variabele op een van beide niveaus, met name het individuele niveau. In de literatuur komt men dit type van modellen zeer veel tegen, en deze staan daar bekend als modellen voor "contextuele analyse". In dit geval gaat het om een enkel model met een afhankelijke variabele. Ook betreft het model slechts één niveau van analyse. Hoewel dan de theorie weliswaar op meerdere niveaus gelijktijdig is geformuleerd, vindt er bij de modelconstructie een reductie tot slechts één niveau plaats.

Het komt overigens ook voor dat een theorie op het individuele niveau wordt geformuleerd, en dat de toetsing daarvan plaats vindt met behulp geaggregeerde gegevens. Geaggregeerde

gegevens zijn gegevens die betrekking hebben op groepsspecifieke verdelingskarakteristieken, zoals het gemiddelde of percentage. Redenen daarvoor zijn er legio: individuele gegevens zijn niet voorhanden of, met name vanwege privacy-redenen niet beschikbaar gesteld, men is gedwongen gebruik te maken van een beperkt computergeheugen, de onderzoeker gelooft zozeer in zijn theorie dat hij niet meer uit is op een adequate modelspecificatie, of eenvoudigweg een researchtraditie. Dit gebruik van geaggregeerde gegevens komt bij economen zeer veel voor. Een dergelijke werkwijze is niet zonder gevaren. Bij aggregatie wordt de binnengroepvariantie weggerandomiseerd, hetgeen dramatische verschuivingen in de coëfficiënten kan teweegbrengen (Robinson, 1953; Blalock, 1964; Hannan, 1971; Kreft en De Leeuw, 1988). Deze verschuivingen staan bekend als de "aggregatie-fout". Daarnaast doen er zich interpretatieproblemen voor: Indien er bij klassen een verband wordt geconstateerd tussen het percentage allochtonen en gemiddelde studieprestatie, dan betekent dat geenszins dat de allochtonen de laagste studieprestaties halen.

De studie van Van der Wolf had wat betreft de probleemstelling en de data gemakkelijk met de MN-benadering benaderd kunnen worden. Het random coëfficiënt-model was in dit geval adequaat geweest. In dat analysemodel zouden eerst de individuele processen kunnen worden gespecificeerd, en daarna zou nagegaan kunnen worden in hoeverre die processen anders lopen op de verschillende typen scholen. Mogelijkerwijs had ook het klassenniveau als een intermediair niveau in de analyse kunnen worden meegenomen. Helaas is dat alles niet gebeurd. Overigens zou ook binnen de MN-benadering de grootte van de steekproef ($N=13$) een probleem zijn gebleven.

Meijnen (1984)

Het tweede voorbeeld van MN-onderzoek dat we hier bespreken is verricht door Meijnen (1984). De probleemstelling bestaat uit vier onderdelen:

a. "In hoeverre kan de statusspecifieke verdeling van intellectuele prestaties bij de aanvang van het lager onderwijs worden toegeschreven aan statusspecifieke socialiseringsprocessen?

b. Zijn er verschillen tussen scholen die relevant zijn voor verschillen in intellectuele prestaties van leerlingen in het algemeen en voor de prestaties van handarbeiderskinderen in het bijzonder?

c. Welke statusspecifieke gezinssocialisatie-kenmerken beïnvloeden statusspecifieke intellectuele prestaties in het lager onderwijs? Welke school- en gezinskenmerken versterken of neutraliseren elkaar?

d. Is het verloop van statusspecifieke intellectuele prestaties in de onderbouw identiek aan die in de bovenbouw? Zijn dezelfde schoolkenmerken in beide fasen relevant?" (p. 7)

De eerste vraag betreft een beschrijving van het proces op het niveau van de leerling. In de tweede vraag wordt tevens de aandacht gericht op de schoolkenmerken als onafhankelijke variabele, waarbij gekeken wordt naar de verschillende categorieën van leerlingen: kinderen van handarbeiders versus kinderen van niet-handarbeiders. Deze vraag kan men uitleggen als gericht op de conditionerende werking van het effect van het (handarbeiders)milieu van de leerling op zijn of haar intellectuele prestatie. De tweede vraag komt derhalve overeen met de MN-benadering.

De derde vraag is gericht op de effecten van variabelen van twee hogere niveaus op het prestatieproces van de leerling: het niveau van de school en het gezin. Het gaat hier om een probleemstelling met drie niveaus. In de vierde vraag, tenslotte, gaat het weliswaar om een vraag met twee niveaus: de leerling en de school, maar de aandacht is daar gericht op verschillende categorieën leerlingen: in de onderbouw en in de bovenbouw. Gezocht wordt derhalve naar een differentieel MN-effect.

Kortom, deze probleemstelling laat een tamelijk complexe uitwerking van de MN-benadering zien.

De structuur van de theorie komt met deze probleemstelling overeen. De theorie die behoort bij de eerste vraag is uitsluitend op het niveau van de leerling gelegen, en betreft de milieuspe-

cifieke socialisatie. De theorie bij de tweede vraag betreft, in overeenstemming met de probleemstelling, twee niveaus. De afhankelijke variabele betreft de intellectuele prestatie van de leerling (geoperationaliseerd als ontwikkelingen in IQ en in leerprestaties), de onafhankelijke variabele van de leerling is zijn milieu, en op schoolniveau zijn dat onder andere schooltype ('modaal', 'cognitief vernieuwend', 'leerlinggericht', 'leerstofgericht' en 'vormingsvernieuwend'), 'moreel van het team' en 'leiderschap van het hoofd'.

Meijnen wil achterhalen in hoeverre schoolvariabelen van invloed zijn op het effect van het milieu op de prestatie. Daarmee gaat het hier om een structuur van een MN-theorie. Aan de theorieën van de twee andere vragen ligt dezelfde structuur ten grondslag, alleen is die hier wat complexer uitgewerkt.

De studie is uitgevoerd bij 446 leerlingen van 24 basisscholen. Het betrof een cohortstudie waarbij de leerlingen in de eerste, de derde en de zesde klas zijn ondervraagd. De ontwikkeling van IQ en van leerprestatie is vastgesteld met behulp van verschillcores. De analyse behorend bij de eerste vraag is gedaan met behulp van padanalyse waarbij allerlei gezinskenmerken worden beschouwd als oorzaken voor de prestatiescores op intelligentie, rekenen en taal. Bij de tweede stap zijn de scholen onderscheiden in verschillende typen (het modale type, het cognitief-vernieuwende type, het leerlinggerichte type, het leerstof-gerichte type en het vormingsvernieuwende type). Eerst is er een éénweg variantieanalyse uitgevoerd. (p. 125). Het blijkt voor de onderbouw dat "... 10 % van de ontwikkelingen in Verbaal IQ aan scholen zijn toe te schrijven ($p < .001$), vervolgens 8 % in het performale IQ ($p = .044$) en 7 % in het algemeen IQ ($p = .092$). Een zwakke maar toch substantiële invloed" (p. 128). Voor de bovenbouw zijn deze cijfers 12 %, 7 % en 9%. (p. 134). Daarbij zijn de gemiddelden vergeleken. Vervolgens zijn de schoolkenmerken aan de leerlingen gehecht en op het niveau van de leerling correlatiecoëfficiënten berekend tussen de schoolkenmerken en de afhankelijke variabelen. Dit is gebeurd voor alle leerlingen gezamenlijk en apart voor de kinderen van handarbeiders, waarbij echter de verschillen in coëfficiënten niet op significantie zijn getoetst.

Omdat het echter hier gaat om een studie vanuit een MN-benadering en een MN-theorie, had de analyse veel beter met behulp van het random coëfficiënt-model uitgevoerd kunnen worden. Bovendien maakt Meijnen gebruik van verschillcores om de prestatieontwikkeling van de leerlingen in kaart te brengen. De ontwikkelingen in IQ en prestatie hadden evenwel evenzeer als een regressie tussen de twee meetmomenten kunnen worden geïndiceerd. In de tweede stap zouden dan de invloed van de schoolvariabelen op deze regressies kunnen worden vastgesteld. De milieugroepen (volgens SES) hadden dan zeer goed als een derde niveau kunnen worden ingebracht. Daarom zou het wenselijk zijn als de studie van Meijnen opnieuw met het random coëfficiënt-model zou worden geanalyseerd opdat meer valide uitkomsten worden verkregen.

De Glopper (1988)

De derde studie waarvoor we hier aandacht vragen is die van De Glopper (1988). In De Glopper's studie wordt verslag gedaan van de nederlandse deelname aan de IEA Written Composition Study. Diverse analyses zijn uitgevoerd; een daarvan is hier van belang. Deze gaat over de verschillen in schrijfprestaties op het niveau van de klas en het schooltype (pp. 130-142).

Het onderzoek had betrekking op de leerlingen uit het derde leerjaar van 107 scholen voor voortgezet onderwijs. Per school nam één klas aan het onderzoek deel: een LTO-, een LHNO, een MAVO-, en een HAVO- of een VWO-klas. Elke leerling maakte drie schrijfofdrachten. Per klas werden de schrijfofdrachten van 12 leerlingen in het onderzoek meegenomen.

Wij richten hier onze speciale aandacht op de relaties tussen de schrijfprestaties, het schrijfonderwijs en de achtergrondvariabelen (Hoofdstuk 6). Hoewel De Glopper erkent dat een analyse met een random coëfficiënt model tot de mogelijkheden behoort, beperken de analyses zich tot één niveau: de klas. "De keuze voor het klasniveau is ingegeven door een tweetal overwegingen. Een eerste overweging heeft betrekking op het doel van de analyses. Voorop stond het achterhalen van de relaties tussen de aanpak van het schrijfonderwijs en de gemiddelde schrijfprestaties van klassen. Klasgemiddelden worden hier beschouwd als de belangrijkste

indicatoren voor de opbrengst van het onderwijs dat de docenten geven. (...) Een tweede overweging is van meer praktische aard en heeft betrekking op het aantal observaties per klas", aldus De Gloppe (p.120).

De volgende prestatievariabelen worden gehanteerd: 'algemene kwaliteit', 'spelling', 'interpunctie' en 'grammatica'. In het onderzoek komen een tiental onderwijsaanbodvariabelen voor: onder andere 'uren Nederlands', 'gelegenheid tot specifieke oefening' en 'frequentie van schrijfopdrachten'. Als achtergrondvariabelen van de klas gelden onder andere de volgende: 'sociaal-ecomische status', 'percentage dialectsprekers' en 'percentage anderstaligen'. De meeste van deze variabelen zijn gebaseerd op de gemiddelden van de scores van de individuele leerlingen per klas.

Nagegaan wordt welk percentage verklaarde variantie van de gemiddelden van de afhankelijke variabele door de afzonderlijke onderwijsaanbod- en achtergrondvariabelen worden gebonden (vermeld in Tabel 6.4, p.129). Vervolgens zoekt De Gloppe met behulp van de vergelijking van de intra-class-correlatiecoëfficiënt ρ met de tussen-leerling variantie naar de binding van de variantie van de afhankelijke variabelen door de klasindeling. Hoewel deze procedure wat omslachtig is (waarom niet meteen de between-class-correlatie-coëfficiënt η^2 berekend?), blijkt daaruit dat het hier toch gaat om de afhankelijke variabele op het individueel niveau (Tabel 6.5 en Tabel 6.6). De procedure leidt tot de volgende conclusie: "(Het) blijkt dat de verschillen tussen klassen bij de algemene kwaliteit van de teksten veruit het grootst zijn en bij de grammatica veruit het kleinst" (p.132). Ten derde worden de product-moment correlaties berekend tussen de onderwijsaanbod- en achtergrondvariabelen met de gemiddelden van de prestatievariabelen (zie Tabel 6.7). Daarna zijn nog enkele analyses uitgevoerd die we hier buiten beschouwing zullen laten.

Een paar zwakke plekken zijn reeds aangeduid. De afhankelijke variabele is niet consequent op het niveau van de klas gelegen, en betreft ook de leerling. Het lijkt zinvol om een analyse inderdaad te starten met te kijken welk deel van de variantie van de individuele prestatievariabelen door de klasindeling wordt 'verklaard'. Daarna zou men verder kunnen gaan zoeken naar klasvariabelen die die 'verklaring' meer inzichtelijk kunnen maken. Het is dan bovendien wenselijk individuele variabelen als SES en sexe in de beschouwing te betrekken. Voorts is het feit dat De Gloppe de gemiddelde schrijfstijl van een klas als een indicator beschouwt als een index voor schrijfonderwijs twijfelachtig. Moet als opbrengst of effect niet in de eerste plaats worden gedefinieerd als datgene wat een docent toevoegt aan wat reeds bij een leerling en in een klas aanwezig is? In dat geval zou de "opbrengstvariabele" gewogen dienen te worden voor dergelijke op beide niveaus aanwezige input-variabelen. Dat leidt ertoe dat het door De Gloppe verworpen random coëfficiënt-model wel adequaat was geweest.

Een tweede type bezwaar betreft niet de keuze van het analyseniveau, maar het gehanteerde model van meting. De analyse van de samenhang van de gemiddelde schrijfstijl per klas met achtergrondvariabelen is een voorbeeld van betekenisverschuiving, de semantische fout, en van de aggregatiefout. De op klassenniveau gevonden correlaties tussen bijvoorbeeld sexe en leeftijd en schrijfstijl kunnen zeer goed door correlaties op het individuele niveau zijn te- weeggebracht.

De Gloppe construeert dus een nieuwe groepsvariabele door aggregatie. Het gaat om een type meetmodel waarin variabelen van meerdere niveaus voorkomen. Een meetmodel voor een geaggregeerde variabele veronderstelt dat het om een groepsvariabele gaat, die op groepsniveau is gedefinieerd. De informatie kan weliswaar worden ontleend aan de individuen, maar dan dient deze ontleening voor te komen in de theorie die behoort bij de theoretische term op het groepsniveau. In geen geval mag de betekenis van een variabele op het groepsniveau zonder meer worden overgenomen (Hüttner, 1986). Daarnaast wordt er door de aggregatie een beperking opgelegd aan de variatie van de scores.

Kortom, de bewaren tegen het gebruik van geaggregeerde en contextuele variabelen zijn zowel theoretisch als statistisch van aard.

PERSPECTIEVEN

Het lijkt erop dat de MN-benadering een vruchtbaar programma is voor het onderzoek van het onderwijs. Zij sluit nauw aan bij de structuurkenmerken van het onderwijssysteem. Bovendien sluit de benadering aan bij de noties over de conditionerende werking van de verschillende variabelen van hoger niveau.

De bovenstaande uiteenzetting over de karakteristieken van het MN-onderzoek is noodzakelijk inleidend van aard. Redenen voor deze inleiding zijn de betrekkelijke onbekendheid van de MN-benadering, de gebrekkige stand van de theorievorming over MN-processen en de nog in ontwikkeling zijnde methodologie. De theorievorming is niet alleen gebaat met goede niveau-specifieke omschrijvingen van de niveaus en de variabelen, maar ook met krachtige noties over de remmende dan wel stimulerende werking van omgevingskenmerken op het individueel gedrag. Hoewel er in het verleden diverse pogingen zijn ondernomen om dergelijke theorieën te formuleren, worden ze veelal in de analyse niet adequaat gehanteerd. De opgave is theorieën met conditiestructuren te formuleren waarbij de mechanismes duidelijk zijn.

Ook de modellen zijn nog volop in ontwikkeling. Over de "power" van de toetsen binnen de random coëfficiënt-modellen is nog weinig bekend. Zoals gezegd, dienen de random coëfficiënt-modellen nog te worden aangevuld met modellen voor padanalyse en met modellen voor ongemeten variabelen.

Toepassingsgebieden van het MN-onderzoek op het onderwijsresearch zijn er legio. We zullen er hier een tweetal noemen.

a. Het onderzoek naar de effectiviteit van het onderwijs. Het bovengenoemde random coëfficiënt-model kan vruchtbaar worden gehanteerd als een algemeen model voor evaluatie van onderwijssystemen en onderwijspraktijken. In dergelijk evaluatie-onderzoek wordt meestal nagegaan in hoeverre er verschillen bestaan tussen de leerlingen die aan een stimulus (bijvoorbeeld een type instructietechnologie, of een groepeeringsnorm) zijn blootgesteld geweest en leerlingen die dat niet zijn. Veelal worden beide categorieën leerlingen gevonden in meerdere klassen. Door het gebruik van het random coëfficiënt-model is het mogelijk te controleren niet alleen voor de verschillen in de samenstelling van die klassen, maar ook voor de individuele aanvangscores. Bovendien is het mogelijk de effecten van de tussen-klasverschillen te verdisconteren. Niet alleen de effecten van stimuli in klassen (Koehler en Van den Eeden, 1987), maar ook van scholen (Roeleveld, Koopman, De Jong en Van den Eeden, 1988) en zelfs van onderwijssystemen (Kreft, 1988) kunnen op deze wijze worden geëvalueerd. Voorbeelden van dergelijke toepassingen van het random coëfficiënt-model op evaluatie-onderzoeken vormen de bijdragen van Mooij en van Terwel en Van den Eeden in dit nummer.

b. Het onderzoek naar schoolloopbanen. Het random coëfficiënt-model biedt de mogelijkheid om naspeuring te doen naar de factoren die eigen zijn aan het onderwijssysteem (die behoren bij de indeling in niveaugroepen, klassen, scholen, schooltypen, scholengemeenschappen of categorale scholen, denominatie) en die van invloed zijn op de ongelijkheid van onderwijskansen van de leerlingen. De afhankelijke variabele bestaat dan uit een of andere output- of doorstromingsvariabele op het niveau van de leerling. Als onafhankelijke variabelen op het niveau van de leerling gelden dan niet alleen zijn eigen geschiktheid, maar ook zijn lidmaatschap van bepaalde sociale categorieën (zoals sociaal milieu, sexe of etniciteit). Nagegaan wordt dan of de input qua leerlingsamenstelling en de beoordelings- en bevorderingsbeslissingen op al de genoemde niveaus van invloed zijn op de schoolloopbaan van de leerlingen. Toepassingen zijn onderzoek naar de effecten van witte en zwarte scholen en het onderzoek naar schoolloopbanen in het voortgezet onderwijs (Roeleveld, Koopman, De Jong en Van den Eeden (1988). Voorts kunnen deze analyses worden toegespitst op achterstandsgroepen (Bosker en Van der Velden, 1989).

Omdat de MN-benadering een programma van onderzoek inhoudt, is daarbij de vraag of zij juiste of onjuiste uitkomsten oplevert irrelevant. Gelet zal dienen te worden op de vruchtbaarheid van de MN-benadering. De beoordeling van deze vruchtbaarheid is een proces dat een lange periode beslaat. Het veronderstelt niet alleen de uitwerking van adequate theorieën waarin

meerdere niveaus gelijktijdig voorkomen en waarin relevante observeerbare variabelen voorkomen. Het veronderstelt tevens de verdere ontwikkeling van modellen en technieken voor de analyses waarmee de geldigheid van deze theorieën kan worden vastgesteld. Ten slotte veronderstelt een dergelijke beoordeling een omvangrijke toepassing van de theorieën en methoden in de dagelijkse onderzoekspraktijk. Dit themanummer beoogt daartoe een bijdrage te leveren.

LITERATUUR

- Aitkin, M. & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *The Journal of the Royal Statistical Society, Series A (General)* 149, part I, 1-43.
- Bertels, K. & Nauta, D. (1974). *Inleiding tot het modelbegrip*. Amsterdam: Wetenschappelijke Uitgeverij.
- Blalock, H.M. (1964). *Causal inferences in non experimental research*. Chapel Hill: The University of North Carolina.
- Bosker, R.J. & Velden, R.K.M. van der (1989). The effects of secondary schools on the education careers of disadvantaged pupils. In: B.P.M. Creemers, T. Peters & D. Reynolds (eds.), *School effectiveness and school improvement*, Lisse: Swets en Zeitlinger.
- Boyd, J.H. & Iversen, G.R. (1979). *Contextual analysis: concepts and statistical techniques*, Belmont, Cal., Wadsworth.
- Dar, Y. & Resh, N. (1986). *Classroom composition and pupil's achievement*. London: Gordon and Breach.
- Eeden, P. van den & Saris, W.E. (1984). Empirisch onderzoek naar multilevel uitspraken. *Mens en Maatschappij*, 2, 165-178.
- Eeden, P. van den & Hauer, J. (red.) (1984). *Analyse van multi-level data: theorie en toepassing*. 's-Gravenhage, SVO (Selecta Reeks).
- Eeden, P. van den & Hüttner, H.J.M. (1984). *Multilevel research*, London, etc.: Sage.
- Glopper, K. de (1988). *Schrijven beschreven: Inhoud, opengingen en achtergronden van het schrijfonderwijs in de eerste vier leerjaren van het voortgezet onderwijs*, 's-Gravenhage: SVO (Selecta reeks).
- Hannan, M.T. (1971). *Aggregation and disaggregation in sociology*. Lexington, Mass.: Heath.
- Hüttner, H.J.M. (1985). *Multilevel onderzoek: de operationalisatie van omgevingskenmerken*. Nijmegen: KUN (Proefschrift).
- Jong, M.J. de (1987). *Herkomst, kennis en kansen: allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs*. Lisse: Swets & Zeitlinger.
- Koehler, J. & Eeden, P. van den (1987). *De invloed van klassamenstelling op schoolloopbanen: een multilevel onderzoek bij het algemeen voortgezet onderwijs in Amsterdam*. Amsterdam: Amsterdams Samenwerkingsverband/Vakgroep Methoden en Technieken VU.
- Koopman, P., Eeden P. van den & Jungbluth, M. (1989). *Opslag en beschikbaarheid van gegevens*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Kreft, G.G. (1988). *Models and methods for the measurement of school effects*. Amsterdam: Universiteit van Amsterdam (proefschrift).
- Kreft, G.G. & Leeuw, J. de (1987). Het schatten van schooleffecten: Problemen bij de keuze van het model en de techniek. *Tijdschrift voor Onderwijs Research*, 12, 75-85.
- Kreft, G.G. & Leeuw, E.D. de (1988). The see-saw effect: a multilevel problem? *Quality and Quantity*, 22, 127-137.
- Leeuw, J. de & Kreft, G.G. (1986). Random coefficient models for multilevel analysis. *Journal of the American Statistical Association*, 74, 311-319.
- Lindenberg, S. (1976). De structuur van theorieën van kollektieve verschijnselen, in W. Arts, S. Lindenberg en R. Wippler (red). *Gedrag en structuur*. Rotterdam. UPR 1-20.
- Marwijk Kooij-von Baumhauer, L. van (1984). *Scholen verschillen: een verkennend onderzoek naar het intern functioneren van 25 scholengemeenschappen voor vwo-havo-mavo* (proefschrift).
- Mensen, Th. & Guldemon, H. (1987). Groepseffecten op individuele leerprestaties: een vergelijking van drie modellen voor multi-level analyse. *Mens en Maatschappij*, 62, 176-199.
- Meijnen, G.W. (1984). *Van zes tot twaalf: Een longitudinaal onderzoek naar milieu- en schooleffecten van loopbanen in het lager onderwijs*, SVO: Flevodruk Harlingen (SVO-reeks 79).
- Merton, R.K. (1957). *Social theory and social structure*. Glencoe: Free Press.
- Mooij, T. (1987). *International multi-level investigation into pupil behavior, achievement, competence, and orientation in educational situations*. 's-Gravenhage/Nijmegen: SVO, ITS (proefschrift).
- Oosthoek, J.K. (1985). *Effecten van de leeromgeving, een sociaalpsychologisch onderzoek naar factoren*

- die de rechtsopvatting van studenten beïnvloeden.* Utrecht: Universitaire Drukkerij RUU (proefschrift).
- Oosthoek, H. & Eeden, P. van den (red.) (1986). *Education from the multi-level perspective: models, methodology and empirical findings.* New York, etc.: Gordon and Breach.
- Raudenbush, S. en A.S. Bryk (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Roeleveld, J., Koopman, P., Jong U. de & Eeden, P. van den (1989). *Schoolloopbanen in Amsterdam.* Amsterdam: Vakgroep M&T, Vrije Universiteit.
- Vos, H. de (1986). Een model van het interactieproces in de schoolklas. *Mens en Maatschappij*, 2, 61, 147-168.
- Winer, B.J. (1971). *Statistical principles in experimental design.* New York, etc.: McGraw-Hill.
- Wolf, J.C. van der (1984). *Schooluitval; een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het regulier onderwijs*, (proefschrift). Lisse: Swets en Zeitlinger.

Manuscript ontvangen 12-12-1989

Definitieve versie ontvangen 13-7-1990

Effecten van gedifferentieerd wiskundeonderwijs: de toepassing van een model voor multilevelanalyse bij curriculumevaluatie

J. Terwel en P. van den Eeden

Rijksuniversiteit Utrecht, Vakgroep Onderwijskunde en Vrije Universiteit, Vakgroep Methoden en Technieken van Sociaal-Wetenschappelijk Onderzoek***

ABSTRACT

The application of a model of multilevel analysis in the case of curriculum evaluation: effects of differentiated instruction in mathematics.

The aim of this article is to describe the effects of a newly developed model of instruction in mathematics in the second grade of Dutch comprehensive schools (LBO-MAVO-HAVO-VWO). Main characteristics of the model are: 'cooperative learning' and 'mathematics in real life situations' (contexts). The database was adopted from an earlier study of Terwel, Herfs, Dekker and Akkermans (1988), and comprises data of 482 pupils in 22 classes. Application of the random coefficient model of multilevel analysis shows that there are significant effects of class-composition. If class-composition is omitted from the model, curriculum-implementation shows a significant effect. No effect of off-task behavior was found. The article shows the fruitfulness of the use of the random coefficient model of multilevel analyses for the assessment of effects in evaluation research.

VRAAGSTELLING

Bij de aandacht voor de totstandkoming van leerresultaten zijn er tenminste drie richtingen te onderscheiden. In de eerste richting ligt het accent op de samenhang tussen leerresultaten en de recrutering van leerlingen. In de tweede richting tracht men leerresultaten te verklaren door de kwaliteit van het curriculum en de instructie.

Daarnaast bestaat er nog een derde richting van aandacht. Die richt zich op de invloed die het onderwijssysteem zelf uitoefent op het leerproces van de leerlingen. Een algemeen centraal element in het onderwijssysteem is dat het is georganiseerd in klassen. Het leerproces zelf vindt bij de leerlingen plaats. Met andere woorden: het onderwijssysteem bezit typisch een geneste structuur van meerdere levels. We spreken van "levels" om een geneste structuur aan te duiden, dit ter onderscheiding van de term "niveaus" in andere betekenissen. In dit artikel, dat zich richt op de situatie in het voortgezet onderwijs, worden twee levels in de analyse betrokken: leerlingen en klassen. In dit verband is de vraag van belang of het leerproces binnen elk van deze levels anders verloopt, en of het antwoord op die vraag verschilt per niveaugroep (zwakke, middelmatige en sterke leerlingen). Het grondidee hierbij is dat de niveaugroep waartoe een leerling behoort en zijn klas een specifieke inhoud kunnen geven aan de leeromgeving van een leerling. Daarmee is gezegd dat de leerresultaten van een leerling mede afhankelijk kunnen zijn van niveaugroep en klas. Het doel van dit artikel is nagaan in hoeverre deze vraag met behulp van empirisch materiaal kan worden beantwoord.

* Vakgroep Onderwijskunde, Postbus 80140, 3508 TC Utrecht.

** Vakgroep Methoden en Technieken van Sociaal-Wetenschappelijk Onderzoek, Koningslaan 22-24, 1075 AD Amsterdam.

De algemene hypothese voor dit artikel is dat het leren van individuele leerlingen onder invloed staat van variabelen op hogere levels, in dit geval de klas. We verwachten dat het individueel leerproces niet een op zichzelf staand gebeuren is, maar afhankelijk is van kenmerken van de klas. Leerlingen van verschillend niveau zullen naar verwachting niet in gelijke mate profiteren van de voordelen van een gunstige classesamenstelling, de curriculuminnovatie en de taakgerichtheid van de klas.

De analyses worden voor elke niveaugroep afzonderlijk uitgevoerd. Men zou kunnen denken dat het hier gaat om drie levels, met de niveaugroep als hoogste level. Dat is evenwel een misvatting, omdat met de drie niveaugroepen niet evenzoveel eenheden zijn die een steekproef uit een of andere populatie vormen, maar wel categorieën van een variabele zijn (en wel behorend bij de leerling). Aan het te hanteren random coëfficiënt model ligt noodzakelijkerwijze een design van getrapte steekproeftrekking ten grondslag, dat niet zou worden gerealiseerd indien de niveaugroep als hoogste level zou worden beschouwd. In een dergelijke situatie wordt wel aangenomen dat er kan worden gegeneraliseerd naar een voorshands onbekende theoretische populatie. Dit gebeurt bijvoorbeeld in onderzoek naar cohorteffecten op schoolverschillen van loopbanen van leerlingen. Een dergelijke aanname van een theoretische populatie van alle mogelijke niveaugroepen is wellicht in principe mogelijk, maar lijkt in dit geval wat ver gezocht. Door een gescheiden analyse van de verschillende niveaugroepen is het in elk geval mogelijk de niveaugroepen apart te karakteriseren.

DATA

De analyse is uitgevoerd met behulp van een dataset, afkomstig van het project Interne Differentiatie (ID 12-16) (Terwel, Herfs, Dekker en Akkermans, 1988). Dat project was gericht op het vaststellen van de effectiviteit van een curriculum voor het wiskundeonderwijs dat is gebaseerd op wiskundig-didactische principes van Freudenthal (1973 a en b, en 1980). Deze principes zijn:

- a. klassikale introducties en reflecties in heterogene klassen;
- b. samenwerking in heterogene subgroepjes;
- c. niveaus in het leerproces (verschillen in oplossingen);
- d. wiskunde in rijke contexten.

Het geheel van deze principes duiden we aan met 'het (wiskunde didactisch) model Freudenthal'. Het curriculum is ontwikkeld door het Instituut voor de Leerplanontwikkeling (SLO).

De gegevens werden verzameld in drie brede scholengemeenschappen, van 482 leerlingen en 12 leraren in 22 tweede klassen, de adviezen van de leerlingen varieerden van LBO tot en met VWO. De voormeting vond plaats in oktober 1985 en de nameting in mei 1986. Naast de gegevens die betrekking hebben op de leerling, zijn ook gegevens over de klas en de niveaugroep in de analyse betrokken.

HYPOTHESEN

Achtergrond voor de hieronder geformuleerde specifieke hypothesen is de verwachting dat de leerwinst onder invloed staat van ten minste twee levels: de leerling en de klas, en dat deze invloed per niveaugroep zal verschillen.

Onze hypothesen zijn te verdelen in twee groepen.

1. De eerste hypothese is dat er een zelfstandig effect is van de voortoets (TOT1) op de natoets (TOT3). Het resultaat van de voormeting is rechtstreeks van invloed op de nameting. Dit individueel effect, op het level van de leerling, duiden we aan met de term "leerwinst". Vandaar dat we die hypothese aanduiden als de leerwinsthypothese. Het kan zijn dat deze leerwinst een

intra-individueel proces is, maar het is daarnaast mogelijk dat zij afhankelijk is van een aantal kenmerken in de onderwijssituatie. Te denken valt hier aan kenmerken gelegen in de klas en kenmerken van de niveaugroep waartoe de leerling behoort.

2. De tweede categorie hypothesen is gelegen op het level van de klas. Deze klassehypothesen nuanceren de leerwinsthypothese door de kenmerken van de klassesituatie erbij te betrekken. Het betreft drie hypothesen: de samenstellingshypothese, de implementatiehypothese, en de hypothese over taakgerichtheid van de klas.

2a. De samenstellingshypothese luidt: hoe hoger het gemiddelde beginniveau van de klas des te hoger is de score op de eindtoets. De variabele classesamenstelling is bepaald door het gemiddelde van de klas op de voortoets (MTOT). De variabele MTOT betreft het level van de klas aan het begin van het schooljaar. Voor de formulering van deze hypothese is aangesloten bij onderzoek naar het effect van klassecompositie op het leren van individuele leerlingen (vgl. Beckermann en Good 1981, Good en Marshall 1984, Oakes 1986, Dar en Resh 1986). Een grondgedachte hierbij is dat het percentage sterke leerlingen in een klas een positief effect heeft op het leren van alle leerlingen in die klas. In klassen met veel zwakke leerlingen verloopt het onderwijsleerproces minder goed en er wordt minder leerstof behandeld dan in klassen met veel sterke leerlingen (vgl. Dreeben en Barr 1987, 34).

2b. De implementatiehypothese luidt: hoe meer het curriculum volgens de bedoelingen is uitgevoerd des te hoger zijn de scores op de eindtoets. De variabele MPERCIA betreft de mate waarin het curriculum in de klas volgens de bedoelingen van de leerplanontwikkelaars is uitgevoerd. MPERCIA is vastgesteld aan de hand van het gemiddelde per klas van de perceptiescore van het curriculum door de leerlingen. De leerlingpercepties zijn gemeten met een vragenlijst voor de perceptie van de onderwijsleersituatie, de zogenoemde PERCIA-schaal (PERceptie van het Curriculum In Actie). Deze vragenlijst voor leerlingen betreft een curriculum-specifieke versie van schalen voor het bepalen van de perceptie van de leeromgeving. Leerlingpercepties zijn als zodanig op te vatten als indicatoren van implementatie. Bovendien zijn er indicaties uit onderzoek dat de gemiddelde perceptie van leerlingen vrij hoog correleert met geobserveerd gedrag van leraren (Fraser en Tobin 1989). Overigens gaat onze voorkeur uit naar observatiegegevens betreffende de implementatie; deze zijn echter niet beschikbaar. Het is niet uit te sluiten dat in de leerlingpercepties houdingsaspecten meespelen. Datgene wat een leerling waarneemt kan beïnvloed zijn door zijn houding ten opzichte van het wiskunde-onderwijs of door zijn specifieke leerbehoeften.

In de PERCIA-vragenlijst zijn verschillende kenmerken van het nieuwe wiskundeonderwijs opgenomen. Deze kenmerken betreffen de leiding en de begeleiding door de leraar, de samenwerking in groepjes, de taakgerichtheid en het didactische principe 'wiskunde in contexten'. Bij keuze en uitwerking van deze begrippen is aansluiting gezocht bij de vakdidactiek, bij leertheorieën en bij resultaten van empirisch onderzoek betreffende effectiviteit van instructie (vgl. Freudenthal 1973 en 1980, Terwel 1984 en 1989).

2c. De taakgerichtheidshypothese is: hoe meer niet-taakgericht gedrag in de klas des te lager de scores op de eindtoets (en omgekeerd) (vgl. Brophy, 1986). We verwachten in alle gevallen een negatief effect op de eindtoetsscores. De variabele NTAJOB is het percentage niet-taakgericht gedrag van de leerlingen in de klas. Deze variabele is gemeten aan de hand van systematische, kwantitatieve observaties in de klas.

Het behoren tot een niveaugroep heeft mogelijk ook invloed op de leerwinst van de leerlingen. We verwachten dat de leerwinst hoger is naarmate een leerling tot een hogere niveaugroep behoort. Het onderwijs volgens het 'model Freudenthal' veronderstelt wellicht kennis en vaardigheden die sterke leerlingen reeds bezitten waardoor zij meer van het onderwijs kunnen profiteren. Bovendien veronderstellen wij dat de sterke leerlingen relatief weinig beïnvloed worden door negatieve factoren in de leeromgeving. Er zijn indicaties uit onderzoek en theorievorming van Dar en Resh (1986) dat leerlingen uit de middelste niveaugroep gevoeliger zijn voor de factor klasse-samenstelling dan sterke leerlingen. Op basis van de verdeling van TOT1 is er een indeling gemaakt in drie niveaugroepen, laag, midden en hoog, conform eerder onder-

zoek (Terwel, Herfs, Dekker en Akkermans 1988). Leerlingen uit een bepaalde niveaugroep ontvangen onderwijs in verschillende klassen. Tussen deze klassen bestaan verschillen in de samenstelling dat wil zeggen verschillen in gemiddelde voorkennis (gemiddelde TOT1). We gaan na of er verschillen tussen de niveaugroepen bestaan in gevoeligheid voor verschillen in MTOT, MPERCIA en NTAKOB.

Voor een beschrijving van de onderzoeksinstrumenten en hun betrouwbaarheid en validiteit verwijzen we naar het oorspronkelijke onderzoeksverslag (Terwel, Herfs, Dekker en Akkermans 1988).

HET MODEL

Het uitgangspunt van onze analyse is de regressie van de natoets (TOT3) op de voortoets (TOT1). Over de effecten van klassevariabelen op deze regressie van TOT3 op TOT1 is hierboven een aantal hypothesen geformuleerd. Deze hypothesen worden getoetst binnen het random coëfficiënt (RC) model voor multilevelanalyse. Dit model is opgebouwd uit twee stappen: in de eerste stap is er sprake van een binnengroepregressie. In de tweede stap worden de uitkomsten daarvan ingevoerd in een tussengroepregressie (De Leeuw en Kreft, 1986, Aitkin en Longford, 1986, Raudenbush en Bryk, 1986, en Goldstein, 1987). In de meest eenvoudige vorm bevat het model twee levels en kan het model in twee stappen worden uitgeschreven, hoewel de toetsing simultaan plaatsvindt.

De twee stappen van het model kunnen als volgt worden uitgeschreven, waarbij hier als groepslevel de klas wordt genomen. In de eerste stap wordt de binnengroepregressie uitgedrukt in de vergelijking:

$$\text{TOT3}_{ij} = a_j + b_j \text{TOT1}_{ij} + z_{ij}$$

waarin:

- i: de individuele leerling ($i = 1 \dots I$)
- j: de klas ($j = 1 \dots J$)
- a_j : het intercept voor klas j
- b_j : de helling voor klas j
- z_{ij} : de storingsterm met variantie s^2 .
- TOT3_{ij} : de wiskundescore einde studiejaar.
- TOT1_{ij} : wiskundescore begin studiejaar.

In de tweede stap is er sprake van een tussengroepregressie van het intercept a_j en de helling b_j op de klassevariabelen.

$a_j = A_0 + A_1 \text{MTOT}_j + A_2 \text{MPERCIA}_j + A_3 \text{NTAKOB}_j + d_j$
 $b_j = B_0 + B_1 \text{MTOT}_j + B_2 \text{MPERCIA}_j + B_3 \text{NTAKOB}_j + e_j$
 met d_j en e_j de storingstermen behorend bij respectievelijk a_j en b_j met varianties t^2 en u^2 . A_0 staat voor een algemene constante, A_m voor het direct effect van een klasvariabele M, B_0 voor het direct individueel effect en B_m voor het effect van een klasvariabele op het effect van TOT1_{ij} op TOT3_{ij} .

Het RC-model (De Leeuw en Kreft, 1986) staat overigens ook bekend als "variance decomposition-model" (Aitkin en Longford, 1986) en "hierarchical linear model" (Raudenbush en Bryk, 1986). Het vertoont overeenkomst met de "separate equation approach" van Boyd en Iversen (1979) en het model van de "tweestaps-procedure" (Van den Eeden en Saris, 1984), in zoverre de coëfficiënten worden geschat. Met behulp van dit model is het mogelijk de klasspecifieke onderwijskansen af te lezen aan de verschillen in de intercepten en de hellingen van een regressie waarin de invloed van TOT1 van leerlingen op hun TOT3 tot uiting wordt gebracht. Bij de analyses is gebruik gemaakt van het VARCL-programma van Longford (1986).

ANALYSE

Alvorens in te gaan op de multilevelanalyse, vermelden we in de Tabellen 1 en 2 de verdelingskenmerken van respectievelijk de individuele en de klassevariabelen.

Bij de toetsing zijn meerdere modellen gehanteerd. In Model 1 wordt alleen de variantie van de afhankelijke variabele (TOT3) vermeld. In Model 2 wordt aangegeven welk deel van deze variantie aan de leerling en welk deel aan de klas kan worden toegeschreven. Met behulp van Model 3 wordt vastgesteld in hoeverre de leerlingvariantie door de voortoets (TOT1) wordt gebonden, in overeenstemming met de leerwinsthypothese. In Model 4 wordt vervolgens nagegaan in hoeverre door de introductie van de variabele TOT1 er een tussenklasvariantie rond de regressie van TOT3 op TOT1 aanwezig is. In Model 5 wordt gekeken in hoeverre de eventuele, laatstgenoemde tussenklasvariantie wordt verklaard door de klassevariabelen MTOT, MPER-

Tabel 1. Kenmerken van de verdelingen van TOT1 en TOT3 en de regressiecoëfficiënten van TOT3 op TOT1 per niveaugroep. Tussen haken is de standaardfout vermeld.

		TOT3	TOT1	toename	regressiecoëfficiënten
NIVEAU-GROEP 1	gemiddelde	29.54	20.62	8.82	a = 8.74 (2.34)
	st.dev	8.98	4.88		b = 1.01 (.11)
	min	2	4.5		R ² = .30
	max	54	26.5		
	cases	197			
	klassen	22			
NIVEAU-GROEP 2	gemiddelde	41.62	30.20	11.42	a = 1.39 (11.19)
	st.dev	10.87	2.26		b = -.34 (.37)
	min	9	27		R ² = .08
	max	75	34.5		
	cases	158			
	klassen	21			
NIVEAU-GROEP 3	gemiddelde	54.93	40.91	14.03	a = 9.68 (5.88)
	st.dev	11.40	5.88		b = 1.11 (.14)
	min	32	35		R ² = .33
	max	86	63		
	cases	127			
	klassen	18			

Tabel 2. Verdelingskenmerken van de klasse-variabelen, alsmede de correlaties tussen de klasse-variabelen.

	gemiddelde	standaard deviatie	minimum	maximum
MTOT1	27.804	5.721	18.231	39.032
MPERCIA	3.305	.206	2.881	3.605
NTAKOB	25.597	24.027	.370	74.620
MTOT1	1.000			
MPERCIA	.570	1.000		
NTAKOB	-.676	-.525	1.000	
	MTOT1	MPERCIA	NTAKOB	

CIA en NTAJOB, overeenkomstig onze gerichtheid op de drie klassehypotheseën. In Model 6, tenslotte, wordt vastgesteld in hoeverre de eventueel resterende interceptvarianties eveneens door deze klassevariabelen worden gebonden. Bij deze toetsingsprocedure wordt telkens gekeken of de modelpassings-parameters DEV van opeenvolgende modellen elkaar verschillen, daarbij rekening houdend met het verschil in aantal vrijheidsgraden. Niet-significante coëfficiënten worden op nul gefixeerd. In alle gevallen is het 95 procents-niveau aangehouden.

We gaan nu in op de gegevens uit de multilevelanalyse. De uitkomsten vermelden de ongestandaardiseerde coëfficiënten. Voor de interpretatie daarvan dient men rekening te houden met de verdelingskenmerken als gemiddelde, minimum en maximum (zie Tabel 1 en Tabel 2). In Tabel 3 zijn de resultaten van de multilevelanalyses vermeld. Niet-relevante uitkomsten zijn weggelaten.

In Model 1 zijn alle vaste coëfficiënten en de random coëfficiënten op groepsniveau weggelaten. Het resultaat is dat alleen de totale variantie van de afhankelijke variabele, s^2 , wordt vermeld. Het valt op in Model 1, dat de te verklaren variantie verschilt per niveaugroep. Zij wordt groter naarmate de niveaugroep hoger is.

In Model 2 wordt de totale variantie van TOT3 uit Model 1 gesplitst in een individueel deel en een klasse-deel (weergegeven door s^2 en t^2), of anders gezegd, in een binnengroepsvariantie en een tussengroepsvariantie. Bij de twee laagste niveaugroepen is er een splitsing van een

Tabel 3. Uitkomst van de multilevelanalyse. Niet-relevante uitkomsten zijn weggelaten. Tussen haakjes is de standaardfout vermeld.

NIVEAUGROEP 1 (Zwakke leerlingen)						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
fixed deel						
individueel effect						
B_0 effect TOT1			.91 (.11)	.91 (.11)	.39 (.11)	.85 (.11)
klasse effect						
interceptvariantie						
verklaard door						
A_1 MTOT						.47 (.15)
A_2 MPERCIA						
A_3 NTAJOB						
hellingvariantie						
verklaard door						
B_1 MTOT					.02 (.01)	
B_2 MPERCIA						
B_3 NTAJOB						
random deel						
s^2 individueel	80.28	63.05*	48.24	48.24	47.97	47.71
t^2 intercept (klas)	—	17.89*	8.06	8.06	4.81	4.50
v^2 helling (klas)						
R^2 individueel			23.5	23.5	23.9	24.3
R^2 intercept			44.7	44.7	73.1	74.9
R^2 helling						
DEV	1423.000	1401.623	1341.757	1341.757	1335.056	1333.400
DEV-verschil		21.377	59.866	0	6.701	8.357
DF-verschil		1	1	1	1	1
met Model		1	2	3	3	4

* Dit absolute getal is uitgangspunt voor berekening van de verklaarde variantie in de volgende modellen.

(Vervolg Tabel 3)

NIVEAUGROEP 2 (Middelmatige leerlingen)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
fixed deel						
individueel effect						
B ₀ effect TOT1			1.03 (.34)	1.03 (.34)		.90 (.19)
klasse effect						
interceptvariantie verklaard door						
A ₁ MTOT						.98 (.19)
A ₂ MPERCIA						
A ₃ NTAKOB						
hellingvariantie verklaard door						
B ₁ MTOT					.31 (.01)	
B ₂ MPERCIA						
B ₃ NTAKOB						
random deel						
s ² individueel	117.303	87.177*	84.104	81.104	84.416	84.360
t ² intercept (klas)		38.609*	30.230	30.230	5.877	5.086
v ² helling (klas)						
R ² individueel			3.7	3.7	3.2	3.3
R ² intercept			21.8	21.8	84.8	86.8
R ² helling						
DEV	1201.216	1183.490	1174.712	1174.712	1157.784	1156.742
DEV-verschil		17.726	8.778	0	16.928	17.970
DF-verschil		1	1	0	1	1
met Model		1	2	3	3	4

NIVEAUGROEP 3 (Sterke leerlingen)

fixed deel						
individueel effect						
B ₀ effect TOT1			1.11 (.14)	1.11 (.14)	1.11 (.14)	1.11 (.14)
klasse effect						
interceptvariantie verklaard door						
A ₁ MTOT						
A ₂ MPERCIA						
A ₃ NTAKOB						
hellingvariantie verklaard door						
B ₁ MTOT						
B ₂ MPERCIA						
B ₃ NTAKOB						
random deel						
s ² individueel	128.844	128.844*	86.874	86.874	86.874	86.874
t ² intercept (klas)						
v ² helling (klas)						
R ² individueel			32.6	32.6	32.6	32.6
R ² intercept						
R ² helling						
DEV	977.453	977.453	927.397	927.397	927.397	927.397
DEV-verschil		0	50.056	0	0	0
DF-verschil		1	1	0	0	0
met Model		1	2	3	3	4

individuele variantie en een klassevariantie van de afhankelijke variabele TOT3. De klassevariantie is voor niveaugroep 1 en 2 respectievelijk 22.1 en 30 percent. Bij de hoogste niveaugroep is er geen significante tussenklassen-variantie. De prestaties van de leerlingen uit deze groep lijken onafhankelijk te zijn van klasse-invloeden.

In Model 3 wordt behalve de random coëfficiënten s^2 en t^2 de vaste coëfficiënt B_0 geschat; s^2 representeert dan de restvariantie op het niveau van de leerling. Model 3 stelt ons in staat de leerwinsthypothese over het effect van de voortoets (TOT1) te toetsen. De leerwinsthypothese voor de individuele leerlingen geldt in elke niveaugroep, zij het dat er opmerkelijke verschillen optreden. In de laagste niveaugroep verklaart de variabele TOT1 23.5 percent van de individuele variantie en bijna de helft van de tussenklassenvariantie (44.7 percent). In de middelste niveaugroep verklaart TOT1 slechts 3.7 percent van de individuele variantie en 21.8 percent van de tussenklasvariantie. In de hoogste niveaugroep is het percentage verklaarde individuele variantie aanzienlijk hoger dan in de twee lagere niveaugroepen: 32.6 percent. De score op de natoets van de leerlingen uit de middelste niveaugroep lijkt nog het minst afhankelijk te zijn van hun score op de voortoets.

In Model 4 wordt vervolgens de random coëfficiënt voor de tussenklasvariantie van de helling van de regressie van TOT3 op TOT1, v^2 , aan het model toegevoegd. In alle niveaugroepen is gebleken dat er in Model 4 geen variantie is tussen de hellingen.

Deze uitkomst impliceert niet dat Model 5 niet voor toetsing in aanmerking komt. In Model 5 wordt vastgesteld welk deel van de tussenklasvariantie van de hellingen wordt verklaard door de klassevariabelen MTOT, MPERCIA en NTAKOB. Dit betekent dat de coëfficiënten B_1 , B_2 en B_3 worden geschat. Bij het onderzoek naar een afwijking van Model 5 ten opzichte van Model 3 zijn we evenwel in een betere uitgangspositie dan bij een dergelijk onderzoek van Model 4. We gebruiken hier theoretische informatie die leert dat de eventuele variatie in hellingen niet toevallig is, maar samenhangt met bepaalde klassevariabelen. De power van een toets van Model 3 (de nulhypothese) tegen Model 5 (de alternatieve hypothese) is dus in het algemeen hoger dan de power van een toets van Model 3 tegen Model 4.

Zoals gezegd, is er evenwel bij geen der niveaugroepen een significante variantie aanwezig van de regressiehellingen van TOT3 op TOT1 tussen de klassen. De verschillen zijn vooral aanwezig tussen de niveaugroepen, maar niet tussen de klassen binnen elke niveaugroep. De toetsing van Model 5 tegen Model 3 evenwel laat in de laagste en de middelste niveaugroep zien dat de variantie in de hellingen tussen de klassen samenhangt met MTOT; in de hoogste niveaugroep is deze samenhang afwezig. Het verband is het sterkst bij de middelste niveaugroep. Dit betekent dat de bovengenoemde hypothesen over de afhankelijkheid van de leerwinst van de klassevariabelen alleen met betrekking tot de classesamenstelling kan worden bevestigd.

In Model 6 toetsen we de drie klassehypothesen. Dit gebeurt door naast de geaccepteerde coëfficiënten ook de coëfficiënten A_1 , A_2 en A_3 in het model ter schatting op te nemen. Alleen bij de twee laagste niveaugroepen leerlingen hangt het gemiddelde van de TOT3-scores van een klas samen met de gemiddelde TOT1-score (MTOT). Dit verband is niet te begrijpen als een gevolg van een aggregatie van het zojuist genoemde verband tussen TOT1 en TOT3 op het individuele level, omdat daarvoor reeds is gecontroleerd bij het toetsen van de leerwinsthypothese. Toen bleek (in Model 3) dat van de variantie in de gemiddelde TOT3 in de eerste niveaugroep 45 percent werd verklaard door introductie van TOT1 en in de middelste niveaugroep voor 21.8 percent. Bij de leerlingen uit de hoogste niveau-groep is de cross-level interactie tussen TOT1 en MTOT in hun effect op TOT3 evenwel afwezig. De andere klassevariabelen MPERCIA en NTAKOB hebben geen significant effect op TOT3.

Dit resultaat is niet strijdig met de uitkomst van Model 5. Zowel Model 5 als Model 6 blijkt in voldoende mate bij de oorspronkelijke gegevens te passen. Omdat wij onze hypothesen hebben ontwikkeld met betrekking tot Model 5 en omdat de percentages verklaarde varianties niet belangrijk van elkaar verschillen geven we aan de uitkomst daarvan de voorkeur.

De bovenstaande uitkomsten laten zien welke verklaring er door het gehanteerde RC-model wordt bereikt. De verhouding tussen de individuele en tussenklassen-variantie is in de twee

laagste niveaugroepen ongeveer 3 : 1 respectievelijk 77.9 versus 22.1 en 74.3 versus 25.7. Bij de laagste niveaugroep is de verklaring van deze varianties in een verhouding van 1 : 3 (dat wil zeggen 24.3 versus 74.9) van de individuele en de tussenklassen-variantie en in de middelste niveaugroep ongeveer 1 : 30 (dat wil zeggen 3.3 versus 86.8). De verschillen in deze niveaugroepen manifesteren zich in de verschillen in de verklaring op de twee betrokken levels. De middelste niveaugroep is in sterke mate afhankelijk van de klasseomgeving. In de middelste niveaugroep is de coëfficiënt voor MTOT .98, in de laagste niveaugroep is deze coëfficiënt .47 terwijl deze coëfficiënt in de hoogste niveaugroep niet significant verschilt van nul. In een contrast hiermee staat de hoogste niveaugroep, waar alleen op het individueel level een verklaring plaatsvindt, die echter sterker is dan in de andere niveaugroepen.

De implementatiehypothese bleek in eerste instantie, in het door ons gekozen model, niet bevestigd te worden. Er bleek echter een samenhang te bestaan tussen MTOT en MPERCIA (zie Tabel 2). Dit duidt op collineariteit waardoor de implementatiehypothese in een model waarin ook MTOT voorkomt niet adequaat kan worden getoetst. Daarom hebben we ook een analyse met Model 6 uitgevoerd naar het effect van MPERCIA zonder dat daarin MTOT was opgenomen. In deze laatste analyse was er wel een significant effect van MPERCIA op de helling, hetgeen overeenkomt met de verwachtingen en de bevindingen van de eerdere studie. In Tabel 4 zijn de uitkomsten vermeld.

Tabel 4. Uitkomst van de multilevelanalyse met weglating van de variabele MTOT in Model 6, voor alle niveaugroepen. De vergelijking vindt plaats met Model 4.

	Niveaugroep 1	Niveaugroep 2	Niveaugroep 3
fixed deel			
individueel effect			
B ₀ effect TOT1	.87 (.11)	1.01 (.34)	
klasse effect			
interceptvariantie			
verklaard door			
A ₂ MPERCIA			
A ₃ NTAKOB			
hellingvariantie			
verklaard door			
B ₂ MPERCIA	8.22 (3.55)	15.56 (6.56)	
B ₃ NTAKOB			
random deel			
s ² individueel	47.857	83.814	86.874
t ² intercept (klas)	5.923	21.470	
v ² helling (klas)			
R ² individueel	24.1	4.9	32.6
R ² intercept	66.1	44.4	
R ² helling			
DEV	1336.734	1169.514	927.397
DEV-verschil	5.023	5.198	0
DF-verschil met Model 4	1	1	1

De uitkomsten in Tabel 4 duiden op een effect van de implementatie van het Model Freudenthal in de laagste en middelste niveaugroepen. Er is geen effect in de hoogste niveaugroep. De sterke leerlingen lijken weinig gevoelig te zijn voor verschillen in implementatie, d.w.z. voor de kwaliteit van de leeromgeving in de klas.

CONCLUSIES

In dit artikel zijn vier hypothesen getoetst tegen de achtergrond van de verwachting dat het leerproces van leerlingen onder invloed staat van variabelen op het hogere level in casu de klas. In het verlengde van deze vier hypothesen zijn de volgende hoofdconclusies te trekken.

De leerwinsthypothese leverde op dat er een zelfstandig effect blijkt te bestaan van de voortoets op de natoets. Datgene wat leerlingen aan wiskundige voorkennis bezitten blijkt rechtstreeks van invloed op hun prestaties na het doorlopen van het curriculum. Deze hypothese is geldig voor elke niveau-groep. De volgende conclusies vullen de conclusie over de leerwinst-hypothese aan door het effect van klassevariabelen erbij te betrekken.

Voor de samenstellingshypothese geldt het volgende: De samenstelling van de klas, tot uitdrukking gebracht door het klassegemiddelde, blijkt het verwachte effect te hebben op de leerwinst van de zwakke en middelmatige leerlingen. Hoe hoger het klassegemiddelde hoe beter hun score op de voortoets wordt omgezet in hun score op de natoets (zie hellingseffect, Tabel 3, Model 5). Daarnaast geldt dat bij de zwakke en middelmatige leerlingen de samenstelling van de klas effect heeft op hun prestaties. Hoe hoger het klassegemiddelde hoe hoger de prestaties van de leerlingen (zie effect op het intercept, Tabel 3, Model 6). Omdat het hier gaat om twee verschillende interpretaties en de eerste ons het meest plausibel voorkomt, geven we de voorkeur aan de eerste. In beide gevallen geldt dat de middelmatige leerlingen het meest gevoelig blijken voor de samenstelling van de klas. Voor de sterke leerlingen gaan de hypothesen echter niet op. De leerlingen in de hoogste niveaugroep lijken individualistisch georiënteerd. Hun prestaties worden niet beïnvloed door de samenstelling van de klas.

De implementatiehypothese bleek aanvankelijk, in het door ons gekozen model, niet bevestigd te worden. Na weglating van de variabele classesamenstelling uit het model (ter vermijding van collineariteit) bleek er echter toch een significant effect van de implementatie van het Model Freudenthal, waarmee de implementatiehypothese en de bevindingen uit de eerdere studie zijn bevestigd.

De hypothese over de taakgerichtheid van de klas blijkt niet bevestigd te worden, ook niet na weglating van MTOT. In de door ons gehanteerde modellen is geen effect aangetoond van de taakgerichtheid van de klas.

DISCUSSIE

We besteden eerst aandacht aan de conclusies betreffende het effect van classesamenstelling. De prestaties van de leerlingen in de hoogste niveaugroep zijn relatief onafhankelijk van de samenstelling van de klas. De twee andere niveaugroepen laten in gelijke mate individuele en klasseverschillen zien. In de middelste niveaugroep kunnen die individueel nauwelijks worden verklaard, maar door de classesamenstelling des te meer. De classesamenstellingshypothese is dus vooral van belang in de middelste niveaugroep.

Interpretaties van deze bevindingen kunnen worden ontleend aan Dar en Resh (1986) en aan Fend (1982). Dar en Resh verklaren de grotere gevoeligheid van de middengroep op basis van een drempelhypothese. Om te kunnen profiteren van een verrijkte leeromgeving moet men een minimum aan kennis en vaardigheden bezitten. Het is juist de middengroep die over de startcompetenties beschikt om de sprong naar een klas met een hoger niveau te kunnen maken. Zij winnen relatief veel in een hogere klas en verliezen overeenkomstig in een zwakke klas. Fend wijst op een tweede interpretatiemogelijkheid, die hij evenwel op schoolniveau heeft geformuleerd. Hij noemt het "differentiellen Förderungseffekt" als een van de belangrijkste en meest consistente gegevens uit het vergelijkende onderzoek naar de Gesamtschule. Dit effect betreft differentiële effecten voor overachieveers en onderachieveers in de Gesamtschule versus het categoriale systeem. Als een overachiever met een matige intelligentie de sprong naar een hoger categoriaal schooltype kan maken heeft hij grote kans te worden meegenomen door de classesamenstelling en daarvan te profiteren. Een qua intelligentie vergelijkbare onderachiever komt in

het categoriale systeem in de negatieve spiraal van het laagste schooltype en zou beter af zijn geweest in de Gesamtschule. Het lijkt aannemelijk te veronderstellen dat deze processen op het niveau van de school ook op het level van de klas optreden.

De uitkomst van onze analyse betreffende de twee andere hypothesen was in eerste instantie onverwacht en enigszins teleurstellend. We sluiten niet uit dat deze beide onverwachte uitkomsten herleidbaar zijn tot problemen betreffende collineariteit omdat implementatie en taakgerichtheid samenhangen met het gemiddelde niveau van de klas. Bovendien is een effect van "restriction of range" mogelijk aanwezig vanwege de indeling in niveaugroepen en klassen. Van belang is ook de overweging dat het hier gaat om een betrekkelijk gering aantal klassen. Dit kleine aantal impliceert een lage power voor het ontdekken van een effect van klassevariabelen op leerlingvariabelen of voor crosslevel interacties. Het niet vinden van een significant resultaat hoeft dus niet te betekenen dat het gezochte effect ook niet bestaat.

Mogelijk is ook de operationalisering van de implementatie, gemeten via de leerlingpercepties, niet adequaat. Voorts kan een rol hebben gespeeld hebben dat in deze analyse alleen projectklassen zijn betrokken. In al deze experimentele klassen werd het 'model Freudenthal' uitgevoerd. Mogelijk was de variantie in implementatie te klein om effecten te kunnen vinden.

In tweede instantie vonden we toch bij de laagste en de middelste niveaugroepen effecten van de implementatie van het Model Freudenthal op de prestaties van de leerlingen op de natoets (Tabel 4). Bij de hoogste niveaugroep is dit effect afwezig.

Uit onze toepassing blijkt het nut van multilevelanalyse bij curriculumevaluatie. In tegenstelling tot de uitkomst van een deel van het oorspronkelijke onderzoek waarbij, in een regressie-analyse op individueel niveau, een gering effect van de implementatie van het curriculum (volgens de leerlingpercepties) werd gevonden, konden we aanvankelijk niet tot een effect besluiten. In een nadere analyse vonden we evenwel een positief effect van de implementatie, hetgeen overeenkomt met de eerdere bevindingen in een voorgaande studie. Voor de overige conclusies van het oorspronkelijke onderzoek betreffende de positieve effecten van het model Freudenthal in vergelijking met een meer traditionele aanpak verwijzen we naar het genoemde eindrapport.

LITERATUUR

- Aitkin, M. & Longford, N. (1986). Statistical modelling issues in class effectiveness studies. *Journal of the Royal Statistical Society*, 149, 1-43.
- Beckerman, T.M. en Good, Th.L. (1981). The classroom ratio of high- and low-aptitude students and its effects on achievement. *American Educational Research Journal*, 18, 317-37.
- Boyd, L.H. & Iversen, G.K. (1979). *Contextual analysis: concepts and statistical techniques*. California: Wadsworth Inc.
- Brophy, J. (1986). *Teacher Effects Research and Teacher Quality*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Dar, Y. & Resh N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, 3, 357-374.
- Dreeben, R. & Barr, R. (1987). An Organizational Analysis of Curriculum and Instruction. In: M. T. Hallinan, *The Social Organization of Schools*. New York: Plenum.
- Eeden, P. van den, & Saris, W.E. (1984). Empirisch onderzoek naar multilevel uitspraken. *Mens en Maatschappij*, 59, 165-178.
- Fend, H. (1982). *Gesamtschule im Vergleich*. Bilanz der Ergebnisse des Gesamtschulversuch. Weinheim und Basel.
- Fraser, B. J. & Tobin, K. (1989). *Combining qualitative and quantitative methods in the study of classroom learning environments*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Freudenthal, H. (1973a). De niveaus in het leerproces en de heterogene leergroep met het oog op de middenschool. In: *Gesamtschule conferentie 1973*. Amsterdam/ Purmerend: APS/ Muuses.
- Freudenthal, H. (1973b). *Mathematics as an educational task*. Dordrecht.
- Freudenthal, H. (1980). *Weeding and Sowing*. Dordrecht/ Boston, Reidel.

- Goldstein, H. (1987) *Multilevel models in educational and social research*. London: Griffin en New York: Oxford University Press.
- Good, T.L. and Marshall, S. (1984). Do students learn more in heterogeneous groups? In: P.L. Peterson, L.C. Wilkinson and M. Hallinan (eds.): *The Social Context of Instruction*. Orlando: Academic Press.
- Leeuw, J. de & Kreft, G.G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11, 57-85.
- Longford, N. (1986). *VARCL-interactive software for variance component analysis. Applications for survey data*.
- Oakes, J. (1985). *Keeping Track, How Schools Structure Inequality*. New Haven and London: Yale University Press.
- Raudenbush, S. & Bryk A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Terwel, J. (1984). *Onderwijs maken. Naar ander onderwijs voor 12-16 jarigen*. Harlingen: Flevodruk. Stichting voor Onderzoek van het Onderwijs SVO-reeks nr.77 (proefschrift).
- Terwel, J., Herfs, P., Dekker, R., en Akkermans W. (1988). *Implementatie en effecten van interne differentiatie*. Den Haag: Stichting voor Onderzoek van het Onderwijs (SVO). Selecta reeks.
- Terwel, J. (1989). Real Maths in Cooperative Groups in Secondary Education. In: N. Davidson. *Cooperative Learning in Mathematics* (p. 228-264). Menlo Park: Addison-Wesley.

Manuscript ontvangen 23-10-1989

Definitieve versie ontvangen 13-7-1990

Effecten van computerprogramma's op ontwikkeling in leesprestaties

T. Mooij

Katholieke Universiteit Nijmegen, Instituut voor Toegepaste Sociale wetenschappen

ABSTRACT

The research question is whether the use of a computer in the class-room affects the *learning results of reading processes of pupils*. A large-scale comparative study of pupils learning to read with and without computer was designed. The pretest was done in November 1988; the posttest occurred in May 1989. In the meantime teachers registered pupils' time-on-task. The multilevel analysis is done with the VARCL-program of Longford. The results show positive learning effects of computer use and some other variables, which are discussed.

1. INLEIDING

Bij de methode voor aanvankelijk leesonderwijs: "Veilig Leren Lezen" (VLL) bestaan diverse materialen die tot doel hebben het leesonderwijs meer te differentiëren (Ten Hove, z.j.) en de methode ook bruikbaar te maken in het speciaal onderwijs (Reijnders, 1980) of in het onderwijs aan tweetalige of allochtone leerlingen (Van de Guchte, 1988). Naast de gebruikelijke leesleerboekjes zijn additionele materialen ontwikkeld zoals: speelleessets, stempeldozen, ideeënboeken, en een liedjesboek met audio-cassettes. In het SVO-project: "Preventie van leesmoeilijkheden" werd bovendien een aantal diagnostische toetsen ontwikkeld (Van Dongen, 1984).

In de onderwijspraktijk blijft echter behoefte bestaan aan leermiddelen met behulp waarvan leerlingen in de klas zelfstandig, op hun eigen niveau, en met onmiddellijke feed-back kunnen oefenen. Omdat dit oefenen mogelijk gemaakt wordt door computerprogramma's, zijn in het kader van de methode VLL negen van dergelijke programma's ontwikkeld. Met elk programma kan een belangrijke deelvaardigheid van het aanvankelijk leesproces geoefend worden.

In vakdidactisch/onderwijskundig opzicht worden de programma's goed ontvangen (zie bijvoorbeeld SCENschrift 1987, 3, p. 17-21; 1987, 5, p. 2-5). Het werken met de computerprogramma's lijkt zowel het leerproces (via de aandachtconcentratie en het taakgedrag) als de leesprestatie (het leereffect) in gunstige zin te beïnvloeden. Deze indruk bestaat ook in andere publikaties over effecten van opname van computerprogramma's in het curriculum: de aandachtconcentratie stijgt; de daadwerkelijk bestede leertijd neemt toe; deelvaardigheden kunnen getraind worden totdat zij in voldoende mate beheerst worden. Een beperking van de beschikbare informatie is echter dat er nog weinig of niet is vastgesteld welke (leer)processen en -effecten zich voordoen bij het gebruik van computerprogramma's in ecologisch valide onderwijsleersituaties (vgl. ook Beishuizen, 1987; Daal, Van der Ley, Bakker, & Reitsma, 1987; Reitsma, 1987; Spaai, Ellermann, & Reitsma, 1986).

Om vast te kunnen stellen of en, zo ja, in welke mate er zich door het gebruik van de nieuwe VLL-computerprogramma's als positief te waarden leereffecten in het aanvankelijk leesonderwijs in groep 3 van het basisonderwijs voordoen, dient er een evaluatie-onderzoek in de onderwijspraktijk uitgevoerd te worden. Hierbij zou de aandacht gericht moeten worden op de vraag welke van de in aanmerking komende instructie- en leerlingkenmerken in welke mate van belang zijn voor de ontwikkeling van de aanvankelijk leesprestatie, in het bijzonder wat de effecten zijn van de computerprogramma's. De vraagstelling van dit onderzoek luidt dan: *welke*

leerling- en instructie-variabelen zijn van belang voor de ontwikkeling van de aanvankelijk leesprestatie in het kader van de methode: "Veilig Leren Lezen", en wat zijn met name de effecten van de computerprogramma's?

In dit artikel bespreek ik de theorie, opzet en resultaten van dit evaluatie-onderzoek. Het project werd financieel gesteund door de SVO (project 8121), de Commissie Programma Evaluatie en de afdeling BO/SP/O van het ministerie van Onderwijs en Wetenschappen. Een gedetailleerd verslag van het onderzoek is opgenomen in het eindrapport (Mooij, 1990).

2. INTERACTIONELE MULTI-NIVEAU THEORIE

2.1. Inleiding

In cognitieve theorieën wordt aangenomen dat een persoon informatie over zijn handelingen en prestaties structureert in *cognitieve kaarten*, die specifiek zijn voor bepaalde delen van de omgeving (Ippel & Elshout, 1990; Kuhl, 1986; Lewin, 1936; Neisser, 1976). De cognitieve structurering is ook van nut bij de verklaring van het waarnemen van een persoon of bij het begrijpen of voorspellen van keuzen uit mogelijke gedragingen. Actuele percepties van kenmerken van de omgeving maar ook de verwachtingen in dit verband kunnen functioneren als motieven, die de handelingen van een persoon sturen (cf. Heckhausen, 1980, 1986). Een cognitieve kaart structureert dan mede de waarderingen van de persoon ten opzichte van de omgeving; omgekeerd beïnvloedt de omgeving de opbouw en structurering van de cognitief-motivationele processen. Het gedrag van een persoon is dan gebaseerd op invloeden vanuit intra-individuele persoonskenmerken, kenmerken van de omgeving respectievelijk de percepties en verwachtingen over de omgeving, en de cognitief-motivationele wisselwerkingen tussen deze soorten kenmerken. Pervin en Lewis (1978) spreken in dit verband over *interactionele processen* waarin kenmerken van de omgeving en kenmerken van de persoon in onderlinge wisselwerking elkaar beïnvloeden.

In de motivatie-theorieën van Bowerman (1978), Harter (1982), Heckhausen (1980, 1986) en White (1959) wordt een intra-individuele handeling-effect-affect opeenvolging geassumeerd. De essentie hiervan is dat een persoon in een bepaalde (soort) situatie iets *doet*, deze activiteit heeft een bepaald *effect*, en dit effect wordt in een bepaalde mate positief of negatief *gewaardeerd* door de persoon. Gebleken bekwaamheid stimuleert dan tot vergelijkbare of iets moeilijkere handelingen in dezelfde of iets moeilijkere situaties; gebleken onbekwaamheid motiveert tot het verrichten van alternatieve handelingen in dezelfde of zo mogelijk andere situaties. Deze *competentie-motivatie* stuurt dus mede de ontwikkeling van de cognitief-motivationele processen en de feitelijke keuzen uit gedragingen ten opzichte van de omgeving.

De competentie-motivatie van een leerling met betrekking tot school zal dan cognitief en motivationeel ook op een schoolvakspecifieke wijze gestructureerd worden (cf. Ames, 1984; Cattell & Child, 1975; Krampen & Mory, 1982; Mooij, 1982, 1987). Het proces tijdens welke gekozen wordt uit verschillende mogelijkheden tot handelen met betrekking tot een specifiek schoolvak en de lessen in dit vak, wordt gedefinieerd als de *schoolvakmotivatie* van een leerling. De schoolvakmotivatie drukt zich onder andere uit in taakgedrag (vgl. Emmer & Evertson, 1980; Evertson, Emmer, & Clements, 1980). *Taakgedrag* doet zich voor indien de handelingen van een leerlinge met betrekking tot de leertaken tot doel hebben de door de leerkracht gestelde opdracht(en) uit te voeren. Het uiteindelijke resultaat van taakgedrag is een bepaalde prestatie in een schoolvak. Een *schoolvakprestatie* is de mate waarin het taakgedrag of de getoonde vaardigheden en bekwaamheden voldoen aan de eisen die vanuit het schoolvak (via de school of de leerkracht) gesteld worden. Kenmerken van de leerling en kenmerken van de onderwijsomgeving zoals de instructie, leerstof en didactische werkvormen hebben dan mede in onderlinge wisselwerking invloed op de ontwikkeling van de schoolvakprestatie (zie ook Starren, Bakker, & Van der Wissel, 1988).

Met betrekking tot de ontwikkeling van de prestatie in het aanvankelijk lezen geven Seegers, Aarnoutse en Mommers (1987) een overzicht van de ontwikkeling van het leren lezen bij een

individuele leerling. Deze ontwikkeling verloopt van het naïef lezen (auditief, visueel en motorisch structureren) via elementaire leeshandelingen (kennis leesrichting, visuele analyse van woorden in grafemen, koppeling van fonemen aan grafemen, auditieve synthese en betekenis herkennen) naar de verkorte leeshandeling (versnelling van woordherkenning). De door deze auteurs gegeven taalkundig-psychologische schets van het ontwikkelingsproces van het leren lezen is voornamelijk cognitief en intra-individueel van karakter. In het kader van de hierboven aangeduide interactionele cognitief-motivationele processen tussen leerling- en omgevingskenmerken worden uitdrukkelijk ook curriculum- of onderwijskenmerken in het aanvankelijk leesproces betrokken. Dit omdat tijdens een aanvankelijk leesles geselecteerde onderwijskenmerken, met name instructie-kenmerken, door de leerkracht gebruikt worden om de aandacht van de leerlinge doelbewust te concentreren op het beoogde leerproces (Starren et al., 1988).

In het kader van de gebruikelijke instructie in de methode VLL is er sprake van het bestaande aanbod van de basisleergang VLL, dat wil zeggen de leesleerboekjes inclusief de additionele leermaterialen zoals speelleessets, stempeldozen, ideeënboeken, liedjesboek met audio-cassettes, en diagnostisch materiaal (Caesar, 1980). Met dit additionele materiaal is reeds een zekere differentiatie naar subgroepniveau en individuele oefening op leerlingniveau mogelijk. De recent ontwikkelde computerprogramma's vormen het nieuwste additionele materiaal met nieuwe mogelijkheden tot differentiatie en individuele stapsgewijze oefening (vgl. ook Roth & Beck, 1987). In vergelijking met de andere instructie- en verwerkings-vormen in het aanvankelijk leesonderwijs, is het oefenen met de computerprogramma's in motivationeel/leertheoretisch opzicht om twee redenen van belang:

- (a) het zou het aanvankelijk lezen als leerproces aantrekkelijker en efficiënter maken, daardoor bevorderend werken op de schoolvakmotivatie, en bijgevolg het taakgedrag van een leerling vergroten. Deze thematiek is theoretisch en empirisch uitgewerkt bij Mooij (1990): de resultaten uit observatie-onderzoek in klassen laten inderdaad zien dat het werken met computerprogramma's samengaat met een grotere mate van taakgedrag;
- (b) vergroting van het taakgedrag kan na verloop van tijd leiden tot een toename van het leereffect in de vorm van een relatief hogere leesprestatie. Hierop wordt nader ingegaan.

2.2. Multi-niveau ontwikkeling van de leesprestatie

Wat betreft de lange-termijn vorderingen in leesprestatie zou verwacht kunnen worden dat in groepen waar leerlingen wel met computerprogramma's oefenen - in vergelijking met groepen waarin dit niet het geval is - de leerlingen een relatief grotere toename in leesvorderingen zullen laten zien. Dit kan verklaard worden doordat - in groepen met computerprogramma's - een hogere schoolvakmotivatie en een grotere mate van taakgericht gedrag zullen leiden tot een hoger actueel leerresultaat; dit hogere leerresultaat motiveert via de competentie-motivatie weer tot volgend hoog leesgericht taakgedrag met volgende hoge resultaten, enzovoorts (vgl. de interactionele prestatie-motivatie theorie bij Heckhausen (1980), Lens (1987) en Mooij (1987)).

In theoretisch opzicht dient deze globale hypothese nader gespecificeerd te worden om de veronderstelde (mogelijke) invloed van het werken met computerprogramma's valide te kunnen vaststellen en andere verklaringen uit te kunnen sluiten. Dit is tevens essentieel omdat de theorie relevant dient te zijn voor de onderwijspraktijk, waar in onderwijsleersituaties steeds verschillende leerling- en instructiekenmerken van invloed kunnen zijn op de leerprocessen van een leerling. Mede vanwege deze ecologische validiteit is er een specificatie nodig naar de verschillende relevante niveaus in onderwijsleerprocessen (Burstein, 1980; Cronbach, 1983). Hier worden de meest relevante niveaus (leerling- en groepsniveau) en verschillende variabelen op elk niveau aan de orde gesteld.

Op het *individuele niveau van de leerling* is de instructie-variabele *hoeveelheid individuele leertijd* of time-on-task van belang (zie Reid, Hopkins, & Holly, 1987): een leerling die meer oefent in lezen, zou vanwege deze grotere mate van tijdbesteding aan het lezen snellere vorderingen kunnen maken dan een leerling die minder oefent. Daarnaast zijn er op dit niveau nog relevante persoonlijke variabelen die een leerling kenmerken, zoals de *begin-leesprestatie* (een hogere begin-leesprestatie kan op zich motiverend zijn dan wel de verklaring vormen voor een

hogere eind-leesprestatie), *sexe* (meisjes zijn veelal iets sneller in lezen dan jongens: zie bijv. Creemers, 1974), en *sociale achtergrond* (kinderen uit een relatief lager sociaal/cultureel/economisch milieu hebben op taalgebied een achterstand in vergelijking met kinderen uit meer bedeelde milieus: zie Tesser, 1986).

Naast deze variabelen op het individueel niveau kunnen ook variabelen *op het niveau van de groep leerlingen* van belang zijn. De meest relevante instructie-variabele is *het al dan niet gebruiken van computerprogramma's* ten behoeve van het aanvankelijk leesonderwijs: het in de groep wel gebruiken van deze programma's zou, vanwege de bij leerlingen ontstane relatief hogere schoolvakmotivatie en het hogere taakgedrag, kunnen leiden tot een relatief hogere leesprestatie van elke leerling in de groep.

Andere relevante instructie-kenmerken en samenstellingskenmerken van de groep leerlingen die tevens verklarend kunnen zijn voor een relatief betere ontwikkeling van de leesprestatie van elke leerling in de groep, zijn de volgende:

1. *leerkrachtmanagement*; dit wordt gedefinieerd als de bekwaamheid van een leerkracht om te realiseren dat elke leerling in de groep kan werken; bovendien toont de leerkracht aan de leerlingen dat hij weet wat de leerlingen doen en in staat is om niet-gewenst gedrag te onderdrukken (Emmer & Evertson, 1980; Tomic, 1985);
2. *interne differentiatie* in de groep. Deze variabele geeft de mate aan waarin de onderwijsleerprocessen voor de leerlingen in de groep geïndividualiseerd verlopen. Naarmate er in een schoolvak gedurende een schooljaar meer gedifferentieerd wordt, zou elke leerling meer in staat kunnen zijn om de eigen prestatie, bekwaamheid en gerichtheid in dit vak (op een positieve wijze) te beïnvloeden;
3. *het aantal leerlingen* in de groep. Glass, Cahen, Smith en Filby (1982) bespreken de literatuur in dit verband en komen tot de conclusie dat zich in relatief kleinere klassen positievere leerprocessen en leereffecten voordoen;
4. *gemiddelde begin-leesprestatie* in de groep: naarmate deze prestatie hoger is, zouden leerlingen elkaar in positieve zin kunnen ondersteunen en zouden de leerprocessen per leerling sneller kunnen verlopen;
5. *heterogeniteit in begin-leesprestatie* tussen de leerlingen in de groep. Het is mogelijk dat naarmate de leerlingen onderling meer verschillen (heterogener zijn) in begin-leesprestatie, de leerprocessen per individuele leerling moeilijker te realiseren zijn door de leerkracht;
6. de *gemiddelde leertijd* die in de groep als geheel of in subgroepen (dus niet individueel) besteed wordt aan aanvankelijk taal/leesonderwijs: een relatief hogere gemiddelde leertijd zal een positief effect kunnen hebben op de vorderingen in individuele leesprestatie van elke leerling in de groep.

Naast de bovenstaande effecten kunnen zich binnen een groep, in samenhang met één of meer groepsvariabelen, in de loop van de tijd *groepsspecifieke ontwikkelingen in de vordering in leesprestatie voordoen*. Een voorbeeld: bij een hogere begin-leesprestatie van een kind laat een leerkracht met een goed en adequaat management dit kind relatief meer vrijheid en zelfstandigheid tijdens het leesproces, hetgeen op den duur een stimulerende invloed kan hebben op de leesontwikkeling van dit kind (zie het praktijkboek van de Ark-Berberis-school en Van Kuyk, 1980). Een leerkracht met een inadequaat management zal een kind met een hogere begin-leesprestatie te sterk bij het klassikale leesproces kunnen blijven betrekken, hetgeen de leesontwikkeling van dit kind relatief zal doen stagneren. Met elk van de groepsvariabelen zijn dergelijke *groepsspecifieke processen* met betrekking tot de begin-leesprestatie denkbaar en zinnig te interpreteren.

De hierboven gespecificeerde processen met betrekking tot de ontwikkeling van de leesprestatie zullen, indien zij van belang zijn, uiteindelijk leiden tot effecten op de uiteindelijke leesprestatie. In de multi-niveau benadering worden deze effecten "fixed" effecten genoemd (Aitkin & Longford, 1986). In het huidige theoretische kader wordt daarnaast ook nog aangenomen dat de variabelen van het individuele niveau (hoeveelheid individuele leertijd, begin-leesprestatie, *sexe* en *sociale achtergrond*) tussen de groepen kunnen verschillen, terwijl deze verschillen nog

niet nader met behulp van een groepsvariabele gespecificeerd kunnen worden. Deze "random" effecten doen zich voor in de vorm van nog niet verklaarde of residuele variantie: als deze random effecten aanwezig zijn, dient in een volgende fase van theorievorming en onderzoek er rekening mee gehouden te worden dat nog andere variabelen dan de thans gespecificeerde relevant kunnen zijn.

Het is mogelijk om de hierboven gepresenteerde twee-niveau theorie met betrekking tot de ontwikkeling van de aanvankelijk leesprestatie uit te breiden met het schoolniveau (vgl. Cronbach, 1983; Lens, 1987; Mooij, 1987). Gezien de recentheid van de ontwikkelingen van de multi-niveau theorie en analyse en de complexiteit van deze thematiek, wordt de gepresenteerde (eerste) versie van de huidige twee-niveau theorie eerst nader onderzocht wat betreft empirische en analyse-technische consequenties en mogelijkheden.

3. METHODE

3.1. Onderzoeksopzet

Het uit te voeren onderzoek is te karakteriseren als een *evaluatie-onderzoek* (vgl. Cronbach, 1983; De Corte, 1979; Hofstee, 1987; Van den Berg, 1987) van het gebruik van computerprogramma's in de instructie van het aanvankelijk leesproces in de onderwijspraktijk. De eenheden van onderzoek zijn leerlingen resp. groepen leerlingen die wel (experimentele conditie) en leerlingen resp. groepen leerlingen die niet (controle-conditie) werken met de MSX-computerprogramma's bij VLL.

Het design is een multi-niveau aanpassing in het kader van een *longitudinaal quasi-veldexperiment* (een: "untreated control group design with pretest and posttest" (zie Cook & Campbell, 1979). Ten eerste wordt in het design onderscheid gemaakt naar de verschillende relevante niveaus met hun eigen variabelen (vgl. ook Cronbach, 1983; Starren et al., 1988). Ten tweede kan ook de *experimentele variabele*, afhankelijk van de te onderzoeken processen of effecten, op verschillende niveaus gelocaliseerd worden. In het huidige onderzoek zijn deze mogelijkheden (a) de hoeveelheid leertijd op individueel- of leerlingniveau en met name de oefentijd met de computer, (b) het werken met de computer in de groep als variabele op groepsniveau: als de computer gebruikt wordt, dan worden eerste instructies e.d. klassikaal gegeven en wordt hij door elke leerling in de groep gebruikt, en (c) ook groepsspecifieke ontwikkelingen zijn hier te definiëren in de zin dat de ontwikkeling van de leesprestatie in een groep beïnvloed wordt door de aard van het computergebruik in de groep. Omdat deze mogelijke quasi-experimentele effecten niet alle tegelijkertijd nagegaan kunnen worden, wordt er voor gekozen de experimentele variabele te definiëren op groepsniveau waardoor het gemiddelde- en groepsspecifieke effect onderzocht kunnen worden. In de discussie worden onder andere resultaten van analyse met de experimentele variabele op individueel niveau aan de orde gesteld.

Naast deze experimentele instructie-variabelen zijn er – conform de theorie in par. 2 – verschillende overige leerlingvariabelen, groepsvariabelen en groepsspecifieke processen die relevant kunnen zijn en waarvan de mogelijke invloed vastgesteld moet worden. In aansluiting bij het gangbare methodologische taalgebruik, kunnen deze "te controleren" variabelen ook aangeduid worden als covariabelen. In Schema 1 staat een overzicht de in het design opgenomen variabelen.

Om organisatorisch-budgettaire redenen vond de voormeting plaats in november 1988 en de nameting in mei 1989. De deelvaardigheden in de programma's 4, 5 en 6 stonden daarom centraal. Het gaat respectievelijk om de programma's: "koppeling van fonemen aan grafemen"; "fonemen op volgorde onthouden"; en "spellingpatronen in eenlettergrepen woorden". Het longitudinale design wordt gekenmerkt door drie fasen, die identiek zijn voor de controle- en experimentele onderzoekseenheden:

- (a) *voormeting* van het technisch lezen, het begrijpend lezen en de spellingvaardigheid bij leerlingen in november 1988;
- (b) van dec. 1988 - april 1989 twee soorten *tussenmeting*:

Schema 1. Twee-niveau variabelenoverzicht.

NIVEAU 1: LEERLING**afhankelijke variabele**

NAMET: leesprestatie tijdens de nameting

instructie-variabele

INTDO: hoeveelheid individuele leertijd

overige leerlingvariabelen

VRMT: begin-leesprestatie tijdens de voormeting

SEXE: geslacht

SES: sociale achtergrond (wegingsfactor leerlinggewicht)

NIVEAU 2: GROEP**experimentele instructie-variabele**

EXPT: niet/wel VLL-computerprogramma's

overige groepsvariabelen

MLKMN: leerkrachtmanagement

MINDF: interne differentiatie

MALLN: aantal leerlingen in de groep

MVRMT: gemiddelde begin-leesprestatie

SDVMT: heterogeniteit in begin-leesprestatie

MSGTDO: gemiddelde leeftijd per (sub)groep

experimenteel groepsspecifiek proces

IEXPT: leesprestatie-ontwikkeling en niet/wel VLL-computerprogramma's

overige groepsspecifieke processen

IVRLK: leesprestatie-ontwikkeling en leerkrachtmanagement

IVRDF: leesprestatie-ontwikkeling en interne differentiatie

IVRAN: leesprestatie-ontwikkeling en aantal leerlingen in de groep

IVRMV: leesprestatie-ontwikkeling en gemiddelde begin-leesprestatie

IVRSD: leesprestatie-ontwikkeling en heterogeniteit in begin-leesprestatie

IVRMT: leesprestatie-ontwikkeling en gemiddelde (sub)groep-leertijd

- registratie van de *hoeveelheid leertijd* per instructie-vorm door de leerkrachten op (1) leerlingniveau en (2) groepsniveau;
- *observatie* in de klas door observatoren (zie hiervoor Mooij, 1990);
- (c) *nameting* bij de leerlingen in april/mei 1989.

3.2. Selectie van experimentele- en controle-onderzoekseenheden

Uit een lijst van alle basisscholen, die tot en met juni 1988 de VLL-computerprogramma's hadden aangeschaft, werden de reguliere basisscholen geselecteerd die (volgens de luchtlijn) binnen een straal van 75 km rondom Nijmegen lagen. Deze scholen vormden de groep scholen die benaderd werden als experimentele school. In augustus 1988 werd uit het totale bestand van reguliere basisscholen in Nederland (N = 8554) een a-selecte steekproef van 700 scholen getrokken. Uit deze steekproef werden verwijderd: (a) de scholen buiten een straal van 75 km (luchtlijn) rondom Nijmegen en (b) de scholen die reeds opgenomen waren in de lijst met scholen met VLL-computerprogramma's (de experimentele scholen). Er resteerden 234 reguliere basisscholen binnen de straal van 75 km rondom Nijmegen. Hieruit werden at random controle-scholen getrokken.

Tabel 1. Deelname van groepen en leerlingen per conditie

	n groepen voormeting	n leerlingen voormeting	n. groepen tussenmeting	n groepen nameting
conditie				
controle	22	388	21	22
experimenteel	27	533	24	27
niet bruikbaar	4	89	3	2
totaal	53	1010	48	51

De experimentele- en controle-scholen werden telefonisch benaderd in augustus en september 1988. Aan de directeur van de school werd een samenvatting gegeven van het doel, de werkwijze, de opbrengst en de verspreiding van de resultaten uit het onderzoek. Het doel werd omschreven als: het verkrijgen van inzicht in de mate waarin instructie-kenmerken (zoals leesleerboekje, speelleeset, stempeldoos, computerprogramma) van invloed zijn op het taakgedrag van een leerling tijdens de les en op de vordering in leesprestatie. Bij de werkwijze werd kort ingegaan op: de voormeting, tijdregistratie, observatie, en nameting (zie verder: Mooij, 1990). Ter vermijding van zelfselectie van leerkrachten werd gevraagd of elke groep 3 binnen de school (leerkracht en alle leerlingen in de groep) mee zou willen doen aan het onderzoek. Er waren slechts enkele basisscholen waarin meer dan één groep 3 aanwezig was. Om deze reden is er verder geen onderscheid gemaakt naar school of groep en worden in de analyse de groepen beschouwd alsof zij onafhankelijk zijn. De onderzoekseenheden zijn dan (1) de groep leerlingen en (2) elke leerling in de groep.

Het totaal aantal groepen dat participeerde aan de voormeting in november 1988 bedroeg 53 (1010 leerlingen). Mede op basis van de observaties in de klas bleken er uiteindelijk 22 groepen leerlingen in de controle-conditie te vallen, terwijl er in 27 groepen gebruik gemaakt werd van de VLL-computerprogramma's tijdens de lessen aanvankelijk lezen. In Tabel 1 is een overzicht opgenomen waarin per conditie worden weergegeven: de aantallen groepen en aantallen leerlingen bij de voormeting, het aantal groepen waarvan de leerkracht deelneemt aan de tussenmeting (tijdregistratie), en het aantal groepen dat deelneemt aan de nameting. De leerlingen in de conditie "niet bruikbaar" worden niet betrokken in de analyse met betrekking tot de resultaten (zie par. 4).

3.3. Operationalisering, meting en dataverzameling

Variabelen op leerlingniveau

Bij de operationalisering van de begin-leesprestatie in november 1988 werd nauw aangesloten bij de in het kader van de methode VLL geconstrueerde toetsen (zie Van Dongen & Mommers, z.j.; Mommers, 1988). Deze toetsen zijn groepsgewijs of individueel afneembaar door de leerkracht.

Het niveau van technisch lezen werd geoperationaliseerd met behulp van:

- de toets voor aanvankelijk lezen (onderverdeeld in drie subtoetsen);
- de snelheidstest voor kennis van grafemen;
- de Caesar-één-minuut-test.

Het niveau van begrijpend lezen werd bepaald met de toets voor begrijpend lezen. Het niveau van de spellingvaardigheid werd gemeten met het woorddictee 1. Mede op grond van de onderlinge correlaties (zie Mooij, 1990) werd besloten de somscores op de vijf toetsen in de voormeting te herleiden tot één indicator van de mate van beheersing van het lezen. De indicator werd

geconstrueerd door (1) de ruwe somscore op elke toets om te zetten in een z-score en (2) de z-scores te sommeren en te delen door hun aantal. De resulterende variabele bestaat uit de gemiddelde score van de z-scores per leerling en wordt *begin-leesprestatie of voormeting* (VRMT) genoemd (zie ook Schema 1). Het gemiddelde (m) hiervan is .010 en de standaardafwijking (sd) is .224 (n = 986).

Sexe (SEXE) wordt gescoord als: jongen (=0) en meisje (=1). De informatie werd verkregen van de leerling, en in geval van ontbrekende scores bij de leerkracht.

Sociale achtergrond (SES) wordt geoperationaliseerd als in het onderzoek met betrekking tot het Onderwijsvoorrangsbeleid. In het kader van dit beleid wordt aan elke leerling volgens een bepaalde wegingsfactor een bepaald leerlingengewicht toegekend. De wegingsfactor wordt gegeven door Hofstee, Schaveling en Koster (z.j., p. 74-76): een leerling uit een "lager" sociaal milieu verkrijgt een hoger leerlingengewicht. De informatie werd verkregen van de school(administratie). In een aantal gevallen werd de medewerking aan de verzameling van deze gegevens geweigerd (om redenen van privacy). Het was niet mogelijk om deze informatie op andere wijzen te achterhalen. Bij de analyse is een gehercodeerde score gebruikt (zie Mooij, 1990).

Vanaf de voormeting tot aan de nameting diende de leerkracht per schooldag een tijdregistratie bij te houden van het onderwijsaanbod per groep en per individuele leerling. Per schooldag was hiervoor een formulier met voorbedrukte categorieën en namen van leerlingen beschikbaar. De leerkracht diende het (geschat) aantal minuten in te vullen, tijdens welke de groep of de leerling bezig was geweest met het uitvoeren of oefenen van een bepaalde lees/taal-activiteit (instructie of begeleiding van de leerkracht, zelfstandig schriftelijk materiaal, oefenen met speel/leermateriaal, oefenen met computerprogramma, en overig). De gegevens met betrekking tot een individuele leerling worden gebruikt ter operationalisering van de *individuele leertijd* (INTD0): het aantal minuten dat een leerling gedurende de onderzoeksperiode gemiddeld per dag besteedt aan individueel werken tijdens de aanvankelijk leesles (zie voor details: Mooij, 1990).

Eind april/begin mei 1989 werd in het kader van de nameting de leesprestatie weer bepaald. Hierbij werd gebruik gemaakt van VLL-toetsen (zie Van Dongen & Mommers, z.j.; Mommers, 1988) en drie CITO-toetsen. De toetsen werden weer afgenomen door de groepsleerkracht. Het niveau van het *technisch lezen* werd bepaald met:

- de snelheidstest voor de kennis van grafemen (zie ook de voormeting);
- de Caesar-één-minuut-test (zie ook de voormeting);
- de toets voor technisch lezen 2 (CITO);

Het *begrijpend lezen* werd gemeten met de toets lees en begrijp 1B (CITO).

De *spellingvaardigheid* werd vastgesteld met het woorddictee 2 (CITO).

Parallel aan de constructie van VRMT werd één indicator geconstrueerd op basis van de vijf scores verkregen tijdens de nameting. De (gewogen) gemiddelde score van de betreffende z-scores is de variabele *eind-leesprestatie of nameting* (NAMET, met $m = -.009$; $sd = .787$; $n = 936$).

Variabelen op groepsniveau

De *experimentele groepsvariabele* EXPT betreft het al dan niet werken met de VLL-computerprogramma's in de groep en wordt geoperationaliseerd als een dichotomie (niet = 0, ja = 1). Bepalend voor de scoring van een groep op deze variabele was de constatering van observatoren in de klas of er daadwerkelijk sprake was van dit gebruik.

De variabelen leerkrachtmanagement (MLKMN), interne differentiatie (MINDF) en aantal leerlingen in de groep (MALLN) zijn geoperationaliseerd met behulp van scoring door observatoren in de klas (zie Mooij, 1990). Leerkrachtmanagement is de somscore van een zes-item schaal (van laag naar hoog; $\alpha = .96$), interne differentiatie de somscore van een vier-itemschaal (van veel naar weinig; $\alpha = .95$), en aantal leerlingen in de groep is het gemiddelde van het aantal leerlingen dat tijdens de geobserveerde lessen aanwezig was. Ook MLKMN en MINDF zijn gemiddelden van de respectievelijke scores uit verschillende lessen.

De gemiddelde begin-leesprestatie (MVRMT) en de mate van heterogeniteit in begin-lees-

prestatie (SDVMT) zijn resp. de gemiddelden en standaardafwijkingen van de scores van de leerlingen in de groep op VRMT. De gemiddelde (sub)groep-leertijd (MSGTD0) is berekend op grond van de informatie van de leerkracht over de tijdbesteding binnen de groep (zie ook bij de variabelen op leerlingniveau). De categorieën hierbij waren: klassikaal (instructie met behulp van bord, letterdoos, werkboekje, werkschrift, leesboekje, een andere methode, en overig) en in subgroepen (niveau-lezen).

Precieze informatie over de (her)scoringen, gemiddelden en standaardafwijkingen per variabele voor de verschillende onderzoekseenheden, en aanvullende informatie over de constructie, betrouwbaarheid en validiteit van de gegevens op leerling- en groepsniveau is opgenomen in Mooij (1990). Wat betreft de betrouwbaarheid kan met enige voorzichtigheid geconcludeerd worden dat er gegevens verkregen zijn die voldoende betrouwbaar zijn, hoewel dit voor de data verkregen uit de tijdregistratie door leerkrachten niet vaststelbaar is. Er zijn geen duidelijke aanwijzingen dat de betrouwbaarheid of (interne) validiteit van de onderzoeksgegevens tekort schiet, hoewel evident is dat hieraan in het kader van contractonderzoek meer aandacht besteed zou moeten kunnen worden.

Tabel 2. Univariante resultaten (gemiddelde, standaardafwijking) per conditie en toetsing op verschil tussen gemiddelden

	controle		experimenteel		F	p
	gem.	sd	gem.	sd		
NIVEAU 1: LEERLING (max n = 921)						
afhankelijke variabele						
NAMET: leespr. nameting	-.10	.88	.06	.71	9.12	<.00
instructie-variabele						
INTD0: indiv. leertijd	12.02	11.94	18.22	11.03	61.77	<.00
overige leerlingvariabelen						
VRMT: begin-leespr.	-.01	.23	.03	.22	5.63	=.02
SEXE: geslacht	.51	.50	.49	.50	.33	ns
SES: sociale achtergrond	1.11	.25	1.14	.25	3.01	ns
NIVEAU 2: GROEP (max n = 49)						
groepsvariabelen						
MLKMN: leerkrachtman.	22.77	2.51	21.83	2.15	1.98	ns
MINDF: interne diff.	14.61	2.09	13.26	1.95	5.54	=.02
MALLN: n lln in groep	16.23	5.97	18.52	6.81	1.53	ns
MVRMT: m begin-leespr.	.01	.13	.02	.13	.09	ns
SDVMT: heter. begin-leespr.	.16	.09	.17	.06	.15	ns
MSGTD0: m (sub)groeptijd	43.38	13.33	43.07	22.27	.00	ns
groepsspecifieke variabelen						
IVRLK	-.11	5.22	.66	5.02	5.03	=.03
IVRDF	-.16	3.46	.36	2.94	5.92	=.02
IVRAN	-.55	5.25	.59	5.49	9.70	<.00
IVRMV	.02	.05	.02	.08	.28	ns
IVRSD	-.01	.07	.01	.07	15.12	<.00
IVRMT	-.54	11.30	.60	12.20	1.92	ns

gem. = gemiddelde; sd = standaardafwijking.

Groepsspecifieke processen

De variabelen die ter verklaring dienen van groepsspecifieke processen met betrekking tot de leesprestatie-ontwikkeling (zie par. 2 en Schema 1) zijn geoperationaliseerd als het produkt van VRMT met de betreffende groepsvariabele. Hierbij zijn de begin-leesprestatie-scores gehanteerd in absolute zin ofwel ongecentreerd (bij Mooij (1990) worden ook resultaten van gecentreerde analyses vermeld).

Random effecten

In par. 2 is de mogelijkheid geopperd dat er ook nog niet-gespecificeerde verschillen tussen groepen of random effecten zouden kunnen optreden. In de analyses zal daarom worden nagegaan of er, na afhandeling van de fixed effecten, nog sprake is van resterende belangrijke random effecten.

4. RESULTATEN

4.1. Inleiding

De univariate resultaten staan in Tabel 2. Hierin worden gegevens over (maximaal) 921 leerlingen en 49 groepen vermeld. Het gaat per variabele om de gemiddelde en standaardafwijking per conditie, en om de resultaten van enkelvoudige toetsing van de verschillen tussen de gemiddelden (zie de F-waarde met $df = 1$ en het bijbehorende significantie-niveau). De variabelen EXPT en IEXPT zijn niet opgenomen omdat de dichotome waarden samenvallen met de conditieverschillen in relatie tot variabele VRMT.

In Tabel 2 komt naar voren dat er tussen de condities een groot verschil is in hoeveelheid individuele leertijd: in de experimentele conditie is deze tijdbesteding 50% omvangrijker dan in de controle-conditie. Op groepsniveau is er echter geen verschil in besteding van groepstijd (MSGTD0). De inzet van de computer lijkt daarmee de individuele tijdbesteding tijdens de aanvankelijk leesles te stimuleren. Daarnaast is er op leerlingniveau tussen de condities een verschil in begin-leesprestatie in november ($p = .02$); dit verschil neemt toe in de loop van het schooljaar (p -waarde eind-leesprestatie $< .00$).

Wat betreft de groepsvariabelen is er tussen de condities in Tabel 2 een verschil in interne differentiatie: in de experimentele conditie wordt er in een groep meer intern gedifferentieerd ($p = .02$). Bij de produktvariabelen zijn van belang de verschillen tussen de condities in leerkrachtmanagement, interne differentiatie, het aantal leerlingen in de groep, en heterogeniteit in begin-leesprestatie.

Er zou nu verder gegaan kunnen worden met multiple-regressie-analyse, of met het succesievelijk berekenen van de gemiddelden van de leerlingvariabelen per groep en het op groepsniveau nagaan van verbanden met behulp van bijvoorbeeld partiële correlaties. Er kan dan echter geen onderscheid gemaakt worden naar aan fixed en random effecten gerelateerde groepsspecifieke processen. Daarom wordt hier de voorkeur gegeven aan multi-level analyse waarin de te verklaren variantie zoveel mogelijk gebonden wordt aan specifieke variabelen per niveau (fixed effects) en er tevens nagegaan kan worden of er per variabele nog tussen-groepen verschillen resteren (random effects: vgl. Van den Eeden, 1987).

De multi-niveau analyse gebeurt met het VARCL-programma (Longford, 1988). Omdat dit programma niet beschikt over faciliteiten om met missing data om te gaan, worden in het navolgende twee verschillende analyses weergegeven: één met inbegrip van SES en één zonder deze variabele.

4.2. Multi-level analyse

In de theorie (par. 2) hebben we gezien dat er relatief veel variabelen van belang kunnen zijn bij de ontwikkeling van de aanvankelijk leesprestatie. Hierdoor wordt de vraag actueel welke van de theoretisch relevante variabelen er nu werkelijk toe doen (zie par. 1). Er zijn meer redenen om na te gaan of de relatief belangrijkste variabelen geselecteerd zouden kunnen worden: (1)

analyse-technisch kan het tegelijk analyseren van alle variabelen in een multi-level model, met name bij de gelijktijdige bestudering van hoofd- en interactie-effecten, tot multi-collineariteitsproblemen leiden (Mooij & Jansen, 1990) en (2) in de praktijk van het onderwijs is het urgent te weten welke aanpak in het aanvankelijk leesonderwijs de relatief "beste" zou zijn, met andere woorden of het pedagogisch en financieel verantwoord zou zijn de computerprogramma's in de klas te gebruiken.

In de analyses wordt het twee-niveau model daarom geanalyseerd conform de volgende stapsgewijze procedure:

1. uitdraai van een nulmodel, dat wil zeggen een model waarin geen predictor-variabelen zijn opgenomen. Met behulp van een dergelijk model wordt zicht verkregen op de variantie die wat betreft NAMET maximaal verklaard kan worden vanuit individueel- en groepsniveau;
2. gelijktijdige inbreng van de vier predictor-variabelen op individueel niveau en stapsgewijze eliminatie van de variabelen met niet-significante coëfficiënten. Een variabele wordt uit de analyse verwijderd indien de verhouding coëfficiënt-schatting : standaardfout kleiner is dan 2. Bij een volgende analyse over de dan nog resterende variabelen wordt erop gelet dat de grootte van het verschil tussen de devianties van de vorige en de actuele analyse niet statistisch significant wordt op 5%-niveau (dit verschil kan getoetst worden met behulp van een chi-kwadraat toets). De bedoeling van deze herhaalde procedure is om zo verantwoord mogelijk op individueel niveau die variabelen te selecteren, die het meest van belang zijn voor de statistische verklaring van de leesprestatie tijdens de nameting. Vergelijking van de residuele variantie op elk niveau met de respectievelijke varianties van het nulmodel geeft informatie over de mate, waarin de uiteindelijk geselecteerde leerlingvariabelen de nameting verklaren;
3. gelijktijdige toevoeging van de experimentele instructie-variabele EXPT en de overige groepsvariabelen aan de nog resterende individuele variabelen, gevolgd door stapsgewijze eliminatie van variabelen op groepsniveau met niet-significante coëfficiënten (vergelijk bij 2). Er worden nu die groepsvariabelen geselecteerd, die het meest van belang zijn in de statistische verklaring van de leesprestatie tijdens de nameting. Variabele EXPT blijft echter steeds in de analyse (om de relatieve invloed van deze experimentele variabele vast te kunnen stellen). Vergelijking van de residuele variantie op groepsniveau met die van de vorige modellen geeft informatie over de mate, waarin de uiteindelijk geselecteerde groepsvariabelen de nameting verklaren;
4. gelijktijdige toevoeging van de groepsspecifieke variabelen aan de geselecteerde individuele- en groepsvariabelen, gevolgd door stapsgewijze eliminatie van (1) niet van belang zijnde groepsspecifieke variabelen en (2) niet meer van belang zijnde individuele- en groepsvariabelen;
5. tenslotte wordt nagegaan of er, gegeven de oplossing onder 4, nog sprake is van random effecten (niet-gespecificeerde verschillen tussen groepen) met betrekking tot de individuele leesontwikkeling.

Bij dit stappenplan zijn nog twee opmerkingen te plaatsen. Ten eerste: tijdens de analyses bleek dat een model waarin zowel EXPT als IEXPT opgenomen waren, niet analyseerbaar was (omdat EXPT in dit onderzoek slechts als een dichotomie geoperationaliseerd kan worden, zijn EXPT en IEXPT statistisch afhankelijk). Om deze reden zijn beide variabelen "gelijkwaardig" en is IEXPT niet in de analyses betrokken.

Ten tweede: wanneer de individuele variabelen niet gecentreerd zijn rondom hun groepsge-middelde, zou het de voorkeur verdienen de opname van groepsspecifieke variabelen (produkt-ofwel interactie-variabelen) steeds vergezeld te laten gaan van de hoofdeffecten (de leerlingvariabele en de groepsvariabele), óók als de coëfficiënten hiervan niet significant zijn. Echter, deze procedure heeft het bezwaar dat opname van al deze afzonderlijke maar veelal gerelateerde variabelen leidt tot (1) een toename in multi-collineariteit en dus vertekening van de analyse-resultaten (Mooij & Jansen, 1990) en (2) een snelle toename van het aantal variabelen, waardoor een analyse slechts zou kunnen plaatsvinden over aanzienlijk grotere aantallen leerlingen en groepen. Omdat hier gepoogd wordt zicht te krijgen op de relatief meest belangrijke variabelen,

Tabel 3. Resultaten van VARCL-analyse inclusief SES (541 ln; 11 controle- en 17 experimentele groepen)

	nulmodel		indiv. model		indiv. + groepsmodel		ind./gr./grspec. model	
	coëf	stft	coëf	stft	coëf	stft	coëf	stft
FIXED EFF.								
overall intercept	-.08		.21		-.65		-.27	
indiv. niv.								
VRMT			2.12	.13	2.06	.13	6.28	.57
SEXE			.13	.05	.13	.05	.12	.05
SES			-.34	.14	-.36	.13	-	-
groepsniv.								
MLKMN					.04	.02	-	-
SDVMT					-1.33	.58	-	-
EXPT					.25	.11	.22	.08
gr.specif.								
IVRAN							-.08	.03
IVRMV							-1.67	.34
IVRSD							-7.16	.92
RANDOM EFF.								
indiv. niv.								
variantie	.53		.35		.35		.28	
groepsniv.								
variantie	.23		.08		.05		.02	
sigma	.48	.07	.29	.05	.23	.04	.16	.04
deviantie	1250.17		1011.08		1000.46		875.50	

mede ten behoeve van volgende verfijndere analyses, wordt hier gewerkt zoals aangegeven in het bovenstaande stappenschema. Elders (Mooij, 1990) zijn onder andere resultaten met gecentreerde variabelen opgenomen.

De eerste stapsgewijze VARCL-analyse wordt uitgevoerd over alle leerlingen die op de in Schema 1 vermelde variabelen (behalve IEXPT) geen missing data hebben. Wegens het relatief grote aantal ontbrekende SES-scores zijn dit 541 leerlingen uit 11 controle-groepen en 17 experimentele groepen. De uiteindelijke resultaten staan in Tabel 3. In de Tabel staat "coëf" voor coëfficiënt en "stft" voor standaardfout.

In de resultaten in Tabel 3 blijkt het volgende. In de kolom "nulmodel" is de maximaal te verklaren leerlingvariantie .53 en de maximaal te verklaren groepsvariantie .23; de verhouding leerlingvariantie : groepsvariantie is dus (ongeveer) 62% : 38%. De sigma (wortel) van de groepsvariantie is .48 met als standaardfout .07 (hetgeen duidt op significantie van de te verklaren residuele groepsvariantie).

Bij het individuele model wordt duidelijk dat, ten opzichte van het nulmodel, de opname van de begin-leesprestatie, de sexe en de SES de individuele residuele variantie doet afnemen van .53 naar .35 en de residuele variantie op groepsniveau van .23 naar .08 (individuele leertijd was in eerste instantie opgenomen maar is daarna verwijderd). Gezien het verschil in devianties (van

1250.17 naar 1011.08) en het verlies van drie vrijheidsgraden is dit een significant verschil tussen de twee modellen, ten gunste van het individuele model. In dit individuele model hebben een hogere begin-leesprestatie, het meisje-zijn en een lagere SES (afkomstig zijn uit een relatief "hoger milieu") een positief effect op de eind-leesprestatie.

De toevoeging van de groepsvariabelen (zie het individuele + groepsmodel) leidt tot een kleine verbetering van de verklaring van de residuele variantie op groepsniveau (van .08 naar .05). Naast de positieve effecten van het leerkrachtmanagement en de experimentele groepsvariabele (gebruik van de computer in de groep), is er tevens een negatief effect van de mate van heterogeniteit in begin-leesprestatie in de groep: leerlingen in groepen met kleinere verschillen in begin-leesprestaties behalen hogere eind-leesprestaties.

De toevoeging van de groepsspecifieke variabelen (zie de laatste kolom in Tabel 3) en de successievelijke eliminatie leidt uiteindelijk weer tot een reductie van de individuele- en groeps-residuele variantie. Dit komt ook duidelijk naar voren in de reductie in deviantie (van 1000.46 naar 875.50), met een gelijkblijvend aantal vrijheidsgraden.

Een volgend model met de mogelijkheid van random-coëfficiënten resulteerde in de constatering dat (significante) random-coëfficiënten afwezig zijn. Daarom is het vierde model in de laatste kolom van Tabel 3 ook het relatief beste model. De resultaten in dit model tonen:

- *positieve effecten van begin-leesprestatie en sexe* (meisjes scoren hoger),
- een positief effect van het *gebruik van de computer in de groep*,
- groepsspecifieke effecten: hoe *kleiner de groep*, of hoe *lager het begin-gemiddelde*, of hoe *homogener de groep qua begin-leesprestatie*, des te *hoger de eind-leesprestatie* resp. hoe beter de leesontwikkeling.

In dit model verklaart deze selectie uit de oorspronkelijke serie variabelen bijna de helft van de oorspronkelijke leerlingvariantie (van .53 naar .28) en is de variantie op groepsniveau gereduceerd van .23 naar .02.

In dit vierde model in Tabel 3 is SES geëlimineerd. Dit zou kunnen komen vanwege een relatief hoge correlatie met de begin-leesprestatie (vgl. de coëfficiënten in Tabel 3). De correlatie tussen SES en VRMT is echter niet bijzonder hoog (Pearson $r = -.23$; $p < .01$), zodat het erop lijkt dat de fixed groepsspecifieke effecten inderdaad alternatief zijn voor en relatief beter zijn dan de modellen zonder deze variabelen. Tevens maakt de eliminatie van SES het mogelijk analyses te verrichten op de variabelen in Schema 1 met uitzondering van SES (en IEXPT). Het belang hiervan is dat, vanwege de missing data op SES, de analyses uitgevoerd kunnen worden over 857 leerlingen in plaats van over 541 leerlingen terwijl het aantal groepen stijgt van 28 naar 49. De vraag is of dit verschillen oplevert in vergelijking met de tot nu verkregen resultaten.

De gehanteerde analyse-procedure is conform die in Tabel 3. De uiteindelijke resultaten komen nauw overeen met die in Tabel 3 en worden hier niet verder vermeld (zie Mooij, 1990).

5. DISCUSSIE

In par. 4.2 zijn positieve leereffecten gevonden van het werken met de VLL-computerprogramma's in de groep, naast de positieve effecten van begin-leesprestatie en sexe (meisjes scoren hoger) en enkele groepsspecifieke effecten: hoe kleiner de groep, hoe lager het begin-gemiddelde en hoe homogener de groep qua begin-leesprestatie, des te beter is de leesontwikkeling. Er bestaan dus aanwijzingen dat, onder constanthouding van de overige genoemde en van belang zijnde variabelen, het in de onderwijspraktijk in de groep kunnen oefenen met de VLL-computerprogramma's van positieve invloed is op de (vordering in) aanvankelijk leesprestatie.

Dit gegeven betekent een ondersteuning van de inzet van computerprogramma's in de praktijk, mede omdat vanwege de grote hoeveelheid verschillende leermaterialen in de methode VLL specifieke effecten van bepaalde onderdelen in deze methode zoals bijvoorbeeld computerprogramma's maar moeilijk te achterhalen zullen zijn: ook deze programma's passen "naad-

loos" in de methode. Bovendien bleek reeds (Mooij, 1990) dat de tijd die besteed wordt aan computerwerk ten dele ten koste gaat van de tijd die beschikbaar is voor speel/leermateriaal, en omgekeerd. Hierbij dient nog opgemerkt te worden dat de leerlingen relatief weinig tijd (kunnen) stoppen in het werken met de computer. Als er, zoals gebruikelijk, maar één computer per groep is, zal deze tijd-besteding per leerling laag blijven (volgens Tabel 2 zes minuten per dag), hetgeen het effect van het werken met deze programma's ook relatief klein zal doen uitvallen.

De hoeveelheid individuele leertijd droeg, in tegenstelling tot de verwachting, niet bij aan de vordering in leesprestatie. Een mogelijke verklaring kan liggen in de aard van de meting van deze individuele instructie-variabele (via de dagelijkse registratie door de leerkracht). Een betrouwbare en valide wijze van meting van een dergelijke variabele zal een grotere onderzoeksinvesteringsvragen dan beschikbaar gesteld werd in het kader van het huidige onderzoek (vanwege de opdrachtgever was meting van deze tijd een voorwaarde voor subsidie-verlening).

In overige analyses (Mooij, 1990) is gepoogd nader zicht te krijgen op de leerprocessen en -effecten. Zo kan het bijvoorbeeld zijn dat een grotere mate van individueel oefenen met de VLL-computerprogramma's de (vordering in) aanvankelijk leesprestatie doet toenemen. Deze veronderstelling werd nagegaan conform de hier beschreven analyse-procedure. De registratie van de computer-oefentijd verliep ook via de leerkracht: misschien was de (on)betrouwbaarheid van deze meting de reden dat er geen (significant) effect van dit oefenen op de vordering in aanvankelijk leesprestatie gevonden werd.

Er zijn echter ook andere verklaringen mogelijk. In de onderzoeksresultaten komt bijvoorbeeld naar voren dat, naarmate een leerling een hogere SES heeft (en dus afkomstig is uit een "lager sociaal milieu"), hij (1) minder individuele leertijd ter beschikking heeft, (2) meer met de computerprogramma's oefent en (3) lager scoort op de begin-leesprestatie en relatief gezien nog lager op de eind-leesprestatie. Met andere woorden: het meer met de computer oefenen komt niet in de leerresultaten tot uiting; eerder lijkt het erop dat de relatieve achterstand ten opzichte van leerlingen uit een "hoger sociaal milieu" groter wordt. Omdat met name de leerlingen uit een "lager sociaal milieu" zowel lagere prestaties behalen als meer oefenen met de computerprogramma's zonder hun prestatie in vergelijking met de andere leerlingen te verbeteren, zouden er vanuit dit individueel oefenen met de computer *in de totale groep onderzochte leerlingen* geen duidelijke effecten gevonden kunnen worden.

Mede om dit na te gaan is, ten tweede, het totaal aantal leerlingen op grond van de absolute en ook de relatieve (gecentreerde) begin-leesprestatie in drie leerlingen-prestatie-groepen verdeeld: laag presterende leerlingen, midden presterende leerlingen en hoog presterende leerlingen (zie Mooij, 1990). Binnen elk van deze zes leerlingengroepen zijn multi-level analyses uitgevoerd die vergelijkbaar zijn met de hier gepresenteerde. Een consistent resultaat was dat in zowel de absolute als de relatieve *hoge* leerlingen-prestatie-groep een positief effect van het gebruik van de computerprogramma's in de groep op de leesontwikkeling gevonden werd. Dit is plausibel in het licht van de hierboven weergegeven onderzoeksresultaten en hun interpretatie.

Ten derde: de begin- en eind-leesprestatie-score bestaan uit conglomeraten van vijf afzonderlijke leestoetsen. Het is mogelijk dat er ten aanzien van de verschillende onderdelen van het aanvankelijk leesproces verschillende leereffecten zijn, die gedeeltelijk tegen elkaar wegvallen bij het gebruik van de combinatie-score. Dit mede omdat in de periode van onderzoek in het bijzonder de programma's 4, 5 en 6 gebruikt werden, die mogelijk ook slechts specifieke effecten teweegbrengen. Ter exploratie hiervan is de stapsgewijze analyse per afzonderlijke leestoets uitgevoerd over steeds alle leerlingen in de 49 groepen (Mooij, 1990). De resultaten laten onder andere zien dat het gebruik van de computer in de groep in het geval van de Caesar-één-minuut toets een significant positief effect heeft. Verder blijkt dat sexe een rol speelt bij de ontwikkeling van het technisch lezen (behalve bij de grafemtoets) en het begrijpend lezen: meisjes behalen hier hogere eind-leesprestaties dan jongens. Bij spellingvaardigheid doet zich dit sexe-effect niet voor.

Geconcludeerd wordt dat de multi-niveau aanpak in theorie en onderzoek van aanvankelijk leesprocessen vruchtbaar werkt en aanleiding kan zijn om "oude" problemen op een nieuwe en adequatere wijze te benaderen. In *theoretisch* opzicht dienen de nu verkregen resultaten nog

eens nader bekeken te worden, en dienen ook de ermee verbonden methodologische en analyse-technische problemen (zoals de relatie tussen hoofd- en interactie-effecten en het optreden van multi-collineariteit) scherper omlijnd en van "hanteerbare" oplossingen voorzien te worden. In elk geval stimuleren de huidige onderzoeksresultaten de verdere uitwerking van het onderwijs-leerproces als een *interactioneel multi-niveau proces*.

LITERATUUR

- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 149, 1-43.
- Ames, C. (1984). Competitive, cooperative and individualistic goal structures: a cognitive-motivational analysis. In R. Ames & C. Ames (Eds.), *Student motivation* (pp. 177-207). New York: Academic Press.
- Ark-Berberisschool & J. van Kuyk (1980). *Zelfstandig werken*. Tilburg: Zwijsen.
- Beishuizen, J.J. (1987). Computer en taalonderwijs. *Didaktief*, 17(9), 25.
- Berg, G. van den (1987). *Effectief evalueren*. Lisse: Swets & Zeitlinger.
- Bowerman, W.R. (1978). Subjective competence: the structure, process and function of self-referent causal attributions. *Journal for the Theory of Social behaviour*, 8, 45-75.
- Burstein, L. (1980). The role of levels of analysis in the specification of education effects. In R. Dreeben & J. A. Thomas (Eds.), *Analysis of educational productivity. Volume I: Issues in microanalysis* (pp. 119-190). Cambridge: MA, Ballinger Press.
- Caesar, F.B. (1980). *Veilig leren lezen. Handleiding 3*. Tilburg: Zwijsen.
- Cattell, R.B., & Child, D. (1975). *Motivation and dynamic structure*. London: Holt, Rinehart and Winston.
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation*. Chicago: Rand McNally.
- Corte, E. de (1979). Objecten, doelen en methodologie van de onderwijspsychologie. *Tijdschrift voor Onderwijsresearch*, 4, 209-218.
- Creemers, H.P.M. (1974). *Evaluatie van onderwijsstijlen binnen het aanvankelijk lezen*. Utrecht: Elin-kwijk.
- Cronbach, L.J. (1983). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Daal, V.H.P. van, Ley, A. van der, Bakker, N.C.M., & Reitsma, P. (1987). Een computergestuurd ortho-didactisch programma voor aanvankelijk lezen (COPAL). *Pedagogische Studiën*, 64, 364-376.
- Dongen, D. van (1984). *Leesmoeilijkheden; naar diagnostiserend onderwijzen bij het leren lezen*. Tilburg: Zwijsen.
- Dongen, D. van, & Mommers, M.J.C. (z.j.). *Lees- en spellingtoetsen voor leerjaar 1 tot 4 van de lagere school*. Nijmegen: Katholieke Universiteit, Vakgroep Interdisciplinaire Onderwijskunde.
- Eeden, P. van den (1987). Theorie en model in het multilevel onderzoek. Conferentiebundel voor de zevende onderwijssociologische conferentie: *Onderwijs en sociale beheersing* (pp. 32-45). Amsterdam: SISWO.
- Emmer, E.T., & Evertson, C.M. (1980). *Effective management at the beginning of the school year in junior high classes*. Austin: University of Texas, Research and Development Center for Teacher Education.
- Evertson, C.M., Emmer, E.T., & Clements, B.S. (1980). *Report of the methodology, rationale and instrumentation of the Junior High Classroom Organization Study*. Austin: University of Texas, Research and Development Center for Teacher Education.
- Glass, G.V., Cahen, L.S., Smith, M.L., & Filby, N.N. (1982). *School class size*. Beverly Hills: Sage.
- Guchte, C. van de (1988). Mondeling Nederlands bij Veilig leren lezen. *Didaktief*, 18(7), 26-27.
- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53, 87-97.
- Heckhausen, H. (1980). *Motivation und Handeln: Lehrbuch der Motivationspsychologie*. Berlin: Springer.
- Heckhausen, H. (1986). Why some time out might benefit achievement motivation research. In J.H.L. van den Bercken, E.E.J. de Bruyn, & Th.C.M. Bergen (Eds.), *Achievement and task motivation* (pp. 7-39). Berwyn and Lisse: Swets & Zeitlinger.
- Hofstee, W.K.B. (1987). Het causaliteitsvraagstuk in evaluatie-onderzoek. In W.K.B. Hofstee (Ed.), *Evaluatiemethodologie* (pp. 93-108). Lisse: Swets & Zeitlinger.
- Hofstee, W.K.B., Schaveling, J., & Koster, K.B. (z.j.). *Beleidsanalytische Verkenning Onderwijs-Voor-rangs-Beleid op Schoolniveau*. Zonder plaats en uitgever.
- Hove, J.R. ten (z.j.). *Toelichting speelleeset*. Tilburg: Zwijsen.

- Ippel, M.J., & Elshout, J.J. (1990) (Eds.), *Training van hogere-orde vaardigheden*. Lisse: Swets & Zeitlinger.
- Krampen, G., & Mory, M. (1982). *Zur kognitiv-emotionalen Verarbeitung schlechter Zensuren bei Schülern*. Trier: Universität Trier, Trierer Psychologische Berichte.
- Kuhl, J. (1986). Motivation and information processing. In R.N. Sorrentino & E.T. Higgins (Eds.), *The handbook of motivation and cognition: foundations of social behavior* (pp. 404-434). Berlin: Springer.
- Lens, W. (1987). Motivatie op school: een theoretische benadering. *Pedagogisch Tijdschrift*, 12, 280-300.
- Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.
- Longford, N.T. (1988). *VARCI Manual*. Princeton: Educational Testing Service.
- Mommers, M.J.C. (1988). Geautomatiseerde kennis grafemen. *Jeugd in School en Wereld*, 72(10), 28-29.
- Mooij, T. (1982). Onderwijsleersituatie en lesondergravend gedrag van lto-leerlingen. In E. Diekerhof (Ed.), *Leren, wat moet je ermee?* (pp. 39-52). Muiderberg: Coutinho.
- Mooij, T. (1987). *Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.
- Mooij, T. (1990). *Leerprocessen en effecten van de computerprogramma's bij: "Veilig Leren Lezen"*. Nijmegen: Instituut voor Toegepaste Sociale wetenschappen.
- Mooij, T., & Jansen, R. (1990). Theory and analysis of multi-level processes and effects: a two-level example on pupil achievement, competence, and orientation. In P. van den Eeden, J. Hox, & J. Hauer (Eds.), *Theory and analysis of multi-level research: convergence or divergence?* Amsterdam: SISWO.
- Neisser, U. (1976). *Cognition and reality*. San Francisco: Freeman.
- Pervin, L.A., & Lewis, M. (1978). Overview of the internal-external issue. In L.A. Pervin & M. Lewis (Eds.), *Perspectives in interactional psychology* (pp. 1-22). New York: Plenum Press.
- Reid, K., Hopkins, D., & Holly, P. (1987). *Towards the effective school*. Oxford: Basil Blackwell.
- Reijnders, H. (1980). *Stap voor Stap lezen. De methode 'Veilig Leren Lezen' bewerkt voor buitengewoon onderwijs en remedial teaching*. Tilburg: Zwijzen.
- Reitsma, P. (1987). *Een sprekende computer als oefenmiddel bij leesmoeilijkheden (Manuscript 598)*. Eindhoven: IPO.
- Roth, S.F., & Beck, I.L. (1987). Theoretical and instructional implications of the assessment of two microcomputer word recognition programs. *Reading Research Quarterly*, 22, 197-217.
- Seegers, G.H.J., Aarnoutse, C.A.J., & Mommers, M.J.C. (1987). *Lezen: een taalkundig-psychologische benadering*. Tilburg: Zwijzen.
- Spaai, G.W.G., Ellermann, H.H., & Reitsma, P. (1986). Effects of two forms of sound feed-back on learning to read words. In: *IPO-annual of progress report, Vol. 21* (pp. 77-85). Eindhoven: IPO.
- Starren, J., Bakker, S.J., & Wissel, A. van der (1988). *Inleiding in de onderwijspsychologie*. Muiderberg: Coutinho.
- Tesser, Ph.T.M. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen: Instituut voor Toegepaste Sociale wetenschappen.
- Tomic, W. (1985). *Docentengedrag en leerresultaten*. Enschede: Febodruk.
- White, R.W. (1959). Motivation reconsidered: the concept of competence. *Psychological Review*, 66, 297-333.

Manuscript ontvangen 22-3-1990

Definitieve versie ontvangen 11-9-1990

De stabiliteit van schooleffecten

J. Roeleveld, U. de Jong en P. Koopman

Stichting Centrum voor Onderwijsonderzoek, Universiteit van Amsterdam*

ABSTRACT

In this article the stability of measures for the effectiveness of schools for secondary education is investigated. Analyses on data of three successive cohorts of students entering eleven of these schools in the city of Amsterdam are reported. First of all, the educational output of the schools show to be fairly stable over the three cohorts. As a second step, multilevel modelling techniques are used to arrive at output measures, which are adjusted for differences in student intake. On these adjusted output measures the schools show more instability from one cohort to the other. Finally, the within school relations between students' initial characteristics and their educational results show considerable variation over cohorts. It is concluded that the findings cast doubt on the adequacy of a concept of effectiveness as an inherent property of schools as entities.

1. INLEIDING

In dit artikel staat de stabiliteit van maten voor de effectiviteit van scholen voor voortgezet onderwijs centraal. Maar allereerst gaan we nader in op de vraag hoe de effectiviteit van scholen onderling kan worden vergeleken.

Gewoonlijk wordt de effectiviteit van een school relatief ten opzichte van andere scholen vastgesteld. Van een aantal scholen kan de effectiviteit worden berekend en scholen die duidelijk boven het gemiddelde zitten worden dan effectieve scholen genoemd en scholen die onder het gemiddelde zitten worden ineffectief genoemd. Scholen worden dus onderling vergeleken aan de hand van hun effectiviteitsscore, terwijl eventueel ook een of ander nationaal gemiddelde of norm, indien bekend, als referentiepunt kan dienen. Het is dan belangrijk dat een dergelijke vergelijking tussen scholen zo eerlijk mogelijk gebeurt. Daarbij kunnen twee benaderingswijzen worden onderscheiden.

In de eerste benaderingswijze worden simpelweg outputmaten van een school berekend. Om de effectiviteit van scholen te bepalen kan men bijvoorbeeld een test afnemen onder de leerlingen, of men kan eenvoudig het percentage gediplomeerden berekenen, enzovoorts. De resultaten van deze berekeningen kunnen dan gebruikt worden om scholen onderling te vergelijken, ten opzichte van hun gemiddelde of ten opzichte van een nationaal gemiddelde. In termen van het veel gebruikte input-proces-output model van scholen (cf. Scheerens et al., 1988) wordt dus de output van scholen gebruikt om scholen onderling te vergelijken. Cuttance noemt dit het 'standaard model' van effectiviteit en vat de kritiek op deze benadering als volgt samen: "Probably the major single problem (.....) is its failure to take account of the intake characteristics of the pupils in schools, thus it cannot indicate which schools are the most effective in terms of the progress their pupils make." (Cuttance, 1987, p. 2).

In de tweede benadering wordt juist wel met de verschillen in aanvangskenmerken van leerlingen op een school rekening gehouden en worden de effectiviteitsmaten van een school op de een of andere manier gecorrigeerd voor deze verschillen. In deze benadering worden scholen dus vergeleken op een outputmaat die is gecorrigeerd voor aanvangsverschillen van de leerlingen en Cuttance (op. cit.) noemt dit dan ook het 'voor aanvangsverschillen gecorrigeerde model' ("intake adjusted model") van effectiviteit.

* SCO, Grote Bickersstraat 72, 1013 KS Amsterdam.

Zo'n statistische correctie gebeurde tot voor kort meestal door op individueel niveau een regressiemodel te schatten, waarin de onderwijsresultaten van leerlingen beïnvloed worden door hun aanvangskenmerken, zoals intelligentie, schoolprestaties, motivatie, sociale achtergrond en dergelijke. Meer recentelijk wordt de effectiviteit van een school echter opgevat als een conditionerende factor, die de effecten van zulke aanvangskenmerken modificeert. Deze effecten kunnen daarmee van school tot school verschillen en de ene school kan voor de ene groep leerlingen (bijvoorbeeld uit hogere sociale milieus) effectiever zijn en de andere school voor een andere groep leerlingen (bijvoorbeeld uit lagere sociale milieus). Wanneer we dergelijke effecten van scholen op het loopbaanproces van hun leerlingen willen vaststellen, dan is een zuiver individueel model waarbij de effecten van aanvangskenmerken voor alle leerlingen, ongeacht hun school, gelijk worden geschat niet langer adequaat. Juist wanneer men op zoek is naar verschillen tussen scholen in effectiviteit zullen ook verschillen tussen de scholen in effecten van aanvangskenmerken in het model moeten worden opgenomen. Een nauwkeurige bepaling van de effectiviteit van scholen, die rekening houdt met de individuele aanvangsverschillen tussen de leerlingen op de scholen, zal daarom uit moeten gaan van de recent ontwikkelde multi-niveau benaderingen, waarin expliciet de nesting van leerlingen binnen scholen is gemodelleerd (Aitkin & Longford, 1986; Raudenbush & Bryk, 1986; Cuttance, 1987).

De hiërarchische structuur van het voortgezet onderwijs in Nederland voegt nog een extra dimensie toe aan het maken van een eerlijke vergelijking tussen scholen. Het Nederlandse voortgezet onderwijs kent verschillende niveaus, van Lbo via Mavo en Havo tot Vwo. Daarnaast verschillen scholen ook nog naar breedte van de aangeboden niveaus, van categoriale scholen tot zeer brede scholengemeenschappen. Als gevolg hiervan bestaat er een heel scala aan typen scholen voor voortgezet onderwijs.

Een goede en eerlijke vergelijking tussen scholen moet rekening houden met deze onderwijsstructuur, omdat anders de verschillen in resultaten van scholen niet zozeer verschillen in effectiviteit weergeven, maar vooral de structuurverschillen tussen scholen weerspiegelen. Kreft (1987) bijvoorbeeld heeft geen rekening gehouden met het feit dat scholen voor voortgezet onderwijs naar niveau en breedte verschillen. Zij concludeert dan ook dat categoriale scholen voor Vwo, gymnasia, de beste resultaten behalen en categoriale Mavo-scholen de slechtste. Maar categoriale Vwo-scholen zijn bedoeld voor de betere leerlingen, beter naar prestatieniveau en motivatie, en categoriale Mavo's zijn bedoeld voor onderwijs aan minder goede leerlingen. Vandaar ook dat de uitkomsten van Kreft nogal tautologisch zijn en eigenlijk alleen de hiërarchische structuur van het Nederlandse voortgezet onderwijs weerspiegelen. Ook Bosker c.s. (Bosker et al., 1988) hebben in één analyse scholen van een verschillend type meegenomen. Zij onderzoeken, net als wij in dit artikel zullen doen, de stabiliteit van uitkomsten van scholen over meerdere jaren en rapporteren een zeer hoge stabiliteit van "algemene schoolkwaliteit", een maat die rekening houdt met de aanvangsverschillen tussen leerlingen. Maar deze stabiliteit is op zijn minst voor een gedeelte afhankelijk van de structuurverschillen tussen scholen van een verschillend type. En zulke structuurverschillen zijn vanzelfsprekend nogal stabiel in de tijd.

Samenvattend concluderen wij, ten eerste, dat ten behoeve van een goede en eerlijke vergelijking tussen scholen voor voortgezet onderwijs scholen van eenzelfde type moeten worden vergeleken. Ten tweede mag niet louter gekeken worden naar de output van scholen, maar moet evenzeer rekening worden gehouden met de aanvangsverschillen tussen de leerlingen, waarbij multi-level benaderingen de aangewezen methode vormen.

2. PROBLEEMSTELLING: DE STABILITEIT VAN EFFECTIVITEIT

In het meeste onderzoek naar schooleffectiviteit wordt vooral gezocht naar schoolkenmerken die de verschillen in effectiviteit tussen scholen kunnen verklaren. De impliciete assumptie van dergelijk onderzoek is dat effectiviteit een relatief stabiel kenmerk van een school is en niet een gevolg van min of meer toevallige processen, die van jaar tot jaar kunnen verschillen.

Omdat veel onderzoek naar schooleffectiviteit is gebaseerd op de schoolloopbanen van leerlingen van één enkel cohort, kan deze assumptie gewoonlijk niet worden getoetst. De resultaten van die paar onderzoeken waarin meerdere leerlingcohorten worden onderzocht doen echter vermoeden dat de genoemde assumptie twijfelachtig is. Goldstein (1987) bijvoorbeeld rapporteert verschillen in effectiviteit tussen cohorten die even groot zijn als de verschillen in effectiviteit tussen scholen. Gray en Jones (1985) geven voorbeelden van verschillen tussen cohorten, zowel met betrekking tot de effectiviteit van leraren als scholen. Als het waar is dat de effectiviteit van een school van jaar tot jaar verschilt, heeft het weinig zin om verschillen tussen scholen te gaan verklaren met schoolkenmerken die verondersteld worden (relatief) stabiel te zijn over jaren, zoals het schoolklimaat.

In dit artikel zullen we onderzoeken of de effectiviteit van een school opgevat kan worden als een stabiel kenmerk van een school. We zullen dat doen door het vergelijken van schattingen voor effectiviteit van scholen voor voortgezet onderwijs aan de hand van gegevens van drie opeenvolgende leerlingcohorten die op die scholen starten in de eerste klas. Hoewel zij niet in dezelfde klassen zitten, volgen de leerlingen uit deze drie cohorten een aantal jaren onderwijs op dezelfde school, met dezelfde directeur, hetzelfde docentencorps, schoolklimaat en dergelijke. Als de effectiviteit van een school inderdaad een inherent en stabiel schoolkenmerk is, zullen over alle drie de cohorten steeds dezelfde scholen al of niet effectief moeten zijn. Als daarentegen wordt gevonden, dat de effectiviteit van een school substantieel verschilt voor leerlingen van verschillende cohorten, zal getwijfeld moeten worden aan de zin van het concept 'schooleffectiviteit' als een inherent kenmerk of eigenschap van een school.

In paragraaf 3 worden de dataset en de variabelen besproken. In paragraaf 4 staat de stabiliteit van outputmaten van scholen centraal. In paragraaf 5 zullen we aandacht schenken aan de stabiliteit van de voor inputkenmerken van leerlingen gecorrigeerde schattingen van effectiviteit. In paragraaf 6 zal dan de stabiliteit van de relatie tussen output en input aan de orde komen. Het artikel wordt in paragraaf 7 afgesloten met een discussie van de resultaten.

3. DATASET EN VARIABELEN

In de analyse zal gebruik worden gemaakt van gegevens uit het Amsterdamse VOCAL-bestand (zie Koopman e.a., 1986). In deze database is onderwijsinformatie opgeslagen van alle leerlingen die tussen 1969 en 1982 in Amsterdam startten in het voortgezet onderwijs. Er zijn gegevens verzameld tot en met 1982 en het VOCAL-bestand bestaat aldus uit 14 langer of korter gevolgde cohorten. Van de leerlingen tot en met cohort 1975 kunnen schoolloopbanen worden gereconstrueerd van minimaal 8 jaren.

In 1986 zijn op een groot aantal scholen uit het VOCAL-bestand aanvullende gegevens verzameld, met name over het beroep van de ouders van leerlingen. Voor de analyses zullen we gebruik maken van gegevens over de cohorten 1973, 1974 en 1975 uit dit aangevulde bestand. De scholen in de analyse zijn van het type scholengemeenschap Havo+Vwo. Van de 16 scholen van dit type hebben er 11 aan de aanvullende dataverzameling meegewerkt met in de drie opeenvolgende cohorten achtereenvolgens in totaal 1.496, 1.562 en 1.568 leerlingen. Het aantal leerlingen in een cohort dat in de eerste klas op een school start varieert van 55 tot 243.

De volgende achtergrondvariabelen van de leerlingen zullen in deze analyse voorkomen:

- Geslacht, SEXE, gecodeerd als 0 = jongen en 1 = meisje.
- Sociaal economische status, SES, gebaseerd op het beroep van de vader van een leerling. Het beroep is gecodeerd volgens de beroepenclassificatie van het CBS (CBS, 1984). Bakker (1987) heeft deze classificatie met behulp van een panel van deskundigen ingedikt tot 83 categorieën. Vervolgens is op een landelijke steekproef van leerlingen die in 1977 zijn gestart in het voortgezet onderwijs, het SMVO-cohort, een homogeniteitsanalyse uitgevoerd op deze 83 categorieën en het opleidingsniveau van de ouders van de leerlingen. Aldus is de variabele SES ratio geschaald van -2 tot en met +5.

Drie kenmerken van de schoolloopbaan tijdens het basisonderwijs zullen worden gebruikt als indicatie van de geschiktheid van een leerling om het Havo of het Vwo succesvol af te ronden:

- Prestatieniveau, PREST, de totaalscore van de gehele CITO-toets. De maximum te behalen score is 300.
- Advies, ADV, het voortgezet onderwijs-advies van het hoofd van de lagere school, gecodeerd van 1 = Lbo tot en met 4 = Vwo.
- Vertraging, VTLO, het aantal jaren vertraging dat een leerling heeft opgelopen op de basis-school.

Met de beschikbare gegevens kan de schoolloopbaan van een leerling in het voortgezet onderwijs beschreven worden in termen van acht opeenvolgende onderwijsposities, OP1 tot en met OP8, volgens de formele evaluatieschaal (FES) van het Nederlandse onderwijssysteem. Deze schaal blijkt zeer valide na vergelijking met uitkomsten van homogeniteitsanalyses (Roeleveld e.a., 1989). In de schaal wordt zowel het gevolgde onderwijstype tot uitdrukking gebracht, als het aantal jaren dat een leerling succesvol heeft afgerond binnen het onderwijstype, inclusief het behalen van een diploma. De onderwijspositie Lbo-1 wordt met 10 punten gewaardeerd, Mavo-1 met 20, Havo-1 met 30 en Vwo-1 met 40 punten. Een brugklas of gezamenlijk leerjaar krijgt het gemiddelde van de samenstellende delen; een Havo-Vwo brugklas krijgt bijvoorbeeld de score 35. Vervolgens wordt elke succesvolle overgang naar een volgende klas gewaardeerd met 10 punten extra. Het behalen van het diploma kan dan ook worden gewaardeerd als een succesvolle overgang. Aan de formele overgangsregels van het Nederlandse onderwijssysteem wordt voldaan, doordat de overstap van bijvoorbeeld Mavo-4 naar Havo-5 gelijk wordt gewaardeerd met het behalen van het Mavo-diploma. Ook het behalen van het Havo-diploma wordt gelijk gewaardeerd als de overstap van Havo-5 naar Vwo-5.

In dit artikel zullen twee onderwijsposities worden gebruikt als schoolloopbaanvariabelen van een leerling:

- De onderwijspositie bij de start van het tweede jaar, OP2.
- De laatste onderwijspositie op de startschool, EIND. Dit is niet noodzakelijk het hoogst bereikte onderwijsniveau van een individuele leerling, omdat een schoolloopbaan op een andere school vervolgd kan worden. Maar het is wel de hoogste positie aan de hand waarvan de resultaten van een school bepaald kunnen worden.

Om de stabiliteit van outputmaten te testen zullen we allereerst per school het percentage gediplomeerden en het gemiddelde van deze laatste onderwijspositie berekenen, over drie cohorten. De cohorten kunnen opgevat worden als een steekproef uit de tijd, zodat met een F-toets nagegaan kan worden of deze gemiddelden al of niet significant van elkaar verschillen. Verder zullen we de correlatie tussen de afzonderlijke outputmaten van de drie opeenvolgende cohorten berekenen. Een hoge correlatie indiceert dat de output van een school in het ene cohort vergelijkbaar is met die in het andere cohort. Een lage of negatieve correlatie indiceert instabiliteit tussen cohorten.

We zullen random coëfficiënt modellen (De Leeuw en Kreft, 1986; Aitkin en Longford, 1986) gebruiken om de schoolspecifieke relaties na te gaan tussen input en output. Door gebruik te maken van deze modellen krijgen we de gewenste voor aanvangsverschillen gecorrigeerde outputmaten. Deze modellen zullen aan de orde komen in de paragrafen 5 en 6.

4. DE STABILITEIT VAN OUTPUTMATEN

Als eerste outputmaat hanteren we het percentage leerlingen dat op de startschool het Havo- en/ of Vwo-diploma behaalt. In Tabel 1 worden voor elke school de slaagpercentages over de drie cohorten weergegeven, alsmede de correlaties tussen deze percentages.

De correlaties in Tabel 1 laten zien dat er een tendens is dat scholen met een hoog slaagpercentage in het ene cohort ook een hoog slaagpercentage hebben in de andere cohorten. Maar

Tabel 1. Percentage leerlingen dat geslaagd is op de startschool; over drie cohorten Havo-Vwo leerlingen.

	1973	1974	1975
school			
1	42.4	45.7	40.4
2	58.7	45.0	45.6
3	57.5	60.2	46.6
4	51.6	52.3	45.7
5	52.8	42.4	47.8
6	45.5	60.0	37.7 **
7	70.9	67.2	61.7
8	40.6	39.6	26.3 **
9	57.3	66.9	60.3
10	55.3	57.1	45.2
11	51.5	54.7	41.7
correlatie tussen de cohorten:			
1973		.61	.69
1974			.86

** = significant op 1%

deze correlatie is verre van perfect en de percentages per school indiceren dat er veel variatie is tussen de cohorten. School 2 bijvoorbeeld, heeft het op één na hoogste slaagpercentage in cohort 1973, maar scoort in de cohorten 1974 en 1975 zowel absoluut als relatief beduidend lager. Wij trekken hieruit de conclusie dat het gevaarlijk is om op basis van slechts één cohort deze outputmaat als evaluatiecriterium te gebruiken.

Tussen leerlingen die geen diploma op hun startschool halen bestaan nog grote verschillen in onderwijsniveau. Sommigen verlaten bijvoorbeeld het onderwijs al na zittenblijven in een 3-Havo klas, terwijl anderen na een overgang naar Vwo-6 de startschool verlaten om op een andere school hun voortgezet onderwijs voort te zetten. Dergelijke verschillen hebben geen invloed op het percentage gediplomeerden, maar komen wel tot uitdrukking in de variabele EIND, de laatst behaalde onderwijspositie op de startschool. De gemiddelde score van de leerlingen van een school op de variabele EIND is daarom een nauwkeuriger maat voor de output van scholen, hoewel dit gemiddelde minder direct te interpreteren is dan een slaagpercentage.

In Tabel 2 worden per school de gemiddelden van EIND over de drie cohorten weergegeven. Om deze tabel goed te interpreteren moet in het oog gehouden worden dat een verschil in gemiddelde tussen scholen van 10 punten kan betekenen dat op de betere scholen leerlingen, gemiddeld, één jaar langer onderwijs volgen op hetzelfde niveau. Of, wat in principe ook mogelijk is, dat de leerlingen evenveel jaren onderwijs op de startschool volgen, maar op Vwo- in plaats van Havo-niveau. In de praktijk zal een verschil in gemiddelde tussen scholen het gevolg zijn van een combinatie van zowel het aantal jaren dat de leerlingen onderwijs volgen op de startschool als ook het niveau waarop zij onderwijs volgen.

Uit Tabel 2 kunnen we lezen dat slechts op één school het gemiddeld behaalde onderwijsniveau significante verschillen tussen de cohorten vertoont. Op alle andere scholen is het verschil tussen de cohorten in de gemiddelden niet significant. Deze stabiliteit komt ook tot uitdrukking in de correlaties tussen de drie cohorten. Deze output van een school over het ene cohort verschilt niet veel van die in de andere twee cohorten. Een volgende illustratie voor deze stabiliteit wordt gegeven in Figuur 1, waarin de relatieve positie van de scholen wat betreft het overall gemiddelde op EIND voor de cohorten 1973 en 1975 zijn geplot.

Tabel 2. Gemiddelde laatst behaalde onderwijspositie op de startschool (EIND); over drie cohorten Havo-Vwo leerlingen.

	1973	1974	1975
school			
1	61.0	65.7	62.2
2	70.1	64.8	68.4
3	68.3	71.4	71.5
4	71.5	73.7	69.0
5	73.9	69.1	71.3
6	79.1	78.4	74.5
7	83.2	79.9	80.4
8	64.6	65.4	57.6 *
9	75.0	77.9	72.9
10	78.0	77.5	77.5
11	72.0	75.5	68.4
correlatie tussen de cohorten:			
1973		.85	.81
1974			.91

* = significant op 5%

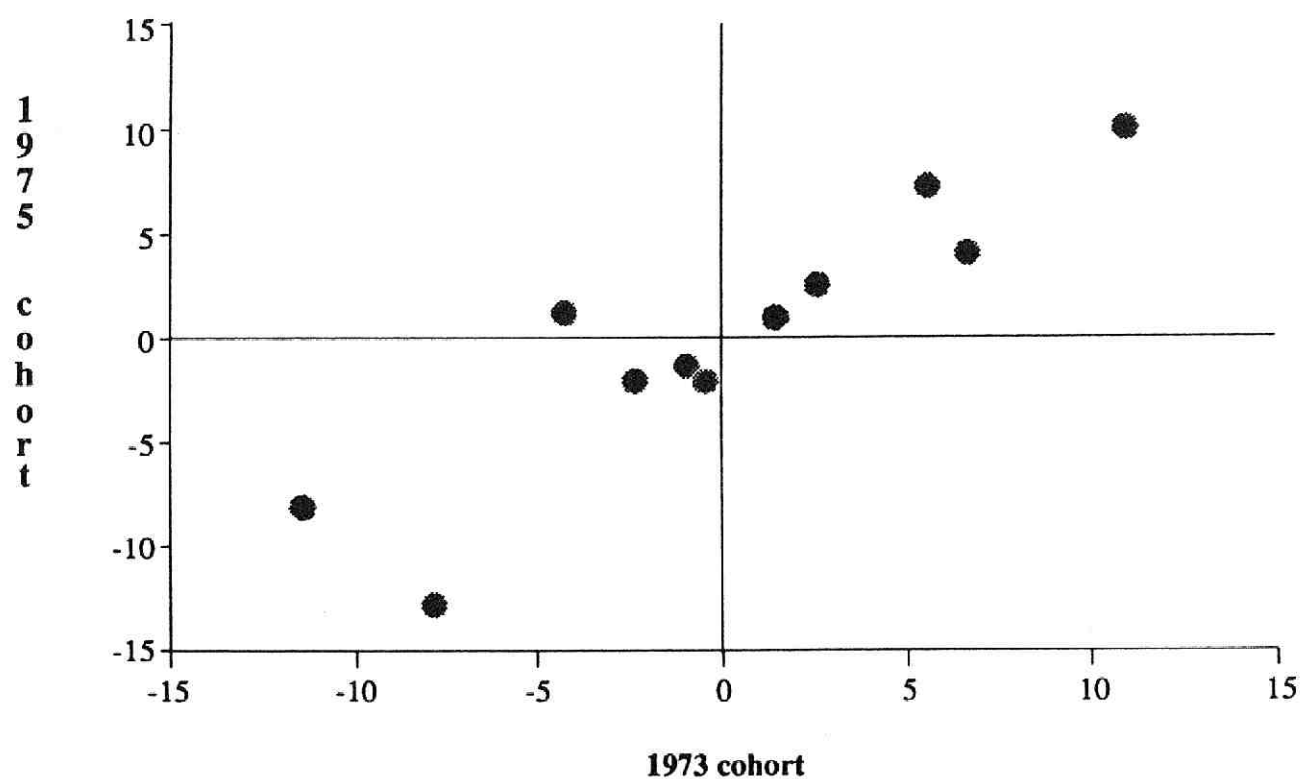


Fig. 1. De stabiliteit van de output van scholen.

Horizontaal de afwijking van het overall gemiddelde van EIND voor het cohort 1973; verticaal de afwijking voor het cohort 1975.

Ook uit Figuur 1 blijkt een redelijke mate van stabiliteit: scholen die in het ene cohort een score hebben boven het gemiddelde, hebben dat ook in het andere cohort.

Onze conclusie luidt dan ook dat, hoewel scholen verschillen van het ene naar het andere cohort wat betreft het percentage leerlingen dat een diploma behaalt, hun output op de gevoeliger maat 'gemiddeld behaalde onderwijspositie op de startschool' redelijk stabiel is. Deze laatste maat lijkt dan ook geschikt om scholen te onderscheiden naar lage en hoge output.

5. STABILITEIT VAN DE VOOR AANVANGSVerschillen GECORRIGEERDE OUTPUT

Zoals in paragraaf 1 beargumenteerd moeten we bij het vergelijken van scholen rekening houden met verschillen in aanvangskenmerken van leerlingen. Maar voordat we hier mee verder gaan, willen we eerst even kort stilstaan bij de stabiliteit van deze leerlingkenmerken van de scholen over de cohorten. In Tabel 3 worden correlaties weergegeven van de gemiddelde sociaal economische status (SES), gemiddeld prestatieniveau (PREST) en gemiddeld advies (ADV) van de leerlingen die op de onderzochte scholen hun voortgezet onderwijs starten.

Over het geheel genomen lijken de gemiddelde aanvangskenmerken van leerlingen op een school redelijk stabiel. De stabiliteit van de gemiddelde sociaal economische status kan er op duiden dat scholen elk jaar leerlingen recruterend uit dezelfde buurten van Amsterdam. De iets grotere stabiliteit van het gemiddelde advies van de leerlinginstroom, vergeleken met het gemiddelde prestatieniveau, is waarschijnlijk het gevolg van de centrale rol die het advies speelt bij het toelatingsbeleid van scholen voor voortgezet onderwijs.

Om te corrigeren voor aanvangsverschillen tussen leerlingen van scholen, moeten de relaties tussen individuele leerlingkenmerken en hun schoolloopbanen worden gemodelleerd. Voor de schoolloopbaan gebruiken we zoals gezegd twee variabelen: de onderwijspositie in het tweede jaar, OP2, en de laatste behaalde onderwijspositie op de startschool, EIND. In OP2 komt de uitkomst van het selectieproces bij de belangrijke overgang na het gemeenschappelijke brugjaar tot uitdrukking. Na een succesvol brugjaar moet op dit moment immers worden gekozen voor òf Havo, òf Vwo. En als het brugjaar niet succesvol is verlopen moet er gekozen worden tussen zittenblijven of afstromen naar een lager onderwijstype. Analyses met als afhankelijke variabele OP2 laten daarmee verschillen tussen scholen zien in succes tijdens de brugperiode.

Het succes van een leerling in de brugperiode is direct van invloed op het uiteindelijke onderwijsresultaat, maar ook na de brugperiode vinden op de scholengemeenschappen nog allerlei selectie- en keuzeprocessen plaats. In de variabele EIND komt het uiteindelijke resultaat van deze processen tot uitdrukking.

Om de relaties tussen de leerlingkenmerken en deze schoolloopbaanvariabelen te modelleren maken we gebruik van random coëfficiënt modellen (RC-modellen; zie De Leeuw & Kreft, 1986; Aitkin & Longford, 1986). In deze multi-niveau modellen worden de individuele onderwijsresultaten van leerlingen geregresseerd op individuele achtergrondkenmerken van die leerlingen. De coëfficiënten van deze regressievergelijking kunnen daarbij van school tot school

Tabel 3. Correlaties tussen cohorten van gemiddelde kenmerken van de instroom van leerlingen.

	SES	PREST	ADV
tussen cohorten			
1973 - 1974	.92	.72	.97
1974 - 1975	.92	.78	.90
1973 - 1975	.93	.89	.91

variëren. In het model kunnen schoolkenmerken worden opgenomen om deze variatie in coëfficiënten tussen scholen te reduceren (zie Raudenbush en Bryk, 1986), maar in dit artikel zullen wij ons beperken tot alleen een schatting van de grootte van de variantie van de coëfficiënten.

Ter illustratie van het achterliggende idee van een dergelijke analyse zullen we in Figuur 2 eerst enkele mogelijke uitkomsten laten zien, die voor de eenvoud slechts één onafhankelijke variabele (namelijk het prestatieniveau PREST) bevatten, die invloed heeft op het individuele onderwijsresultaat van een leerling.

In Figuur 2-A is er slechts één regressielijn. De relatie tussen het prestatieniveau van een leerling en zijn of haar resultaten is hetzelfde op alle scholen. In Figuur 2-B lopen drie parallelle regressielijnen, één voor elke school. De helling van deze lijnen zijn gelijk, maar de drie scholen verschillen in intercept. Deze verschillen in intercept geven verschillen tussen scholen weer wat betreft het gemiddelde bereikte onderwijsniveau, na controle voor aanvangsverschillen in prestatieniveau. Naarmate de regressielijn hoger ligt, en het intercept dus groter is, des te effectiever de school is voor alle leerlingen, ongeacht hun individuele aanvangsniveau. In Figuur 2-C tenslotte verschillen scholen niet alleen in intercept, maar ook in helling. De ene school is effectiever voor de betere leerlingen, de andere school is effectiever voor leerlingen met een lager prestatieniveau. In geval van Figuur 2-C kunnen we niet langer spreken van 'de' effectiviteit van een school maar is er sprake van differentiële effecten van scholen voor verschillende groepen leerlingen.

Om voor aanvangsverschillen gecorrigeerde outputmaten van scholen te verkrijgen gebruiken we in deze paragraaf het type model zoals dat is weergegeven in Figuur 2-B. Alleen de intercepten kunnen dan tussen scholen variëren. In de volgende paragraaf zullen we nagaan of het noodzakelijk is om ook de hellingen te laten variëren, zoals in Figuur 2-C.

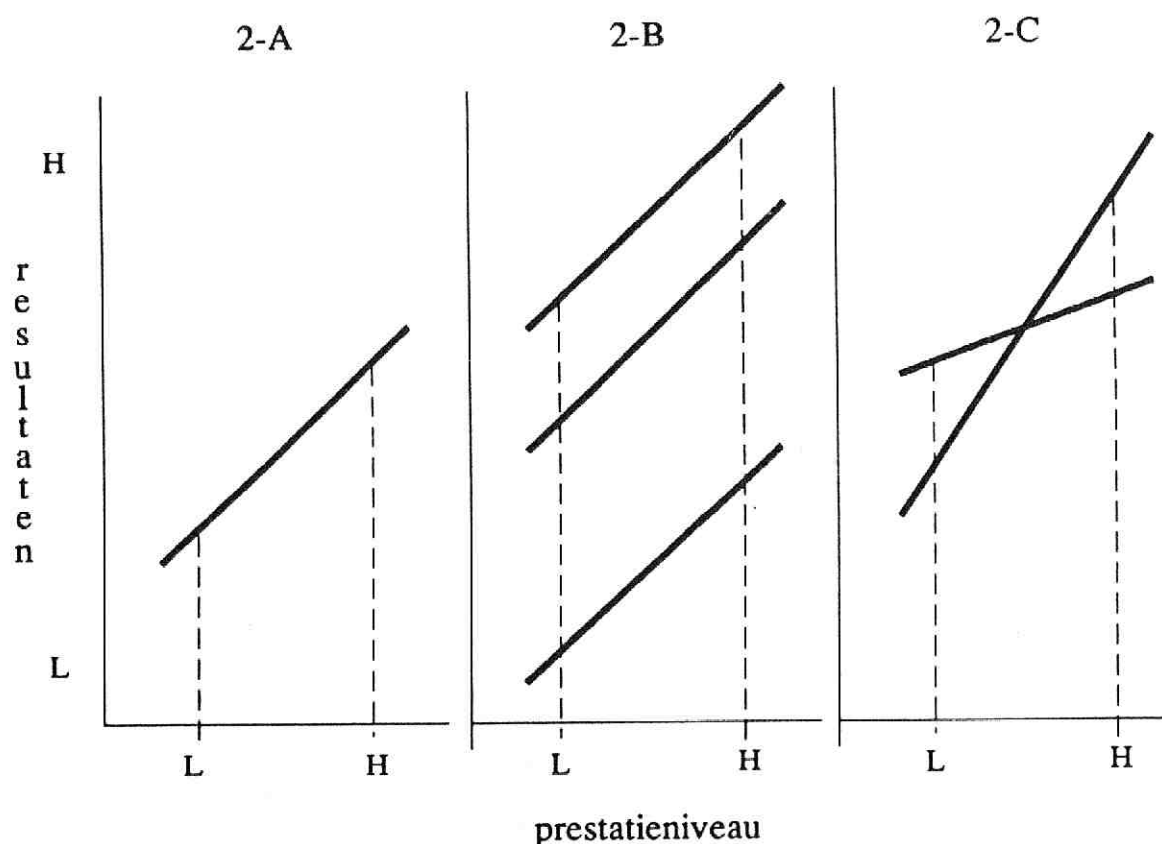


Fig. 2. Drie voorbeelden van schoolspecifieke effecten van prestatieniveau op onderwijsresultaten.

Als afhankelijke variabelen nemen we de twee schoolloopbaanvariabelen. Eerst wordt de regressie van de onderwijspositie in het tweede leerjaar, OP2, op de individuele achtergrondmerken geanalyseerd. Dit model maakt het mogelijk om verschillen tussen scholen in hun 'brugklaseffect' te onderzoeken, gecorrigeerd voor de aanvangsverschillen tussen de leerlingen.

Daarna wordt de laatst behaalde onderwijspositie op de startschool, EIND, zowel geregresseerd op de onderwijspositie in het tweede leerjaar, OP2, als op de individuele achtergrondvariabelen. Met dit tweede model kan het effect van scholen op de onderwijsresultaten na de brugklas worden geanalyseerd.

Dit laatste model kunnen we in de volgende vergelijking uitschrijven, voor leerling i in school j :

$$\begin{aligned} \text{EIND}_{ij} = & b_{0j} + b_{1j}\text{SEXE}_{ij} + b_{2j}\text{SES}_{ij} + b_{3j}\text{VTLO}_{ij} + \\ & b_{4j}\text{PREST}_{ij} + b_{5j}\text{ADV}_{ij} + b_{6j}\text{OP2}_{ij} + e_{ij} \end{aligned} \quad (1)$$

(e_{ij} met variantie s^2)

In het algemene model kunnen zowel het intercept b_{0j} als de regressiegewichten b_{kj} uit deze vergelijking verschillend zijn voor elke school; de coëfficiënten worden verondersteld random variabelen te zijn met variantie t_k^2 (voor $k = 0$ tot 6). In deze paragraaf starten we echter onze analyse met de veronderstelling dat de regressiecoëfficiënten gelijk zijn voor elke school en dat alleen de intercepten b_{0j} kunnen variëren (analoog aan Figuur 2-B). In dit geval wordt vergelijking (1) uitgebreid met de volgende vergelijking:

$$b_{0j} = B_0 + u_j \quad (u_j \text{ met variantie } t_0^2) \quad (2)$$

In deze vergelijking wordt gesteld dat het intercept van vergelijking (1) geschreven kan worden als de som van een algemeen gemiddelde B_0 en een schoolspecifieke afwijking u_j van dit gemiddelde. De analyses met dit model zijn uitgevoerd met het programma VARCL (Longford, 1986).

De resultaten voor de eerste afhankelijke variabele OP2 zijn weergegeven in Tabel 4. In het bovenste gedeelte worden schattingen voor de algemene (fixed) coëfficiënten gegeven, met

Tabel 4. Algemene coëfficiënten en de variantie componenten in het RC-model voor OP2; over drie cohorten Havo-Vwo leerlingen.

		1973		1974		1975	
algemene coëfficiënten							
intercept	5.8	(.7)		6.2	(.6)	10.7	(.6)
SEXE	0.8	(.4)		1.2	(.4)	1.8	(.4)
SES	0.3	(.2)		0.5	(.2)	0.0	(.2)
VTLO	-0.8	(.6)		-1.3	(.7)	-1.7	(.6)
PREST	0.11	(.01)		0.11	(.01)	0.09	(.01)
ADV	2.7	(.4)		2.3	(.4)	2.0	(.4)
variantie componenten							
leerlingen (s^2)	38.8			40.3		39.1	
intercepten (t_0^2)	4.5	(.5)		3.3	(.4)	7.9	(.6)
totaal % verklaarde variantie							
	22			21		19	

tussen haakjes de standaardfouten. (Deze coëfficiënten zijn ongestandaardiseerd en kunnen daardoor niet zonder meer met elkaar worden vergeleken als gevolg van de schaalverschillen tussen de variabelen; zie paragraaf 3.) In het volgende gedeelte van Tabel 4 zijn de schattingen voor de varianties van de twee random gedeelten weergegeven, met voor de variantie van het intercept tussen haakjes de standaardfout van de wortel uit de variantie (sigma). In het laatste gedeelte van de tabel staat het totale percentage van de variantie in de afhankelijke variabele OP2 weergegeven dat verklaard wordt door het model.

In het algemene gedeelte van het model zien we enkele reeds bekende effecten in schoolloopbaanmodellen in het voortgezet onderwijs (zie bijv. Tesser 1981; Dronkers, 1989): positieve effecten voor meisjes, positieve effecten van sociaal economische status, prestatiescore en advies en een negatief effect van de vertraging die een leerling in het basisonderwijs heeft opgelopen. Hoewel niet alle algemene coëfficiënten statistisch significant zijn, verwijderen we ze niet uit het model. In de eerste plaats willen we hetzelfde model analyseren over de drie cohorten. En in de tweede plaats willen we zo veel mogelijk variantie laten verklaren door de individuele leerlingvariabelen. Tabel 4 laat echter zien dat slechts een klein deel van de variantie in OP2 door het model verklaard wordt. In alle drie cohorten is de variantie van de intercepten tussen de scholen significant. Dit betekent dat er verschillen tussen scholen overblijven in de gemiddelde onderwijspositie van de leerlingen na het brugjaar, ook als wordt gecontroleerd voor de individuele leerlingkenmerken.

Het wordt dan interessant of hetzelfde ook geldt voor de outputmaat EIND. Op dezelfde manier als in Tabel 4 worden in Tabel 5 de resultaten voor deze afhankelijke variabele gepresenteerd.

De variantie in de laatst behaalde onderwijspositie van de leerlingen op hun startschool, EIND, is veel groter dan de variantie in hun positie in het tweede leerjaar. Tevens blijkt dat het model een veel groter deel van deze variantie verklaart. Enkele van de algemene coëfficiënten zijn niet significant, maar we kiezen er wederom voor ze niet uit het model te verwijderen. Andere algemene coëfficiënten variëren sterk van cohort tot cohort, bijvoorbeeld het effect van het prestatieniveau, achtereenvolgens .07, .14 en .17, en advies, met achtereenvolgens een coëfficiënt van 1.9, 1.1 en 4.8. Er lijkt sprake te zijn van een toenemend effect van PREST en ADV op de schoolloopbaan in het voortgezet onderwijs. Dit wijst op het bestaan van een periode-effect: een systematische verandering tussen de cohorten (Hagenaars, 1985).

Tabel 5. Algemene coëfficiënten en de variantie componenten in het RC-model voor EIND; over drie cohorten Havo-Vwo leerlingen.

	1973	1974	1975
algemene coëfficiënten			
intercept	-62.6 (1.3)	-77.8 (1.2)	-95.4 (1.6)
SEXE	-0.3 (1.1)	0.5 (1.1)	3.5 (1.1)
SES	0.4 (0.6)	-0.4 (0.5)	0.3 (0.6)
VTLO	-3.7 (1.7)	-4.4 (1.9)	-1.8 (1.8)
PREST	0.07 (.04)	0.14 (.03)	0.17 (.03)
ADV	1.9 (1.1)	1.1 (1.0)	4.8 (1.0)
OP2	2.56 (.09)	2.50 (.08)	2.32 (.05)
variantie componenten			
leerl. (s^2)	306.7	303.2	288.5
intercept (t_0^2)	13.8 (1.0)	11.2 (0.9)	24.9 (1.2)
totaal % verklaarde variantie			
	55	56	56

Onze aandacht gaat in deze paragraaf echter vooral uit naar de variantie tussen de intercepten van de scholen. Tabel 5 laat zien dat deze variantie significant is in elk cohort. We kunnen wederom concluderen dat er verschillen in output blijven tussen scholen (gemiddelde EIND), ook als we rekening houden met de leerlingkenmerken. Dit betekent dat niet alle verschillen in output tussen de scholen het gevolg zijn van de aanvangsverschillen van de leerling van een school. De verschillen tussen de scholen in hun voor aanvangsverschillen van leerlingen gecorrigeerde output zijn significant.

De laatste vraag in deze paragraaf is of deze voor aanvangsverschillen gecorrigeerde output van scholen hetzelfde is over drie cohorten. Om deze vraag te beantwoorden hebben we het VARCL-programma gebruikt om per school zogenoemde 'posterior means' te berekenen voor de random (vrij geschatte) intercepten. Deze 'posterior means' zijn de schattingen voor u_j in vergelijking (2), de afwijking van de schoolspecifieke intercepten van het algemene intercept. In analogie met de twee-dimensionale weergave van Figuur 2-B geven de 'posterior means' de afstanden weer tussen de parallelle regressielijnen en kunnen zij worden beschouwd als een schatting van de voor aanvangsverschillen gecorrigeerde output van een school (zie Aitkin en Longford, 1986). In Tabel 6 zijn de 'posterior means' voor de random intercepten in de drie cohorten weergegeven.

Uit Tabel 6 blijkt dat voor sommige scholen de 'posterior means' aanzienlijk verschillen van cohort tot cohort, terwijl de uitkomsten van andere scholen veel stabiel zijn. De correlaties tussen de cohorten zijn lager dan die van de niet gecorrigeerde output (zie Tabel 2), maar ze zijn niet erg laag of negatief. Dit betekent dat er een zekere stabiliteit is in de voor aanvangsverschillen gecorrigeerde output van scholen. Toch laat Tabel 6 eveneens zien dat het gevaarlijk is om de effectiviteit van een *specifieke* school vast te stellen enkel en alleen aan de hand van één cohort. Scholen 1 en 6, bijvoorbeeld, zouden op basis van cohort 1973 geclassificeerd worden als 'even ineffectief'. Op basis echter van cohort 1975 zou school 1 geclassificeerd worden als een gemiddelde school, en school 6 als een zeer ineffectieve school. Zo ook is school 3 effectief als alleen naar cohort 1973 gekeken wordt, en ineffectief als naar het 1975-cohort gekeken wordt. Deze instabiliteit van de voor aanvangsverschillen gecorrigeerde output van scholen kan verder geïllustreerd worden met Figuur 3, waarin de 'posterior means' van de cohorten 1973 en 1975 zijn geplot.

Tabel 6. De 'posterior means' van de random intercepten van de scholen in het RC-model voor EIND; over drie cohorten Havo-Vwo leerlingen.

	1973	1974	1975
school			
1	-4.3	-1.3	0.0
2	2.3	-1.1	3.7
3	2.9	1.3	-2.3
4	1.5	3.5	0.6
5	-1.1	-2.3	-4.4
6	-3.8	-2.4	-7.3
7	3.5	2.3	7.1
8	-0.4	-0.4	2.0
9	2.7	5.8	7.3
10	-6.2	-5.3	-6.6
11	2.8	-0.2	0.0
correlatie tussen de cohorten:			
1973		.76	.72
1974			.76

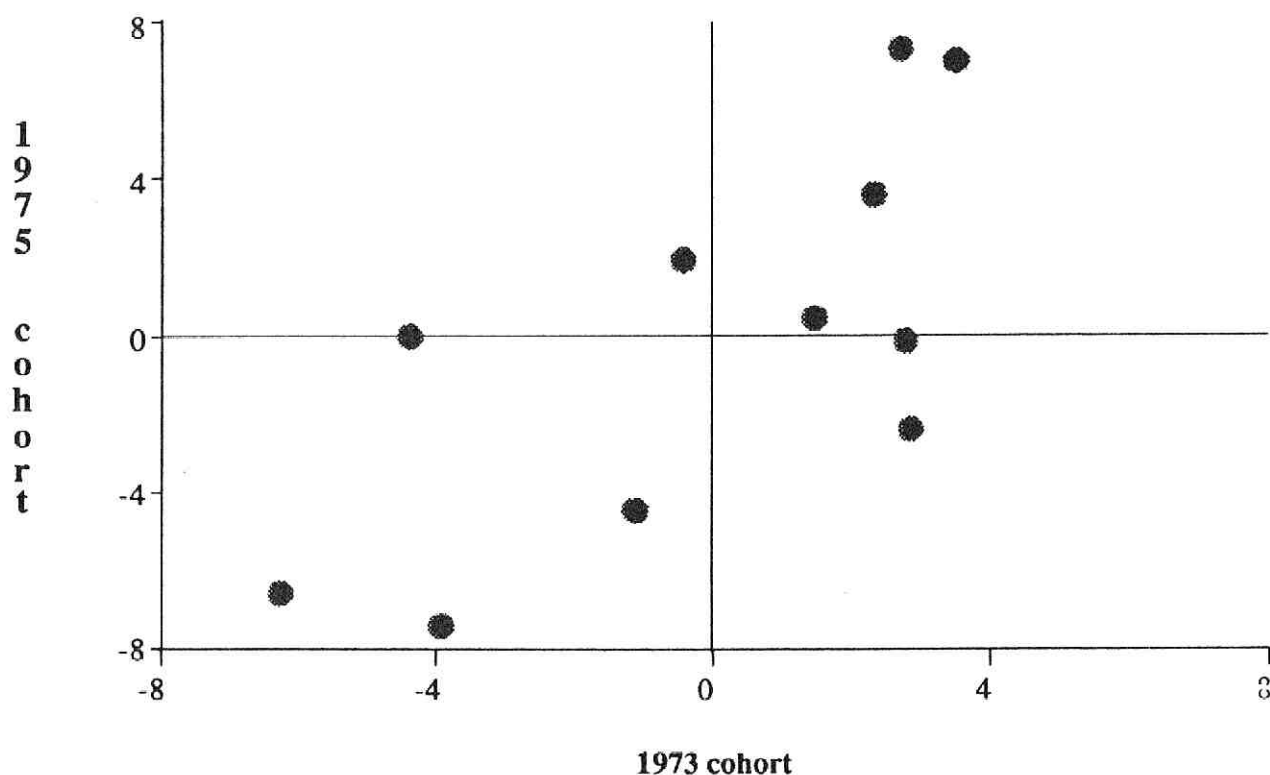


Fig. 3. De stabiliteit van de voor aanvangsverschillen gecorrigeerde output van scholen. Horizontaal de 'posterior means' van de random intercepten voor cohort 1973; verticaal de posterior means voor cohort 1975.

In de rechter helft van Figuur 3 zien we 6 scholen die gekarakteriseerd kunnen worden als (relatief) effectief op basis van het cohort 1973. Als we echter afgaan op het 1975 cohort, zijn twee van deze scholen gemiddeld effectief en kan één school zelfs ineffectief genoemd worden. De drie resterende scholen kunnen als consistent effectief gekarakteriseerd worden.

Hoewel er dus wel enige stabiliteit is in de voor aanvangsverschillen gecorrigeerde output van scholen, is deze stabiliteit te laag om op basis van de gegevens van één cohort leerlingen te kunnen rechtvaardigen dat een specifieke school gekarakteriseerd mag worden als effectief of ineffectief.

6. DE STABILITEIT VAN DE RELATIES TUSSEN INPUT EN OUTPUT

De resultaten van paragraaf 5 geven een complex beeld te zien en dit zou nog kunnen toenemen wanneer de assumptie uit het model, dat er geen verschillen in hellingen tussen de scholen bestaan, niet juist blijkt te zijn.

Om deze assumptie te testen hebben we modellen doorgerekend waarin *alle* regressiecoëfficiënten uit vergelijking (1) vrij mogen variëren tussen de scholen (een 'totaal model'). Het VARCL-programma geeft voor elk doorgerekend model een maat voor de passing bij de data. Het verschil tussen de maten van twee modellen volgt een chi-kwadraat verdeling waardoor de verbetering van de passing als gevolg van het vrij geschat laten worden van één of meer coëfficiënten geëvalueerd kan worden. In Tabel 7 geven we de verschillen in chi-kwadraat weer van de modellen waarin alle hellingen en de intercepten vrij zijn gelaten (totaal model), ten opzichte van het eenvoudiger model uit paragraaf 5, waarin alleen de intercepten vrij waren en de hellingen niet (het nul model). Verder geven we voor elk cohort het verschil in de chi-

Tabel 7. De verbetering van de passing van zowel RC-modellen met vrij geschatte hellingen en intercepten (totaal model) als een optimaal RC-model, beiden vergeleken met een nul model met alleen vrije intercepten; over drie cohorten Havo-Vwo leerlingen.

	1973	1974	1975
afhankelijke OP2	verschil in chi ² df	verschil in chi ² df	verschil in chi ² df
totaal model	11.2 10	22.7 10	55.4 10
optimaal model (met vrije hel- lingen van:)	9.0 2 PREST	23.8 4 PREST ADV	55.8 4 PREST ADV
afhankelijke EIND	verschil in chi ² df	verschil in chi ² df	verschil in chi ² df
totaal model	7.3 12	3.6 12	15.3 12
optimaal model (met vrije hel- lingen van:)	6.3 2 ADV	— geen	14.9 4 PREST ADV

kwadraat van een optimaal model, met slechts enkele vrij te schatten parameters meer dan in het nul model van paragraaf 5. Na het verschil in chi-kwadraat wordt ook het verschil in aantal vrijheidsgraden tussen deze modellen en het nul model worden weergegeven.

Uit de vergelijking tussen de totale en de optimale modellen blijkt dat de verbeterde fit van de totale modellen ook verkregen kan worden met meer spaarzame modellen met slechts een of twee vrije hellingen. Opvallend is dat in het optimale model van EIND niet de hellingen van OP2 vrij gelaten hoeft te worden, maar die van variabelen die de resultaten van de basisschoolperiode weergeven. De optimale modellen, die exploratief passend gemaakt zijn, zijn niet gelijk voor alle cohorten, maar de gegevens in Tabel 7 lijken er op te duiden dat scholen kunnen verschillen in de hellingen van PREST en ADV. Dit zou betekenen dat de afhankelijkheid van de schoolloopbaan van een leerling in het voortgezet onderwijs van zijn of haar resultaten in het basisonderwijs (PREST en ADV), van school tot school kan variëren.

Om de stabiliteit van deze verschillen over de drie cohorten nader te onderzoeken, hebben we analyses uitgevoerd met een model voor EIND, waarin zowel de intercepten als de hellingen voor PREST en ADV vrij zijn geschat voor de scholen. In dit model zijn de vergelijkingen (1) en (2) uit paragraaf 5 uitgebreid met de vergelijkingen:

$$b_{4j} = B_4 + v_j \quad (v_j \text{ met variantie } t_4^2) \quad (3)$$

$$b_{5j} = B_5 + w_j \quad (w_j \text{ met variantie } t_5^2) \quad (4)$$

In vergelijking (3) wordt gesteld dat de regressiecoëfficiënt b_{4j} uit vergelijking (1) geschreven kan worden als de som van een algemene (fixed) coëfficiënt B_4 en een schoolspecifieke afwijking v_j . Vergelijking (4) kan op dezelfde wijze geïnterpreteerd worden.

Voor deze vrij geschatte hellingen kunnen 'posterior means' berekend worden, die schattingen zijn voor v_j en w_j . Gecombineerd met de schattingen van de algemene coëfficiënten B_4 en B_5 levert dit de schoolspecifieke regressiecoëfficiënt van PREST en ADV op. In Tabel 8 worden deze coëfficiënten voor de drie cohorten weergegeven.

Tabel 8. Schoolspecifieke regressiecoëfficiënten van PREST en ADV in het RC-model voor EIND; voor drie cohorten Havo-Vwo leerlingen.

school	b_{4j} (PREST)			b_{5j} (ADV)		
	1973	1974	1975	1973	1974	1975
1	.067	.140	.111	1.76	2.75	-1.05
2	.062	.137	.167	1.30	1.09	-2.67
3	.061	.146	.177	-.05	.56	2.53
4	.063	.158	.180	12.70	.49	3.05
5	.066	.132	.145	-2.96	1.28	3.74
6	.066	.133	.230	-.30	1.49	4.52
7	.061	.152	.132	-1.13	.90	-3.59
8	.063	.140	.116	4.97	1.04	-.57
9	.062	.164	.099	-1.07	-.04	-.10
10	.070	.120	.262	8.14	.66	5.60
11	.061	.140	.221	-.34	-.01	7.10
correlatie tussen de cohorten:						
1973		-.71	.41		-.10	.16
1974			-.60			-.32

Als we allereerst kijken naar de coëfficiënten van PREST, zien we dat er niet veel variatie is tussen de scholen in de cohorten 1973 en 1974. Vandaar dat de variaties van deze coëfficiënten niet opgenomen behoeften te worden in de optimale modellen voor EIND in Tabel 7. Voor het 1975-cohort was het vrij laten van deze coëfficiënten wel nodig voor het optimale model en in Tabel 8 zien we dat er een veel grotere variatie is tussen de scholen in dit cohort. Als we kijken naar de coëfficiënten per school zien we dat deze tussen cohort 1973 en 1974 ongeveer zijn verdubbeld, terwijl het beeld voor cohort 1975 meer divers is: voor sommige scholen wordt de coëfficiënt groter en voor andere scholen kleiner. Het periode-effect van een toename van het belang van PREST voor de schoolloopbaan lijkt daarmee niet op alle scholen op te treden. De conclusie uit Tabel 8 is dat de coëfficiënten van PREST niet alleen tussen de scholen verschillen, maar ook voor cohorten op dezelfde school.

Hetzelfde geldt in nog sterkere mate voor de coëfficiënten van ADV. Tabel 8 laat veel onsystematische variatie in deze coëfficiënten zien, die zelfs voor sommige scholen in sommige cohorten resulteert in een negatief effect van ADV op EIND.

Een mogelijke verklaring van deze instabiliteit is dat de hellingen van PREST en ADV sterk covariëren (zgn. bouncing beta's). Wanneer echter soortgelijke modellen berekend worden met alleen de hellingen van PREST vrij over de scholen en de hellingen van ADV gefixeerd, dan blijkt er een vergelijkbare instabiliteit in de coëfficiënt van PREST per school op te treden (met correlaties tussen de cohorten van -.40 tot +.30). Een andere mogelijkheid is om deze instabiliteit op te vatten als ruis: als onsystematische variaties, die verder geen specifieke verklaring behoeven.

Wij menen echter dat de analyses wel degelijk duiden op een werkelijke instabiliteit in de schoolspecifieke effecten van leerlingkenmerken op schoolloopbanen en dat daarvoor verklaringen moeten worden gezocht. De processen waardoor leerlingen met bepaalde aanvangskenmerken eindigen op bepaalde onderwijsposities op deze scholen voor voortgezet onderwijs zijn niet stabiel, maar variëren van cohort tot cohort. Als op een school bij de ene instroomgroep de aanvangsverschillen tussen de leerlingen worden omgezet in grote verschillen in uiteindelijke onderwijsresultaten, dan hoeft dat voor een volgende instroomgroep nog helemaal niet zo te zijn.

Ondersteuning voor onze opvatting dat de instabiliteit een reëel fenomeen is vinden wij in de ervaringen van directeuren van scholen voor voortgezet onderwijs, die wij in de loop van ons project geïnterviewd hebben. Enkele van deze directeuren hadden het over de onvoorspelbaarheid van de schoolloopbanen van bepaalde klassen of groepen leerlingen. Zij benadrukten dat soms zelfs zeer veelbelovende leerlingen, als gevolg van meer of minder toevallige groepsprocessen, 'de kolder in de kop krijgen' en hun interesse en motivatie voor het leren op school in het geheel kunnen verliezen.

7. DISCUSSIE

De uitkomsten van de in dit artikel gepresenteerde analyses leiden tot de volgende conclusies:

- De niet voor aanvangsverschillen gecorrigeerde output van scholen is vrij stabiel over de tijd.
- Er is meer instabiliteit in de wel voor aanvangsverschillen gecorrigeerde output van scholen. De grotere stabiliteit van de 'ruwe' output, in vergelijking met de gecorrigeerde output, lijkt erop te wijzen dat er een tendens bij scholen bestaat om, ongeacht de kwaliteit van de instroom in een bepaald jaar, een soort vaste taakstelling of norm te behalen. Mogelijk werken hier mechanismen, die vergelijkbaar zijn met 'de wet van Posthumus' bij het toekennen van cijfers aan leerlingen.

De instabiliteit van de voor aanvangsverschillen gecorrigeerde output kan er de oorzaak van zijn dat het zo moeilijk blijkt om een verklaring voor verschillen in effectiviteit tussen scholen te vinden in termen van verschillen in schoolkenmerken. De meeste van deze schoolkenmerken blijven immers onveranderd gedurende een zekere tijdsspanne, maar de (relatieve) effectiviteit fluctueert wel over diezelfde periode.

We concluderen dat het identificeren van 'exceptioneel effectieve of ineffectieve' scholen op basis van de resultaten van alleen één instroomgroep niet vrij van risico is. Een betere benadering hiervoor kan zijn om specifieke scholen te identificeren die *consistent* effectief of ineffectief zijn voor een aantal opeenvolgende instroomgroepen.

- Met betrekking tot de relatie tussen leerlingkenmerken en schoolloopbanen is er een grote instabiliteit in de schoolspecifieke relaties.

Dit betekent dat het niet goed mogelijk is om een onderscheid te maken tussen 'egalitaire' scholen, waar zowel de in het begin betere als de minder goede leerlingen ongeveer dezelfde schoolloopbanen volgen, en 'selektieve' scholen, waar de in het begin betere leerlingen het duidelijk beter doen dan de minder goede leerlingen. Het eerder genoemde 'wet van Posthumus - mechanisme' lijkt zo te werken dat scholen een vrij constante output hebben, maar dat de vraag welke leerlingen de eindstreep halen niet goed te beantwoorden valt: in het ene jaar zullen dat vooral de betere leerlingen zijn, maar in het volgende jaar kan dat ook een groot deel van de mindere leerlingen zijn.

De resultaten van onze analyses zetten vraagtekens bij het nut van een concept van effectiviteit als een inherente eigenschap van scholen als een entiteit. Waarschijnlijk heeft effectiviteit meer van doen met processen op een lager niveau, bijvoorbeeld het niveau van de klas, die verschillen binnen dezelfde school, of zelfs met specifieke groepsprocessen van leerlingen binnen een klas. Als dit inderdaad het geval is, dan zijn er consequenties voor het onderwijsbeleid, omdat algemene schoolkenmerken makkelijker te manipuleren zijn dan specifieke processen die zich binnen een klas afspelen.

LITERATUUR

- Aitkin, M. & Longford N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 149, 1-43.
- Bakker, B. (1987). *Onderwijs en de kans op werkloosheid*. (Paper gepresenteerd op ORD'87, Groningen).
- Bosker, R.J., Guldemond, H., Hofman R.H. & Hofman W.H.A. (1988). *De stabiliteit van schoolkwaliteit*. Groningen: RION.
- CBS (1984). *Beroepenclassificatie 1984*. 's-Gravenhage: Staatsuitgeverij.
- Cuttance, P. (1987). *Modelling variation in the effectiveness of schooling*. Edinburgh: Centre for Educational Sociology.
- Dronkers, J. (1989). *The yield of Dutch educational attainment research; a retrospective view of the last decennium*. In: Bakker, B.F.M., J. Dronkers & G.W. Meijnen (eds). *Educational opportunities in the welfare state*. Nijmegen: ITS.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin & Co.
- Gray, J. & Jones B. (1985). Combining quantitative and qualitative approaches to studies of school and teacher effectiveness. In: D. Reynolds (ed). *Studying school effectiveness*.
- Hagenaars, J.A.P. (1985). *Loglineaire analyse van herhaalde surveys; panel-, trend- en cohortonderzoek*. Tilburg: KUB.
- Koopman, P., Eeden P. van den & Jong U. de (1986). Categoriele Mavo-scholen en schoolloopbanen in Amsterdam. In: W.J. Nijhof & E. Warries (eds). *De opbrengst van het onderwijs*. Lisse: Swets & Zeitlinger.
- Kreft, G.G. (1987). Heeft de Mammoetwet succes gehad? In: G.W. Meijnen, A.F.M. Nieuwenhuis & J.L. Peschar (eds). *Selectie en kwalificatie in het onderwijs*. Lisse: Swets & Zeitlinger.
- Leeuw, J. de & Kreft I. (1986). Random coefficient models for multi level analysis. *Journal of Educational Statistics*, 11, 57-86.
- Longford, N. (1986). VARCL-interactive software for variance component analysis. Applications for survey data.
- Raudenbush, S. & Bryk A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Roeleveld, J., Koopman, P., Jong U. de & Eeden P. van den (1989). *Schoolloopbanen in Amsterdam*. Amsterdam: vakgroep M&T/VU (Eindverslag SVO-project nr. 6595; een bewerkte versie zal gepubliceerd worden).
- Scheerens, J., Stoel, W.G.R., Vermeulen C.J.A.J. & Pelgrum W.J. (1988). *De haalbaarheid van een onderwijsindicateurenstelsel voor het basis- en voortgezet onderwijs*. Enschede: OCTO.
- Tesser, P.T.M. (1981). *Schoolloopbanenonderzoek in Nederland. Een trendstudie naar ontwikkelingen in het empirisch onderzoek tussen 1960 en 1980*. Nijmegen: ITS.

Manuscript ontvangen 23-10-1989

Definitieve versie ontvangen 13-7-1990

Statistische aspecten van multi-niveau onderzoek

Roel J. Bosker en Tom A.B. Snijders

Vakgroep Onderwijsorganisatie en -management, Universiteit Twente en Vakgroep Statistiek en Meettheorie, Rijksuniversiteit Groningen en Vakgroep Empirisch Theoretische Sociologie, Rijksuniversiteit Utrecht

ABSTRACT

The random coefficient hierarchical model for multilevel research is reviewed. The random effect model is treated with the associated generalized least squares and empirical Bayes estimation methods and with procedures for testing the regression parameters and the (co)variance components. Approximate formulas for standard errors in two-level models with random intercepts are given and it is discussed how these can be used to choose optimal sample sizes at both levels.

1. INLEIDING

In onderwijsonderzoek zijn tot voor kort effecten van scholen op het functioneren van leerlingen veelal onderzocht door hetzij te werken met naar het schoolniveau geaggregeerde gegevens, hetzij de analyse uit te voeren op het leerlingniveau, zonder rekening te houden met het feit dat de leerlingen in scholen zijn gegroepeerd. Dat aan beide werkwijzes grote problemen kleven, met name met betrekking tot generalisatiekwesties, is door diverse auteurs uiteengezet (bijvoorbeeld Kref, 1987). In het onderstaande zullen we een tegenwoordig veel gebruikt statistisch multi-niveau model presenteren, waarmee de genoemde problemen omzeild kunnen worden.

Uitgangspunt is dat er een tweetraps-steekproef¹ is getrokken: eerst een steekproef van scholen en vervolgens binnen elke school een steekproef van leerlingen (Moors & Mulwijk, 1975). In dit geval zijn dus leerlingen genest onder scholen. Gemakshalve zullen we dit statistisch model behandelen aan de hand van een denkbeeldig onderzoek. Hierin wordt onderzocht of scholen verschillen in hun effect op het schoolsucces van leerlingen, rekening houdend met verschillen in intelligentie en milieu van herkomst. Bovendien wordt onderzocht of scholen in hun effect op schoolsucces interacteren met het herkomstmilieu: zijn sommige scholen vergelijkenderwijs effectiever voor leerlingen uit het lager milieu, terwijl andere scholen het schoolsucces van de leerlingen uit het hoger milieu relatief gunstig beïnvloeden?

Onderzoek naar schooleffecten: algemene werkwijze

Het effect van een school definiëren we als datgene wat een school toevoegt aan hetgeen bij een leerling op die school verwacht had mogen worden op basis van vaardigheden en kennis zoals die aanwezig waren bij entree van die leerling in de betreffende school. De essentie van deze definitie bestaat hieruit, dat we pas van een effect van de school kunnen spreken, als we er redelijk zeker van zijn dat met alternatieve verklaringen voor schoolverschillen op het niveau van de leerling zelf rekening is gehouden. De benaderingswijze die hiervoor zal worden gevolgd kan men zich als volgt opgebouwd denken (cf. Abalos e.a., 1985).

R.J. Bosker, Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.
T.A.B. Snijders, VSM-FPPSW, Rijksuniversiteit Groningen, Oude Boteringestraat 23, 9712 GC Groningen; ETS-FSW, Rijksuniversiteit Utrecht, Heidelberglaan 2, 3584 CS Utrecht.

1. We voorspellen voor alle leerlingen hun prestaties met de relevante begingegevens, zoals IQ en milieu van herkomst.

2. Deze voorspelling is nooit perfect: sommige leerlingen blijken boven het voorspelde te scoren en andere juist eronder. De eerste groep leerlingen wordt wel aangeduid als 'overachieveers' en de tweede groep als 'underachieveers'. Er kan natuurlijk ook nog een (kleine) groep leerlingen zijn die precies volgens verwachting presteert.

3. In het onderwijs mogen we ervan uitgaan dat sommige scholen systematisch meer 'overachieveers' hebben dan andere scholen. Ware dat niet het geval dan zouden we het onderzoek naar effecten van scholen ogenblikkelijk kunnen staken. Met andere woorden: ook op het niveau van scholen zou men kunnen spreken van scholen die boven of juist onder verwachting presteren. Daarom worden in het model ook school-effecten opgenomen. Individuele en school-effecten worden simultaan geanalyseerd.

2. HET HIËRARCHISCH MODEL VOOR MULTI-NIVEAU ONDERZOEK

Het random coëfficiënt-model

Het random coëfficiënt-model² levert een efficiënte methode om schooleffecten te schatten; het wordt behandeld in o.a. de artikelen van Aitkin en Longford (1986) en van Kreft en de Leeuw (1987), het boek van Goldstein (1987), en het overzichtsartikel van Raudenbush (1988). Het kan worden gebruikt voor data waarin meerdere niveaus van nesting aanwezig zijn; wij zullen in dit artikel alleen het twee-niveau geval behandelen. Dit model heeft in zijn eenvoudigste vorm de volgende gedaante (stochastische variabelen zijn steeds onderstreept>):

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \epsilon_{ij}, \quad (1)$$

waarin leerling i in school j prestatiescore y_{ij} heeft die voorspeld wordt uit de achtergrondvariabele x met het gewicht (regressie-coëfficiënt) β_1 .

Dat deze voorspelling niet perfect is blijkt uit het schoolspecifieke intercept β_{0j} en het leerlingsspecifieke residu ϵ_{ij} . Wat (1) doet afwijken van het conventionele regressiemodel is het schoolspecifieke intercept. Dit kan men zich als volgt opgebouwd denken:

$$\beta_{0j} = \beta_0 + u_{0j}, \quad (2)$$

Het schoolspecifieke intercept is dus opgebouwd uit een algemeen intercept β_0 en een schoolspecifieke afwijking u_{0j} . Het eigene van het random coëfficiënt-model bestaat erin, dat van u_{0j} wordt aangenomen dat het een stochastische variabele is. In model (1)+(2) speelt het toeval dus op twee plaatsen een rol: op leerlingniveau (ϵ_{ij}) en op schoolniveau (u_{0j}). Dit weerspiegelt de gemaakte assumptie dat er sprake is van een tweetraps- steekproef. Vanwege deze hiërarchische structuur noemt men dergelijke modellen wel hiërarchische lineaire modellen. Men zou ook model (1)+(2) kunnen gebruiken zonder de assumptie dat de u_{0j} stochastische variabelen zijn; dan zijn het van school tot school verschillende intercepten, vaste doch in eerste instantie onbekende getallen. Dit kan het vaste coëfficiënt-model worden genoemd³. Het aantal onbekende parameters dat bij het vaste coëfficiënt-model moet worden geschat is veel groter dan dat voor het random coëfficiënt-model: iedere school geeft aanleiding tot een extra parameter u_{0j} , terwijl bij het random coëfficiënt-model als parameter niet de individuele waarden u_{0j} optreden maar alleen hun variantie. Omdat het aantal parameters bij het vaste coëfficiënt-model relatief groot is ten opzichte van de hoeveelheid data, leidt dit model tot instabiele parameterschattingen. Dit kan worden beschouwd als een gevolg van "overfitten" ofwel kanskapitalisatie. Bij het random coëfficiënt-model leveren de scholen informatie niet alleen over zichzelf maar ook over elkaar, omdat ze alle uit dezelfde populatie afkomstig zijn. Die informatie over elkaar bestaat uit informatie over de waarde van het algemeen intercept β_0 en de waarde van de variantie van u_{0j} , en gaat deze kanskapitalisatie tegen.

We nemen aan dat u_{0j} normaal verdeeld is met gemiddelde 0 en variantie σ_1^2 , terwijl ϵ_{ij} normaal verdeeld is met gemiddelde 0 en variantie σ^2 . Om een zinvol onderscheid tussen school- en leerlingniveau te verkrijgen nemen we aan dat de schoolresiduen u_{0j} en de leerlingresiduen ϵ_{ij} onderling onafhankelijk zijn.

De interactie tussen school en de achtergrondvariabele x , bijvoorbeeld SES, waarmee we onderzoeken of leerlingen uit het lager milieu op sommige scholen relatief (namelijk t.o.v. de leerlingen uit het hoger milieu binnen die school) beter af zijn dan op andere scholen, kunnen we opvatten als de mate waarin een school verschillen tussen leerlingen naar milieu van herkomst omzet naar prestatieverschillen. Dit kan worden uitgewerkt door een uitbreiding van formules (1)+(2). We wijzigen (1) in

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}, \quad (3)$$

waarin β_{1j} de schoolspecifieke regressiecoëfficiënt van y op x (SES) is. Naarmate deze coëfficiënt groter is, zijn er op de school grotere verschillen tussen leerlingen uit verschillende milieus. Analooq aan (2) kan nu ook β_{1j} worden gesplitst volgens

$$\beta_{1j} = \beta_1 + u_{1j}. \quad (4)$$

Met het bovenstaande hebben we een model geschetst aan de hand waarvan verschillen tussen scholen, hetzij naar (voor instroomverschillen gecorrigeerde) gemiddelde output, hetzij naar de relatie van milieu van herkomst (of sexe, IQ e.d.) met schoolsucces, opgespoord kunnen worden. Een volgende vraag is hoe deze verschillen tussen scholen statistisch voorspeld kunnen worden uit verschillen in schoolkenmerken. Daartoe nemen we in de vergelijkingen in (2) en (4) als voorspeller het schoolkenmerk z op met bijbehorende coëfficiënt γ_0 respectievelijk γ_1 :

$$\beta_{0j} = \beta_0 + \gamma_0 z_j + u_{0j}, \quad (5)$$

$$\beta_{1j} = \beta_1 + \gamma_1 z_j + u_{1j}. \quad (6)$$

Substitutie van (5) en (6) in (3) levert nu de volgende complete uitdrukking voor het random coëfficiënt-model:

$$y_{ij} = \beta_0 + \gamma_0 z_j + \beta_1 x_{ij} + \gamma_1 z_j x_{ij} + u_{0j} + u_{1j} x_{ij} + \epsilon_{ij}. \quad (7)$$

Wanneer vergelijking (6) wordt toegepast op de geschatte regressiecoëfficiënten levert dit wat Burstein (1980) de 'slopes as outcomes' benadering noemde. Hierbij wordt dus een scheiding gemaakt tussen eerst een analyse binnen elke school apart, en vervolgens een analyse waarbij de scholen de analyse-eenheden vormen. De random coëfficiënt-benadering is anders doordat deze scheiding niet wordt gemaakt: model (7) wordt in één keer geanalyseerd.

We zien dat vergelijking (7) veel weg heeft van een conventioneel regressiemodel met een interactie tussen het leerlingkenmerk x en het schoolkenmerk z . Alleen de residu-structuur is gecompliceerder. Dit is een direct gevolg van het feit dat de intercepten en de hellingen opgevat worden als uitkomsten van een steekproef uit een populatie. Dit leidt ook tot een andere covariantie-structuur. De y -waarden voor leerlingen van *verschillende* scholen hebben covariantie 0; maar voor twee aselekt gekozen verschillende leerlingen i en i' van *dezelfde* school j volgt uit (7) de covariantie

$$\text{cov}(y_{ij}, y_{i'j}) = x_{ij} x_{i'j} \text{ var}(u_{1j}) + \sigma_1^2.$$

Deze covariantie komt tot stand door hun gemeenschappelijke omgeving; de grootte hangt mede af van hun individuele kenmerken x .

Om de interpretatie van de regressiecoëfficiënten te vergemakkelijken is het aan te raden,

dat de verklarende variabelen worden genomen als deviatiescores. Dit kan als volgt worden toegelicht. Indien de leerlingvariabelen x (bijvoorbeeld SES-scores) zijn gedefinieerd als afwijkingsscores ten opzichte van de schoolgemiddelden, is de gemiddelde x -waarde op school j , dus $\bar{x}_j$, voor elke j gelijk aan 0. Door in (7) in te vullen $x_{ij} = 0$ ziet men dat dan $\beta_0 + \gamma_0 z$ de regressielijn is van y op z voor de gemiddelde leerling per school; dus de regressielijn die men zou krijgen na aggregatie van de leerlingvariabelen tot schoolgemiddelden. Indien de schoolvariabelen z ook afwijkingsscores zijn ten opzichte van het gemiddelde, dus $\bar{z} = 0$, dan is $\beta_0 + \beta_1 x$ de regressielijn van y op x voor de gemiddelde school. Cronbach (1976) en Boyd & Iversen (1979) gaan uitgebreider in op de wenselijkheid om de leerlingvariabelen per school te centreren; wij komen hier in paragraaf 4 op terug. Wanneer men werkt met deviatiescores zullen er ook geen problemen optreden t.g.v. eventuele collineariteit tussen x_{ij} en z_j enerzijds, en $z_j x_{ij}$ anderzijds. In dit artikel nemen we verder aan dat in elk geval de x -variabelen afwijkingsscores binnen scholen zijn, zodat hun school-gemiddelde steeds 0 is.

3. PARAMETERSCHATTING IN HET RANDOM COËFFICIËNT MODEL

De parameters die optreden in het random coëfficiënt-model (7), eventueel uitgebreid tot een model met meerdere x - of z -variabelen, kunnen worden geschat of getoetst volgens de in de genoemde literatuur behandelde methoden. Deze zijn geïmplementeerd in enkele computerprogramma's: HLM (Bryk e.a., 1986), VARCL (Longford, 1988), ML (Goldstein en Rasbash, 1988). Deze methoden kunnen worden opgevat als zogenaamde gegeneraliseerde kleinste kwadraten (GLS, Generalized Least Squares) methoden, waarbij rekening wordt gehouden met de speciale covariantiestructuur ten gevolge van de twee-traps steekproeftrekking, en waarbij de hierbij horende varianties en covarianties uit de data worden geschat. In model (7) zijn de parameters het intercept β_0 , de regressie-coëfficiënten γ_0 , β_1 en γ_1 , de covariantiematrix van de optredende random school-effecten (u_{0j} , u_{1j}), en de residuele leerlingvariantie σ^2 . (De noodzaak om met $\text{cov}(u_{0j}, u_{1j})$ rekening te houden wordt toegelicht in de volgende paragraaf.) De γ -parameters hebben betrekking op de systematische effecten van de schoolvariabelen z . De covariantiematrix van (u_{0j} , u_{1j}) geeft aan, hoe groot de resterende 'onverklaarde' schoolverschillen nog zijn.

Vaak wil men uiteraard ook informatie hebben over individuele scholen; dus over de 'toevalsafhankelijke' regressiecoëfficiënten β_{0j} en β_{1j} . Het meest eenvoudig is het schatten van deze school-parameters met de gewone kleinste kwadratenmethode (OLS, 'Ordinary Least Squares') op grond van de gegevens voor de betreffende school. De met deze methode geschatte regressiecoëfficiënten geven we aan met $\hat{\beta}_{hj}$ ($h = 0, 1$, enz.). Het is echter bekend dat deze schatters meestal zeer instabiel zijn (d.w.z. hun standaardfouten zijn groot), vooral als de steekproefomvang n_j op school j klein is. Een veelal betere benadering is mogelijk door, in het kader van het hiërarchische model (7), ook relevante informatie van andere scholen te gebruiken; de invloed van deze andere informatie zal afnemen, naarmate de steekproefomvang n_j toeneemt. De zo verkregen schattingen zullen we aangeven met $\hat{\beta}_{hj}^*$. De hierbij gebruikte technieken (zie bijvoorbeeld Raudenbush en Bryk, 1986 en Goldstein, 1986) kunnen als volgt worden geïnterpreteerd: de schatter $\hat{\beta}_{hj}^*$ is een gewogen gemiddelde van enerzijds de met OLS geschatte regressiecoëfficiënt $\hat{\beta}_{hj}$ voor de betreffende school en anderzijds de gemiddelde geschatte regressiecoëfficiënt voor alle scholen. De gewichten hangen af van de data: hoe kleiner de standaardfout van de OLS-schatter $\hat{\beta}_{hj}$ is, en hoe groter de verhouding σ_j^2/σ^2 (tussen-scholen : binnen-scholen variantie), des te groter is het gewicht voor de schatter $\hat{\beta}_{hj}$ van school j zelf. De zo verkregen schatters worden wel 'shrinkage' schatters genoemd. Deze benaming betreft het inkrimpen van de afwijking tussen de geschatte regressiecoëfficiënt voor school j , en het gemiddelde van deze coëfficiënten voor alle scholen tesamen. Men kan dit ook beschouwen als een regressie naar het gemiddelde. Een andere voor deze techniek gebruikte term is 'empirisch Bayes', afgekort EB. Dit doelt erop dat er een kansverdeling voor de regressiecoëfficiënten wordt gepostuleerd, nl. een verdeling over de populatie van scholen; en dat er empirische informatie over deze kansver-

deling gebruikt wordt, in tegenstelling tot de subjectivistische Bayesiaanse benadering waarbij de onderzoeker a priori geformuleerde subjectieve informatie gebruikt. In overeenstemming met de in de Bayesiaanse statistiek gebruikte begrippen worden de schatters β_{hj}^* wel a posteriori waarden voor β_{hj} genoemd.

Een van de voordelen van deze EB schattingsmethode bestaat hieruit, dat ongelijke steekproefomvang n_j per school geen enkel probleem opleveren. Met ongelijke steekproefomvang wordt automatisch rekening gehouden: voor de scholen met kleinere n_j wordt bij de schatting van β_{hj} sterker 'gekrompen' in de richting van het gemiddelde voor alle scholen samen. Het is dus onnodig om de steekproefomvang per school ongeveer gelijk te trekken door scholen met kleine n_j uit de analyse weg te laten, of van scholen met grote n_j sommige leerlingen weg te laten⁴.

4. COVARIANTIECOMPONENTEN

In het voorgaande hebben we de random intercepten en random hellingen gescheiden behandeld. In de praktijk zullen deze twee echter enige covariantie hebben. Dat is theoretisch gezien in sommige gevallen ook plausibel. Zo zijn er theorieën dat scholen die hun leerlingen gemiddeld naar een hoger niveau brengen, ook tot meer milieuafhankelijke ongelijkheid zullen leiden. In dit geval correleren β_{0j} en β_{1j} positief. Met de interpretatie van deze correlatie moeten we echter zeer voorzichtig zijn. De onderstaande grafiek kan dit wellicht snel inzichtelijk maken. In de figuur zijn drie scholen afgebeeld. Bij deze scholen is er sprake van heterogene binnengroepsregressie: de regressie van schoolsucces op IQ verschilt van school tot school. Er zijn echter geen interceptverschillen wanneer we kijken naar de y_1 -as. Zouden we nu bij elke IQ-score 10 punten optellen, dan verschuift de y -as 10 punten naar links: de y_2 -as in de figuur. Op deze as vinden we wel interceptverschillen. School 1 die het zo goed lijkt te doen heeft zelfs een negatief intercept. Binnen deze populatie van 3 scholen is de correlatie tussen β_{0j} en β_{1j} negatief. Trekken we echter 10 punten van de IQ-score af, dan verschuift de y -as 10 punten naar rechts: de y_3 -as. Hierop heeft school 1 wel een positief intercept; nu is binnen deze populatie van 3 scholen de correlatie tussen β_{0j} en β_{1j} positief. Omdat het gemiddelde 100 van het IQ een willekeurige keuze is (we hebben te maken met een interval- en niet een ratio-schaal), zijn deze operaties wel toegelaten. Zijn er nu wel of geen schooleffecten? Een oplossing voor dit ogenschijnlijk merkwaardige probleem is in het verre verleden aangedragen door Cronbach (1976) en later nog eens verfijnd door Boyd en Iversen (1979). Het voorstel is de x_{ij} -scores te centreren rond hun schoolgemiddelden $\bar{x}_{.j}$:

$$x'_{ij} = x_{ij} - \bar{x}_{.j}$$

Deze deviatiescores x'_{ij} representeren het IQ van een leerling als afwijking van het schoolgemiddelde. Interceptverschillen tussen scholen zijn nu verschillen tussen de gemiddelde leerlingen van die scholen. Men kan tegenwerpen dat het weinig zinvol is twee leerlingen met elkaar te vergelijken die weliswaar de gemiddelde leerling van hun school zijn, maar onderling sterk verschillen in hun absolute IQ-score (bijvoorbeeld 85 versus 115). De oplossing voor dit probleem is, dat in de regressievergelijking ook de $\bar{x}_{.j}$ -scores worden opgenomen als een verklarende variabele op schoolniveau. In dat geval corrigeren we voor deze tussen-scholen verschillen en kunnen we dus vergelijkbare leerlingen vergelijken door het school-hoofdeffect te definiëren als de waarde van de geschatte regressielijn voor de betreffende school in het algemeen gemiddelde $\bar{x}$.⁵

Alhoewel we tot nu toe steeds modellen hebben behandeld met maximaal één van school tot school verschillende (random) helling, is het goed mogelijk modellen met meerdere random hellingen te specificeren. In dat geval moeten we, analoog aan het vooraangaande, beducht zijn op een correlatie tussen deze random hellingen. Ook deze samenhang is theoretisch aanneme-

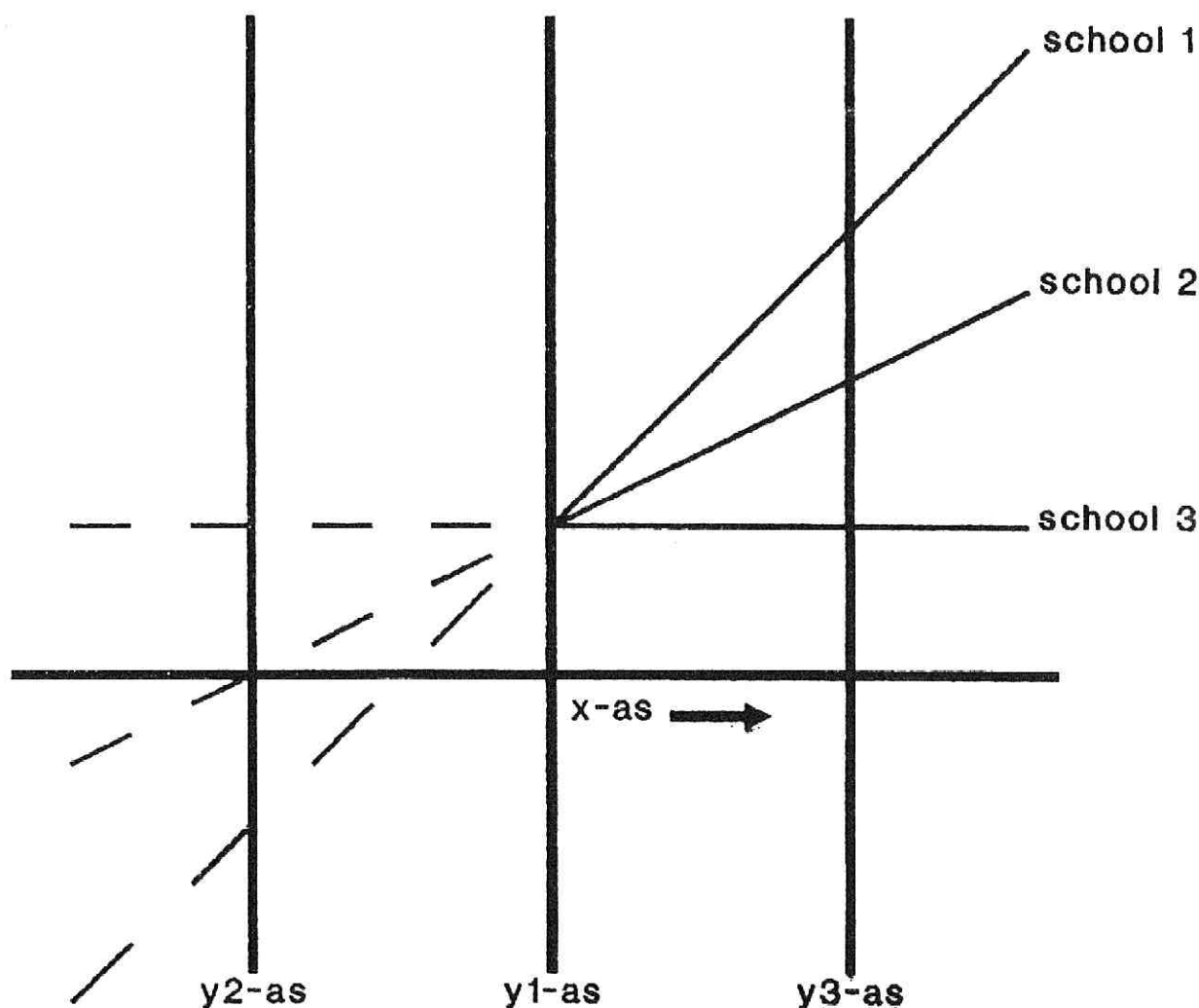


Fig. 1. Covariantie tussen intercept en helling.

lijk: scholen die sterk selecteren op IQ en dus een steile IQ-helling hebben, vertonen wellicht een minder steile SES-helling. Zijn we echter primair geïnteresseerd in de SES-helling, dan ontstaan er interpretatie-problemen, omdat het random SES-effect verdisconteerd kan zijn in het random IQ-effect. De beste oplossing voor dit probleem is een adequate modelspecificatie: door het introduceren van de volgens de theorie juiste voorspellers op schoolniveau van de random IQ- en SES-effecten (in casu: het opnemen van de geschikte interactietermen in model (7): $IQ_{ij} * z_{1j}$, $IQ_{ij} * z_{2j}$, $SES_{ij} * z_{3j}$, etc.) kunnen we de variantie in de random effecten trachten te binden, waarmee de covariantie tussen de random hellingen afneemt en bovendien de random effecten en hun covariantie theoretisch verklaard kunnen worden.

5. TOETSEN IN HET RANDOM COËFFICIËNT MODEL

In het random coëfficiënt model (7) zijn als onbekende parameters opgenomen de regressiecoëfficiënten β_0 , β_1 , γ_1 en γ_2 en de (co)variantie-componenten σ^2 , $\sigma^2_1 = \text{var}(u_{0j})$, $\text{var}(u_{1j})$ en $\text{cov}(u_{0j}, u_{1j})$. Naast het schatten ervan is het ook van belang om te toetsen of dergelijke parameters 0 zijn, en soms ook of ze een andere vooraf gespecificeerde waarde hebben en (voor regressie-coëfficiënten) of sommige aan elkaar gelijk zijn.

Voor regressiecoëfficiënten kan dit betrekkelijk eenvoudig gebeuren met een t-toets. We geven in deze bespreking een regressiecoëfficiënt in het algemeen aan met β . De in paragraaf 3 genoemde software geeft naast de schatter $\hat{\beta}$ ook de standaardfout hiervan: $SE(\hat{\beta})$, zodat de t-waarde $T(\beta) = \hat{\beta}/SE(\hat{\beta})$ kan worden berekend. Vanwege het twee-traps karakter van de steekproef heeft $T(\beta)$ onder de nulhypothese " $\beta = 0$ " niet exact een t-verdeling; maar als de steekproefomvang (vooral het aantal scholen) niet te klein is dan zal de standaardfout zijn gebaseerd op een stabiele variantieschatting, en zal de kansverdeling van $T(\beta)$ toch goed door een t-verdeling of standaardnormale verdeling worden benaderd. Als een grove benadering (die meestal te laag en dus voorzichtig uitvalt) van het aantal vrijheidsgraden kan men $N-(p+q+1)$ nemen, waarin N het aantal scholen is, $p+1$ het aantal β -parameters en q het aantal γ -parameters, en dus $p+q+1$ het totaal aantal regressiecoëfficiënten. Indien $N-(p+q+1)$ voldoende groot is, kan men $T(\beta)$ toetsen in de standaardnormale verdeling.

Voor de variantiecomponenten kan deze methode niet worden toegepast. Hiervoor geldt namelijk dat het quotiënt

$$(\text{schatter} - \text{populatiewaarde})/\text{standaardfout}$$

slechts dan ongeveer een t-verdeling heeft, wanneer de populatiewaarde van de betreffende variantiecomponent voldoende sterk (d.w.z. meerdere standaardfouten) van 0 verschilt. Dus zeker voor de nulhypothese dat een variantiecomponent 0 is, kan geen gebruik worden gemaakt van een t-toets. Hier kan echter het principe van de likelihood-ratio toets (besproken in elk leerboek over theoretische statistiek) gebruikt worden, en wel op de volgende wijze. Beschouw een bepaald model, zoals (7), dat wordt gekarakteriseerd door de erin opgenomen vrije parameters; het model geven we aan met M_1 en het aantal vrije parameters met r_1 . Voor (7) bijvoorbeeld is $r = 8$ (vier regressiecoëfficiënten, de varianties van u_{0j} , u_{1j} en ϵ_{ij} en de covariantie tussen u_{0j} en u_{1j}). Voor gegeven data wordt de deviance van het model, aan te geven met $D(M_1)$, gedefinieerd als -2 maal de natuurlijke logaritme van de gezamenlijke kansdichtheid van de waarnemingen, waarbij de volgens de maximum likelihood-methode geschatte parameterwaarden worden ingevuld; zie bijvoorbeeld McCullagh en Nelder (1983: 24 e.v.) of Dobson (1983: 40). Stel dat er voor een aantal (één of meer) parameters in model M_1 moet worden getoetst, of ze gelijk zijn aan 0. Het sub-model van M_1 dat wordt verkregen door deze parameters 0 te stellen, geven we aan met M_0 , en het aantal overblijvende vrije parameters met r_0 . Neem als voorbeeld dat in model (7) wordt getoetst of de heterogeniteit van de regressie op x volledig kan worden verklaard door de schoolvariabele z . Dit betekent dat de term $u_{1j}x_{ij}$ uit (7) kan worden weggelaten, ofwel dat $\text{var}(u_{1j})$ en $\text{cov}(u_{0j}, u_{1j})$ gelijk aan 0 worden gesteld. Het resulterende model kan worden aangegeven als

$$y_{ij} = \beta_0 + \gamma_0 z_j + \beta_1 x_{ij} + \gamma_1 z_j x_{ij} + u_{0j} + \epsilon_{ij} \quad (8)$$

met $r_0 = 6$. Ook voor model M_0 kan de deviance worden berekend. Uit de definitie van de deviance en de maximum likelihoodschatter kan worden afgeleid, dat $D(M_0) \geq D(M_1)$. Er geldt nu dat, indien model M_0 juist is, het verschil $D(M_0) - D(M_1)$ een chi-kwadraat verdeling heeft met $r_1 - r_0$ vrijheidsgraden. (Strikt genomen geldt deze kansverdeling alleen asymptotisch, voor een groot aantal scholen N ; ook voor minder grote N is de chi-kwadraat verdeling hier echter een goede benadering.) Hieruit volgt dat het verschil in deviance gebruikt kan worden als toetsgrootte voor de nulhypothese dat M_0 het juiste model is, tegen de alternatieve hypothese dat M_1 juist is maar M_0 niet. Het aantal vrijheidsgraden van de chi-kwadraat verdeling is het aantal parameters waarvan wordt getoetst of ze een bepaalde waarde (bijvoorbeeld 0) hebben; zoals uit het gegeven voorbeeld blijkt, moeten hierbij de covarianties tussen random effecten niet worden vergeten.

Het kan soms voorkomen, dat men meerdere regressiecoëfficiënten of covariantiecomponenten tegelijk wil toetsen; bijvoorbeeld bij het toetsen van het effect van een nominale verklarende variabele met k niveaus, die in het regressiemodel door $k-1$ dummy-variabelen wordt

gerepresenteerd. Een dergelijke toets kan op precies dezelfde manier met het verschil van deviances worden uitgevoerd.

6. KEUZE VAN STEEKPROEFOMVANGEN

Computerprogramma's zoals HLM, ML en VARCL geven voor de in dit artikel besproken modellen parameterschattingen met hun standaardfouten. Om echter een onderzoek te plannen, bijvoorbeeld voor het bepalen van het aantal benodigde scholen en leerlingen per school, heeft de onderzoeker de behoefte aan een vooraf gemaakte raming van de te verwachten standaardfouten. Hiervoor worden in het onderstaande formules gegeven voor het geval dat er sprake is van een random (dus per school verschillend) intercept β_{0j} , maar niet van random regressiecoëfficiënten. Een deel van het bijzondere van multi-niveau problematiek blijft dan buiten schot, want de binnen-groeps regressies zijn dus homogeen. Deze beperking wordt hier gemaakt om de formules eenvoudig te houden. Bewijzen voor de hier gegeven formules, evenals formules voor het meer algemene geval dat er wel random regressiecoëfficiënten zijn, worden gegeven in Snijders en Bosker (1989)⁶.

We veronderstellen een voor alle N scholen gelijk aantal onderzochte kinderen, aangegeven met n . Het aantal opgenomen leerlingvariabelen wordt aangegeven met p , het aantal schoolvariabelen met q . We kunnen het model uitdrukken als

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \gamma_1 z_{1j} + \dots + \gamma_q z_{qj} + u_{0j} + e_{ij} \quad (9)$$

De vector van regressiecoëfficiënten $(\beta_0, \beta_1, \dots, \beta_p, \gamma_1, \gamma_2, \dots, \gamma_q)'$ wordt geschat door de in paragraaf 3 genoemde GLS schatter, geïmplementeerd in de genoemde computerprogramma's. Deze schatter geven we aan met $(\beta_0^*, \beta_1^*, \dots, \beta_p^*, \gamma_1^*, \gamma_2^*, \dots, \gamma_q^*)'$. Het is gewenst om de steekproefomvang N (aantal scholen) en n (aantal leerlingen per school) zo te kiezen, dat de standaardfouten van deze schatters klein zijn. Om daar iets over te zeggen, hebben we informatie nodig over de covariantiematrix van de geschatte vector $(\beta_0^*, \beta_1^*, \dots, \beta_p^*, \gamma_1^*, \gamma_2^*, \dots, \gamma_q^*)'$. Hiervoor geldt het volgende. De schatters $(\beta_1^*, \dots, \beta_p^*)'$ en $(\beta_0^*, \gamma_1^*, \gamma_2^*, \dots, \gamma_q^*)'$ zijn bij benadering (voor grote N) ongecorrleerd. (De schatter β_0^* voor het intercept doet dus mee met de schatters voor de regressiecoëfficiënten van de schoolvariabelen.) De covariantiematrix van $(\beta_1^*, \dots, \beta_p^*)'$ wordt bij benadering (voor grote N) gegeven door

$$\text{cov}(\beta_1^*, \dots, \beta_p^*) \approx \frac{\sigma^2}{nN} (\Sigma^B)^{-1}, \quad (10)$$

waarin Σ^B de binnen-scholen covariantiematrix is van de leerling-variabelen x_1 t/m x_p . De gebruikelijke assumptie wordt gemaakt dat deze covariantiematrix dezelfde is voor elke school. De covariantiematrix (10) is hier dezelfde uitdrukking als geldt voor de situatie van een enkelvoudig aselechte steekproef leerlingen van omvang nN . Voor de vector van geschatte regressiecoëfficiënten van schoolvariabelen $(\beta_0^*, \gamma_1^*, \gamma_2^*, \dots, \gamma_q^*)'$ is de covariantiematrix bij benadering (voor grote N) gelijk aan

$$\text{cov}(\beta_0^*, \gamma_1^*, \gamma_2^*, \dots, \gamma_q^*) \approx \frac{1}{N} (\sigma_1^2 + \sigma^2/n)(\Sigma^T + \mu'\mu)^{-1}, \quad (11)$$

waarin μ de vector van gemiddelden en Σ^T de tussen-scholen covariantiematrix is van de variabelen $(x_0, z_1, \dots, z_q)'$; de hier met het intercept β_0 corresponderende x_0 is een 'dummy-variabele', voor alle leerlingen gelijk aan 1, die dus meedoet met de schoolvariabelen. Dezelfde formule als (11) zou worden verkregen wanneer er zou worden gewerkt met de schoolgemiddelden $\bar{y}_j$ in plaats van de individuele leerlinggegevens y_{ij} , en de regressiecoëfficiënten zouden worden geschat met de OLS methode na aggregatie tot schoolniveau.

De standaardfouten van de geschatte regressiecoëfficiënten zijn de wortels uit de diagonaal-elementen van (10), resp. (11). Aan deze formules is o.a. het volgende te zien. Veronderstel dat een onderzoek wordt gepland en dat op grond van kostenoverwegingen het totale aantal onderzochte leerlingen (dus nN) een vast gegeven is. Indien de onderzoeker geïnteresseerd is in de regressiecoëfficiënten β_0 en γ_h en niet zozeer in bijvoorbeeld de intercepten per school afzonderlijk $\beta_0 + u_{0j}$, dan wordt de kleinst mogelijke standaardfout verkregen door N zo groot mogelijk, en n gelijk aan 1 te kiezen. Het is dan uiteraard niet mogelijk om σ^2 te schatten; daarvoor zou men n minstens gelijk aan 2 moeten nemen.

De veronderstelling dat kostenoverwegingen leiden tot een bepaalde vaste of maximale waarde voor de totale steekproefomvang nN , is hier eigenlijk niet realistisch: de kosten van een onderzoek zullen behalve van het totaal aantal leerlingen ook van het totaal aantal scholen afhangen. Laten we aannemen dat de kosten van het onderzoeken van een leerling op een reeds onderzochte school gelijk zijn aan f , en dat de extra kosten voor het onderzoeken van een school gelijk zijn aan cf . De variabele kosten van het onderzoek zijn dan dus gelijk aan $N(n+c)f$. We zijn weer geïnteresseerd in een steekproefopzet die bij gegeven totale kosten leidt tot een minimale standaardfout voor β_0^* en γ_h^* . Veronderstel een budget van kf voor variabele kosten, voor een of ander gegeven getal k . Hieruit volgt dan $N(n+c)f = kf$, dus $N(n+c) = k$. Als N is vastgesteld volgt hieruit n , en vice versa. Door in (11) in te vullen $N = k/(n+c)$ kan men zien dat de varianties van β_0^* en γ_h^* onder deze veronderstellingen van n afhankelijk zijn via de evenredigheidsfactor

$$(n+c)(\sigma_1^2 + \sigma^2/n). \quad (12)$$

Deze factor is minimaal (zonder rekening te houden met de noodzakelijke heeltalligheid van n), wanneer n gelijk is aan de optimale waarde

$$n_{\text{opt}} = (\sigma \sqrt{c})/\sigma_1.$$

Hieruit volgt dat voor $n = n_{\text{opt}}$ de kleinste standaardfout bij gegeven kosten, en ook de kleinste kosten bij een gegeven standaardfout worden behaald. Kwalitatief is het uiteraard duidelijk, dat n_{opt} des te groter zal zijn, naarmate c groter is, en ook naarmate de binnen-school variantie groter is ten opzichte van de tussen-scholen variantie. Opmerkelijk is, dat dit optimale aantal leerlingen per school niet afhangt van het totale budget of de gewenste standaardfout; deze laatsten bepalen natuurlijk wel de omvang van N .

Om de formule voor n_{opt} te gebruiken is het nodig, een raming te maken van de verhouding tussen de residuele variantie op leerlingniveau σ^2 en de residuele variantie op school-niveau σ_1^2 . Daar zal altijd een flinke onzekerheidsmarge in zitten. Hoeveel hebben de standaardfouten te lijden, als in het onderzoek niet precies de optimale waarde van n wordt gebruikt? Het antwoord kan worden geleverd door bestudering van de functie (12), die immers de evenredigheidsfactor is via welke de varianties van β_0^* en γ_h^* van n afhangen. Ter illustratie gaan we er van uit, dat de verhouding tussen de binnen-scholen en de tussen-scholen variantie $\sigma^2:\sigma_1^2 = 9:1$ is. Dat deze verhouding realistisch is blijkt bijvoorbeeld uit het onderzoek van Coleman e.a. (1982), Willms (1987), Roeleveld (1987) en Bosker e.a. (1988). In figuur 2 wordt de grafiek gegeven van formule (12) als functie van n voor de waarden $c = 10$, $\sigma^2 = 0.9$, $\sigma_1^2 = 0.1$. Hierbij is $n_{\text{opt}} = 9.49$; onder de restrictie dat n heeltallig is bereikt (11) zijn minimumwaarde van 3.8 voor $n = 9$ en $n = 10$. Uit de figuur is duidelijk dat het minimum niet heel scherp is; voor $6 \leq n \leq 15$ is de functie kleiner dan of gelijk aan 4.

De formules (10) en (11) kunnen ook worden gebruikt voor het bepalen van het minimale aantal scholen N , nodig om te zorgen dat de standaardfout van een regressiecoëfficiënt onder een bepaalde waarde zal liggen. Hiervoor zijn ramingen nodig van de in de formules optredende gemiddelden en binnen- en tussen-scholen covariantiematrices, evenals van de tussen-leerling en tussen-school varianties σ^2 en σ_1^2 . Door de formules te berekenen voor verschillende praktisch haalbare waarden van N verkrijgt men dan een indruk van de te verwachten standaardfou-

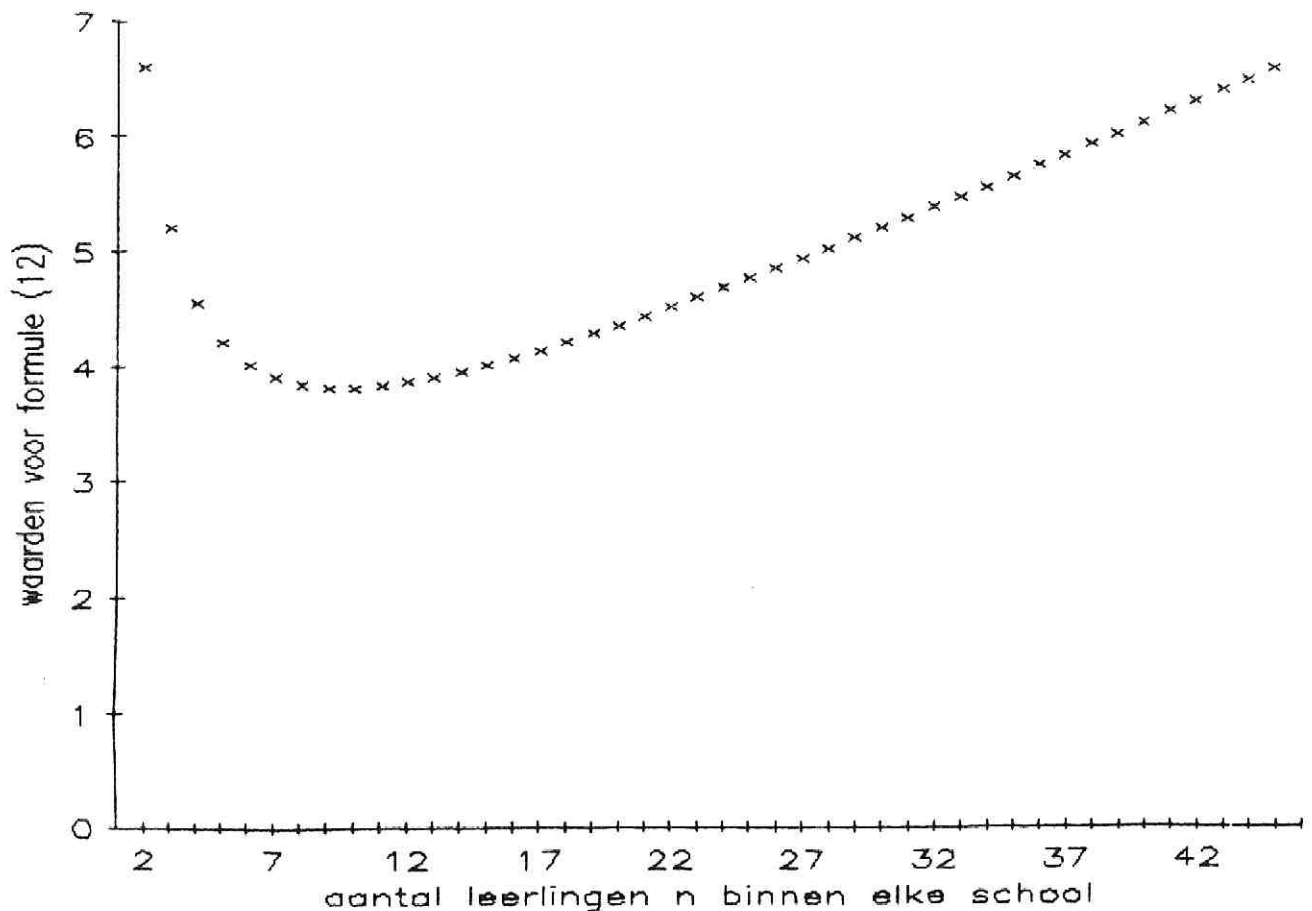


Fig. 2. De grafiek van $(n+c)(\sigma_1^2+\sigma^2/n)$ voor $c = 10$, $\sigma^2 = 0.9$, $\sigma_1^2 = 0.1$

ten. Ter illustratie geven we een voorbeeld⁷. We nemen weer $c = 10$, $\sigma_1^2 = 0.1$, $\sigma^2 = 0.9$ en veronderstellen dat er één schoolvariabele is, die is gestandaardiseerd tot gemiddelde 0 en variantie 1. Verder veronderstellen we dat er wordt gewerkt met de optimale waarde $n = n_{\text{opt}} = (\sigma \sqrt{c})/\sigma_1$, en bekommeren we ons er niet om dat dit niet een geheel getal is. Neem aan dat we primair geïnteresseerd zijn in de standaardfouten van β_0^* en γ_1^* , dus het gemiddelde intercept en de regressiecoëfficiënt voor de schoolvariabele. Omdat de boven ingevoerde dummyvariabele x_0 gemiddelde 1 en variantie 0 heeft, is $\Sigma^T \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} =$ en $\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Invullen in (11) levert nu als diagonaalelementen $0.19/N$. In grafiekvorm kunnen we nu N voorstellen als functie van de gewenste standaardfout van β_0^* en γ_1^* . Hierbij beperken we ons tot standaardfouten tussen 0.03 en 0.12, die relevant zijn voor regressiecoëfficiënten in de buurt van 0.10. Dit omdat het al mooi is als schoolvariabelen 10% van de variantie tussen scholen, dus 1% van de totale variantie kunnen binden (cf. Bosker en Scheerens, 1989), hetgeen overeenkomt met een gestandaardiseerde regressiecoëfficiënt van 0.10.

Voor dit simpele design kan het volgende worden geconcludeerd (een conclusie die uiteraard niet zonder meer naar andere designs kan worden gegeneraliseerd). Willen we een effect met een gestandaardiseerde regressiecoëfficiënt van 0.1 met een redelijke kans ontdekken met een klein aantal, zeg $N = 40$ scholen, dan zal het significantieniveau α ruim gekozen moeten worden, bijvoorbeeld $\alpha = 10\%$ bij eenzijdige toetsing. Bij strenge toetsing ($\alpha \leq 1\%$) is een groot aantal scholen nodig ($N > 75$). Natuurlijk kunnen we ook trachten het design te verbeteren volgens de daarvoor geldende overige methodologische regels (bijvoorbeeld, waar dit relevant kan worden uitgevoerd, gestratificeerde steekproeftrekking van scholen of het gebruik van meer covariaten).

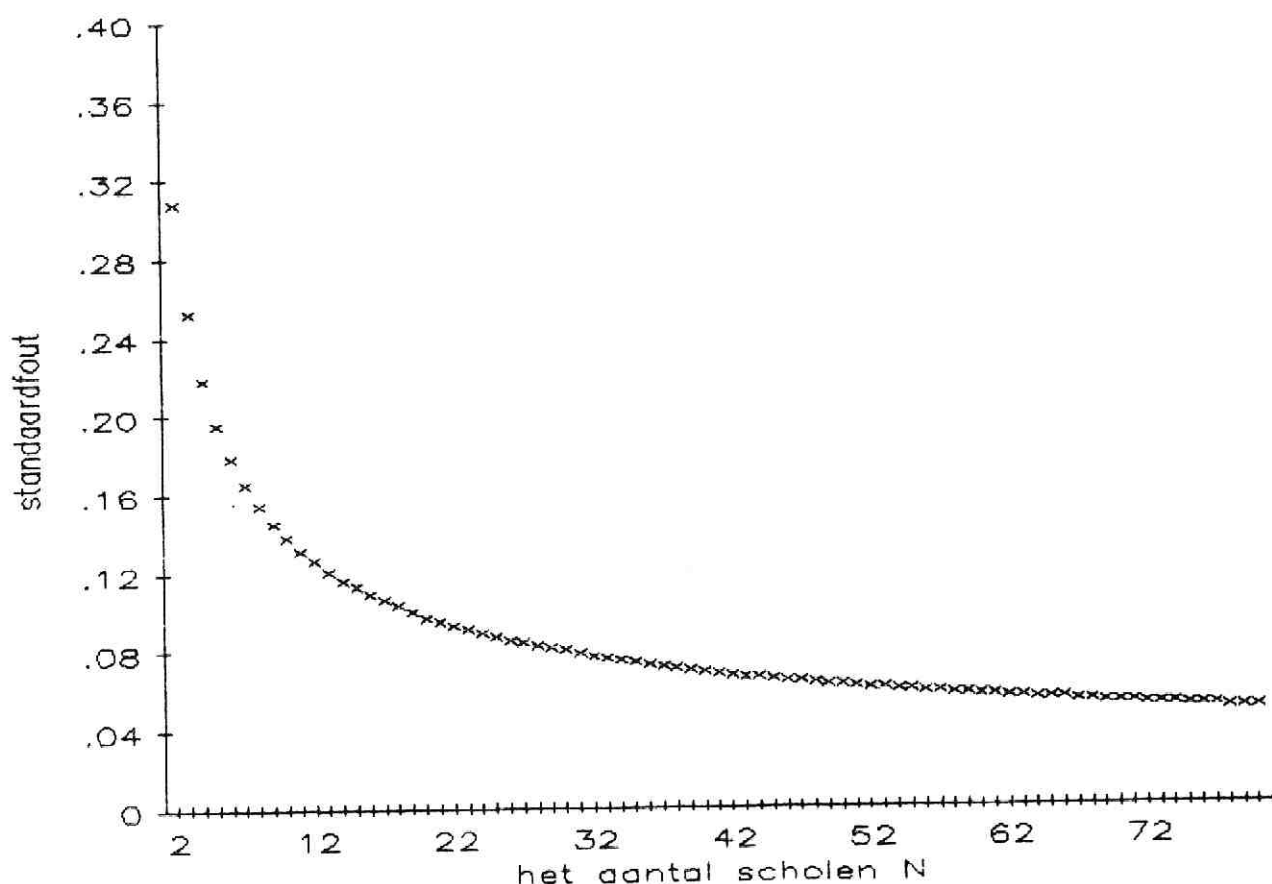


Fig. 3. Het aantal scholen N als functie van de gewenste standaardfout.

7. NABESCHOUWING

In het voorgaande is het random coëfficiënt model uitvoerig toegelicht. De getrouwe toepasser van regressie- of variantie-analytische technieken vraagt zich wellicht af wat nu de meerwaarde van het hiërarchisch model is boven deze meer gangbare technieken.

Regressie-analyse kan men in multi-niveau onderzoek op drie manieren "fout" toepassen (een meer uitgebreide discussie wordt gegeven door Aitkin en Longford, 1986). Aggregatie van de leerlingkenmerken naar het schoolniveau en vervolgens een regressie-analyse op dit niveau, heeft als nadeel dat de y_{ij} -scores (bijvoorbeeld prestaties) inadequaat gecorrigeerd zijn voor controle-variabelen (bijvoorbeeld IQ). Bovendien kunnen we na aggregatie de interactie-effecten van leerling- en schoolkenmerken niet meer onderzoeken. Disaggregatie van de schoolkenmerken naar het leerlingniveau zonder ermee rekening te houden dat de leerlingen binnen scholen zijn genest, heeft als belangrijk bezwaar dat de standaardfouten van vooral de regressie-coëfficiënten behorend bij de schoolvariabelen verkeerd geschat worden, omdat de assumptie van onafhankelijke waarnemingen onjuist is.

Er is ook een twee-staps regressie-analyse mogelijk (zie Willms, 1984 en van den Eeden en Saris, 1984): eerst worden de y_{ij} -scores per school voorspeld met de leerlingvariabelen. De schoolspecifieke intercepten en regressiecoëfficiënten worden vervolgens in een tweede stap voorspeld met de schoolvariabelen ('slopes as outcomes'). Deze methode gaat er van uit dat de intercepten en regressiecoëfficiënten van school tot school verschillen, maar zonder te veronderstellen dat ze onderling samenhangen. (In het random coëfficiënt model wordt samenhang gepostuleerd door te veronderstellen dat de scholen een steekproef uit een populatie van scho-

len vormen.) De intercepten en regressiecoëfficiënten zijn dus 'vaste' en niet 'random' effecten. Het nadeel is tweeledig: 1) er worden gigantisch veel parameters geschat met alle risico van kanskapitalisatie (dus grote standaardfouten); 2) we kunnen de uitkomsten niet generaliseren naar andere scholen.

Een standaard variantie-analyse met nesting van leerlingen onder scholen, en met covarianten op leerling-niveau, is hier ook een techniek waaraan nadelen kleven. De schoolspecifieke intercepten worden gemodelleerd als vaste in plaats van 'random' effecten, waartegen dezelfde bezwaren gelden die hierboven tegen de twee-staps regressie-analyse zijn aangevoerd. Bovendien worden homogene binnengroepsregressies verondersteld (bijvoorbeeld de relatie SES - prestaties is voor alle scholen hetzelfde). Deze assumptie is strijdig met veel theorieën over onderwijseffecten, waarin juist de schoolafhankelijkheid van de relatie tussen prestaties en kenmerken van de leerling (men kan hier denken aan het Onderwijs Voorrangs Beleid, dat o.a. beoogt de binnen-schoolse kloof tussen lager en hoger milieu leerlingen in de loop der tijd te verkleinen) centraal staat.

Een aantrekkelijk aspect van het in dit artikel behandelde 'random coefficient' model is onzes inziens dat 'shrinkage' technieken worden gebruikt, zodat zowel ongebalanceerde designs als uitbijters ('outliers') goed behandeld kunnen worden. Men zou kunnen stellen dat het een nadeel is, dat de techniek door de Bayesiaanse context leidt tot onzuivere schattingen van schooleffecten. Deze vertekening ('bias') is echter voorzichtig van aard (conservatief in de statistische betekenis van dit woord): hij trekt de geschatte schooleffecten in de richting van het gemiddelde schooleffect, hetgeen in overeenstemming is met de methodologische eis de alternatieve hypothese (dat er verschillen tussen scholen bestaan) met enig wantrouwen te beschouwen. In het licht van dit ingebouwd wantrouwen is nadere bestudering van het onderscheidingsvermogen van toetsen in het 'random coefficient' model, waartoe we in paragraaf 6 een aanzet hebben gegeven, allerminst een overbodige luxe.

NOTEN

1. De Engelse benaming voor deze werkwijze is 'two-stage sampling' (Cochran, 1977).
2. Er zijn vele benamingen voor het model dat we in dit artikel zullen beschrijven. In de engelstalige literatuur komt men onder meer tegen: 'mixed linear model', 'multi-level model', 'hierarchical linear model', 'variance component model', 'random coefficient model' en 'varying parameter model'.
3. De benaming "vaste coëfficiënt-model" wordt ook wel gebruikt voor het model waarbij de schooleffecten verwaarloosd worden (oftewel gelijk aan 0 gesteld) en dus een gewone regressie-analyse wordt uitgevoerd.
4. Niet alle multi-niveau software is bestand tegen een extreem laag aantal leerlingen per school. HLM bijvoorbeeld vereist varianties en covarianties binnen scholen die ongelijk aan 0 zijn. In VARCL is dit echter geen vereiste.
5. Afhankelijk van de overige in het model opgenomen regressoren bestaat in sommige gevallen het verschil tussen het al dan niet werken met binnen-groep deviatiescores alleen uit een verschil in parametrisering, met daardoor de mogelijkheid van interpretatieverschillen, van in wezen hetzelfde model; terwijl in andere gevallen het wezenlijk verschillende statistische modellen betreft. Een discussie over voors en tegens van centreren wordt gevoerd door Raudenbush (1989a, b), Longford (1989) en Plewis (1989).
6. Op dit moment wordt een project uitgevoerd om standaardfouten en onderscheidingsvermogens te onderzoeken in designs met zowel random intercepten als random regressiecoëfficiënten. De in heterogene binnengroepsregressies geïnteresseerde lezer zal dus nog enig geduld moeten oefenen voordat richtlijnen over optimale steekproefgroottes voor dergelijke designs bekend zijn.
7. Het lijkt merkwaardig om N te berekenen uit een formule waarin N groot wordt verondersteld. Men moet bij dit laatste echter denken aan $N \geq 10$ waarbij $n \geq 5$.

LITERATUUR

- Abalos, J., Jolly, J. & Johnson, R. (1985). *Statistical methods for selecting merit-schools*. AERA-paper, Chicago.

- Aitkin, M. & Longford, N.T. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society, Series A (General)* 149, 1-43.
- Anderson, L.W. & Mandeville, G.K. (1986). *Towards a solution of problems inherent in the identification of effective schools*. AERA-paper, San Francisco.
- Bosker, R.J., Guldemon, H.G., Hofman, R.H. & Hofman, W.H.A. (1988). *Kwaliteit in het voortgezet onderwijs*. RION: Groningen.
- Bosker, R.J. & Scheerens, J. (1989). Issues in the interpretation of the results of school-effectiveness research. *International Journal of Educational Research*, 13, 741-751.
- Boyd, L.H. & Iversen, G.R. (1979). *Contextual Analysis: concepts and statistical techniques*. Wadsworth: Belmont.
- Bryk, A.S., Raudenbush, S.W., Seltzer, M., & Congdon, R.T., (1988). *HLM*. Chicago: University of Chicago.
- Burstein, L. (1980). The analysis of multilevel data in educational research, *Review of Research in Education*, 8, 185-233.
- Cochran, W.G. (1977). *Sampling techniques*, 3d ed. Wiley, New York.
- Coleman, J.S., Hoffer, T. & Kilgore, S. (1982). *High School Achievement*. New York: Basic Books.
- Cronbach, L.J. (1976). *Research on classrooms and schools: formulation of questions, design and analysis*. Occasional Paper. Stanford: Stanford Evaluation Consortium.
- Dobson, A. (1983). *An introduction to statistical modelling*. London: Chapman and Hall.
- Eeden, P. van den & Saris, W.E. (1984). Empirisch onderzoek naar multilevel uitspraken. *Mens en Maatschappij*, 59, 164-178.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Charles Griffin & Co.
- Goldstein, H. & Rasbash, J. (1988). *Multilevel software version 88.3*. London: University of London.
- Kreft, G.G. & Leeuw, J. de (1987). Het schatten van schooleffecten. Problemen bij de keuze van het model en de techniek. *Tijdschrift voor Onderwijsresearch*, 12, 75-85.
- Kreft, G.G. (1987). *Models and methods for the measurement of schooleffects*. Dissertatie. Amsterdam.
- Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.
- Longford, N.T. (1988). *VARCL: Software for variance component analysis of data with hierarchically nested random effects (maximum likelihood)*. Manual. Princeton, N.J.: Educational Testing Service.
- Longford, N.T. (1989). To center or not to center. *Multilevel Modelling Newsletter*, 1.3, 7, 11.
- Moors, J.J.A. & Mulwijk, J. (1975). *Steekproeven, een inleiding tot de praktijk*. Amsterdam: Elsevier.
- Plewis, I. (1989). Comment on 'Centering' predictors in multilevel analysis'. *Multilevel Modelling Newsletter*, 1.3, 6, 11.
- Raudenbush, S.W. & Bryk, A.S. (1986). A hierarchical model for studying schooleffects. *Sociology of Education*, 59, 1-17.
- Raudenbush, S.W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13, 85-116.
- Raudenbush, S.W. (1989a). 'Centering' predictors in multilevel analysis. *Multilevel Modelling Newsletter*, 1.2, 10-12.
- Raudenbush, S.W. (1989b). A response to Longford and Plewis. *Multilevel Modelling Newsletter*, 1.3, 8-11.
- Roeleveld, J. (1987). Verschillen in resultaten tussen Nederlandse scholen voor voortgezet onderwijs. In J. Scheerens en W.G.R. Stoel (red.): *Effectiviteit van onderwijsorganisaties*. Lisse: Swets & Zeitlinger.
- Snijders, T.A.B. & Bosker, R.J. (1989). *Standard errors and sample sizes in two-level designs*. Heymans Bulletin HB-89-984-EX, Psychologische Instituten R.U. Groningen (aangeboden ter publikatie).
- Willms, J.D. (1984). School effectiveness within the public and private sector. *Evaluation Review*, 8, 1, 113-135.
- Willms, J.D. (1987). *A multi level approach for analyzing school effectiveness scores*. Edinburgh: CES.

Enige literatuur over multi-niveau onderzoek

(Opmerking: het onderstaande overzicht van recente literatuur over modellen voor multilevel analyse is bedoeld als een eerste kennismaking. Bij de keuze van de referenties is vooral gelet op teksten die een overzicht bieden, die toepassingen geven op het gebied van de onderwijswetenschappen en die betrekking hebben op de Nederlandse situatie. Daarmee is deze lijst uiteraard onvolledig.)

Aitkin, M., en N. Longford, 1986, Statistical modelling issues in school effectiveness studies. *The Journal of the Royal Statistical Society. Series A (General)*, 149, 1-43.

In dit artikel wordt op een systematische manier een overzicht gegeven van de verschillende deelmodellen van het random coëfficiënten-model. De modellen worden telkens geïllustreerd met toepassingen ontleend aan de Londense ILEA-studies.

Bock, D.R. (red.), 1989, *Multilevel analysis of educational data*, San Diego etc. Academic Press.

Deze bundel bevat een aantal bijdragen over statistische aspecten en die samenhangen met de rondom coëfficiënten-modellen. In enkele gevallen zijn illustraties van toepassingen toegevoegd.

Bosker, R.J., en R.K.W. van der Velden, 1989, Schooleffecten en rendementen. In: J. van Damme en J. Dronkers (red.), *Jongeren in school en beroep*. Amsterdam/Lisse, Swets en Zeitlinger, 25/40.

In deze bijdrage wordt een multilevel analyse uitgevoerd op een longitudinaal databestand dat verschillende scholen bevat. Het model wordt aangewend in evaluatie-onderzoek.

Boyd, L.H. en G.K. Iversen, 1979, *Contextual analysis: concepts and statistical techniques*. Wadsworth, Cal.: Wadsworth.

Dit inmiddels klassieke handboek geeft zonder veel statistiek een systematisch overzicht van een aantal technische problemen die zich voordoen bij multilevel analyse. Het kan worden beschouwd als een aardige voorbereiding op de random coëfficiënt-modellen.

Eeden, P. van den, 1986, *Aspects of Multilevel Research*. Amsterdam, Vrije Universiteit.

Een uiteenzetting van methodologische en theoretische uitgangspunten van multilevel onderzoek, aangevuld met voorbeelden over toepassingen op analyse en operationalisering.

Eeden, P. van den, en J. Hauer (red.), 1984, *Analyse van multiniveau-data: Theorie en toepassing*. 's-Gravenhage, SVO.

Deze bundel bevat een reeks van bijdragen van Nederlandse auteurs over multilevel analyse. De uitgebreide inleiding is van de hand van L. Burnstein.

Eeden, P. van den, en H.J.M. Huettner, 1982, *Multilevel Research*. Beverly Hills/London, Sage.

Een schets van de uitgangspunten voor multilevel onderzoek, betrekking hebbend op de probleemstelling, theorie, operationalisering en analyse. Het boekje bevat een zeer uitgebreide bibliografie.

Erbring, L., en A.A. Young, 1979, Individuals and social structure: contextual effects and endogeneous feedback. *Sociological Methods and Research*, 7, 396-430.

Een overzicht van het 'endogene feedbackmodel' voor multilevel analyse, en een toepassing ervan.

Goldstein, H., 1987, *Multilevel models in educational research*. London, Griffin.

Dit boekje geeft beknopt een systematisch overzicht van de random coëfficiënten-modellen.

Hannan, M.T., 1971, *Aggregation and disaggregation in Sociology*. Lexington, Mass., DC Heath.

Met een minimum aan statistiek wordt in dit inmiddels klassieke boek de 'aggregatie-fout' uiteengezet en verklaard.

Huettner, H.J.M., 1985, *Multilevel onderzoek: de operationalisatie van omgevingskenmerken*. Nijmegen, Katholieke Universiteit.

In dit proefschrift wordt de operationalisatie van omgevingskenmerken in descriptieve en in normatieve zin behandeld.

Koehler, S., en P. van den Eeden, 1987, De invloed van klassamenstelling op schoolloopbanen: een multilevel onderzoek bij het algemeen voortgezet onderwijs in Amsterdam. Amsterdam, Vrije Universiteit, vakgroep Methoden en Technieken.

Een toepassing van het random coëfficiënten-model op de effecten van kenmerken van klassen op de schoolloopbanen van leerlingen.

Koopman, P., P. van den Eeden, P. Jungbluth, 1989, Opslag en beschikbaarheid van gegevens: een voorstudie voor de evaluatie van de invoering van het voortgezet basisonderwijs. (SVO project 7149). Amsterdam, SCO.

Een beschrijving van hoe multilevel gestructureerde gegevens te manipuleren en op te slaan in een database.

Kreft, G.G., 1987, Models and methods for the measurement of schooleffects. Diss. Universiteit van Amsterdam.

In dit proefschrift wordt van het random coëfficiënten-model gebruik gemaakt bij de evaluatie van de mammoetwet in Nederland.

Kreft, G.G., en J. de Leeuw, 1987, Het schatten van schooleffecten problemen bij de keuze van het model en de techniek. Tijdschrift voor Onderwijsresearch, 12, 75/85.

Het model met wisselende intercepten en vaste hellingen wordt uiteengezet en geïllustreerd.

Leeuw, J. de en G.G. Kreft, 1986, Random co-efficient models for multilevel analysis. Journal of Educational Statistics, 11, 57/85.

Dit artikel bevat een algemene statistische uiteenzetting van de random coëfficiënten-modellen en een vergelijking van verschillende toetsingsprocedures.

Mooij, T., 1987, Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations. 's-Gravenhage, SVO/Nijmegen, ITS.

Een multilevel analyse op twee jaargangen leerlingen van het LBO. Nagegaan wordt het effect van klas- en instructiekenmerken van de vakken Nederlands, Engels en lichamelijke oefening op het leerproces van de leerlingen.

Oosthoek, J.K., 1985, Effecten van de leeromgeving, een sociaalpsychologisch onderzoek naar factoren die de rechtsopvatting van studenten beïnvloeden. Utrecht, Universitaire Drukkerij.

In dit proefschrift wordt een multilevel analyse uitgevoerd, gericht op de effecten van kenmerken van de Nederlandse universitaire opleidingen in de rechten op de rechtsopvatting van de studenten.

Oosthoek, H., en P. van den Eeden, (red.), 1984, Education from the multi-level perspective: Models, methodology and empirical findings. New York, etc. Gordon and Breach.

Deze bundel bevat zowel overzichtsartikelen als toepassingen van multilevel analyses op het gebied van de onderwijswetenschappen.

Raudenbush, S., en A.S. Bryk, 1986, A hierarchical model for studying school effects. Sociology of Education, 59, 1/17.

Op een heldere wijze wordt een statistische uiteenzetting gegeven van het random coëfficiënten-model. Dit model wordt vervolgens toegepast op data van de studie van Coleman, Hoffer en Kilgore.

Roeleveld, J., P. Koopman, U. de Jong en P. van den Eeden, 1988, Schoolloopbanen in Amsterdam: Leerlingen op Amsterdamse scholen voor algemeen voortgezet onderwijs in de jaren zeventig. Amsterdam, vakgroep Methoden en Technieken (Rapport SVO 6595).

Het gaat hier om een verslag van een grootschalig multilevel onderzoek naar de effecten van diverse schoolkenmerken bij vier cohorten Amsterdamse leerlingen. In de analyses is niet alleen rekening gehouden met cohort, maar ook met schooltype (MAVO, HAVO, VWO), schoolstructuur (categoriale scholen of scholengemeenschappen) en denominatie.

Saldern, M. von, (red.), 1986, Mehrebenenanalyse: Beitrage zur Erfassung hierarchischer strukturierter Realitaet. Belz, Weinheim und München.

Een bundeling van bijdragen over toepassingen van multilevel modellen.

Multi-niveau programmatuur: enige gegevens

Er zijn op dit moment vier computerprogramma's beschikbaar, die specifiek bedoeld zijn voor het soort multi-niveau analyses die in deze aflevering van TOR aan de orde komen.

Deze programma's zijn:

- VARCL (Variance Component Analysis; auteur: N. Longford)
- HLM (Hierarchical Linear Modelling; auteurs: S. Raudenbush en A. Bryk)
- ML (Multilevel; auteurs: J. Rashbash, R. Prosser en H. Goldstein)
- GENMOD (GENeral MODelling; auteurs: B. Hermalin en A.F. Anderson)

Van deze vier programma's zijn verschillende versies in omloop, afhankelijk van het aantal analyseniveaus dat ze kunnen behappen. Van VARCL bestaan de versies VARCL-2, VARCL-3 en VARCL-9. Van HLM bestaat HLM-2, en HLM-3 is aangekondigd. De opvolger van ML-2, te weten ML-3, wordt gedistribueerd per december 1989. GENMOD wordt, voor zover ons bekend, in Nederland nauwelijks gebruikt.

Een geheel nieuw programma, Multipath (auteur: J. de Leeuw) is aangekondigd. Met dit programma kan pad-analyse met multilevel data worden gedaan. De verwachting is dat dit programma in de loop van 1990 op de markt zal komen.

Hoewel de verschillende programma's al zeer goed bruikbaar zijn is geen van de programma's definitief klaar. Nog regelmatig worden verbeteringen aangebracht of uitbreidingen toegevoegd.

Multilevel modelling newsletter

Om op de hoogte te blijven van de ontwikkelingen op het gebied van de multilevel analyse is een abonnement op de Multilevel Modelling Newsletter aan te bevelen. U kunt zich opgeven bij:

Bob Prosser
c/o Math Statistics & Computing Department
Institute of Education/University of London
20 Bedford Way
London WC1H 0AL

Vergelijking van de programma's

Een onderzoek naar de vergelijking van VARCL, HLM, ML-2 en GENMOD bevindt zich momenteel in een eindfase (Kreft, 1990). In dit onderzoek wordt aandacht besteed aan de volgende punten:

- het algoritme
- hoe moet de set-up van de data er uit zien
- hoe wordt er omgegaan met missing values
- welke opties heeft het programma
- hoe zien de output-files er uit
- schatting van de betrouwbaarheid en de daaraan verbonden statistische maten
- gebruikersvriendelijkheid
- door welke unieke functies en eigenschappen onderscheidt het programma zich
- welke beperkingen heeft het programma

Hoewel gebruik makend van verschillende algorithmes gaven de programma's dezelfde uitkomsten, al heeft Kreft enkele onverwachte uitzonderingen gevonden. Deze uitzonderingen betroffen zogenoemde 'bugs' in de programma's die merendeels eenvoudig te herstellen bleken.

Afgezien van de uitkomsten verschillen de programma's op marginale onderdelen: snelheid, geheugenbeslag, presentatie van de resultaten, etc. ML-2 wijkt af van de overige programma's in zoverre ML-2 is ingebed in het statistisch programmapakket NANOSTAT. Dit heeft tot ge-

volg, dat het prepareren van de datasets en het eventuele transformeren (hercoderen) van gegevens als het ware binnen hetzelfde programma plaats kan vinden. Dit heeft tevens tot gevolg dat de uitkomsten van de analyses weergegeven kunnen worden met de grafische mogelijkheden van NANOSTAT. Bij de overige programma's moeten de preparatie van de datasets en de transformaties van de data plaatsvinden in een afzonderlijk statistisch programma, zoals SPSS en SAS. Evenzo moet bij deze programma's voor een eventuele grafische presentatie van resultaten gebruik worden gemaakt van andere programma's.

Prijzen, bestellingen en hardware

HLM

HLM-2 en HLM-3 kosten beiden ongeveer \$ 300.- en zijn te bestellen bij:

Scientific Software
1369 Neitzel Road
Mooresville
Indiana 46158-9312
U.S.A.

Van HLM-2 zijn mini- en mainframe versies beschikbaar onder VM/CMS, UNIX en VMS. Ook is er een MS-DOS versie, waarvoor 640K, een hard disk en een mathematische co-processor nodig is.

VARCL

VARCL-3 kost \$ 800.- en is te bestellen bij:

N. Longford
21-T Rosedale Road
Educational Testing Service
Princeton
New Jersey 08541
U.S.A.

VARCL is oorspronkelijk ontworpen op een VAX onder VMS. Voor ter zake kundigen is het eenvoudig VARCL te implementeren op welk mainframe dan ook. Van VARCL-2 en VARCL-3 zijn ook versies voor AT-computers beschikbaar.

ML

Voor diegenen die ML-2 voor £ 50.- hebben aangeschaft is ML-3 gratis. Voor snelle bestellers kost ML-3 eveneens £ 50.-. De definitieve prijs van ML-3 zal in het voorjaar van 1990 worden bepaald. ML-3 is te bestellen bij:

Bob Prosser
c/o Math Statistics & Computing Department
Institute of Education
University of London
20 Bedford Way
London WC1H 0AL

ML-3 is ontworpen voor AT-computers met een harddisk. Er is een versie beschikbaar voor AT's met mathematische co-processoren. Er bestaan concrete plannen voor VAX/VMS versie.

Literatuur

Kreft, I., Leeuw J. de & Kim, K. (1990). Comparing four different statistical packages for hierarchical linear regression: GENMOD, HLM, ML-2 and VARCL. University of Los Angeles, versie 2.0, UCLA-rapport #310.

Ontvangen 13-7-1990

Notities en Commentaren

ONDERWIJSONDERZOEKERS GEVEN ZICH BLOOT

N.A.J. Lagerweij en J.F.M. Claessen

Rijksuniversiteit Utrecht en Open Universiteit, Heerlen

Onderwijskundigen denken verschillend over hun invloed en rol in het publieke debat over onderwijsvraagstukken. Ook over de betekenis en functies van onderwijsonderzoek voor het onderwijs lopen de meningen onder onderwijskundigen uiteen. Deze situatie is niet alleen typisch voor onderwijskundigen, het is ook kenmerkend voor de sociale wetenschappen in het algemeen.

Onlangs verscheen een rapport over schaalvergroting in het basisonderwijs (PSB, 1990). De reacties erop waren talrijk en in veel gevallen fel en afwijzend. Het gaat hier duidelijk om een in ons land gevoelig onderwerp.

Ook onder onderwijskundigen bleken de meningen niet onverdeeld gunstig. Creemers en Hoebe gaven in het vorig nummer van dit blad commentaar onder de titel: "Onderwijsonderzoek te kijk gezet". De inhoud is een duidelijke illustratie van verschillen in opvattingen onder onderwijskundigen. Het bijzondere van hun reacties is, dat zij hun subjectieve en normatieve opinie tot maat maken voor anderen en voor de onderwijskunde in het algemeen. Daarmee overschatten zij hun relatieve positie onder onderwijskundigen en maken zij ten onrechte gewag van het op het spel staan van reputaties van collega's.

Ongeacht de verschillen in opinies onder onderwijskundigen, steeds blijft, zo leert ons het verleden, bescheidenheid geboden omtrent de doorwerking van onderwijskundige ideeën en onderzoeksresultaten in onderwijspolitiek en onderwijspraktijk. Onderwijskundigen mogen volgens Creemers en Hoebe gehoor geven aan een verzoek van de staatssecretaris van Onderwijs en Wetenschappen om mee te werken aan een advies over schaalvergroting in het onderwijs. Blijkens hun commentaar hadden zij zelf wel enkele beleidsadviezen kunnen verzamelen. De staatssecretaris vroeg echter niet aan deze RION-onderzoekers of zij zouden willen meewerken aan een advies. Hij nodigde daartoe een tweetal andere onderwijskundigen uit, te weten Lagerweij en Claessen. Op persoonlijke titel schreven deze twee een nota, die als bijlage werd opgenomen in het rapport: 'Schaal en kwaliteit in het basisonderwijs'. Het gehele rapport kwam mede onder hun verantwoordelijkheid tot stand.

Creemers en Hoebe zijn van mening, dat de onderwijskundige argumentatie in het rapport "een onaanvaardbaar laag empirisch gehalte en een onaanvaardbaar hoog speculatief gehalte heeft." Een dergelijke beschuldiging is gebaseerd op de opvatting dat meer belang gehecht moet worden aan empirische evidentie dan aan theoretische speculatie en creativiteit. Onze reactie kan kort zijn: wij delen in geen enkel opzicht deze normatieve zienswijze van Creemers en Hoebe.

Het siert hen, dat zij bescheidenheid prediken als het gaat om het empirisch onderbouwen van beleidsadviezen. Onderwijsonderzoek heeft maar weinig houvast te bieden bij bijvoorbeeld het bepalen van de inhoud van de basisvorming, of het vergroten van de zorgbreedte, of het invoeren van milieu-educatie of internationale aspecten in het onderwijs. Wie daarbij vasthoudt aan hetgeen uit empirisch onderzoek tot dan toe bekend is, zal meestal tot weinig steun biedende uitspraken kunnen komen.

Onderwijskunde is meer dan onderwijsonderzoek. Wetenschap is meer dan wat de cijfers ons leren. Wie echter een technologische wetenschapsopvatting heeft, zal al gauw afhaken bij wat dan in zijn ogen 'speculatie' heet. Wij verhelen geenszins dat de onderwijskundige bijlage van het rapport niet en ook niet grotendeels gebaseerd is op reeds verricht onderzoek. Dat onderzoek

is er immers niet, omdat de in het rapport beschreven 'optimale grote school' (nog) niet bestaat. We hebben een redenering opgesteld, waarbij we moesten blijven binnen het totale budget dat voor het basisonderwijs beschikbaar is. De vraag die we ons daarbij gesteld hebben luidt: 'Wat is de optimale en de minimale schoolgrootte in relatie tot maatstaven van kwaliteit in het basisonderwijs?' Vervolgens zijn we nagegaan welke trends er te vinden zijn in het denken over kwalitatief goed basisonderwijs voor de komende decennia.

Onze voorstellen zullen niet voor eenieder vanzelfsprekend zijn. Voor velen bleken zij zelfs verwerpelijk en onaanvaardbaar te zijn. Dat was te voorzien. In onze nota staat op pagina 88 de volgende passage: "Het ligt niet in de lijn van de verwachting dat we in deze notitie tot antwoorden komen, die voor iedereen in onderwijsland aanvaardbaar en overtuigend zijn. De problematiek van de optimale schoolgrootte is namelijk niet op empirische of objectieve wijze alleen tot een oplossing te brengen. Onvermijdelijk spelen op lokaal en nationaal niveau normatieve overwegingen en keuzen een rol als het gaat om het aangeven van een maat voor de basisschool. Dit normatieve karakter wordt zichtbaar in het hiernavolgende, in het bijzonder bij de onderwijskundige, economische en schoolorganisatorische onderbouwing van de voorstellen. Op grond van andere opvattingen en waarden kan men ook tot andere keuzen komen. Onze bedoeling is een redenering te volgen die ervan uitgaat dat de optimale schoolgrootte leidt tot onderwijs van goede kwaliteit voor alle leerlingen die naar het basisonderwijs gaan."

Ook Creemers en Hoebe zijn eigenlijk wel voor schaalvergroting, zoals valt af te leiden uit hun voorstellen voor beleidsalternatieven. Zij komen met enkele andere voorstellen voor onderwijsbeleidsmakers voor de besteding van door schaalvergroting vrijgekomen gelden. Wij verschillen daarmee kennelijk niet van mening over wenselijkheid en mogelijkheid van schaalvergroting in het basisonderwijs. Helaas beperken zij overwegingen van kwaliteit alleen tot leerprestaties. Het is een trend in onderwijsonderzoek alle veranderingen in het onderwijs af te meten aan de vraag, of zij bijdragen aan verhoging van de leerprestaties. Deze te enge opvatting over kwaliteit van het onderwijs wordt in het rapport niet gevolgd. Kwaliteit van onderwijs omvat behalve de bereikte leerresultaten ook de nagestreefde doelen, de gekozen leerinhoud, de wijze waarop er in scholen en klassen gewerkt wordt en tenslotte onder welke voorwaarden kinderen en leerkrachten hun werk moeten doen.

Wij kunnen ons voorstellen dat ook Creemers en Hoebe van mening zijn dat een goede basisschool:

- een vrijgestelde directeur heeft, die de school in haar geheel kan leiden, organisatorisch en onderwijskundig;
- een conciërge en een administratieve kracht bezit, die zorg dragen voor ondersteuning van de schoolleiding;
- zorg op maat verleent aan alle leerlingen en de huidige uitstroom naar speciaal onderwijs heeft beperkt;
- doet wat de wet vraagt: een ononderbroken schoolloopbaan garanderen voor alle leerlingen, daarbij zorg dragend voor optimale vormen van individualiserend onderwijs;
- leerkrachten bezit, die zodanig worden ingezet dat recht wordt gedaan aan hun voorkeuren en specialisaties binnen een gericht loopbaanbeleid.

Zo'n wenselijke basisschool krijgt alleen kans als er sprake is van een zekere omvang (250 à 450 leerlingen). Bovendien moet het basisonderwijs ook betaalbaar blijven. Een kleine school is veel duurder per leerling dan een grote school.

Welke empirische evidenties moeten deze keuzen voor een goede basisschool nog onderbouwen? Er is eerder verbeelding nodig om deze school te bepleiten dan gegevens uit onderwijsonderzoek zoals Creemers en Hoebe op het RION (laten) verrichten. Het is positief dat deze Groningse onderwijskundigen bescheiden zijn in hun verwachtingen het onderwijs op basis van hun onderzoeksresultaten te kunnen veranderen. Het is te betreuren dat zij angstig zijn om creatief onderwijskundig mee te denken over de school in de toekomst.

Over de kans op verwerkelijking van de voorstellen in het rapport kunnen we kort zijn. Wij hebben ons niet laten leiden door motieven van haalbaarheid. Onze bijdrage was gericht op het presenteren van een ideaal model voor de basisschool, die binnen het gegeven budgettaire kader

alle kansen krijgt de beste kwaliteit onderwijs te leveren, die past binnen hetgeen de wet thans voorschrijft. We hebben vanuit onderwijskundige argumentatie willen bijdragen aan de maatschappelijke meningsvorming betreffende de schoolgrootte. Over de graad van wetenschappelijkheid en over reputatie van onderwijsonderzoek maken wij ons even druk als Creemers en Hoebe. Alleen ontbreekt het aan objectieve maatstaven om te bepalen wie hier te kijk gesteld wordt.

LITERATUUR

- Creemers, B.P.M. & Hoebe, W.Th.J.G. (1990). Onderwijsonderzoek te kijk gezet. Commentaar op het rapport "Schaal en kwaliteit in het basisonderwijs". *Tijdschrift voor Onderwijsresearch*, 15, 234-241.
- Projectgroep Schaalvergroting Basisonderwijs (1990). *Schaal en kwaliteit in het basisonderwijs*. Zoetermeer: Ministerie van Onderwijs & Wetenschappen.

Ontvangen 3-10-1990

De meerwaarde van een goede probleemaanpak

Marcel V.J. Veenman en Jan J. Elshout
Universiteit van Amsterdam

ABSTRACT

The relation between intellectual ability, novice problem solving behavior and learning in two different learning environments was investigated. High and low intelligent novices in the domain of physics were assigned to either a structured or an unstructured learning condition in order to master major principles of electricity theory. In the structured learning environment the subject matter was presented in a fixed sequence from qualitative to quantitative reasoning about relations between concepts in the domain. Furthermore, students were guided towards understanding electricity theory by conducting pre-set experiments in simulated electrical circuits. The unstructured approach allowed for unguided learning by discovery. The quality of problem solving behavior was assessed by an analysis of tape-recorded think-aloud protocols. Intelligence as well as quality of problem solving approach correlated positively with learning measures, but contrary to our predictions the correlation between intelligence and problem solving approach was not significant. In fact, problem solving approach contributed independently to learning. No learning effects due to structuredness of learning environment or interaction effects between intelligence and learning environment were found.

INLEIDING

De literatuur laat er weinig twijfel over bestaan dat er grote verschillen zijn tussen beginners en experts in het oplossen van natuurkunde-problemen (Chi, Glaser & Rees, 1982). Niet alleen maken beginners meer fouten en hebben zij meer tijd nodig om tot een oplossing te komen, tevens bestaan er grote kwalitatieve verschillen tussen beiden. Ten eerste beschikken experts niet alleen over méér kennis dan beginners, de kennis waarover zij beschikken is ook beter georganiseerd (Chi et.al., 1982), minder 'boekig' van aard (Elshout, 1985) en daardoor direct inzetbaar. Verder verrichten experts, alvorens tot handelen over te gaan, een uitgebreide kwalitatieve analyse van de probleemstelling (Chi, Feltovich & Glaser, 1981). Als gevolg van deze kwalitatieve analyse representeren experts de natuurkunde-problemen direct in termen van relevante principes en wetmatigheden, waardoor schema's met voorwaartse oplossingsstrategieën worden geactiveerd (Chi et.al., 1981; Anderson, 1985). Beginners vertonen daarentegen de neiging om direct en bovendien weinig systematisch te handelen (Elshout, 1985, 1987a). Zij categoriseren een probleemstelling aan de hand van oppervlakkige kenmerken van het probleem, hetgeen resulteert in een gebrekkige probleemrepresentatie waarop uitsluitend 'zwakke' probleemoplossingsstrategieën als means-ends-analysis kunnen worden toegepast (Chi et.al., 1981; Anderson, 1985).

Ook tussen beginners onderling kunnen echter verschillen bestaan. Ondermeer Elshout introduceerde de notie van de 'expert-beginner' (Elshout, 1985); een beginner die, hoewel nieuwkomer in een bepaald domein, zich expertachtig gedraagt door een meer effectieve probleemaanpak te vertonen. In tegenstelling tot 'echte' beginners, die ten prooi vallen aan het beginnerssyndroom (Elshout, 1987b), zouden expert-beginners zich beter oriënteren alvorens te handelen, meer systematisch te werk gaan, minder slordig zijn en meer evalueren tijdens het probleemoplossingsproces. Bovendien toonden Chi, Bassok, Lewis, Reimann & Glaser (1989) aan dat goede studenten meer elaboreren, hetgeen hun begrip van de gepresenteerde natuurkunde-principes ten goede komt. Dit effectieve aanpakgedrag zou één van de 'gereedschappen' zijn die

de student zélf inbrengt in de leersituatie om het leerproces te structureren. Het beschikken over dit 'gereedschap' is volgens Veenman, Elshout & Bierman (1989b) een manifestatie van intelligentie.

Eerder rapporteerden Elshout & Veenman (1990) de resultaten van een onderzoek naar de relatie tussen intelligentie, beginnersgedrag en leren in het domein van de warmteleer. Zij lieten hoog en laag intelligente proefpersonen met een natuurkunde-deficiëntie gedurende ruim een uur werken in een computergesimuleerd warmteleer-laboratorium, waarbij hardop-denk protocollen werden afgenomen. De resultaten op verschillende leermaten waren eenduidig: hoog intelligente proefpersonen bleken significant meer te leren dan laag intelligente proefpersonen. Opmerkelijker was echter dat de hoog intelligente proefpersonen ook in de twee weken ná het experiment nog een significante progressie boekten, terwijl de laag intelligente proefpersonen terugvielen naar hun oorspronkelijke niveau. De hardop-denk protocollen werden geanalyseerd op verschillende aspecten van het aanpakgedrag, zoals oriëntatie-activiteiten en de handelings-systematiek. Hoewel hun povere probleemaanpak alle proefpersonen kenmerkte als typische beginners, bleken de hoog intelligente proefpersonen een significant betere aanpak te vertonen dan de laag intelligente proefpersonen. De aanpakcores correleerden significant met de diverse leermaten, maar wanneer intelligentie werd uitgepartieerd, benaderden deze correlaties evenwel de nul. Omdat de spreiding in aanpakcores binnen de beide intelligentie-groepen tamelijk gering was, kon niet worden aangetoond dat een goede probleemaanpak een surplus-waarde heeft voor het leerproces.

Onderhavig onderzoek is een vervolg op het onderzoek van Elshout & Veenman. De relatie tussen intelligentie, beginnersgedrag en leren is opnieuw doel van studie, maar nu in het nabij gelegen domein van de electriciteitsleer. Een tweede aspect van dit onderzoek betreft de opzet van de gesimuleerde leeromgeving. In het onderzoek van Elshout & Veenman (1990) werd de helft van de proefpersonen toegewezen aan een ongestructureerde leerconditie, waarin zij warmteleer-principes dienden te ontdekken door het zelfstandig bedenken en uitvoeren van experimenten in het gesimuleerde laboratorium. De andere helft werd toegewezen aan een gestructureerde leerconditie. In deze conditie werden de proefpersonen aangezet tot het formuleren van hypothesen, die zij middels het uitvoeren van vooraf bepaalde, 'leerzame' experimenten in het laboratorium konden toetsen. Dit Socratisch experimenteren, een leren-door-doen variant op de Socratische dialoog (Collins & Stevens, 1983), zou misconcepties en misvattingen moeten uitlokken en corrigeren. De resultaten van Elshout & Veenman vertoonden echter geen leereffect ten gevolge van verschillen in gestructureerdheid van de leeromgeving. Als verklaring voor het uitblijven van zo'n leereffect opperden zij dat het uitsluitend structureren van de experimentatie met name de zwakkere studenten onvoldoende steun biedt. Deze studenten hebben moeite met de abstractiegraad van de leerstof en zouden wellicht meer gebaat zijn bij een gestructureerde representatie van de leerstof (Veenman, Balk & Bierman, 1989a).

White & Frederiksen (1987, 1990) hebben een theoretisch model ontwikkeld voor de opbouw van de leerstof in het domein van de electriciteitsleer. Zij stellen daarbij de werkwijze van experts, die alvorens mathematische operaties uit te voeren een probleem eerst uitgebreid analyseren in kwalitatieve termen (Chi et.al., 1981), als voorbeeld. Het traditionele natuurkunde-onderwijs besteedt volgens White & Frederiksen nauwelijks aandacht aan deze kwalitatieve, conceptuele analyse en presenteert electriciteitsprincipes vrijwel direct als mathematische relaties en wetmatigheden. Zo'n kwantitatieve benadering kan problemen opleveren. Een student die net de wet van Ohm (bijvoorbeeld in de vorm van $V = I * R$) heeft geleerd, kan bij gebrek aan kennis van de causale relaties tussen spanningsverschil, stroomsterkte en weerstand concluderen dat wanneer de weerstand (R) twee keer zo groot wordt, het spanningsverschil (V) dus ook twee keer zo groot wordt. De notie dat zowel spanningsverschil als weerstand de onafhankelijke variabelen zijn die de afhankelijke variabele stroomsterkte bepalen, is dan afwezig. Het leren van een algebraïsche of symbolische syntaxis blijkt ook in andere contexten, zoals wiskunde of computer-programmeren (Perkins & Simmons, 1988), geen garantie voor het uitbannen of voorkomen van misvattingen op grond waarvan tot een verkeerde conclusie kan worden besloten. Om dergelijke misvattingen te voorkomen, opteren White & Frederiksen voor een simula-

tie-leerprogramma dat de student laat beginnen met een elementaire vorm van kwalitatief redeneren over circuitgedrag. Dit zogenaamde nulde-orde niveau redeneren wordt uitgelokt door vragen over fundamentele aspecten van circuitgedrag als "Zal het lampje wel of niet branden?" en "Staat er een spanningsverschil over het lampje?". Hiermee worden relevante concepten op concrete wijze bij de student geïntroduceerd. Vervolgens kunnen vragen van meer complexe, maar nog steeds kwalitatieve aard aan de student worden voorgelegd. Op dat 'eerste-orde niveau' wordt in een vraag bijvoorbeeld een variabele gevarieerd: "Wordt de stroomsterkte groter of kleiner als men de weerstand in een circuit vermindert?". De student leert dan causale verbanden te leggen tussen de verschillende concepten. Pas wanneer het kwalitatief redeneren over circuitgedrag vlekkeloos wordt beheerst, kan men de student kwantitatieve vragen voorleggen als: "In welke mate gaat het lampje feller of zwakker branden wanneer de bronspanning wordt verdubbeld?". Deze kwantitatieve analyse van een circuit zou dan een logisch vervolg vormen op de kwalitatieve analyse-methoden, die de student reeds meester is. Een dergelijke opbouw van de leerstof van kwalitatief naar kwantitatief redeneren zou met name de zwakke, minder intelligente student tijdens het Socratisch experimenteren in een gesimuleerde leeromgeving de noodzakelijke conceptuele ondersteuning bieden (Elshout & Veenman, 1990).

Onderhavig onderzoek is derhalve enerzijds bedoeld om de waarde te bepalen van de structureringsmethoden die de student zelf inbrengt in het leerproces (de probleemaanpak) als predictor voor leren, en anderzijds om het leereffect te bepalen van de structuur die de leeromgeving aanbrengt in het leerproces (Socratisch experimenteren en de opbouw van de leerstof). Bij leereffecten door de leeromgeving wordt een Aptitude-Treatment-Interactie tussen intelligentie en gestructureerdheid van de leeromgeving verwacht.

METHODE

Proefpersonen

De proefpersonen waren eerstejaars psychologiestudenten die, analoog aan Elshout & Veenman (1990), werden geselecteerd op een hoge of lage intelligentie en een minimum aan natuurkunde-vooropleiding. Daartoe werd bij alle eerstejaars studenten een testbatterij afgenomen van zes intelligentie-tests, die elk een vaardigheid uit het Structure-of-Intellect model van Guilford (1967) vertegenwoordigde. Geselecteerd werden studenten van wie de samengestelde score op deze intelligentie-tests ongeveer een standaarddeviatie afweek van het gemiddelde ($M=17.75$, $sd=4.33$, $N=376$). Studenten met een extreem lage intelligentie-score (meer dan 1.5 sd onder het gemiddelde) werden overigens niet in het onderzoek opgenomen. Voorts werden uitsluitend studenten geselecteerd, die geen natuurkunde-eindexamen hadden afgelegd en hooguit 3 jaar natuurkunde-onderwijs hadden genoten. Omdat de probleemaanpak van proefpersonen werd gescoord op grond van hardop-denken-protocollen, een tijdrovende en arbeidsintensieve methode, konden relatief weinig proefpersonen in het onderzoek worden opgenomen. De 8 hoog intelligente en 11 laag intelligente proefpersonen werden per intelligentie-niveau random toegewezen aan de experimentele condities. Van de 11 laag intelligente proefpersonen werden er 6 aan de ongestructureerde conditie en 5 aan de gestructureerde conditie toegewezen, terwijl de beide experimentele condities ieder 4 hoog intelligente proefpersonen bevatte. De vier experimentele groepen vertoonden geen duidelijke verschillen in sexe, leeftijd of in het aantal jaren natuurkunde-onderwijs.

Leercondities

Voor het creëren van de beide leeromgevingen werd een simulatie van een elektrisch circuit geïmplementeerd in Course-of-Action, een geavanceerde auteurstaal voor Macintosh-computers (Veenman et al., 1989a). In het circuit van dit 'electriciteits-laboratorium' konden de proefpersonen een lampje, een schakelaar of verschillende weerstanden plaatsen en met behulp van diverse meters de spanning of stroomsterkte op verschillende punten meten. Dit laboratorium kende

twee varianten. De eerste variant, die werd aangeboden in het eerste deel van de cursus, werd gevormd door een enkelvoudige schakeling met één lampje of één weerstand, terwijl de variant in het tweede cursusdeel een seriële schakeling betrof met twee lampjes of weerstanden naast elkaar.

In de ongestructureerde leerconditie ontvingen de proefpersonen, na een korte inleiding in de noodzakelijke electriciteitsbegrippen, de opdracht om in het laboratorium met een enkelvoudige schakeling uit te zoeken "hoe spanningsverschil, stroomsterkte en weerstand met elkaar samenhangen". Zij kregen daarbij geen feedback op de experimenten die zij in het laboratorium uitvoerden en op de conclusies die zij daaruit trokken. Dit eerste deel werd bij het verlaten van het laboratorium afgesloten met een lijstje van relevante formules. In het tweede deel van de ongestructureerde leerconditie kregen de proefpersonen dezelfde opdracht voor een seriële schakeling. Ook aan het eind van dit deel werd hen een lijstje met relevante formules gepresenteerd. Feitelijk werd de proefpersonen in de ongestructureerde conditie gevraagd om zelfstandig enkele relevante electriciteitsprincipes (de wet van Ohm, de wet van Kirchhoff en het concept electrisch vermogen) te ontdekken.

De gestructureerde leerconditie werd gevormd door een gestructureerde opbouw van de leerstof in combinatie met Socratisch experimenteren. Zowel het eerste deel met de enkelvoudige schakeling als het tweede deel met de seriële schakeling was opgebouwd volgens de hierboven beschreven hiërarchie van kwalitatief redeneren naar kwantitatief redeneren (White & Frederiksen, 1987, 1990). De concepten werden op het nulde-orde niveau één voor één geïntroduceerd, op het eerste-orde niveau met elkaar in relatie gebracht en tenslotte werden de wetmatigheden in een mathematisch kader geplaatst. Daarbij werd de proefpersoon telkens de gelegenheid geboden van Socratisch experimenteren. Voorafgaand aan elk experiment diende de proefpersoon als antwoord op een vraag een hypothese te genereren, die vervolgens kon worden getoetst in het electriciteitslaboratorium. Het laboratorium was daartoe ingericht met uitsluitend het noodzakelijke instrumentarium voor het betreffende experiment. Na elk experiment diende de proefpersoon de gestelde hypothese te evalueren, waarbij een incorrecte evaluatie opnieuw gelegenheid bood tot experimenteren. Tot slot volgde telkens een uitleg waarin de leerstof nog eens op een rijtje werd gezet.

De proefpersonen werden gevraagd om tijdens het doorlopen van het leerprogramma hardop te denken. De proefleider was geïnstrueerd om uitsluitend in te grijpen bij calamiteiten of bij lange stiltes om het hardop-denken te stimuleren. Time-on-task werd niet constant gehouden, maar werd gemeten voor postexperimentele controle. De hardop-denken-protocollen werden op de band opgenomen en eventuele notities van de proefpersonen werden bewaard om in samenhang met de protocollen te worden beoordeeld. Deze protocollen werden nadien door drie 'blinde' beoordelaars geanalyseerd door al argumenterend tot een gezamenlijk oordeel te komen. Deze beoordelingsmethode ontbeert weliswaar de bepaling van een interbeoordelaarsbetrouwbaarheid, maar heeft eerder (zie Elshout & Veenman, 1990) tot bevredigende resultaten geleid. De protocol-analyses waren gericht op beoordeling van de mate van voorkennis, zich voordoende misvattingen, diverse aanpakmaten (oriëntatie-activiteiten, systematische ordelijkheid, nauwkeurigheid, elaboratie- en evaluatie-activiteiten) en het bereikte leerniveau.

Misconceptions

In een pretest-posttest design werd kort voor en direct na het experiment een misconceptievragenlijst afgenomen, bestaande uit 22 items die random op de computer werden aangeboden. Elk item bevatte een stelling, waarin bekende misconcepties uit de literatuur (Cohen, Eylon & Ganiel, 1983; Shipstone, 1985; Brna, 1988) waren verwerkt. Deze misconcepties zijn het 'stroomverbruik-model' (stroom wordt door een lampje 'opgegeten'), het kano-model (sequentieel redeneren door in gedachten met de stroom mee te 'varen'), het 'clashing currents' model (electriciteitsdeeltjes ontmoeten elkaar in een lampje en uit die botsing ontstaat licht), het niet differentiëren tussen spanningsverschil en stroomsterkte, en de misvatting dat een grotere weerstand een groter vermogen heeft (meer weerstand betekent harder 'duwen' en dus meer energieverbruik). De proefpersoon diende elke stelling te beoordelen op correctheid, waarbij elk juist

antwoord een punt toevoegde aan de score. Dezelfde vragenlijst werd twee weken later als retentie-toets afgenomen.

Opgaven

Aan het eind van de beide cursusdelen werd de proefpersoon een aantal kwantitatieve opgaven per computer aangeboden. De opgaven na het eerste deel waren 10 eenvoudige sommen waarin de wet van Ohm en de berekeningsformule voor elektrisch vermogen moest worden toegepast in een enkelvoudig elektrisch circuit. Alle proefpersonen, dus ook die in de ongestructureerde leerconditie, kregen na elk antwoord feedback over de juiste oplossingsmethode. De 6 opgaven na het tweede cursusdeel waren complexer van aard dan die van het eerste deel. Ook bij deze opgaven dienden de wet van Ohm en de vermogen-formule te worden toegepast, maar nu in een circuit met een seriële schakeling. Het oplossen van deze sommen deed een groter beroep op conceptuele kennis en vereiste meer denkstappen, bijvoorbeeld omdat men de bronspanning moest onderscheiden van de spanning over één van de weerstanden.

De opgaven werden volgens twee methoden gescoord. De eerste methode betrof een objectieve opgave-scoring van het aantal juist opgeloste sommen. De tweede scoring geschiedde aan de hand van de hardop-denken protocollen, waarbij de opgaven-protocollen los van en voorafgaand aan de cursus-protocollen werden geanalyseerd. In deze protocolscoring van de opgaven op een vierpuntsschaal werd vooral gelet op de correctheid van de oplossingsmethode. Een rekenfoutje werd de proefpersoon dan minder zwaar aangerekend.

Statistische analyses

Over alle maten uit het onderzoek werden variantie-analyses uitgevoerd met intelligentie en leerconditie als onafhankelijke factoren. Bij de misconceptie-scores werd verder een variantie-analyse met herhaalde meting uitgevoerd. Tevens werden alle maten in een correlatieve analyse opgenomen.

RESULTATEN

Misconcepties

De betrouwbaarheid van de misconceptie-vragenlijst was laag (voor pretest, posttest en retentietest respectievelijk .42, .42 en .53). Sommige items met een hoge itemrestcorrelatie op de pretest bleken op de posttest een lage, soms zelfs negatieve itemrestcorrelatie te scoren en vice versa. Items, die dezelfde misconceptie zouden meten, bleken onderling wisselend positief en negatief te correleren. Kortom, het optreden van misconcepties was niet consistent.

De variantie-analyses op de herhaaldelijk afgenomen vragenlijst lieten uitsluitend bij de pretest een significant intelligentie-effect zien van $F(1,15)=5.47$ ($p<.04$). Hoog intelligente proefpersonen gaven gemiddeld meer correcte antwoorden op de pretest dan laag intelligente proefpersonen (zie Tabel 1). Analyses van de posttest en retentietest vertoonden echter geen significante effecten (zie Tabel 4). Overigens viel één laag intelligente proefpersoon van de ongestructureerde conditie uit op de retentietest. De variantie-analyse voor herhaalde metingen gaf een significant effect op de herhaalde meting met $F(2,28)=3.91$ ($p<.04$), maar de gemiddelden per toets (voor pretest, posttest en retentietest respectievelijk 11.00, 12.06 en 12.33; met standaarddeviaties variërend van 2.0 tot 2.3) tonen dat dit leereffect niet groot was.

In de hardop-denken protocollen van de gestructureerde leerconditie bleek slechts één misvatting sterk vertegenwoordigd. Deze misvatting betrof de afname van stroomsterkte na een lampje of weerstand, maar werd in slechts een enkel geval expliciet verklaard met het stroomverbruikmodel. In mindere mate trad de conceptueel diepe misconceptie voor het voetlicht, dat een grotere weerstand méér energie verbruikt. Uitsluitend het vermogen van een oneindig grote weerstand, zoals een open schakelaar waardoor er geen stroom meer kan lopen, confronteerde proefpersonen met hun misconceptie. De resterende misvattingen bleken evenwel van oppervlakkige aard te zijn en werden slechts incidenteelesignaleerd. Overigens kon men in de protocollen van de ongestructureerde leerconditie nauwelijks misvattingen bespeuren.

Tabel 1. Celgemiddelden (en standaarddeviaties) van Pretest, Posttest en Retentie-test.

	Pretest		Posttest		Ret.test	
	LIQ	HIQ	LIQ	HIQ	LIQ	HIQ
Ongestruct.	10.83 (1.84)	12.25 (2.87)	11.33 (3.01)	11.75 (1.71)	10.80 (2.78)	12.75 (2.50)
Gestruct.	9.2 (2.17)	12.25 (0.96)	11.80 (1.30)	13.25 (0.96)	12.60 (1.67)	13.50 (1.00)
Te zamen.	10.09 (2.07)	12.25 (1.98)	11.55 (2.30)	12.50 (1.51)	11.70 (2.36)	13.13 (1.81)

Opgaven

Hoewel de beide scoringsmethoden ongeveer dezelfde resultaten opleverden, bleek de betrouwbaarheid van de protocol-methode een fractie hoger dan die van de objectieve scoringsmethode. De respectievelijke alpha's waren voor het eerste deel .66 en .63, en voor het tweede deel .82 en .80. Derhalve werd gekozen voor de protocolgescoorde opgaven.

Een variantie-analyse over de opgaven van deel 1 toonde een significant intelligentie-effect met $F(1,15)=11.41$ ($p<.005$), waarbij de hoog intelligente proefpersonen gemiddeld hoger scoorden dan de laag intelligente (zie Tabel 2). Een zelfde intelligentie-effect met $F(1,15)=10.78$ ($p=.005$) werd gevonden in de opgaven van deel 2. Ook hier scoorden de hoog intelligente proefpersonen gemiddeld hoger dan de laag intelligente (zie Tabel 2). Intelligentie correleerde significant met zowel Opgaven deel 1 ($r=.64$) als met Opgaven deel 2 ($r=.65$).

In deel 1 noch in deel 2 van de opgaven konden leerconditie-effecten of interactie-effecten worden bespeurd (zie Tabel 4). De gemiddelde scores laten zelfs geen trend in de voorspelde richting zien (zie Tabel 2).

Voorkennis en Leerniveau

De mate van schoolse voorkennis waarvan de proefpersonen in de protocollen blijk gaven, was van een zeer laag niveau en ongeveer gelijk voor alle proefpersonen. Voorkennis werd beoordeeld op een schaal van 1 tot 10 en de gemiddelde score was 2.26. Bijna alle proefpersonen scoorden een 2, slechts drie van hen scoorden een 3 of een 4.

Het bereikte leerniveau werd gebaseerd op de mate waarin de proefpersonen blijk gaven van cumulatief leren tijdens het experiment. Bij het beoordelen van de protocollen werd erop gelet

Tabel 2. Celgemiddelden (en standaarddeviaties) van opgaven deel 1 en deel 2.

	Deel 1.		Deel 2.	
	LIQ	HIQ	LIQ	HIQ
Ongestruct.	18.83 (5.95)	25.25 (3.40)	6.83 (3.06)	10.25 (6.85)
Gestruct.	15.20 (3.63)	22.50 (2.65)	5.20 (4.82)	15.25 (1.89)
Te zamen	17.18 (5.15)	23.88 (3.18)	6.09 (3.83)	12.75 (5.37)

of de proefpersonen de relevante concepten, grootheden en eenheden beheersten, of Kirchoff's wet werd toegepast, of de wet van Ohm correct werd gehanteerd en of het begrip electrisch vermogen werd begrepen. In het bijzonder werd nagegaan of het geleerde tijdens eerdere blokken in het programma correct werd toegepast in de daarop volgende blokken. Deze stapeling van kennis is immers essentieel voor de leerhiërarchie van White & Frederiksen (1987, 1990). Leerniveau werd eveneens op een tienpuntsschaal beoordeeld alsof het een proefwerkcijfer betrof. Een variantie-analyse over de leerniveau-scores vertoonde wederom een significant intelligentie-effect met $F(1,15)=5.81$ ($p<.03$). De hoog intelligente proefpersonen scoorden gemiddeld een 6, terwijl de laag intelligente proefpersonen een onvoldoende van 4.1 incasseerden (zie Tabel 3). Intelligentie correleerde .59 met het protocolgescoorde leerniveau. Zowel het leerconditie- als het interactie-effect was niet significant (zie Tabel 4). Hoewel er sprake was van een duidelijke vooruitgang ten opzicht van het voorkennisniveau, moet toch worden geconcludeerd dat het bereikte leerniveau niet hoog was.

Aanpak

De aanpakmaten werden aan de hand van de protocollen elk apart gescoord op een vierpuntsschaal. Oriëntatie-activiteit werd beoordeeld op de mate waarin de proefpersoon vaststelde wat gegeven was en wat werd gevraagd, een mentaal model bouwde van de vraag, voorspellingen genereerde en eventueel gedachten-experimenten uitvoerde. De beoordeling van systematische ordelijkheid was gebaseerd op de mate van planning en de systematische uitvoering van een plan. Nauwkeurigheid had betrekking op de ordelijkheid van aantekeningen, het consequent en correct benoemen van grootheden en eenheden, en de afwezigheid van slordigheden (bijvoorbeeld bij het aflezen van de meters, het afronden van getallen etc.). Elaboratie werd beoordeeld op de mate waarin de proefpersoon zelf verklaringen genereerde, op de leerstof reflecteerde door het geven van omschrijvingen in eigen woorden en de mate waarin verbanden werden gelegd tussen

Tabel 3. Celgemiddelden (en standaarddeviaties) van Leerniveau.

	LIQ	HIQ
Ongestruct.	3.75 (1.89)	5.25 (2.22)
Gestruct.	4.50 (1.23)	6.75 (1.04)
Te zamen	4.09 (1.59)	6.00 (1.79)

Tabel 4. F-waarden van de variantie-analyses op de leermaten. Voor alle F-waarden geldt $df=(1,15)$, behalve voor de retentie-test waar geldt $df=(1,14)$.

Leermaat	IQ	Leerconditie	Interactie
Posttest	0.95	1.05	0.29
Retentie-test	1.98	1.59	0.27
Opgaven deel 1	11.41 ($p<.005$)	2.47	0.05
Opgaven deel 2	10.78 ($p=.005$)	0.67	2.62
Leerniveau	5.81 ($p<.03$)	2.09	0.23

Tabel 5. Correlaties van de leermaten met aanpak, intelligentie en met aanpak waarbij intelligentie is uitgepartieerd.

Leermaten	Aanpak	Intelligentie	Aanpak (part.cor.)
Posttest	.42*	.25	.36
Retentie-test	.17	.30	.10
Opgaven deel 1	.42*	.64**	.26
Opgaven deel 2	.68**	.65**	.63**
Leerniveau	.82**	.59**	.80**

* significantie-niveau $p < .05$; ** significantie-niveau $p < .01$

verschillende leeronderdelen. Tot slot werden evaluatie-activiteiten gescoord op de mate van 'monitoring' tijdens het uitvoeren van experimenten, het controleren van uitkomsten en het valideren van conclusies.

Hoewel de proefpersonen duidelijk beginnersgedrag vertoonden - de oriëntatie-activiteiten hadden bijvoorbeeld allerm minst het karakter van een diepe kwalitatieve analyse (Chi et.al., 1981) - waren de individuele verschillen groot. Deze verschillen waren echter nauwelijks intelligentie-gerelateerd. Variantie-analyses over de afzonderlijke aanpakmaten vertoonden geen significante intelligentie-effecten (hooguit een trend bij elaboratie van $F(1,15)=3.41$, $p < .09$) en de correlaties met intelligentie waren overwegend laag (voor oriëntatie .30, systematiek .27, nauwkeurigheid -.08, elaboratie .54 en voor evaluatie .30). De meeste correlaties van de aanpakmaten, uitgezonderd nauwkeurigheid, met de diverse leermaten waren echter substantieel (variërend van .28 tot .78).

De samengestelde aanpakmaat, een sommatie van de afzonderlijke vijf maten, liet hetzelfde beeld zien. De variantie-analyse vertoonde geen significant intelligentie-effect ($F(1,15)=1.71$, $p < .22$) en de correlatie met intelligentie bedroeg .37. Interessant zijn de correlaties met de leermaten en vooral die correlaties waarbij intelligentie is uitgepartieerd (zie Tabel 5). Juist bij de meer complexe leermaten (Opgaven deel 2 & Leerniveau) bleek aanpak een surplus-waarde bovenop intelligentie te hebben voor de mate van leren.

Time-on-task

De verschillen in time-on-task tussen de beide leercondities en tussen de beide intelligentie-condities waren aanzienlijk. De proefpersonen in de gestructureerde condities brachten gemiddeld 208 minuten in het laboratorium door, terwijl de proefpersonen in de ongestruceerde condities het leerprogramma in gemiddeld slechts 118 minuten hadden doorlopen ($F(1,15)=66.79$, $p=.0001$). Een dergelijk verschil tussen de beide leercondities is het gevolg van de uitgebreide serie opdrachten die proefpersonen in de gestructureerde conditie moesten uitvoeren. Laag intelligente proefpersonen hadden verder significant meer tijd nodig dan hoog intelligente proefpersonen (gemiddeld 175 versus 141 minuten, $F(1,15)=10.56$, $p < .006$). Het interactie-effect was overigens niet significant ($F(1,15)=1.39$, $p=.26$). Time-on-task bleek echter nauwelijks positief gecorreleerd aan de leermaten (correlaties varieerden van zelfs -.50 tot maximaal .30), een indicatie dat time-on-task geen belangrijke rol speelt als storende variabele.

DISCUSSIE

De lage betrouwbaarheden van de misconceptie-vragenlijsten alsmede het grillige patroon van de itemcorrelaties vormen een indicatie voor de instabiliteit van misconcepties bij beginners in het domein van de electriciteit. Bovendien laten de kwalitatieve analyses van de hardop-denken-protocollen zien dat de meeste misconcepties slechts incidenteel optreden, terwijl verklaringen

in termen van foutieve, diepe modellen veelal ontbreken. Een mogelijke verklaring is dat beginnende studenten bij gebrek aan domeinkennis (zie het voorkennis-niveau), niet over dergelijke foutieve, diepe modellen beschikken zoals door diverse auteurs in Gentner & Stevens (1983) wordt verondersteld. Pas wanneer een ignorante proefpersoon door de vraagstelling gedwongen wordt om ad hoc een keuze te maken, zoals bij de misconceptie-vragenlijst of in de gestructureerde leerconditie, ontstaan er oppervlakkige 'vergissingen' die bij een volgende gelegenheid gemakkelijk kunnen worden vervangen door andere 'vergissingen'. Vermoedelijk komen de vergissingen van beginners ten dele voort uit onduidelijkheden in het alledaags taalgebruik, zoals het geval is bij stroom en energie ('stroomverbruik') en elders bijvoorbeeld bij warmte en temperatuur (Elshout & Veenman, 1990).

Geenszins verrassend is dat intelligentie een belangrijke predictor blijkt te zijn voor de mate van leren op de kwantitatieve opgaven en het bereikte leerniveau, waarbij moet worden opgemerkt dat de intelligentie-effecten enigszins worden geflatteerd door gebruikmaking van extreme groepen. Opvallend daarentegen is wél dat aanpakgedrag (uitgezonderd elaboratie) relatief laag correleerde met intelligentie, terwijl aanpak anderzijds positief correleerde met diverse leer-maten. Bovendien bleek aanpak een meerwaarde te hebben bovenop intelligentie als predictor voor de mate van complex leren. De opvatting dat hoog intelligente studenten zouden beschikken over een betere probleemaanpak is dus niet zonder meer houdbaar. In het onderzoek van Elshout & Veenman (1990) werd deze relatie tussen intelligentie en probleemaanpak wél geconstateerd; dit onderzoek wijst in een andere richting. Voorlopig is deze strijd nog onbeslist. Dit onderzoek toont evenwel aan dat een goede probleemaanpak een zelfstandige bijdrage aan het leerproces kan leveren, met name bij complex leren.

Leereffecten ten gevolge van verschillen in leeromgeving konden wederom niet worden bespeurd. Ondanks de combinatie van Socratisch experimenteren met een speciale opbouw van de leerstof én ondanks de enorme verschillen in time-on-task, bleken proefpersonen in de gestructureerde conditie niet beter te presteren dan de proefpersonen in de ongestructureerde conditie. De verwachting van White & Frederiksen (1987, 1990) dat een training in kwalitatief redeneren het presteren op kwantitatief niveau zou bevorderen, werd dus niet bewaarheid. Evenmin is de verwachting uitgekomen dat laag intelligente proefpersonen meer baat zouden hebben bij deze in hoge mate gestructureerde leeromgeving (Elshout & Veenman, 1990). Het geringe aantal proefpersonen kan het verkrijgen van significante interacties evenwel hebben bemoeilijkt. Toch vertonen de opgaven-celgemiddelden van laag intelligente proefpersonen geen trend in de verwachte richting van een Aptitude-Treatment-Interactie en gaan de resultaten op bereikt leerniveau evenmin in de voorspelde richting. Een voorzichtige conclusie luidt dan ook dat de kwaliteiten die de student zélf in het leerproces inbrengt (intelligentie en probleem-aanpak), het leerproces méér ten goede komen dan externe structurering van de leertaak in een computergesimuleerde leeromgeving.

Door 'uitval' van de misconceptie-vragenlijsten valt het echter niet helemaal uit te sluiten dat (laag intelligente) proefpersonen in de gestructureerde leerconditie wél beter presteerden op kwalitatief niveau. De resultaten van bereikt leerniveau wijzen weliswaar niet in deze richting, maar een objectieve leermaat zou steviger in de schoenen staan. Een vervolgonderzoek betreft daarom het repliceren van dit experiment met een objectieve, kwalitatieve leermaat. Daartoe is op dit moment een electriciteitsleervragenlijst ontwikkeld met voldoende interne validiteit ($\alpha=.78$, $n=440$). Een andere onderzoekslijn die eveneens wordt gevolgd, behelst het verrichten van soortgelijk onderzoek in andere domeinen dan de natuurkunde. Een dergelijk experiment in het domein van de statistiek staat nu in de steigers.

LITERATUUR

- Anderson, J. R. (1985). *Cognitive psychology and its implications*. New York: Freeman.
- Brna, P. (1988). Confronting misconceptions in the domain of simple electrical circuits. *Instructional Science*, 17, 29-55.

- Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In: Sternberg, R.J. (Ed.) *Advances of the psychology of human intelligence*. (Vol. 1). Hillsdale: Erlbaum.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P. & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Cohen, R., Eylon, B. & Ganiel, U. (1983). Potential difference and current in simple electrical circuits: A study of students' concepts. *American Journal of Physics*, 51(5).
- Collins, A. & Stevens, A. L. (1983). A cognitive theory of inquiry teaching. In: C. M. Reigeluth (Ed.) *Instructional design theories and models: an overview of their current status*. Hillsdale: Erlbaum.
- Elshout, J. J. (1985). Een goed begin is het halve werk: over determinanten van effectief beginnersgedrag. In: Mönks, F.J. & Span, P. (Red.) *Hoogbegaafden in de samenleving*. Nijmegen: Dekker & v.d. Vegt.
- Elshout, J. J. (1987a). Probleemoplossen als context voor leren probleemoplossen. *Nederlands tijdschrift voor de psychologie*, 42, 344-353.
- Elshout, J. J. (1987b). Problem solving and education. In: E. de Corte, H. Lodewijks, R. Parmentier & P. Span (Eds.) *Learning and instruction*. Oxford: Pergamon Books Ltd/Leuven: University Press.
- Elshout, J. J. & Veenman, M. V. J. (1990). Wat doet intelligentie voor studenten? In: Ippel, M. J. & Elshout, J. J. (Red.) *Training van hogere orde denkvaardigheden*. Lisse: Swets & Zeitlinger.
- Gentner, D. & Stevens, A. L. (1983). *Mental models*. Hillsdale: Erlbaum.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Perkins, D. N. & Simmons, R. (1988). Patterns of misunderstanding: an integrative model for science, math, and programming. *Review of Educational Research*, 58 (3), 303-326.
- Shipstone, D. M. (1985). On children's use of conceptual models in reasoning about current electricity. In: Duit, R., Jung, W. & Rhöneck, C. von (Eds.) *Aspects of understanding electricity*. Kiel: IPN.
- Veenman, M. V. J. & Balk, V. A. & Bierman, D. J. (1989a). Dr. Stat, computer ondersteund onderwijs in de statistiek. *Tijdschrift voor het economisch onderwijs*, 3, 65-70.
- Veenman, M. V. J. & Elshout, J. J. & Bierman, D. J. (1989b). Structured and unstructured approaches to simulation: What do the students think? *Proceedings of the Sixth Canadian Symposium on Instructional Technology*. Halifax, Canada, 457-461.
- White, B. Y. & Frederiksen, J. R. (1987). Qualitative models and intelligent learning environments. In: R. W. Lawler & M. Yazdani (Eds.) *Artificial intelligence and education, Volume one: Learning environments and tutoring systems*. Norwood: Ablex Publishing Corporation.
- White, B. Y. & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 42, 99-157.

Manuscript ontvangen 1-5-1990

Definitieve versie ontvangen 18-7-1990

Het gebruik van kleurendruk bij zoektaken in schriftelijk studiemateriaal*

Bernadette H.A.M. van Hout Wolters en Willeke M.J. Kerstjens
*Universiteit van Amsterdam** en Open Universiteit Heerlen*

ABSTRACT

Research relating to the design of written instructional materials gives relatively little attention to the use of colour. Yet text books are frequently printed in colour. In the study reported here the systematic use of colour in thematically distinct parts of an instructional text facilitated the process of searching that text for specific information. Furthermore, students valued the use of colour in instructional texts positively. Subjects were freshmen at a nurses' training college.

INLEIDING

Er zijn weinig onderzoeksgegevens bekend op het gebied van kleurgebruik in schriftelijk studiemateriaal, hoewel daarin toch regelmatig met kleurendruk wordt gewerkt. Omdat kleurendruk duur is, wordt hierdoor de prijs van het studiemateriaal aanzienlijk verhoogd. Voor het ontwerpen van schriftelijk studiemateriaal zou kennis van de effecten van kleurgebruik heel wenselijk zijn. In dit artikel richten we ons niet op het gebruik van kleur in illustraties, maar op kleurgebruik in de tekst zelf.

In eerder onderzoek van de auteurs (Van Hout Wolters, Kerstjens & Verhagen, 1990b) werd nagegaan wat het effect is van kleurgebruik in een studietekst met als doel de structuur van de tekst te verhelderen. De tekst die in dit onderzoek werd gebruikt, had reeds een overzichtelijke indeling en lay-out. Door thematisch onderscheiden tekstgedeelten een verschillende achtergrondkleur te geven, zou de studietekst nog aan overzichtelijkheid kunnen winnen, waardoor het onthouden van de tekstinhoud verbeterd zou kunnen worden. Ter vergelijking kreeg de helft van de proefpersonen (HBO-Verpleegkunde) een zwart-witte tekst met dezelfde indeling en lay-out. Er bleek geen significant verschil te zijn in de herinnering van tekstgedeelten tussen de groep met de zwart-witte tekst en de groep met de kleuren-tekst. Wel werd het kleurgebruik positief gewaardeerd door de studenten.

Het hier beschreven onderzoek sluit bij dit onderzoek aan. Eén van de vragen die naar aanleiding van dit onderzoek rees, was, of kleur als structureringsmiddel wellicht wél het opzoeken van informatie in de studieteksten zou kunnen vergemakkelijken. Het zoekend lezen vergt namelijk een andere aanpak dan het begrijpend en studerend lezen: het is meer doelgericht en selectief, waarbij men vaker grote delen kan overslaan (Guthrie & Mosenthal, 1987). Als het zoekend lezen door kleurgebruik verbeterd zou kunnen worden, zou dit voor de onderwijspraktijk van belang zijn, omdat studenten in studieboeken, (computer)handleidingen en naslagwerken vaak informatie moeten opzoeken.

Over begrijpend en studerend lezen is aanzienlijk meer onderzoek verricht dan over *zoekend lezen*, ook al wordt het zoekend lezen al tientallen jaren als aparte leesstrategie onderscheiden (zie bijv. Van Hout Wolters, 1971). Pas sinds het midden van de jaren tachtig krijgt het zoekend lezen wat meer aandacht van onderzoekers. Dit onderzoek richt zich echter meer op het zoeken

* Het beschreven onderzoek werd uitgevoerd toen beide auteurs werkzaam waren aan de Universiteit Twente, Faculteit Toegepaste Onderwijskunde.

** Instituut voor de Lerarenopleiding, Universiteit van Amsterdam, Herengracht 256, 1016 BV Amsterdam.

van informatie in 'documents' (tabellen, grafische presentaties, landkaarten, dienstregelingen e.d.), dan op het zoeken van informatie in lopende tekst.

Guthrie, een belangrijke vertegenwoordiger van dit onderzoek, beschreef in 1988 een 'cognitive model of search' (Guthrie, 1988). Dit model gaat ervan uit dat als men specifieke informatie opzoekt in 'documents', men de volgende cognitieve processen doorloopt: *doelformulering*, *categorie-selectie* (selectie van een doel-relevant deel van het materiaal), *extractie van informatie* (localisatie van gegevens die nodig zijn om het doel te bereiken), *integratie* (combinatie van de gevonden informatie met eerder gevonden informatie of met het doel), en *recycling* (het nogeens doorlopen van de eerste vier componenten totdat aan de doelvereisten is voldaan). Uit onderzoek van Dreher en Guthrie (1988) en Guthrie en Britten (1989) blijkt dat deze vijf proces-componenten niet alleen onderscheiden kunnen worden bij het zoeken van informatie in 'documents', maar ook in lopende tekst.

Wat betreft *kleurgebruik* bij het opzoeken van informatie is weliswaar onderzoek bekend (o.a. Carter & Cahill, 1976; Christ & Corso, 1983; Luder & Barber, 1984), maar praktisch al dit onderzoek heeft betrekking op niet-tekstueel materiaal (met name numeriek en grafisch materiaal). Een voorbeeld van een dergelijk onderzoek is het onderzoek van Luder en Barber (1984) betreffende het vinden van informatie over brandstofvoorziening op gekleurde cockpit-beeldschermen. Uit veel van dit niet-tekstuele onderzoek komt kleur naar voren als een efficiënte code bij zoektaken: de zoektijden kunnen verkort worden door het gebruik van kleur.

Een verklaring voor de kortere zoektijden door het gebruik van kleur kan gezocht worden in Gestaltvorming (Carter & Cahill, 1976) of segmentatie (Davidoff, 1987). Alle items van dezelfde kleur worden waargenomen als een groep (= Gestalt), die perceptueel op de voorgrond komt te liggen. Het oog beweegt zich binnen deze groep en zoekt het gevraagde item. In plaats van alle losse items afzonderlijk te bekijken, zoals het geval zou zijn als er zwart-wit materiaal gebruikt was, hoeft men alleen te zoeken binnen één groep van één kleur. Hierdoor kan de zoektijd verkort worden. In termen van het juist beschreven model van Guthrie (1988) kan men zeggen, dat door kleurgebruik de categorieën in het materiaal verhelderd worden, waardoor de categorie-selectie en daarmee de extractie van informatie versneld wordt.

Het enige beschikbare onderzoek dat betrekking heeft op kleurgebruik bij zoektaken in *studieteksten* wordt beschreven door Wright en Lickorish (1988). De daar gebruikte studieteksten bestonden uit 2800 woorden en waren 22 pagina's lang. Ze waren onderverdeeld in vijf gedeelten, die werden afgebakend door 'natuurlijke breekpunten' in de tekst. De vijf gedeelten waren op een verschillende kleur papier gekopieerd. De controle-tekst was op wit papier gekopieerd. Er werden geen significant kortere zoektijden gevonden bij het zoeken naar specifieke informatie in de gekleurde teksten. Wel werd bij de gekleurde teksten de tweede helft van de 10 vragen significant sneller beantwoord dan de eerste helft: een mogelijk oefeneffect binnen de kleur-conditie. Een punt van kritiek op het onderzoek van Wright & Lickorish is dat de kleuren bij hen weliswaar gebruikt werden om vijf tekstgedeelten te onderscheiden, maar dat de kleuren geen betekenis hadden. Het is met name niet duidelijk wat onder 'natuurlijke breekpunten' in de tekst werd verstaan. Bovendien kwam één kleur twee keer voor. De structureringsfunctie van de kleur was dus onduidelijk.

Van belang bij kleurgebruik is het feit of er nadrukkelijk op de aanwezigheid en de betekenis van de kleur gewezen wordt. Uit onderzoek van Lyczak (1976) is gebleken dat kleurgebruik alleen in dat geval positief werkt. Uit onderzoek van Luder & Barber (1984) met betrekking tot zoektaken blijkt zelfs dat de kleurtoevoeging storend werkt voor proefpersonen die zich er niet van bewust zijn dat het kleurgebruik systematisch toegepast is.

Vaak zijn mensen meer gemotiveerd om bij zoektaken met gekleurd materiaal te werken dan met zwart-wit materiaal (o.a. Greenskin & Fleming, 1984; Narborough-Hall, 1985). Ook blijkt uit een overzichtsstudie van Lamberski (1980) dat bij verbale taken kleurcodes in het algemeen geprefereerd worden boven andere coderingssystemen.

In het hier beschreven onderzoek werd het effect nagegaan van het gebruik van kleurendruk op het zoeken van informatie in schriftelijk studiemateriaal. Het gebruikte onderzoeksmateriaal had reeds een overzichtelijke indeling en lay-out (zie Materiaal), omdat hiervan positieve effecten

bekend zijn (zie bijv. Hartley, 1986) en het geen extra drukkosten met zich meebrengt. Vanwege de extra kosten van kleurendruk, zou alleen een extra effect van kleurgebruik, toegevoegd aan de gebruikelijke effecten van een overzichtelijke indeling en lay-out, interessant zijn.

In tegenstelling tot bij Wright en Lickorish (1988) vervulde de kleur een duidelijke structureringsfunctie. Elke kleur had een eigen betekenis. Doordat een tekstgedeelte met een bepaalde achtergrondkleur gedrukt was, wisten de studenten, zonder naar de inhoud van het tekstgedeelte te kijken, wat het behandelde thema was. Ze werden ook expliciet op het kleurgebruik en de betekenis daarvan geattendeerd. Tevens werd nagegaan in hoeverre de studenten het kleurgebruik bij deze zoektaak waardeerden. Aansluitend bij het voorafgaande werden de volgende hypothesen geformuleerd:

- Door het systematisch kleuren van thematisch onderscheiden gedeelten in een studietekst wordt de inhoudelijke structuur zodanig verhelderd, dat het opzoeken van informatie in de gekleurde tekst sneller gaat dan in een zwart-witte tekst met dezelfde inhoud, indeling en lay-out.
- Het systematisch kleuren van thematisch onderscheiden tekstgedeelten wordt door studenten positief gewaardeerd.

METHODE

Proefpersonen

Als proefpersonen fungeerden 42 eerstejaars studenten van een HBO-opleiding Verpleegkunde. Ze deden vrijwillig mee aan het onderzoek. Als vergoeding kregen ze een boekenbon van f. 10,—.

De studenten werden willekeurig toegewezen aan de kleur- en zwart-wit conditie ('random groups design'). Om te controleren of deze at random samengestelde groepen aan elkaar gelijkwaardig waren, werden de volgende gegevens van de studenten verzameld: geslacht, vooropleiding, studieresultaten en voorkennis met betrekking tot het behandelde onderwerp in de studietekst.

Materiaal

De studietekst die in dit onderzoek werd gebruikt, was getiteld "Neurologische aandoeningen" en bestond uit 1100 woorden en 4 A4-pagina's. Deze tekst was ook gebruikt in voorafgaand onderzoek (Van Hout Wolters, Kerstjens & Verhagen, 1990b). In de tekst kwamen zeven neurologische aandoeningen aan de orde, waarbij steeds vier thema's/aspecten konden worden onderscheiden:

- definitie of omschrijving van de neurologische aandoening
- oorzaak van de neurologische aandoening
- symptomen van de neurologische aandoening
- behandeling van de neurologische aandoening.

In de experimentele conditie kregen deze vier thema's elk een eigen achtergrondkleur: de naam of omschrijving lichtrood, de oorzaak lichtblauw, de symptomen lichtgroen en de behandeling grijs. Bovenaan elke pagina stond een uitleg bij de kleurcodering. In een schriftelijke toelichting bij de tekst werd eveneens gewezen op de kleuren: deze zouden de studenten kunnen helpen bij het opzoeken van antwoorden in de tekst. In de controle-conditie kreeg de studietekst de gebruikelijke witte achtergrondkleur; deze tekst had verder precies dezelfde inhoud, indeling en lay-out. Wat de indeling betreft: de tekst bevatte zeven paragrafen met daarbinnen subparagrafen, met als titel/subtitel de benaming van een neurologische aandoening. Kenmerkend voor de lay-out was: het gebruik van twee kolommen, korte alinea's met witregels daartussen, en nummering van subparagrafen. In de schriftelijke toelichting bij de controle-conditie werd uiteraard niet gewezen op de kleur, maar op de losse alinea's: deze zouden de studenten kunnen helpen bij het opzoeken van de antwoorden in de tekst.

Meting van de afhankelijke variabelen

De afhankelijke variabelen waren:

- het aantal beantwoorde vragen in een vastgestelde tijd
- de waardering van het kleurgebruik.

Er waren 28 open vragen geconstrueerd, waarop de antwoorden in de tekst opgezocht moesten worden. De constructie van de vragen was uitgevoerd met behulp van een toetsmatrix, zodat ze gelijkmatig over de zeven neurologische aandoeningen gespreid waren. Bij elke neurologische aandoening werden vragen gesteld over de vier thema's/aspecten: definitie of omschrijving, oorzaak, symptomen, en behandeling. In de vragen stond niet expliciet vermeld op welk aspect ze betrekking hadden; de proefpersoon moest dat zelf bepalen. (Een voorbeeldvraag: 'Een patient heeft ernstige pijn in zijn gezicht en motorische uitvalsverschijnselen. Op welke aandoening kan dat wijzen?'). Gemeten werd hoeveel vragen de proefpersonen binnen een vastgestelde tijd beantwoord hadden en hoeveel vragen ze goed hadden beantwoord. Voor dit laatste werden de antwoorden door twee onafhankelijke beoordelaars gescoord, aan de hand van een door de onderzoekers opgesteld scoringsvoorschrift. Goede antwoorden kregen één punt en bij een onvolledig of foutief antwoord werd een half punt respectievelijk nul punten gegeven. Als er verschil van mening was over de beoordeling, werden het antwoord en de twee beoordelingen opnieuw bekeken door een derde beoordelaar, die een nieuwe en bindende beoordeling gaf.

Om te meten in hoeverre de studenten het kleurgebruik en de opgedragen taak waardeerden, werd een vragenlijst opgesteld met 5-punts-schaal vragen en een open vraag.

Procedure

Aan het begin van de onderzoeksbijeenkomst werden de studenten at random verdeeld in een experimentele en een controle-groep. Ze kregen de opdracht om de toelichting die voorafging aan de tekst goed te lezen (zie Materiaal) en de tekst in vijf minuten globaal door te nemen, zodat ze wisten hoe de tekst was opgebouwd en waar ze informatie terug zouden kunnen vinden. Meegedeeld werd dat ze daarna open vragen zouden krijgen waarop ze de antwoorden in de tekst op moesten zoeken. Bij het globaal doornemen en de zoektaak mochten geen strepen in de tekst worden gezet of aantekeningen worden gemaakt. Voor het opzoeken van de antwoorden kregen ze 14,5 minuut. Deze tijd was vastgesteld na proefonderzoek. Na het beantwoorden van de vragen vulden de studenten de genoemde waarderings-vragenlijst in betreffende het kleurgebruik en de opgedragen taak. Ook werden enige achtergrondgegevens van de studenten verzameld (geslacht, vooropleiding, studieresultaten, voorkennis). De hele onderzoeksbijeenkomst duurde ongeveer één uur.

RESULTATEN

Controle op gelijkwaardigheid van experimentele en controle-groep

Eerst werd nagegaan of de groepen door randomisatie gelijkwaardig waren samengesteld wat betreft de volgende achtergrondgegevens van de studenten: geslacht, vooropleiding, studieresultaten en voorkennis. Er werden geen significante verschillen gevonden tussen de experimentele en de controle-groep. Beide groepen konden dus als gelijkwaardig worden beschouwd.

Totaal aantal beantwoorde vragen

Met een t-toets werd nagegaan of het aantal beantwoorde vragen in de kleur-conditie en de zwart-wit-conditie van elkaar verschilde (zie Tabel 1). Het bleek dat er in de kleur-conditie significant meer vragen waren beantwoord dan in de zwart-wit-conditie.

Tabel 1. Totaal aantal beantwoorde vragen en t-toets data.

	n	M	SD	t(40)	p (eenzijdig)
kleur	20	15.20	0.59	2.20	0.02
zwart-wit	22	13.05	0.76		

Aantal goed beantwoorde vragen

De hogere snelheid bij het opzoeken van de antwoorden in de kleur-conditie zou onnauwkeurigheid tot gevolg kunnen hebben, die tot uitdrukking kan komen in een groter aantal gemaakte fouten. Om na te gaan hoeveel vragen er fout waren beantwoord, werden de antwoorden daarom beoordeeld door twee onafhankelijke beoordelaars. De interbeoordelaars-betrouwbaarheid was .99. Er werd een gecorrigeerde score vastgesteld: van het aantal beantwoorde vragen werd het aantal fout beantwoorde vragen afgetrokken.

Ook als uitgegaan werd van dit gecorrigeerde aantal beantwoorde vragen, bleek er een significant verschil te zijn tussen de kleur- en zwart-wit conditie. In de kleur-conditie waren significant meer vragen goed beantwoord dan in de zwart-wit-conditie (zie Tabel 2).

Tabel 2. Aantal goed beantwoorde vragen en t-toets data.

	n	M	SD	t(40)	p(eenzijdig)
kleur	20	13.90	2.56	1.73	0.046
zwart-wit	22	12.25	3.51		

Waardering van het kleurgebruik

Om inzicht te krijgen in de waardering van het kleurgebruik waren 5-punts-schaal vragen gesteld aan de experimentele groep. De resultaten zijn in Tabel 3 weergegeven.

Tabel 3. Waardering van het kleurgebruik in de studietekst.

	M	SD
Dat de studietekst op gekleurde alinea's gedrukt was, vond ik:		
heel fijn (1) - heel vervelend (5)	1.55	0.83
heel handig (1) - heel lastig (5)	1.20	0.52
heel verhelderend (1) - heel verwarrend (5)	1.45	0.69
heel noodzakelijk (1) - heel overbodig (5)	2.45	0.76
Doordat de studietekst gekleurd was, kon ik me bij het opzoeken van de antwoorden veel beter concentreren (1) - veel slechter concentreren (5)	1.60	0.68
Doordat de studietekst gekleurd was, was ik veel meer gemotiveerd (1) - veel minder gemotiveerd (5) om de antwoorden op te zoeken	2.35	0.75
Ik heb bij het opzoeken van de antwoorden in de studietekst heel veel (1) - helemaal niet (5) gebruik gemaakt van de kleuren	1.40	0.60
Ik kon de kleuren in de studietekst heel goed (1) - heel slecht (5) onderscheiden	1.05	0.22

Uit deze tabel kan geconcludeerd worden dat de studenten het kleurgebruik in de studietekst in het algemeen positief waardeerden.

Uit de antwoorden op de *open vraag* ('Wat was volgens jou de bedoeling van het kleurgebruik in de studietekst?'), bleek dat de studenten het kleurgebruik in de tekst een structureringsfunctie toekenden, waardoor het opzoeken van antwoorden versneld kon worden. Ook werd uit de antwoorden op deze vraag duidelijk hoe het kleurgebruik inwerkte op hun zoekgedrag: men zocht alleen binnen een gedeelte met één achtergrond-kleur.

Waardering van de taak

Naast de vragen over het kleurgebruik in de kleur-conditie werden aan de studenten in beide condities nog enkele algemene 5-punts-schaal vragen gesteld over de opgedragen zoektaak. Hierbij bleek er één significant verschil te zijn tussen de kleur- en zwart-wit-conditie: de studenten in de kleur-conditie vonden de beschikbaar gestelde tijd om de tekst globaal door te nemen (5 min.) ruimer dan de studenten in de controle-conditie ($M_{\text{kleur}}=2.45$, $M_{\text{zwart-wit}}=3.23$, $t(40)=2.62$, $p=.01$).

De kleur- en zwart-wit-conditie bleken niet significant te verschillen m.b.t. de andere vragen over de zoektaak. Voor beide condities gold het volgende: men vond het opzoeken van de antwoorden in de studietekst in het algemeen redelijk leuk, het opzoeken kostte hen weinig moeite, men voelde zich niet erg onder druk staan, men stond neutraal tegenover het meedoen aan het onderzoek (het midden tussen 'veel zin' en 'geen zin'), men vond de studietekst interessant, en men vond het niet storend dat men niet mocht onderstrepen en aantekeningen mocht maken. Geen enkele student had ingevuld dat hij of zij kleurenblind was.

CONCLUSIES EN DISCUSSIE

De belangrijkste resultaten kunnen als volgt omschreven worden:

- In de kleurconditie werden binnen de vastgestelde tijd significant meer vragen beantwoord dan in de zwart-wit conditie. Ook als het aantal beantwoorde vragen gecorrigeerd werd door bij beide condities het aantal fout beantwoorde vragen van het totale aantal af te trekken, waren de verschillen significant.
- Het kleurgebruik werd in het algemeen positief gewaardeerd; het werd met name als handig, verhelderend en fijn ervaren.

Concluderend kan gesteld worden dat beide hypothesen voor dit onderzoek worden bevestigd. De hypothese ten aanzien van de verhoging van de zoeksnelheid bij gekleurde teksten wordt zelfs bevestigd als er een 'harde' maat gebruikt wordt, namelijk als het aantal foute antwoorden van het totale aantal beantwoorde vragen wordt afgetrokken. Dit resultaat is opmerkelijk, omdat het zoeken naar een antwoord waarbij dat antwoord niet wordt gevonden, de studenten wel tijd kost. Waarschijnlijk is dat laatste ook de reden dat het verschil tussen beide condities bij deze 'harde' maat weliswaar significant, maar minder groot is.

In dit onderzoek werden de effecten van kleur nagegaan bij gebruik van een korte tekst met een overzichtelijke indeling en lay-out. Vanwege de hoge kosten van kleurendruk is het immers voor de onderwijspraktijk alleen interessant als er een extra effect van kleurgebruik wordt gevonden, toegevoegd aan de gebruikelijke effecten van een heldere indeling en lay-out. Het is echter te verwachten dat bij langere en minder overzichtelijke teksten het verschil tussen de kleur- en zwart-wit conditie groter zal zijn.

Uit eerder onderzoek bleek reeds dat kleurgebruik de prestatie op niet-tekstuele zoektaken kan versnellen (zie Inleiding). Uit het hier beschreven onderzoek blijkt dat ook de zoektijden bij tekstuele zoektaken verkort kunnen worden als kleur als structureringsmiddel wordt gebruikt. Net als bij niet-tekstuele taken kan een verklaring voor dit verschijnsel gezocht worden in Gestaltvorming of segmentatie, zoals uit de antwoorden van de studenten op de vragenlijst bleek. De studenten zoeken alleen binnen een bepaalde kleur (een categorie, in termen van

Guthrie, 1988) en slaan de rest over zonder naar de inhoud van de tekst te kijken. Het tekstgedeelte waarbinnen gezocht wordt om het gevraagde te vinden, wordt daardoor veel kleiner, waardoor de zoektijden verkort worden. Aan de procescomponenten 'categorie-selectie' en 'extractie van informatie' (Guthrie, 1988) wordt bij gekleurd materiaal dus minder tijd besteed dan bij zwart-wit materiaal.

Dat kleur een functie kan vervullen bij het verduidelijken van de structuur van een tekst, blijkt ook uit het feit dat de studenten in de kleur-conditie de beschikbaar gestelde tijd om de tekst globaal door te nemen significant ruimer vonden dan de studenten in de controle-conditie. Blijkbaar versnelde de kleurtoevoeging het zicht op de tekststructuur.

Uit eerder onderzoek van de auteurs (Van Hout Wolters, Kerstjens & Verhagen, 1990b) bleek dat kleurtoevoeging geen significant effect had op de herinnering van de tekstinhoud. Bij het hier beschreven onderzoek met betrekking tot zoektaken blijken de kleurtoevoegingen wel effectief te zijn. Dit verschil wordt waarschijnlijk veroorzaakt door een verschil in cognitieve taken: het zoekend lezen is meer doelgericht en selectief (Guthrie & Mosenthal, 1987). Uitgaand van de vijf proces-componenten die Guthrie (1988) beschrijft, onderscheidt het zoekend lezen zich van het begrijpend en studerend lezen met name in de component 'categorie-selectie': dit speelt bij het zoekend lezen een belangrijker rol. Door het kleuren van de categorieën kan deze categorie-selectie versneld worden.

Een ander verschil tussen het vorige en het hier beschreven onderzoek is, dat bij het vorige onderzoek wat minder expliciet is gewezen op de functie van het kleurgebruik. Bij het hier beschreven onderzoek werd de aandacht van de studenten gericht op de aanwezigheid en functie van het kleurgebruik door middel van een extra schriftelijke toelichting.

Voor de *onderwijspraktijk* is er uit de resultaten van dit onderzoek af te leiden dat, mits de kleur een duidelijke structureringsfunctie heeft en er gewezen wordt op de aanwezigheid en de betekenis ervan, het gebruik van kleur voor zoektaken in schriftelijk studiemateriaal zinvol kan zijn. De redenen zijn dat het zoekproces op deze manier versneld kan worden en kleur in het algemeen positief gewaardeerd wordt door de gebruikers. Deze eigenschappen van kleurgebruik zullen echter niet altijd opwegen tegen de extra drukkosten voor kleurendruk. Dat zal mogelijk wel het geval zijn, als het schriftelijk materiaal expliciet bedoeld is om er informatie in op te zoeken, zoals bij (computer)handleidingen en naslagwerken.

LITERATUUR

- Carter, M.C., & Cahill, R.C. (1976). Color Code Size for Searching Displays of different Density. *Human Factors*, 18 (3), 273-280.
- Christ, R.E., & Corso, G.M. (1983). The effects of extended practice on the evaluation of visual display codes. *Human Factors*, 25 (1), 71-84.
- Davidoff, J. (1987). The role of colour in visual displays. *International Review of Ergonomics*, 1, 21-42.
- Dreher, M.J., & Guthrie, J.T. (1988). *Cognitive processes in textbook search tasks*. Paper presented at the National Reading Conference.
- Greenskin, J.J., & Fleming, R.A. (1984). The use of colour in command control status boards. In Gibson, G.P. (ed.). *Colour codes versus monochrome electronic displays*, Proceedings of a NATO workshop, London: HMSO.
- Guthrie, J.T. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly*, 23, 178-199.
- Guthrie, J.T., & Britten, T. (1989). *Effects of text structure and strategy type on cognitive components of search*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Francisco.
- Guthrie, J.T., & Mosenthal, P. (1987). Literacy as multidimensional: Locating information and reading comprehension. *Educational Psychologist*, 22, 279-297.
- Hartley, J. (1986). Planning the typographical structure of instructional text. *Educational Psychologist*, 21, 315-333.
- Hout Wolters, B.H.A.M. van (1971). Verbetering van leesgewoonten: Een pleidooi voor onderzoek. *Levende Talen*, 282, 686-709.

- Hout Wolters, B. van, & Kerstjens, W. (1990a). *Searching tasks and the use of colour in instructional texts*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Boston.
- Hout Wolters, B. van, Kerstjens, W., & Verhagen, P. (1990b). The use of colour to structure instructional texts. In S. Dijkstra, B.H.A.M. van Hout Wolters, & P.C. van der Sijde, *Research on instruction: Design and effects*. Englewood Cliffs, NJ: Educational Technology.
- Lamberski, R.J. (1980). *A comprehensive review of the methodology and findings in color investigations*. Denver, Colorado: Association for Educational Communications & Technology.
- Luder, C.B., & Barber, P.J. (1984). Redundant color coding on airborne CRT displays. *Human Factors*, 26 (1), 19-32.
- Lyczak, R.A. (1976). Learning to read: The redundant cues approach. *Journal of Educational Psychology*, 68 (2), 157-166.
- Narborough-Hall, C.S. (1985). Recommendations for applying colour coding to air traffic control displays. *Display*, 131-137.
- Wright, P., & Lickorish, A. (1988). Color cues as location aids in lengthy texts on screen and paper. *Behavior and Information Technology*, 7 (1), 11-30.

Manuscript ontvangen 20-9-1990

Definitieve versie ontvangen 20-11-1990

Sekseverschillen in ruimtelijke vaardigheid: andere ervaringen?

Margot Taal

*Faculteit der Psychologie, Universiteit van Amsterdam**

ABSTRACT

The assumption that sex-related differences in spatial abilities are related to sex-differentiated experiences with spatial activities has been empirically supported with undergraduate and college students. The purpose of the present study was to assess such experiential effects during childhood on the spatial abilities of young adolescents. Three-hundred-and-fifteen thirteen-to-fifteen-year-old adolescents participated in this study. They were presented with a simple version of the questionnaires employed in the previous studies, involving two questions about male and female construction activities during childhood. The spatial abilities of the children of both sexes were assessed by the Dutch revised version of the Differential Aptitude Test (DAT'83) and the Embedded Figures Test. The results indicated that (a) more girls than boys had experiences with cross-sex construction activities, and (b) participation in male and female construction activities in childhood correlated with spatial abilities in young adolescence.

INLEIDING

Volgens Thurstone (1944) is ruimtelijke vaardigheid één van de primaire mentale vaardigheden. Op een aantal maten voor ruimtelijke vaardigheid blijken jongens gemiddeld hoger te scoren dan meisjes. De vraag is of deze verschillen biologisch bepaald zijn of verklaard kunnen worden uit een verschil in relevante ervaringen.

Volgens Maccoby en Jacklin (1974) is het sekseverschil in 'ruimtelijk inzicht' ondubbelzinnig vastgesteld naast dat in verbale vaardigheid, wiskundevaardigheid en agressie. Volgens Maccoby en Jacklin beschikken mannen en jongens in de puberteit en daarna, maar niet daarvoor, over betere analytische en non-analytische, visueel ruimtelijke vaardigheden dan hun vrouwelijke leeftijdgenoten. Zij ontleen hun conclusies aan een zogenaamde 'box score' benadering. Daarbij worden de aantallen geconstateerde, statistisch significante, sekseverschillen van alle onderzoeken over hetzelfde onderwerp bij elkaar opgeteld evenals de aantallen overeenkomsten tussen de seksen. Er wordt in deze benadering geen rekening gehouden met de grootte van de verschillen. Een nieuw ontwikkelde techniek voor meta-analyse maakt gebruik van de effectgrootte d . Deze is gedefinieerd als de verhouding van het verschil tussen de gemiddelden tot de 'gepoolde' standaardafwijking ($d = x\text{-gem.}/sd$). Volgens Cohens (1969) criteria is een d van 0.20 klein, een d van 0.50 middelgroot en een d van 0.80 groot. Linn en Hyde (1989) concluderen op grond van hun onderzoek met deze techniek a) dat sekseverschillen in ruimtelijke vaardigheid de laatste vijftien jaar afnemen en b) dat sekseverschillen verdwijnen door training.

Linn en Petersen (1985) onderscheiden drie typen ruimtelijke vaardigheden. Op het terrein van de ruimtelijke waarneming ($d=.44$) en de mentale rotatie ($d=.73$) zijn jongens beter dan meisjes en mannen beter dan vrouwen. Alleen bij ruimtelijke visualisatie wordt geen sekseverschil van enig belang gevonden ($d=.13$).

Tot de tests die 'ruimtelijke visualisatie' meten behoort onder meer de subtest 'ruimtelijk inzicht' van de DAT (Bennett, Seashore, & Wesman, 1974). In tegenstelling tot Linn en Peter-

Met dank aan Yvonne Uittenbogaard en Janneke de Vries voor de verzameling van de gegevens.

* Fac. Psychologie, UvA, Vakgr. Ontwikk. Psych., Roetersstraat 15, k 1016, 1018 WB Amsterdam.

sens bevindingen worden op de subtest Ruimtelijk Inzicht van de DAT in Nederland (DAT'83-RI) wel sekseverschillen aangetroffen. Uit de normgegevens (Evers & Lucassen, 1984) blijkt bijvoorbeeld dat in de brugklas, tweede klas LBO en de tweede en derde klas MAVO het gestandaardiseerde verschil (d) tussen de scores van jongens en meisjes respectievelijk 0.27, 0.59, 0.51 en 0.41 zijn. In de meest heterogene groep, de brugklas is het sekseverschil het kleinst. Dat komt overeen met de teneur van de sekseverschillen op het gebied van wiskunde (Hyde, Fennema, & Lamon, 1990): Hoe heterogener de groep, des te kleiner de verschillen. Te oordelen naar de grootte van de standaarddeviaties hebben we althans in deze klassen niet met een artefact van doen. In de brugklas blijkt de standaarddeviatie even groot te zijn als in de hogere klassen.

Waarom verschillen meisjes en jongens in ruimtelijke vaardigheden? Hoewel gespeculeerd wordt over mogelijk biologische factoren in het ontstaan van genoemde sekseverschillen, duidt de bevinding dat het sekseverschil op perceptuele en 'closure'snelheidstests na training verdwijnt (Lohman, 1987) op het belang van oefening en ervaring. Deze bevinding sluit aan bij de opvatting dat sekseverschillen in ruimtelijke vaardigheid gesocialiseerd worden. Jongens krijgen ander speelgoed dan meisjes, ze worden meer gestimuleerd aan sport te doen en hen wordt meer bewegingsvrijheid gelaten (Smith, 1986, Newson & Newson, 1986 en Colley, 1986).

Newcombe, Bandura, & Taylor (1983) rapporteren onderzoek bij eerstejaars studenten waarin ruimtelijke vaardigheid (DAT-RI) gerelateerd wordt aan relevante activiteiten. Zij onderscheiden veertig mannelijke, twintig neutrale en eenentwintig vrouwelijke activiteiten die een beroep doen op ruimtelijke vaardigheden. Dat zijn vaardigheden die nodig zijn voor het oplossen van opgaven van de subtest Ruimtelijk Inzicht van de DAT. In hun onderzoek moesten de proefpersonen op een zes-puntsschaal aangeven in welke mate ze wekelijks aan elk van deze activiteiten deelnemen. Overall is er een correlatie van DAT-RI-scores met ruimtelijke activiteiten van $r=.33$ ($p<.05$), voor vrouwen alleen is $r=.40$ ($p=.05$) en voor mannen alleen $r=.18$ (n.s.). Voorbeelden van ruimtelijke activiteiten die correleren met DAT-RI-scores zijn glasblazen, modelvliegtuigen bouwen, timmeren, fotograferen, disco dansen en kleren maken. Van de mannelijke activiteiten correleren er bij de vrouwen zeven positief met de DAT-RI, bij de mannen drie; van de neutrale evenals de vrouwelijke activiteiten correleren er bij de vrouwen drie en bij de mannen twee positief.

Signorella, Krupa, Jamison en Lyons (1986) gebruiken in hun onderzoek bij eerstejaars studenten een verkorte versie van de activiteitenlijst van Newcombe et al., met tien mannelijke, tien vrouwelijke en tien neutrale ruimtelijke activiteiten. Ze relateren de geselecteerde activiteiten aan een ruimtelijke vaardigheid die volgens Linn en Petersen tot 'ruimtelijke waarneming' behoort (horizontaliteit à la Piaget). Alleen bij vrouwen blijken mannelijke activiteiten positief gecorreleerd te zijn met de horizontaliteitstaak ($r=.50$, $p<.01$), bij mannen is dat niet het geval ($r=.10$, n.s.). Hoewel uit de resultaten van deze onderzoeken geen conclusies kunnen worden getrokken over de oorzakelijkheid van de relatie, kan men de resultaten wel zo interpreteren dat het ondernomen hebben van bepaalde ruimtelijke activiteiten samengaat met bepaalde ruimtelijke vaardigheden.

Dat slechts een klein aantal activiteiten correleert met ruimtelijke vaardigheid zou kunnen komen doordat te weinig gezocht is naar activiteiten die kinderen doen. Juist op jonge leeftijd is de invloed van ervaring met bepaalde activiteiten op de ontwikkeling van mentale vaardigheid groot. Volgens de literatuur over socialisatie in de jeugd jaren komen er al vroeg sekseverschillen voor in constructie-activiteiten. Ook een aantal activiteiten die in de lijst van Newcombe et al. met de DAT-RI correleren zijn constructie-activiteiten. De vraagstelling in het huidige onderzoek luidt: Gaat ervaring met constructie-activiteiten in de jeugd gepaard met ruimtelijke vaardigheid vroeg in de adolescentie? Als er iets in de biologische uitrusting van meisjes de ontwikkeling van ruimtelijke visualisatie zou verhinderen, dan zou men vanaf het begin van de puberteit een vergroting van de achterstand verwachten. Als echter gebrek aan ervaring met relevant constructiespel debet is aan die achterstand dan varieert ruimtelijke vaardigheid veeleer met die ervaring.

In het huidige onderzoek kiezen we als mannelijke constructie-activiteit modelbouwen van

vliegtuigen of auto's, bijvoorbeeld met lego of blokken en als vrouwelijke activiteit het maken van (poppen)kleden van patroon. Het huidige onderzoek maakt gebruik van retrospectieve rapportage en deze leent zich niet voor de nauwkeurigheid waarmee Newcombe et al. (zespuntsschaal) en Signorella et al. (zevenpuntsschaal) de activiteiten van hun proefpersonen registreerden. De rapportage beperkt zich tot de vraag of proefpersonen dergelijke activiteiten wel of niet minstens tien keer gedaan hebben.

Om ruimtelijke vaardigheid te registreren bij jonge adolescenten maken we gebruik van gangbare Nederlandse tests. Dat zijn de genoemde subtest 'Ruimtelijk Inzicht' van de DAT'83 (Evers & Lucassen, 1984) en de ruimtelijke vaardigheidentest van Witkin, Oltman, Raskin en Karp (1971), de 'Embedded Figures Test' (EFT). Alleen van de DAT'83 zijn Nederlandse normen bekend en worden er seksverschillen vermeld.

Is er sprake van één type ruimtelijke vaardigheid bij deze verschillende tests? Volgens Linn en Petersen (p. 1484) behoren de test voor Verborgen Figuren (EFT) en de subtest Ruimtelijk Inzicht van de Differentiële Aanleg Test (DAT-RI) tot de categorie 'spatial visualization'. Zoals onder andere Caplan, MacPherson en Tobin (1985) opmerken bestaan er veel verschillende indelingen van ruimtelijke vaardigheden. In een recent overzicht van Lohman (1987) bijvoorbeeld wordt de EFT ingedeeld bij de factor 'Flexibility of Closure', de DAT-RI wordt niet genoemd. Overigens meent Lohman dat, als deze factor wordt gemeten door de EFT, 'the factor is virtually indistinguishable from the general visualization factor' (p.186).

De vraag naar de samenhang tussen vrouwelijke en mannelijke constructie-activiteiten in de jeugd en ruimtelijke vaardigheid vroeg in de adolescentie wordt als volgt beantwoord. De prestaties van middelbare scholieren op twee klassikale tests, de DAT'83-RI en de EFT, worden in de eerste en derde klas gerelateerd aan gerapporteerde ervaringen met constructiespel.

METHODE

Proefpersonen

waren 315 leerlingen, 119 van de eerste (gem. 12;8 jaar; s.d. 9 mnd) en 125 van de derde klas (gem. 15;0 jaar; s.d. 9 mnd) van het LBO voor land- en tuinbouw (n=244) en 71 van de eerste klas LTS voor Mode en Confectie in de Randstad (gem. 12;10 jaar; s.d.10 mnd). Op het LBO waren bijna evenveel jongens als meisjes vertegenwoordigd (47 versus 53 %), op de LTS zaten maar twee jongens (3%).

Testmateriaal

Het testmateriaal bestond uit de DAT'83-RI en de EFT.

Bij de DAT'83-RI, krijgt de proefpersoon een reeks figuren voorgelegd. De eerste daarvan wordt de proefpersoon voorgesteld als een twee-dimensionaal patroon. Eén van vier volgende (drie-dimensionale) figuren zou van dit patroon gemaakt kunnen worden. De proefpersoon moet beslissen welke dat is. Er zijn 60 opgaven en elke goede oplossing levert 1 punt op.

Bij de EFT staan per opgave twee figuren naast elkaar, een eenvoudige vorm en een ingewikkelde figuur. Deze eenvoudige vorm maakt in dezelfde grootte deel uit van de ingewikkelde figuur. Opdracht aan de proefpersoon is de eenvoudige figuur daarin met potlood over te trekken. Er zijn 25 opgaven, waarvan de oplossingen dichotoom gescoord worden.

Ervaring met constructie-activiteiten. Er werden twee vragen gesteld over constructie-activiteiten in de kinderjaren. A. Heb je vaak (= minstens tien keer) (poppen)kleding gemaakt? (ja/nee). B. Heb je vaak (= minstens tien keer) aan modelbouwen gedaan (bv. lego, blokken/ auto's, vliegtuigen)? (ja/nee). De scores variëren van 0 (geen enkele activiteit) tot 2 (kledingmaken én modelbouwen).

Procedure

Aan alle leerlingen werden klassikaal de DAT'83-RI en de EFT afgenomen. Aan alle leerlingen werden ook de twee vragen naar hun constructie-activiteiten voorgelegd.

RESULTATEN

Hebben de jongens een grotere ruimtelijke vaardigheid dan de meisjes? Ja, de testcores van de leerlingen van de eerste klas LBO geven een klein sekseverschil te zien ten gunste van de jongens (zie tabel 1). De correlaties tussen sekse en ruimtelijke vaardigheid zijn voor de DAT'83-RI $r=.18$ ($p<.05$) en voor de EFT $r=.20$ ($p<.01$)

Wordt het sekseverschil groter in de derde klas? Nee, in de derde klas hebben de jongens nog wel een kleine voorsprong bij de DAT'83-RI ($r=.16$, $p<.05$), maar niet meer bij de EFT ($r=.07$, n.s.). Het sekseverschil wordt dus niet groter, maar kleiner.

Hebben de jongens meer ervaring met ruimtelijke activiteiten dan de meisjes? Ja, de seksen verschillen kwa ervaring (zie Tabel 1) zowel in de eerste klas ($r=.35$; $p<.01$) als in de derde ($r=.30$; $p<.01$). In klas 1 heeft slechts de helft van de meisjes ervaring met enige vorm van constructiespel tegenover 85 procent van de jongens, maar weinig jongens hebben andere spelervaring dan die met modelbouwen, in tegenstelling tot de meisjes. Het verschil in ervaring in klas 3 is nagenoeg hetzelfde.

Om op de LBO de invloed van ervaring met constructiespel te kunnen evalueren, verdelen we spelervaring in tweeën: Geen of wel ervaring (met bouwen, naaien, of beide). Variëren de ruimtelijk inzicht-scores naar gelang ervaring met constructiespel? Spelervaring in de jeugd gaat inderdaad gepaard met grotere ruimtelijke vaardigheid. In klas 1 (zie Tabel 2) geldt dat zowel de DAT'83-RI ($r=.26$, $p<.01$) als de EFT ($r=.21$, $p<.05$). De samenhang van DAT'83-RI met ervaring in constructiespel is in klas 1 dus groter dan die met sekse ($r=.18$). In klas 3 (zie Tabel 3) gaat ervaring wel gepaard met hogere DAT'83-RI-scores ($r=.28$, $p<.01$), maar correspondeert daar niet meer met de EFT ($r=.07$). Ook in deze klas maakt voor de DAT'83-RI-scores ervaring meer verschil uit dan sekse ($r=.16$).

In klas 1 verschilt de samenhang van ervaring met de EFT en in klas 3 die van de DAT'83-RI voor meisjes en jongens. Wellicht vanwege de geringere variatie in type ervaring is de samenhang bij de jongens veel kleiner. Bij de meisjes is de relatie met de EFT $r=.40$ ($p<.01$) en met de DAT'83-RI $r=.35$ ($p<.01$), maar bij de jongens zijn die relaties allebei $r=.11$ (n.s.).

Tabel 1. Ruimtelijke vaardigheid en percentages spelervaring; klas 1 en 3 LBO.

	geen ervaring	naaien	bouwen	beide ervaringen	DAT-RI gem.	s.d.	EFT gem.	s.d.
klas 1:								
meisjes(n=54)	46	13	32	9	18.5	6.6	11.2	4.2
jongens(n=63)	14	3	79	3	21.3	8.8	13.1	4.9
klas 3:								
meisjes(n=60)	37	8	33	22	26.4	10.6	15.8	5.5
jongens(n=64)	13	0	86	2	29.8	11.0	16.5	4.8

Tabel 2. Ruimtelijke vaardigheid, gelet op spelervaring; klas 1 LBO.

	ervaring:	DAT'83-RI		EFT	
		wel	niet	wel	niet
meisjes	n.	29	25	29	25
	gem.	19.8	17.0	12.7	9.4
	s.d.	7.4	5.2	4.1	3.6
jongens	n.	53	9	54	9
	gem.	22.5	16.9	13.1	14.6
	s.d.	9.0	6.4	5.1	3.8

Tabel 3. Ruimtelijke vaardigheid, gelet op spelervaring; klas 3 LBO.

ervaring:		DAT'83-RI		EFT	
		wel	niet	wel	niet
meisjes	n.	37	23	37	23
	gem.	29.5	21.8	16.7	14.4
	s.d.	9.9	10.3	5.4	5.6
jongens	n.	56	8	56	8
	gem.	30.3	26.6	16.5	16.8
	s.d.	11.5	6.3	5.0	3.5

Tabel 4. Ruimtelijke vaardigheid gelet op ervaring; meisjes klas 1 LTS.

ervaring:		geen (n=14)	naaien (n=13)	bouwen (n=20)	combinatie (n=20)
DAT'83-RI	gem.	15.3	18.9	16.8	23.4
	s.d.	3.1	7.3	4.2	9.2
EFT	gem.	10.1	12.7	13.9	17.1
	s.d.	4.5	5.2	5.9	6.2

Bij de dichotomisering van de ervaring met constructiespel gaat met name de informatie over de rol van de afzonderlijke ervaringen verloren: Gaat een combinatie van ervaringen gepaard met grotere ruimtelijke vaardigheid dan ervaring met enkel bouwen of naaien? De LTS-sters (zie Tabel 4) hebben een grotere variatie in ervaring met constructiespel in vergelijking met de meisjes LBO ($r=.22$; $p<.01$). Op de LTS treffen we minder meisjes aan zonder enige ervaring (22%) dan op het LBO (46%) en meer met ervaring in én modelbouwen én patroonnaaien. Getrichotomiseerd in geen, één of een combinatie van ervaringen, blijkt dat ook de ervaringen van de LTS-meisjes een redelijk sterke samenhang vertonen met ruimtelijke vaardigheid (DAT-RI én EFT: $r=.41$, $p<.01$). Ervaring met naaien blijkt gepaard te gaan met ongeveer even grote ruimtelijke vaardigheden als ervaring met bouwen, maar ervaring met patroon-naaien én modelbouwen gaat gepaard met de grootste ruimtelijke vaardigheid.

Valt de grotere ruimtelijke vaardigheid van de leerlingen in de derde klas toe te schrijven aan grotere spelervaring? Nee, ondanks het vrijwel even grote aantal leerlingen in klas 3 als in klas 1 met ervaring in constructiespel, hebben de derdeklassers grotere ruimtelijke vaardigheid dan klas 1. Dat geldt zowel de leerlingen met ervaring (klas-DAT'83-RI: $r=.40$; EFT: $r=.35$) als die zonder (DAT'83-RI: $r=.37$; EFT: $r=.42$). En het geldt zowel de meisjes (DAT'83-RI: $r=.41$; EFT: $r=.42$) als de jongens (DAT'83-RI: $r=.40$; EFT: $r=.34$, $p<.01$). De toename in het ruimtelijk visualisatievermogen kan dus niet worden toegeschreven aan toenemende ervaring met constructiespel.

DISCUSSIE

Zowel de meisjes als de jongens van de derde klas hebben grotere ruimtelijke vaardigheid dan de eersteklassers. Als er iets in de biologische uitrusting van meisjes de ontwikkeling van ruimtelijke visualisatie zou verhinderen, dan zou men juist op de leeftijd dat de geslachtsrijpheid ontstaat een vergroting van de achterstand verwachten. Het tegendeel blijkt het geval.

Het sekseverschil in één van de manieren van visualiseren (EFT) verdwijnt zelfs, het andere (DAT'83-RI) blijft ongeveer even klein. De toename in het visualisatievermogen kan niet worden toegeschreven aan toenemende ervaring met constructiespel, want die blijft ongeveer gelijk. Bovendien zijn zowel de leerlingen met ervaring als die zonder, beter dan de eersteklassers. Het lijkt waarschijnlijker dat tussen klas 1 en klas 3 de ruimtelijke vaardigheid van beide seksen vergroot wordt door ervaringen met wiskunde en andere probleemoplossende activiteiten op school.

In tegenstelling tot wat recent in de Amerikaanse literatuur wordt gerapporteerd, treffen we in het huidige onderzoek op de DAT'83-RI in beide klassen van het LBO sekseverschillen aan en op de EFT in de eerste klas. De verschillen houden het midden tussen 'klein' en 'middelgroot'. Of de jongens beter zijn in ruimtelijke visualisatie dan de meisjes, blijkt enigszins afhankelijk van de manier waarop die gemeten wordt.

In overeenstemming tot wat in de literatuur over sekserolontwikkeling wordt gerapporteerd blijken de jongens zich met andere constructie-activiteiten bezig te houden dan de meisjes. Er zijn vrijwel geen jongens te vinden die wat aan patroonnaaien doen, wat het zicht op de rol van uiteenlopende typen ervaringen met constructiespel bemoeilijkt. Hoewel we ervan uitgingen dat patroonnaaien een typische meisjes bezigheid is en modelbouwen typisch voor jongens, lijkt dat op de LBO maar in beperkte zin het geval. Heel wat meisjes daar doen aan modelbouwen en er zijn er eigenlijk maar weinig die (poppe)kleden maken. Typerender voor de meisjes lijkt dat er maar weinig zijn met ervaring in enig constructiespel, terwijl modelbouwen in zoverre typisch iets voor jongens is dat ze het bijna allemaal doen. Voor een aantal meisjes lijkt de barrière te zijn overwonnen om met jongensspeelgoed te spelen, maar nog voor te weinig en bij de jongens bestaat er nog geen interesse voor meisjesspeelgoed. Voor beide seksen lijkt dat geen gunstige zaak. Naaiervaring blijkt namelijk gepaard te gaan met betere ruimtelijke visualisatie en een combinatie van ervaring in patroonnaaien en modelbouwen levert de beste visualisatie. De leerlingen zonder ervaring met constructiespel staan altijd onderaan de ranglijst.

De relaties tussen ervaring met constructiespel en de DAT'83-RI zijn wat kleiner dan die in het onderzoek van Newcombe et al. werden gevonden. Opvallend gelijk zijn de sekseverschillen in de relaties tussen constructiespel en DAT'83-RI-scores: Het lijkt voor jongens geen, maar voor meisjes des te meer verschil uit te maken of ze ervaring hebben met constructiespel of niet. Ondanks het zeer eenvoudige instrument om ervaring met constructiespel te peilen, levert dit onderzoek sterke aanwijzingen voor het nut van ruimtelijke ervaring voor visualisatievermogen. Of moeten we oorzaak en gevolg andersom interpreteren? Maken de goede visualiseerders een hobby van het soort activiteiten dat een beroep doet op hun visualisatievermogen? De correlatieve gegevens in dit onderzoek lenen zich helaas niet voor gevolgtrekkingen in termen van oorzaak en gevolg. Wel is het grotere vermogen tot ruimtelijke visualisatie in de derde klas een aanwijzing dat deze vaardigheid te leren valt.

Om meer duidelijkheid te krijgen over de oorzakelijke rol van ervaring met constructiespel in de ontwikkeling van ruimtelijke visualisatie, zal het beoefenen van het spel experimenteel kunnen worden gestimuleerd. Dat geldt zowel het patroonnaaien bij jongens als het modelbouwen bij (meer) meisjes. En het ruimtelijk visualisatievermogen van jongere kinderen zou evenals hun interesse in constructie-activiteiten longitudinaal gevolgd kunnen worden.

LITERATUUR

- Bennett, G.K., Seashore, H.G., & Wesman, A.G. (1974). *Differential Aptitude Test*. New York: Psychological Corporation.
- Caplan, P.J., MacPherson, G.M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? *American Psychologist*, 40, 786-799.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Colley, A. (1986). Sex roles in leisure and sport. In D.J. Hargreaves, & A.M. Colley, *The psychology of sex roles* (pp 233-249). London: Harper & Row.

- Evers, A., & Lucassen, W. (1984). *DAT'83. Voorlopige handleiding*. Lisse: Swets & Zeitlinger.
- Hyde, J.S., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139-155.
- Linn, M.C., & Hyde, J.S. (1989). Gender, mathematics, and science. *Educational Researcher*, 18, 17-27.
- Linn, M.C., & Petersen, A.C. (1985). *Child Development*, 56, 1479-1498.
- Lohman, D.F. (1987). Spatial abilities as traits, processes, and knowledge. In R.J. Sternberg (ed.) *Advances in the psychology of human intelligence*, vol. 4 (pp 181-248). Hillsdale NJ: Erlbaum.
- Maccoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford Un. Press.
- Newcombe, N., Bandura, M.M., & Taylor, D.G. (1983). Sex differences in spatial ability and spatial activities. *Sex roles*, 9, 377-386.
- Newson, J., & Newson, E. (1986). Family and sex roles in middle childhood. In D.J. Hargreaves, & A.M. Colley, *The psychology of sex roles* (pp 142-158). London: Harper & Row.
- Signorella, M.L., Krupa, M.H., Jamison, W., & Lyons, N. (1986). A short version of a spatial activity questionnaire. *Sex Roles*, 14, 475-479.
- Smith, P. (1986). Exploration, play and social development in boys and girls. In D.J. Hargreaves, & A.M. Colley, *The psychology of sex roles* (pp 118-141). London: Harper & Row.
- Thurstone, L.L. (1944). *A factorial study of perception*. Chicago, IL: Un. of Chicago Press.
- Witkin, H.A., Oltman, P.K., Raskin, E., & Karp, S.A. (1971). *Manual for the Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists Press.

Manuscript ontvangen 18-9-1990

Definitieve versie ontvangen 22-11-1990

Notities en Commentaren

VARIATIE IN SCHOOLGEMIDDELDEN OP DE CITO-EINDTOETS BASISONDERWIJS

Th. A. van Batenburg*

RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen

ABSTRACT

In this article two main topics are discussed. The first issue is the stability of school effectiveness as indicated by CITO standardized achievement tests in grade 8 of Dutch elementary education. For two consecutive years this stability is estimated to be about .6. Because of this rather low stability of the school means it is concluded that the impact of the so-called 'effective school variables' cannot be substantial. After all, these school characteristics will remain highly stable from year to year. The second issue concerns the impact of the socio-economic and ethnic composition of a school on mean achievement. It is shown that the percentage of working class children and children from ethnic minority groups is negatively related with mean achievement. However, this relationship does not seem to be linear. When the percentage rises above 80% for working class children and ethnic minorities together, or when the percentage of ethnic minority children alone rises above 50%, the results decline rapidly.

STABILITEIT VAN SCHOOLEFFECTEN

De jaren tachtig zouden wat het onderwijs betreft gekarakteriseerd kunnen worden als de jaren van de hoop. Ondanks dat grote innovatieprogramma's als OSM en GEON niet de verwachte resultaten opleverden, heeft men de moed niet verloren. Dit zou een gevolg kunnen zijn van de eind jaren zestig ontstane "effectieve scholen beweging". Dat scholen verschillen in effectiviteit spreekt vooral beleidsmakers aan omdat in deze "beweging" juist die variabelen centraal staan die beleidsmatig gemakkelijk te beïnvloeden zijn. Er is dan ook een onderzoeksstroom gaan vloeien waarbij een nog steeds uitdijend universum aan variabelen van "effectieve scholen" ontdekt wordt. Overzichten waarbij deze variabelen weer in een aantal factoren worden samengevat kunnen gevonden worden bij Blom, Brandsma en Stoel (1986), Fraser, Walberg, Welch, en Hattie (1987), Mortimore e.a. (1988) en Brandsma en Knuver (1989). Vooral onderzoek uitgevoerd in Amerika laat zien dat schoolklimaat-, schoolorganisatie-, en schoolleidervariabelen van invloed zijn op leerprestaties. Nederlands onderzoek en vooral dat van Brandsma en Knuver (1989) geeft een genuanceerder beeld van de effecten van schoolvariabelen. Er worden wel effecten van schoolvariabelen gevonden, echter groot zijn deze niet.

De schoolvariabelen verklaren na controle voor leerlinginstroomkenmerken ongeveer 12% van de variantie in leerlingprestaties. Belangrijke variabelen hierbinnen zijn: een regelmatige evaluatie van de leerlingprestaties, een groepering in rijen, een planning volgens het rooster en het geven van huiswerk. De hoge verwachtingen die beleidsmakers hebben, worden hiermee gedeeltelijk de grond ingeboord: de variabelen die zij kunnen manipuleren, blijken slechts een kleine bijdrage te kunnen leveren. De vraag rijst of enige hoop voor beleidsmakers nog wel op zijn plaats is. In dit artikel wordt allereerst een bijdrage geleverd aan de discussie over het belang van schoolvariabelen voor het verklaren van leerlingprestaties. Dit wordt gedaan door de stabiliteit van schoolgemiddelden van leerlingprestaties te bepalen.

* Met dank aan drs. H.P. Brandsma en dr. H. Kuyper.
RION, Postbus 1286, 9701 BG Groningen.

Onder stabiliteit wordt verstaan de correlatie tussen gemiddelde leerprestaties van scholen in verschillende schooljaren. Geargumenteed kan worden dat een lage stabiliteit van schoolgemiddelden het belang van schoolvariabelen doet afnemen.

Een hoge stabiliteit is een fundamentele voorwaarde omdat, indien criteriumvariabelen niet stabiel zijn het weinig zin heeft naar voorspellers te zoeken. Wanneer de rangorde van scholen naar gemiddelde leerlingprestaties van jaar tot jaar varieert, komen de relaties tussen min of meer stabiele schoolkenmerken en effectiviteitsvariabelen op losse schroeven te staan (Brandsma & Knuver, 1989).

De uitkomsten van eerder onderzoek laten een wisselend beeld zien. Rutter e.a. (1979) rapporteren een hoge stabiliteit terwijl Rowan, Bossert en Dungen (1983) een geringe mate van stabiliteit rapporteren. Tussen leerjaren binnen scholen worden correlaties van .80 gerapporteerd (Willms, 1987; Goldstein, 1987). Bosker, Guldemon, Hofman en Hofman (1989) schatten dat de stabiliteit ergens tussen de .33 en .89 ligt. Op grond van deze en andere onderzoeksresultaten (bv. Roeleveld & De Jong, 1989) concluderen Brandsma en Knuver (1989) dat de stabiliteit van schoolgemiddelden van leerlingprestaties redelijk is. In hun eigen onderzoek vinden zij stabiliteitscorrelaties van .60 voor taal en .62 voor rekenen. Deze correlaties vinden zij op een redelijke stabiliteit wijzen hetgeen naar mijn mening onjuist is. Met minder dan 40% gebonden variantie kan beter gesproken worden van een vrij lage stabiliteit.

Aangenomen mag worden dat de stabiliteit van schoolkenmerken zeer hoog is. Een schoolklimaat en een schoolorganisatie zal niet op stel en sprong veranderen, een schoolleider wordt niet vaak nieuw aangesteld. Een school staat meestal in een buurt met vrij constante kenmerken zodat de instroom over een periode van zo'n drie jaar ook niet sterk zal veranderen. Wanneer gemiddelde leerprestaties van basisscholen over drie leerjaren stabiel zijn zou dit een argument kunnen zijn voor effectiviteitsvariabelen. In dit artikel worden de relaties in gemiddelde schoolscores op de CITO-eindtoets tussen drie schooljaren gerapporteerd.

De sociaal-etnische samenstelling van de school

In het westen van het land bestaat een zekere ongerustheid over het verschijnsel "zwarte scholen". Eind 1984 zijn er alleen al in Den Haag, Rotterdam en Utrecht 32 basisscholen met meer dan 80% buitenlandse leerlingen (Everts e.a., 1987). Wanneer de Surinaamse leerlingen worden meegerekend dan komt dit aantal veel hoger uit. Op scholen waar steeds meer allochtone leerlingen komen, ontstaat een soort "white flight": ouders van Nederlandse leerlingen halen hun kinderen van school wanneer er in hun ogen te veel allochtone kinderen op komen. Dit vanwege de problemen die deze leerlingen met zich mee zouden brengen. Aangetoond is dat allochtone leerlingen een grote achterstand in de Nederlandse taal hebben (De Jong, 1987). Daarnaast spelen ook allerlei sociaal-emotionele problemen een rol. Scholen met veel allochtone leerlingen staan vaak in de "sociaal-zwakkere buurten" waarin ook de autochtone leerlingen minder kansen hebben.

Op scholen in dergelijke buurten hebben zwakkere leerlingen maar weinig gelegenheid om zich op te trekken aan goede klasgenoten omdat die er meestal nauwelijks zijn. Bovendien zal bij een hoge concentratie allochtone leerlingen de spreekfrequentie van het Nederlands laag zijn. Dit is zeker het geval in kleine groepen met weinig autochtone leerlingen. Momenteel tracht de overheid de onderwijskansen van allochtone en arbeiderskinderen te verbeteren door middel van het inzetten van extra leerkrachten. Arbeiderskinderen wegen bij de formatieberekening voor 1.25 "normaal" kind en allochtone leerlingen voor 1.90.

In dit artikel wordt aangegeven wat de gemiddelde schoolscores op de CITO-toets zijn van scholen die verschillen in het percentage leerlingen met een gewicht van 1.90 en met een gewicht van 1.25.

Databestanden

Voor dit onderzoek zijn gegevens gebruikt uit zes datasets:

1. De gegevens van het ministerie van O&W over leerlingenaantallen, leerlingen met een gewicht 1.25 (arbeidersleerlingen) en gewicht 1.90 (allochtone leerlingen) in het schooljaar

- 1985-1986 (meetpunt september). Het schooljaar 1985-1986 wordt verder aangeduid als 1986.
2. De O&W gegevens voor het schooljaar 1986-1987 waarbij drie meetmomenten zijn gemiddeld. Dit schooljaar wordt verder aangeduid als 1987.
 - 3, 4, 5. De schoolgemiddelden op de CITO-eindtoets in 1986, 1987 en 1988.
 6. De gegevens van het zogenaamde KGF-project met data over klassegrootte en klassemblesamenstelling (zie Van Batenburg, Koster & Ros, 1989).

Deze zes datasets werden gekoppeld aan de hand van de postcode en het straatnummer van de school.

RESULTATEN

Stabiliteit

In Tabel 1 worden de correlaties gepresenteerd tussen de schoolgemiddelden op de CITO-toets in de verschillende schooljaren. Deze schoolgemiddelden zijn gebaseerd op scores van minimaal 10 leerlingen per school. Hiermee vallen de scholen af die maar enkele leerlingen testen en de scholen die in één van de drie schooljaren niet aan de CITO-toets meededen. De volgende afkortingen worden in de tabel gebruikt: ta86= de gemiddelde taalscore in 1986; rek86= idem voor rekenen in 1986; inf= informatieverwerking.

De stabiliteit varieert tussen de .59 en .68. De stabiliteit van de rekenscores is relatief het hoogst, die van informatieverwerking het laagst. De stabiliteit van taal-, reken- en informatieverwerkingscores is in opeenvolgende jaren iets hoger dan wanneer er een jaar tussen zit. De correlaties tussen de vakken binnen een schooljaar zijn met waarden boven de .80 vrij hoog.

De correlaties van het aantal leerlingen op school, het aantal leerlingen met een gewicht 1.90 en het aantal leerlingen met een gewicht 1.25 tussen twee schooljaren (1986 en 1987) zijn respectievelijk: .98, .98 en .96. Hieruit blijkt dat de schoolsamenstelling in twee opeenvolgende schooljaren nagenoeg gelijk is. De correlaties hebben echter betrekking op grotendeels dezelfde leerlingen: ieder jaar komt er meestal slechts één klas bij en gaat er slechts één naar het voortgezet onderwijs. De CITO-gegevens hebben betrekking op groep 8 terwijl de Ministerie-gegevens betrekking hebben op de gehele school. Daarom is nagegaan in hoeverre een groep een representatieve deelverzameling is van de school. In Tabel 2 worden de correlaties gepresenteerd tussen de verschillende aantallen kinderen op school en het aantal leerlingen met een bepaald gewicht in groep 8.

Tabel 1. Stabiliteit van schoolgemiddelden op de Eindtoets Basisonderwijs.

	ta86	rek86	inf86	ta87	rek87	inf87	ta88	rek88	inf88
ta86	—								
rek86	.83	—							
inf86	.87	.80	—						
						N=2382			
ta87	.65	.60	.61	—					
rek87	.63	.68	.59	.84	—				
inf87	.62	.58	.62	.90	.83	—			
ta88	.60	.54	.57	.64	.59	.61	—		
rek88	.55	.60	.52	.59	.67	.56	.82	—	
inf88	.58	.53	.59	.62	.58	.62	.90	.83	—

Tabel 2. Correlaties tussen aantal leerlingen in groep 8 en aantal leerlingen in de hele school.

leerlinggewicht	1.0	1.25	1.90
1986	.81	.81	.81
1987	.81	.78	.80

De correlaties tussen de indicatoren voor klasse- en schoolsamenstelling liggen rond de .80. De klas- (groep 8) en schoolsamenstelling komen redelijk overeen. Geconcludeerd kan worden dat de schoolprestatiescores minder stabiel zijn dan de indicatoren voor de sociale samenstelling van de school.

Schoolprestaties naar schoolsamenstelling

Dat arbeiderskinderen (met gewicht 1.25) en allochtone kinderen (1.90) gemiddeld minder goede resultaten behalen is algemeen bekend. Hoe groot deze verschillen zijn is minder bekend. In Tabel 3 wordt een overzicht gepresenteerd van de gemiddelde percentielscores per school waarbij het percentage 1.25 en 1.90 leerlingen oploopt.

In Tabel 3 vallen een aantal zaken op. Ten eerste wat betreft het percentage allochtone leerlingen de sterke afname van prestatiescores naarmate er meer allochtone leerlingen op school zijn. Dit blijkt ook uit de negatieve correlaties in Tabel 4 tussen prestaties en het percentage 1.25 en 1.90 leerlingen op school. Op zich is het gegeven dat een hoog percentage 1.25 en 1.90 leerlingen de prestaties drukt natuurlijk al bekend, maar de afname lijkt niet lineair te zijn.

Tabel 3. Gemiddelde percentielscores op de Eindtoets Basisonderwijs uitgesplitst naar sociale samenstelling van de school.

	1986				%1.90 leerlingen			
	%1.25 leerlingen				TA	REK	INF	N
	TA	REK	INF	N	TA	REK	INF	N
0- 5%	58	56	58	180	52	51	52	1611
5-10%	53	51	53	48	49	47	49	312
10-15%	54	53	55	76	46	46	47	125
15-20%	52	50	51	97	43	42	44	98
20-25%	51	50	52	148	43	43	42	60
25-30%	52	50	52	185	40	41	41	37
30-35%	49	48	49	189	39	40	40	31
35-40%	51	49	51	209	42	44	40	19
40-45%	50	49	49	194	39	39	38	19
45-50%	49	48	48	216	33	34	35	17
50-55%	47	47	47	232	37	35	34	4
55-60%	47	47	47	171	33	37	31	12
60-65%	47	48	48	142	27	27	27	9
65-70%	45	45	44	110	26	26	29	10
70-75%	46	46	46	90	33	31	34	4
75-80%	44	45	44	49	27	27	30	4
80-85%	45	44	43	34	18	20	17	3
85-90%	50	44	41	7	21	24	20	6
90-95%	43	47	47	5	48	53	37	1

Tabel 3 (vervolg)

	1987				1988			
	TA	1.25 REK	INF	N	TA	1.90 REK	INF	N
0- 5%	58	57	58	188	53	52	52	1571
5-10%	47	46	46	34	49	49	49	321
10-15%	52	51	52	77	45	45	45	137
15-20%	52	51	52	102	43	43	43	77
20-25%	53	52	52	149	43	43	43	69
25-30%	50	49	50	180	42	42	42	49
30-35%	51	51	51	193	37	39	38	38
35-40%	49	49	50	215	40	44	41	21
40-45%	50	50	50	209	29	32	30	15
45-50%	48	48	48	241	34	39	38	13
50-55%	47	47	47	208	31	32	31	19
55-60%	47	48	47	179	33	34	31	12
60-65%	47	47	46	139	32	36	34	9
65-70%	44	44	44	112	24	25	23	10
70-75%	47	46	45	68	14	20	15	2
75-80%	47	47	46	50	30	32	29	5
80-85%	44	46	43	22	10	16	12	7
85-90%	45	44	42	12	19	26	20	5
90-95%	39	43	39	4	28	37	28	2

	1988				1988			
	TA	1.25 REK	INF	N	TA	1.90 REK	INF	N
0- 5%	60	58	61	188	54	53	54	1571
5-10%	49	49	50	34	51	50	51	321
10-15%	54	53	54	77	47	47	48	137
15-20%	52	53	53	102	45	45	46	77
20-25%	54	54	54	149	46	45	45	69
25-30%	52	51	52	180	44	43	43	49
30-35%	52	51	51	193	42	43	41	38
35-40%	51	51	51	215	40	45	41	21
40-45%	51	51	52	209	37	37	36	15
45-50%	50	49	50	241	34	38	34	13
50-55%	48	48	48	208	35	36	35	19
55-60%	50	50	49	179	30	32	29	12
60-65%	49	50	48	139	29	31	29	9
65-70%	47	47	47	112	28	28	28	10
70-75%	49	48	48	68	26	29	29	2
75-80%	45	46	45	50	41	40	38	5
80-85%	42	44	43	22	19	20	20	7
85-90%	40	39	39	12	21	26	20	5
90-95%	44	53	49	4	34	37	31	2

TA = Taal,
 REK = Rekenen,
 INF = Informatieverwerking,
 N = Aantal scholen

Tabel 4. Correlaties tussen percentage 1.25 + 1.90 leerlingen en gemiddelde prestaties per school.

		TA	1986 REK	INF	TA	1987 REK	INF	TA	1988 REK	INF
%86	1.25	-.26	-.19	-.29	-.24	-.20	-.26	-.24	-.19	-.26
%86	1.90	-.41	-.34	-.40	-.46	-.38	-.46	-.40	-.34	-.43
%87	1.25	-.25	-.18	-.27	-.23	-.20	-.25	-.23	-.18	-.26
%87	1.90	-.40	-.34	-.40	-.46	-.38	-.46	-.41	-.34	-.43

Een zeker percentage allochtonen drukt de gemiddelde prestaties niet erg, wordt dit percentage te groot dan nemen de gemiddelden snel af. Bij meer dan 50% allochtone leerlingen worden de gemiddelde schoolprestaties sterk slechter. Eén school (in 1986) met bijna uitsluitend allochtone leerlingen vormt hierop een opvallende uitzondering, de prestaties zijn er goed, vooral op rekenen. Het betreft hier dan ook een internationale elite school. Andere onverwachte storingen in het verloop van de gemiddelden zijn mogelijk toe te schrijven aan het lage aantal scholen in de categorieën met een hoog percentage 1.25 en 1.90 leerlingen. Bij lage aantallen heeft een uitschieter meer invloed op het gemiddelde dan bij hoge aantallen.

Het "verslechterings-effect" kan eerder optreden wanneer er van beide groepen leerlingen (1.25 en 1.90) veel leerlingen op school aanwezig zijn.

In Tabel 5 worden de schoolgemiddelden op de CITO-eindtoets in oplopend percentage van 1.25 en 1.90 leerlingen samen gepresenteerd.

Tabel 5. Gemiddelde percentagescores op de Eindtoets Basisonderwijs uitgesplitst naar het percentage 1.25 + 1.90 leerlingen van de school.

% 1.25+ 1.90	1986			1987			1988			N86	N87-88
	TA	REK	INF	TA	REK	INF	TA	REK	INF		
0- 5%	58	56	58	58	58	58	60	59	61	173	180
5-10%	61	59	62	55	54	53	60	58	60	26	16
10-15%	59	56	59	59	56	58	58	58	59	44	43
15-20%	56	55	56	56	55	57	57	58	58	72	64
20-25%	54	53	55	55	53	54	54	53	55	93	101
25-30%	54	53	54	55	53	55	57	56	56	123	109
30-35%	53	51	52	54	53	53	54	53	54	137	135
35-40%	52	51	52	51	51	52	53	52	53	168	182
40-45%	51	50	52	52	52	53	53	52	53	208	198
45-50%	51	50	50	53	52	51	54	53	54	196	195
50-55%	49	48	48	50	49	49	52	51	51	221	220
55-60%	50	48	49	48	48	48	50	50	50	188	210
60-65%	48	48	48	47	48	48	50	50	49	160	166
65-70%	45	46	45	45	45	46	44	48	47	160	153
70-75%	45	45	44	45	46	45	46	47	47	119	120
75-80%	44	44	45	43	44	43	44	44	44	103	99
80-85%	41	41	40	39	40	39	41	43	42	76	62
85-90%	39	37	37	34	36	34	37	36	37	38	44
90-95%	33	35	35	30	32	30	33	35	31	45	49
95-100%	26	29	26	25	30	25	28	32	27	32	36

TA = Taal,
 REK = Rekenen,
 INF = Informatieverwerking,
 N = Aantal scholen

Het verloop van de gemiddelden in Tabel 5 wijst er op dat bij meer dan 80% 1.25 en 1.90 leerlingen samen de gemiddelde schoolprestaties sterker teruglopen.

DISCUSSIE

De stabiliteit in schoolgemiddelden op de CITO-eindtoets basisonderwijs lijkt aan de lage kant met correlaties tussen de .59 en .68. Hieruit kan worden afgeleid dat schoolvariabelen van weinig waarde zijn voor het voorspellen van leerprestaties. Uit effectiviteitsonderzoek blijkt dat schoolvariabelen hooguit zo'n 12% variantie kunnen verklaren. Al naar gelang het onderzoek zijn dit telkens andere variabelen. Dit verschijnsel kan op de aanwezigheid van kanskapitalisatie wijzen. Van schoolorganisatorische, schoolklimaat- en schoolleiderschapsvariabelen e.d. kan dus niet al te veel verbetering verwacht worden. Beleidsmatig ingrijpen op schoolvariabelen heeft niet veel zin. Wanneer uit onderzoek naar voren komt dat deze variabelen wel verklarende waarde hebben, dan werken ze vermoedelijk als "stand in" voor achterliggende meer op het onderwijsproces gerichte variabelen. De in Nederland gevonden variabelen lijken er op te wijzen dat "ouderwets onderwijs" dat gericht is op de basisvaardigheden het meest effectief is. Wanneer leerkrachten daarnaast hoge verwachtingen van hun leerlingen hebben en voor een prettige sfeer weten te zorgen is dat mooi meegenomen.

Een mogelijke verklaring voor het feit dat in Nederland weinig effectieve schoolvariabelen worden gevonden en bijvoorbeeld in Amerika wel, kan in de variantie in kwaliteit van de scholen gelegen zijn. Met name uit Amerika komen berichten over zeer penibele situaties op scholen in achterbuurten. In Nederland zijn wat betreft criminaliteit, wapenbezit en drugsverslaving duidelijk minder problemen op de basisschool dan in Amerika. In Nederland is vermoedelijk de range van zeer slechte naar zeer goede scholen kleiner.

Boven de 50% 1.90 leerlingen op school en bij meer dan 80% allochtone en arbeiderskinderen worden de prestaties veel slechter. Scholen kunnen blijkbaar een fors percentage leerlingen aan met een min of meer problematische achtergrond. Wordt echter een bepaalde grens overschreden dan nemen de prestaties zienderogen af. Het is de vraag of het beleid hier wel op moet ingrijpen. Het vervoeren van leerlingen naar andere scholen is kostbaar en heeft mogelijk ook onbedoelde negatieve bijeffecten.

LITERATUUR

- Batenburg, Th. A. van, Koster, D., & Ros, A. (1989). Kleinere groepen, betere prestaties? In J. Scheerens & J.C. Verhoeven (Red.), *Schoolorganisatie, beleid en onderwijskwaliteit*. Lisse: Swets & Zeitlinger.
- Brandsma, H.P., & Knuver, A. (1989). *Basisschoolkenmerken als determinanten van het functioneren van leerlingen. Deelrapport II resultaten*. Groningen: RION.
- Blom, M.J.T., Brandsma, H.P., & Stoel, W.G.R. (1986). *Effectieve scholen in het basisonderwijs*. Groningen: RION, Instituut voor Onderwijsonderzoek.
- Bosker, R.J., Guldemon, H., Hofman, R.H., & Hofman, W.H.A. (1989). De stabiliteit van schoolkwaliteit. In J. Scheerens & J.C. Verhoeven (Red.), *Schoolorganisatie, beleid en onderwijskwaliteit*. Lisse: Swets & Zeitlinger.
- Everts, H., Golhof, A., Stassen, P. e.a. (1987). *De cultureel-etnische situatie op OVB-scholen*. Utrecht: Vakgroep onderwijskunde.
- Fraser, B.J., Walberg, H.J., Welch, W.W., & Hattie, J.A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 147-186.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin & co.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen*. Lisse: Swets & Zeitlinger.
- Mortimore, P., e.a. (1988). *School matters: The junior years*. Somerset: Open Books.
- Roeleveld, J., & Jong, U. de (1989). *Evaluating effectiveness of secondary schools in the Netherlands: models and stability*. Paper presented on the ICSE. Rotterdam.
- Rowan, B., Bossert, S.T. & Dwyer, D.C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, 12 (4), 24-31.

- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours. Secondary schools and their effects on children*. London: Open Books.
- Willms, J.D. (1987). *The stability of schooleffects: a four year longitudinal study*. CES Edinburgh (unpublished paper).

Ontvangen 9-11-1989

Boekbespreking

F.M.E. Gerritsen, *VTO taalscreening 3- tot 6-jarigen*. Lisse: Swets & Zeitlinger, 1988.

Deze Leidse dissertatie is het verslag van een onderzoek dat moest leiden tot een 'taalscreeningsinstrumentarium' dat in staat stelt in hun ontwikkeling bedreigde drie- tot zesjarigen op te sporen. Het instrumentarium zou geschikt moeten zijn voor gebruik in de jeugdgezondheidszorg en dus gehanteerd moeten kunnen worden op consultatiebureaus en schoolartsendiensten. Ik zal eerst het boek kort samenvatten en daarna op een paar punten ingaan.

Het boek bestaat uit 16 hoofdstukken. In hoofdstukken 1 en 2 worden doel en praktischeisen uiteengezet. Het instrumentarium moet van de kinderen uit de beoogde leeftijdsgroepen de tien procent selecteren die het minst gevorderd is in taal- en spraakontwikkeling. Screening moet gebeuren naar taalinhoud, -vorm en gebruik. Het onderzoek mag niet meer dan 10 minuten in beslag nemen.

In hoofdstuk 3 wordt uitgelegd dat het instrumentarium bestaat uit het eigenlijke Taalscreeningsinstrument (TSI), een Oudervragenlijst (OVL) en een Schoolvragenlijst (SVL). De TSI's, een voor elke leeftijdsgroep van 3, 4 of 5 jaar, bestaan uit items van gangbare taal-, intelligentie- en ontwikkelingstests die naar schatting door 90% van de kinderen goed zouden worden beantwoord. De vragenlijsten vragen naar taalgedrag en taalontwikkeling en naar enkele achtergrondgegevens. Nadat op grond van de uitkomsten van vooronderzoek TSI's en vragenlijsten waren verbeterd werden ze gebruikt in een onderzoek, dat in hoofdstuk 4 wordt beschreven, om gegevens te verzamelen over 1565 kinderen. Van hen werden 893 door psychologiestudenten onderzocht op speelzaal of kleuterschool (het schoolbestand), de anderen (het praktijkbestand) door de beoogde gebruikers (schoolartsen, wijkverpleegkundigen en anderen) op school, thuis of op het consultatiebureau.

De hoofdstukken 5 tot en met 8 gaan over de analyse van de TSI's. Eerst werden de items verwijderd die de interne consistentie of homogeniteit van het instrument aantastten of moeilijk scoorbaar waren. De analyses over de resterende items laten een sterke overeenkomst zien tussen school- en praktijkbestand in gemiddelde totaalscore, principale componentenstructuur van de items en betrouwbaarheid, een sterke samenhang tussen totaalscore en leeftijd, maar een zwak verband met het opleidingsniveau van de ouders en geen effect van het geslachtsverschil. Er blijken proefleider-effecten te zijn als kinderen woorden of zinnen moeten nazeggen. Bij het praktijkbestand lijkt er een effect te zijn van het beroep (arts etc.) van degene die het TSI afneemt, maar deze interpretatie is niet zeker door contaminatie met andere variabelen. Door die overeenkomst was het gerechtvaardigd om de beide bestanden samen te voegen om de somscore te bepalen die de 10% zwaksten moet afzonderen.

De alpha-coëfficiënten van de TSI's variëren van .87 tot .92. Een principale componenten-analyse over de items geeft twee dimensies, een algemene ('taalbegrip') en een bipolaire die de items in twee groepen opdeelt en 'geheugen' (korte- versus lange-termijn) genoemd wordt. De TSI-norm, dat is de somscore die de steekproef opdeelt in de laagste 10% tegenover de overigen, hangt samen met beide dimensies.

De hoofdstukken 9 tot en met 11 behandelen de analyse van de vragenlijsten. Zijn de TSI's bedoeld om voornamelijk gegevens te verzamelen over het taalbegrip van het kind,

de vragenlijsten zouden de noodzakelijke aanvulling geven over zijn niveau van taalgebruik thuis en op school. Elke lijst bestaat uit vragen die het taalgedrag betreffen en vragen naar extra informatie over kind en milieu. Factoranalyses over de taalgedragsvragen resulteren in twee factoren die benoemd worden als 'taalvaardigheid' en 'communicatie'; soms is er nog een derde factor, 'het sociaal-emotionele aspect van de taal'. De TSI-score hangt het meest samen met de taalvaardigheidsfactor. Ook voor de somscores op de OVL- en SVL-vragen die het hoogst op die factor laadden werden normgrenzen bepaald. Deze normen leveren echter teveel 'zwakke' kinderen op; daarom moeten de OVL- en SVL-scores altijd 'in samenhang worden gezien met de TSI-norm'. De correlatie tussen SVL-norm en TSI-norm is groter dan die tussen OVL- en TSI-norm.

Hoofdstuk 12 behandelt aspecten van de concurrente criteriumgebonden validiteit van de TSI-norm: De overeenstemming met het dichotome logopedisten-oordeel (goed/slecht) over kinderen die bij een audiologisch centrum zijn aangemeld wordt 'zeer groot' genoemd (de correlatie varieert rond .70). Leerlingen van een school voor slechthorende en moeilijk sprekende kinderen behaalden gemiddeld zelfs lagere TSI-totaalscores dan de kinderen die bij het audiologisch centrum waren aangemeld.

Hoofdstuk 13 doet verslag van een follow-up van 55 risico-kinderen. Die waren geselecteerd op grond van een voorlopige norm, nog niet met de TSI-normen die in hoofdstuk 7 worden behandeld. De taalontwikkeling van 36 van deze risico-kinderen blijkt volgens de later bepaalde TSI-normen 'goed' te zijn. De correlatie tussen de eerste en de tweede afname, ongeveer twee jaar later, van het TSI is niet laag .58. In geval van de SVL is hij slechts .37. In interviews meldden de ouders van deze 55 kinderen nogal eens taal-, spraak- en schoolproblemen. De TSI-totaalscore bij tweede afname blijkt ook weer sterker samen te hangen met de taalfactor in de SVL dan met die in de OVL. Daaraan wordt de conclusie verbonden dat ouder en leerkracht het (taal)gedrag van kinderen op verschillende wijze waarderen.

Hoofdstuk 14 bevat een aantal aanbevelingen voor praktijk en onderzoek. De hoofdstukken 15 en 16 bestaan uit samenvattingen.

De centrale vraag die men zich stelt bij lezing van het boek dat ik in het voorgaande samenvatte, is natuurlijk: Is het ontwikkelde instrument geschikt voor het beoogde doel? Het boek maakt het de lezer echter moeilijk deze vraag te beantwoorden. Dat komt omdat dat doel slechts kort- en vaagweg is aangegeven en het daardoor niet duidelijk is wat dat doel is. Volgens blz. 4 gaat het om de 'vertraagde/gestoorde taal- en spraakontwikkeling'. Maar op dezelfde bladzijde wordt ook gewag gemaakt van 'risicogroepen'. Op blz. 131 wordt als doel genoemd 'het opsporen van in hun ontwikkeling bedreigde 3- tot 6-jarigen'. Volgens blz. 175 moet echter 'taalvertraging' worden opgespoord. Zoals ik deze termen opvat is er tenminste één onderscheid in aan te brengen, namelijk tussen feitelijke vertraging en risico of dreiging. In het eerste geval is er al wat 'aan de hand'; in het tweede geval gaat het erom of er in de toekomst wat aan de hand zal zijn. In het eerste geval zou het instrument abnormale taalontwikkeling moeten kunnen detecteren; in het tweede geval is het voorzien van negatieve ontwikkeling het beoogde doel. In het eerste geval is inhouds- en constructvaliditeit het primaire criterium, eventueel de predictie van het concurrente expert-oordeel, in het tweede predictieve validiteit in de letterlijke zin. Nu zou het kunnen zijn dat de schrijfster ervan overtuigd is dat het risico van een abnormale taalontwikkeling op latere leeftijd het best kan worden afgeleid uit abnormale taalontwikkeling op een vroeger tijdstip. Beide doelen, detectie en predictie, zouden dan met een en hetzelfde instrument gediend kunnen zijn. Het lijkt erop dat dit de bedoeling was. Immers, de overeenstemming met het gelijktijdige logopedisten-oordeel werd bepaald, en ook werd de latere taalontwikkeling van een aantal risico-kinderen onderzocht. Maar dat dit standpunt (latere taalachterstand wordt het best voorspeld door vroege taalachterstand) een redelijk

standpunt is in het licht van de onderzoeksliteratuur wordt door de schrijfster niet aannemelijk gemaakt.

Dezelfde vaagheid en onduidelijkheid kenmerkt het boek ook op andere punten. Ik licht er een paar uit. Het TSI zou de taalontwikkeling moeten bepalen naar inhoud, vorm en gebruik, zo wordt in een inleidend hoofdstuk gesteld. Of en waarom aanvankelijk de verzamelde items aan deze voorwaarde voldoen, komt overigens niet aan de orde. Het stellen van deze voorwaarde houdt -naar ik veronderstel- in dat taalvaardigheid een multidimensioneel concept is: in ieder geval de capaciteiten op het terrein van inhoud, vorm en gebruik kunnen min of meer onafhankelijk van elkaar variëren. Niettemin werden de items geselecteerd op homogeniteit en dus toch alsof ze gebruikt moeten worden om een unidimensionele vaardigheid te meten. Bovendien fungeerde de samenhang met de algemene factor in de TSI als criterium voor de selectie van OVL- en SVL-vragen waarvoor vervolgens een norm is bepaald. Indien taalvaardigheid beschouwd moet worden als iets dat in meerdere opzichten kan variëren, dan betekent restrictie tot items die nauw gerelateerd zijn aan één enkele factor verhoging van de betrouwbaarheid in de zin van interne consistentie, maar een beperking van de inhouds- en constructvaliditeit. Er is, met name in verband met de vaardigheid in tweede-talen, een discussie gaande over unidimensionaliteit versus opdeelbaarheid van taalvaardigheid die ook hier van belang is (bijv. Carroll, 1983; Oller, 1983; Sang & Vollmer, 1980). Als de beslissingen die inzake het onderzoek genomen zijn voortvloeien uit een standpunt in deze materie dan had het de lezer geholpen als de schrijfster daarover explicieter was geweest.

Een ander voorbeeld van onduidelijkheid over de overwegingen die aan het onderzoek ten grondslag liggen betreft de kwaliteit van het instrumentarium voor screeningsdoelinden. Op blz. 4 leest men dat een taalscreening alleen in globale zin aangeeft of er sprake is van vertraging of stoornis; verder diagnostisch onderzoek moet de aard en de omvang ervan bepalen. Het lijkt dan voor de hand te liggen dat de eerste toets op adequaatheid van een screeningsinstrument is of bij aansluitend diagnostisch onderzoek het aantal valse negatieven en positieven binnen de perken blijkt te blijven. Die toets blijft echter achterwege. Er wordt niet uitgelegd waarom.

De OVL's en SVL's waren natuurlijk niet bedoeld als criteriumtoetsen ter bepaling van de adequaatheid van de screening; men kan ze kwalijk als diagnostische onderzoeksmiddelen betitelen. Bovendien blijken deze vragenlijsten veeleer bedoeld om informatie te verwerven ter 'aanvulling' op het TSI (bijv. blz. 23-24, 123); ze moeten dan uiteraard niet dezelfde informatie geven als het TSI.

De lezer ziet zich echter met een nieuw probleem geconfronteerd: wat moet onder dat 'aanvullen' verstaan worden? Op blz. 24 wordt gesteld dat de TSI's informeren over de taalkennis op onderscheiden taalcomponenten. Ouders zouden veeleer een idee hebben van de globale taalvaardigheid van hun kind. Hun gegevens vormen daardoor een welkome aanvulling. Dat dat kan is zeker niet uitgesloten, maar ik kan me een instrumentariumgebruiker voorstellen die graag zou willen weten welke aanvullende conclusies verbonden moeten worden aan de constatering dat een kind met een beneden normale TSI-score door de ouders in taalvaardigheid positief wordt beoordeeld of omgekeerd. Men kan zich ook voorstellen dat de gebruiker neigt naar het alternatief, dat er iets mis is met een van beide of allebei de informatiebronnen. Hoe de schrijfster denkt over die alternatieve interpretatie heb ik niet uit het boek kunnen opmaken.

De schrijfster lijkt overigens zelf te twijfelen aan het gewicht dat bij een beslissing moet worden gehecht aan de vragenlijsten. Als ze regels geeft voor 'bewaking' van een bepaald kind (blz. 20), dan is het uitgangspunt de TSI-score en niet het oordeel van ouders of leerkrachten (hetgeen mij een merkwaardig en weinig functioneel standpunt lijkt dat geen recht doet aan hun betrokkenheid en ongerustheid). Als ze kinderen selecteert voor een

follow-up-studie dan is het uitgangspunt weer het TSI, niet de OVL of SVL, laat staan een combinatie.

Het is overigens opmerkelijk dat bij de identificatie van een risico-kind niet geadviseerd wordt nadere diagnostiek uit te voeren, zoals uit de boven gereleveerde kenschetsing van screening zou volgen, maar het te heronderzoeken met de screeningstoets. Ook dit advies wordt niet toegelicht.

Iets anders dat ten onrechte niet is toegelicht is het volgende. Er waren twee groepen proefpersonen, het schoolbestand en het praktijkbestand. De TSI-scores van beide bestanden werden met elkaar vergeleken voordat ze werden samengenomen voor verdere analyse. Het argument was dat kinderen wellicht anders reageren op psychologiestudenten dan op artsen en wijkverpleegsters. Maar voor de analyse van de OVL- en de SVL-data werden de gegevens van beide bestanden kennelijk samengenomen zonder dat getoetst was of ook daar ook sprake was van overeenstemming. Het argument dat gehanteerd werd bij de analyse van de TSI-scores lijkt me echter in nog sterkere mate op te gaan voor de ouders en de leerkrachten die de vragenlijsten invulden. Zij zullen de betreffende volwassenen immers beter kunnen differentiëren naar functie en status en, eventueel rekening houdend met hun eigen belangen, hun beantwoording daarop afstemmen.

Voor de follow-up-studie waren kinderen geselecteerd die op het TSI een lage score hadden. Aan hun ouders werd om medewerking gevraagd. 'De meeste ouders reageerden positief. ... Soms overheerste de verbazing: 'hij/zij praat toch heel goed!' maar meestal was men juist bereid mee te doen ...' (blz. 146). Ik leid hieruit af -maar daarover ben ik niet zeker, de beschrijving is zeer beperkt- dat aan de ouders verteld was dat de lage score van hun kind op het TSI de reden voor selectie was. De meewerkende ouders werden vervolgens geïnterviewd. Het lijkt me geen wonder als bij zo'n werkwijze '... de ouders meermalen een problematisch verlopende taal- en spraakontwikkeling vermelden' (blz. 163).

Een ernstig gebrek van dit boek vind ik bovendien het gemis aan wetenschappelijke fundering van het onderzoek op een aantal relevante punten. Het theoretisch kader wordt in slechts acht bladzijden behandeld. Bovendien komen er slechts enkele algemene onderwerpen aan de orde: *Wat is taal? Het belang van de taal voor de ontwikkeling van het kind. Hoofdzaken van de normale taalontwikkeling van 3 tot 6 jaar.* Specifieker onderwerpen die voor het onderwerp van de dissertatie van belang zijn worden niet besproken. Zo mis ik een uiteenzetting over het begrip 'screening' waaruit de kwaliteitseisen voor het te ontwikkelen screeningsinstrument en de opzet van het onderzoek konden worden afgeleid. Dat er over dit onderwerp relevante zaken gezegd kunnen worden hebben Cools & Hermans (1977) laten zien en ik neem aan dat de ontwikkeling op dit terrein sinds 1977 niet heeft stilgestaan. Een ander onderwerp dat hier niet onbehandeld had mogen blijven is de persistentie van taalproblemen: als een kind op jonge leeftijd taalproblemen heeft, is daarvan op latere leeftijd nog iets van te merken aan het taalgedrag? (De schrijfster beperkt zich tot de consequenties op korte termijn voor de sociale en cognitieve ontwikkeling.) En het omgekeerde: hebben taalproblemen op latere leeftijd hun voorboden in vroegere taalproblemen? Het nut van het beoogde screeningsinstrument is immers slechts beperkt als de gedetecteerde problemen merendeels vanzelf overgaan of als veel taalproblemen pas op latere leeftijd ontstaan. De onderzoeksliteratuur (bijv. Aram, Ekelman & Nation, 1984; Bernstein & Stark, 1985; Rhea & Cohen, 1985; Stark, 1984) die het mogelijk maakt iets over dit onderwerp te zeggen dat tot rechtvaardiging of nuancering van de probleemstelling had kunnen dienen, wordt in dit boek niet behandeld. Een ander onderwerp waarover ik graag wat meer had gelezen, is dat van de dimensionaliteit van taalgedrag. Het uitgangspunt van de schrijfster is dat taalvaardigheid multidimensioneel gekarakteriseerd dient te worden. Daarnaast handelt ze echter niet, heb ik eerder al gesteld,

want ze selecteert de onderdelen van het onderzoeksinstrumentarium op hun samenhang met een algemene factor. Dat laatste zal bij de constructie van het TSI door gangbare psychometrische overwegingen zijn geïnspireerd. Kennelijk verkeerde de schrijfster op dit punt in een dilemma. De eerder genoemde discussie over dit onderwerp, het multi- dan wel unidimensionele karakter van taalvaardigheid (zie Rost (1989) voor een recenter bijdrage), hoeft de schrijfster echter niet onbekend te zijn; ze verwijst immers naar Van Bon (1984) waarin deze discussie aan de orde is en waarin onderzoeksresultaten worden gepresenteerd die een multidimensionele opvatting nuanceren. De uitkomsten, een multifactoriële structuur, zowel als het dilemma waarin de auteur zich geweten moet hebben, hadden het gerechtvaardigd de literatuur over dit onderwerp te behandelen en aan de discussie bij te dragen.

Het zal duidelijk zijn dat ik dit geen goed boek vind. Er staat teveel niet in en wat erin staat is vaak niet goed te begrijpen.

REFERENTIES

- Aram, D.M., Ekelman, B.L. & Nation, J.E. (1984). Preschoolers with language disorders. *Journal of Speech and Hearing Research*, 27, 232-244.
- Bernstein, L.E. & Stark, R.E. (1985). Speech perception development in language-impaired children: a 4-year follow-up study. *Journal of Speech and Hearing Disorders*, 50, 21-30.
- Bon, W.H.J. van (1984). *Aspecten van taalvaardigheid*. Lisse: Swets & Zeitlinger.
- Carroll, J.B. (1983). Psychometric theory and language testing. In J.A. Oller (Ed.), *Issues in language testing research*. Rowley: Newbury House.
- Cools, A.T.M. & Hermans, J.M.A. (1977). *Vroegtijdige onderkenning van problemen in de ontwikkeling van kinderen*. Amsterdam: Swets & Zeitlinger.
- Oller, J.W. (1983). A consensus for the eighties? In J.A. Oller (Ed.), *Issues in language testing research*. Rowley: Newbury House.
- Rhea, P. & Cohen, D.J. (1984). Outcomes of severe disorders of language acquisition. *Journal of Autism and Developmental Disorders*, 14, 405-421.
- Rost, D.H. (1989). Reading comprehension: skill or skills. *Journal of Research in Reading*, 12, 87-113.
- Sang, F. & Vollmer, H.J. (1980). Modelle linguistischer Kompetenz und ihre empirische Fundierung. In R. Grotjahn & E. Hopkins (Red.), *Empirical research on language teaching and language acquisition*. Bochum: Brockmeyer.
- Stark, R.E. (1984). Four-year follow-up study of language impaired children. *Annals of Dyslexia*, 34, 49-68.

W.H.J. van Bon,
 Instituut voor Orthopedagogiek,
 K.U. Nijmegen



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

Greetje P.C. van der Werf, RION,
Postbus 1286, 9701 BG Groningen,
tel. 050-636657

Jrg. 14, nr. 5/6

1990

Inhoudsopgave

Congressen, Symposia, Cursussen, Trainingen	376
Ontvangen publikaties	382

Congressen, Symposia, Cursussen, Trainingen

CORPORATE TRAINING FOR EFFECTIVE PERFORMANCE

Research in Human Resource Development at University of Twente, Enschede, The Netherlands, September 25-27, 1991.

Proposal Submission Deadline: March 1, 1991

INTRODUCTION

We are pleased to announce that the University of Twente, Enschede, in The Netherlands, will host the Second International Research Conference on Corporate Training & Development. The first Conference took place also at this University and it resulted in publication of the book 'Strategic Human Resource Development' edited by M. Mulder, A.J. Romiszowski, P.C. van der Sijde (Amsterdam: Swets & Zeitlinger, 1990). Papers will be published as proceedings, and will be considered for publication in an International journal and a book. Participants and presenters are expected from universities, business, industry, and training institutions. Training and development managers, instructors, researchers, consultants and policy makers are invited to present a research paper or a referenced substantial study or to attend the conference.

Corporate Training and Development

Human resources are an important contribution to the performance and the development of organizations. Although this is widely accepted, there are many issues to be solved to establish effective training and development practice. Economic barriers are sometimes preventive in upgrading people in organizations, as training and development is an economic risk for both the individual and the organization. Social, cultural, technological, and demographic factors are also at stake when organizations wish to integrate learning results with corporate performance. Corporate training and development has become an increasingly important tool in organizations, as there are no other ways to acquire a sufficiently skilled work force for the 21st century.

Management, as well as the training and development professional has many problems to resolve: organise, manage, plan, implement and evaluate training and development practices. Are there tested theories, models, approaches, and guidelines to serve as effective and efficient means? What implications do these have for effective performance in organizations?

Studies in all these areas pertaining to sectors of economy such as banking, insurance, public administration, engineering, communications, and public services are welcomed at this conference. Papers in the subject areas management, accounting, sales, engineering, health care, services, and office practice are also welcome. Papers on national and international studies regarding participation and expenditures for training and development are particularly welcome as these studies provide benchmarks for individual corporations as to whether they provide an appropriate level of training and development in their organization.

Special emphasis is placed on referenced *empirical studies and qualitative inquiries* which are designed to enhance the knowledge base of corporate training and development.

The program is structured around the following topics:

I. Training and development in changing environment

Papers contributing to questions about training and development programs in relation to a changing environment. Implications for human assets of innovations may be presented.

II. Instructional design, instructional media and effective performance

If training and development has to become an integral part of technological innovation, the course or instructional plans for business and industry have to be devised pro-actively. Papers related to this issue can be presented in this track.

III. Effectiveness of training programs

Theories, models and strategies accompanied by empirical research on evaluation and (cost-) effectiveness of training and development programs for the organization in general, or specific jobs or branches specifically, may be presented within this track.

IV. (Inter)national studies

National surveys on training and development expenditure, participation, levels, content domains by which comparisons within and between nations may be given.

Call for Papers

Paper proposals are invited for the conference tracks mentioned before. There will also be a free track, in which papers can be presented that are more or less loosely related to the main theme of the conference. We also encourage proposals for poster sessions, demonstrations, symposia, and workshops. Proposals will be selected by an evaluation committee chaired by prof. dr. W.J. Nijhof.

The deadline for submitting proposals is March 1, 1991

Please observe the following guidelines for submitting proposals.

Papers/postersessions:

A paper or poster proposal should consist of an abstract of not more than 100 words, and a synopsis of the proposed paper or poster of not more than two pages. The abstract will be printed in the final program of the conference.

Demonstrations/workshops:

A proposal for demonstrations/workshops should consist of a summary of the demonstration or workshop, including a description of the type of involvement of the attendants of the sessions, of not more than two pages. Add an abstract of not more than 100 words. This abstract will be printed in the final program of the conference.

Symposia:

A symposium is centered around one theme and should combine at least three papers which are presented and discussed in relation to each other. The proposal should consist of an abstract of the symposium of not more than 100 words as well as an abstract and synopsis per paper in the symposium.

Conference sessions for presentations are 45 minutes each.

Further Instructions for Submitting a Proposal

Please send three copies of each of the following documents when submitting a proposal.

1. Paper Proposal Cover Sheet, containing

- last name, first name, title
- institution
- mailing address
- city/zipcode
- state/country
- phone (home; work)
- fax
- title of contribution
- co-authors
- indication of the conference track I, II, III, IV or Free
- the kind of proposal: paper, poster, workshop, demonstration, symposium

2. *Summary.* A 2- to 3-page (8 1/2 x 11 paper) single-spaced summary, for use in evaluation of the proposal. The summary should contain:
 - objectives
 - perspective/theoretical framework
 - methods/techniques
 - data source
 - results/conclusions/point of view
 - educational or scientific importance of the study
3. *Abstract.* An abstract of 100 words maximum that will be printed in the program.
4. *Envelopes.* Three self-addressed business-size envelopes for acknowledgement of receipt of proposal, notification of the decision of the proposal selection committee, and notification of scheduled time if accepted.

Send proposals before March 1, 1991 to

Martin Mulder

University of Twente, Department of Education

P.O. Box 217, 7500 AE Enschede, The Netherlands

Telephone international + -31-53-89 36 52 or -31-53-893578

Telefax international + -31-53-35 65 31

The completed paper is required on July 25, 1991.

ONDERWIJSRESEARCHDAGEN '91

1. Inleiding

De jaarlijkse Onderwijs Research Dagen (ORD) zullen in 1991 op 16 en 17 mei worden gehouden in Amsterdam. Tijdens deze dagen zullen Nederlandse en Belgische onderwijsonderzoekers met elkaar en met gebruikers van onderzoeksresultaten praten over hun onderzoek. Traditiegetrouw gaat het daarbij niet alleen over de uitvoering en de resultaten van onderzoek, maar ook over de condities waaronder onderzoek wordt uitgevoerd. Voor veel onderzoekers zijn de ORD dan ook een belangrijk moment voor reflectie over het eigen vak en voor contacten met vakgenoten uit andere instellingen. Dat gebeurt tijdens de geplande ORD-activiteiten, maar Amsterdam biedt ook talrijke mogelijkheden voor informele contacten.

In navolging van de ORD '90 in Nijmegen zullen de deelnemers vóór de ORD rond 1 mei 1991 een bundel krijgen toegezonden met de papers van één thema naar keuze. Andere bundels kunnen tegen extra betaling worden verkregen. Ook in Amsterdam wordt getracht om de discussie tussen presentator en toehoorders goed uit de verf te laten komen.

2. Papersessies, symposia, cursussen, presentaties

Tijdens de ORD staan de presentaties van papers op een zestal thema's centraal. Daarnaast is er de mogelijkheid om papers te presenteren in symposia en om cursussen en andere presentaties te organiseren.

Papersessies. Papersessies zullen worden gepresenteerd in het kader van een van de zes thema's of een subthema. Elk (sub)thema wordt ingeleid door de voorzitter in het licht van de stand van zaken op het betreffende onderzoeksgebied. Bij de presentatie van papers zal de nadruk liggen op het aangaan van een discussie tussen onderzoeker en toehoorders. Daarbij wordt ervan uitgegaan dat de toehoorders het in de bundel opgenomen paper hebben bestudeerd. Een discussiant

zal een aanzet tot de discussie leveren. Er zal een beperkt aantal vrije papers worden toegelaten, met een voorkeur voor papers die, in de vorm van een symposium, een samenhangend geheel vormen.

Symposia. Niet bij een (sub)thema behorende papers kunnen worden gepresenteerd in een symposium. Een symposium bestaat uit vier of vijf paperpresentaties met een inleiding door een symposiumvoorzitter. Ook bij de symposia opent een discussiant de discussie. Voor symposia in het kader van één van de zes thema's dient overlegd te worden met de organisatiecommissie of met de coördinatoren.

Cursussen. Tijdens de ORD is er de mogelijkheid een beperkt aantal cursussen te organiseren.

Presentaties. Behalve voor de reguliere marktpresentaties is er op de ORD ook gelegenheid om programma's en andere wetenswaardigheden te demonstreren.

3. Thema's van de ORD

Elk jaar worden op de ORD vier vaste en enkele vrije thema's aan de orde gesteld. Voor de ORD '91 heeft de organisatiecommissie de volgende zes thema's gekozen, met een voorkeur voor bijdragen op de erbij genoemde subthema's:

Onderwijsleerprocessen

- taalvaardigheid
- rekenen, wiskunde en probleemoplossen
- motivationele en affectieve processen
- zelfinstructie, zelfregulatie en metacognitieve processen

Curriculum

- effectieve scholen
- nieuwe vakinhouden
- zorgverbreding/speciaal onderwijs

Onderwijs en Samenleving

- etniciteit
- onderwijs en arbeidsmarkt
- vrouwen en onderwijs

Methodologie en Technologie

- methodologie
- technologie

Beleid en Organisatie

- maatschappelijke context van beleid
- evaluatie van beleid
- management en organisatie

Hoger Onderwijs en Volwasseneneducatie

- hoger onderwijs
- volwasseneneducatie

4. Inschrijving

Inschrijving dient te geschieden via het inschrijfformulier. Door het invullen en ondertekenen verplicht men zich tot het betalen van het inschrijfgeld en de kosten van aangevraagde extra's. Inschrijving is alleen mogelijk voor beide congresdagen en is verplicht voor iedereen (toehoorders, discussianten, coördinatoren, voorzitters, paperpresentatoren, etc.).

Het inschrijfgeld is inclusief de kosten van koffie, thee en lunches op beide dagen en van één bundel congrespapers.

Het inschrijfgeld is bij inschrijving voor 15 maart 1991:

- VOR-leden f 200,--
- niet VOR-leden f 250,--
- studenten en aio's f 100,--

Na inzending van het formulier wordt een faktuur toegestuurd. Het verschuldigde bedrag moet op 15 april door het congressecretariaat zijn ontvangen.

Bij inschrijving tijdens de ORD is het inschrijfgeld:

- VOR-leden f 300,--
- niet VOR-leden f 300,--
- studenten en aio's f 125,--

Data

- 15-03-91 sluiting van voorinschrijving
- 15-04-91 betalingen voor voorinschrijving moeten door het congressecretariaat zijn ontvangen
- 01-05-91 verzending van congresstukken aan deelnemers
- 16/17-05-91 congresdagen

Adres Congressecretariaat ORD '91, SCO, Grote Bickersstraat 72, 1013 KS Amsterdam, tel.: 0.20 - 5550315/300.

5. Hotelaccommodatie

De hotelreserveringen voor de ORD '91 worden verzorgd door Wagonslits Reizen, 1012 JS Amsterdam, tel.: 020 - 241361, telefax: 020 - 235107. Wegens de grote te verwachten "bollandruk" op de congresdagen, die net voor Pinksteren vallen, wordt aangeraden snel hotelruimte te reserveren. De uiterste datum hiervoor is in ieder geval 1 maart 1991, daarna is er geen garantie dat gevraagde accommodatie beschikbaar zal zijn.

PRIJSVRAAG OVER HET SCHOOLBOEK

Ter gelegenheid van het eerste lustrum van de leerstoel, die bekleed wordt door prof. dr. N.L. Dodde, schrijven de Maatschappij tot Nut van 't Algemeen en de Vakgroep Onderwijskunde van de Rijksuniversiteit Utrecht *een prijsvraag* uit. Het onderwerp luidt:

"Het schoolboek als middel tot onderwijsvernieuwing in verleden en toekomst"

Voor het inzenden van teksten worden uitgenodigd studenten en pas afgestudeerden (na januari 1986) van het HBO en het WO. Behalve nieuwe teksten kunnen ook mogelijk reeds (na 1986) verschenen werkstukken ingediend worden.

De jury die de inzendingen zal beoordelen bestaat uit:

- prof.dr. N.L. Dodde (onderwijskundige)
- prof.dr. N.A.J. Lagerweij (onderwijskundige)
- dr. H.A.M. Franssen (onderwijskundige)
- drs. L. Erades (namens Maatschappij tot Nut van 't Algemeen)
- J. Ahlers (journalist)
- J. Timmer (educatieve uitgever)

De inzendingstermijn sluit 1 juli 1991.

Werkstukken met een maximale omvang van 30 pagina's verzenden aan:

Vakgroep Onderwijskunde
T.n.v. dr. H.A.M. Franssen
Postbus 80.140
3508 TC Utrecht

Het beschikbare prijzengeld is in totaal f. 5.000,-.

Dit bedrag is als volgt verdeeld: eerste prijs f. 2.500,-, tweede prijs f. 1.500,- en derde prijs f. 1.000,-. Het toekennen van de tweede en derde prijs is mogelijk gemaakt door bijdragen van respectievelijk de uitgeverij Malmberg in Den Bosch en de uitgeverij Educaboek te Culemborg.

De prijswinnaars worden bekend gemaakt op een symposium op 22 november 1991 ter gelegenheid van het lustrum in Utrecht over het thema "De rol van de leraar in de school in verleden en toekomst". De produkten van de prijswinnaars zullen worden opgenomen in een boek dat ter gelegenheid van de lustrumviering zal verschijnen. Daarin worden tevens de teksten van de voordrachten opgenomen.

De leerstoel is ingesteld vanwege de Maatschappij tot Nut van 't Algemeen aan de Rijksuniversiteit te Utrecht. In 1986 aanvaardde de heer Dodde de leeropdracht die luidt: "Geschiedenis van het onderwijs, in het bijzonder de geschiedenis van het Nederlandse onderwijs in relatie tot maatschappelijke ontwikkelingen, ideeën over volksontwikkeling en gevoerd onderwijsbeleid." De leerstoel is ondergebracht bij de vakgroep Onderwijskunde van de Faculteit der Sociale Wetenschappen, Rijksuniversiteit Utrecht.

Ontvangen publikaties

- ARBO (1990). *Opmaat tot samenspel*. Advies over de binding tussen regulier en speciaal onderwijs. Zeist: ARBO.
- Berg, R. van den (red.) (1990). *Bouwstenen voor Evaluatie voor de praktijk van onderwijs en begeleiding*. Houten/Antwerpen: Bohn Stafleu Van Loghum.
- Boei, F. (1990). *Ontwikkelingen in declaratieve kennis. Een onderzoek naar veranderingen in kennis-schemata van eerstejaars PABO-studenten*. Proefschrift RU Leiden. De Lier: Academisch Boeken Centrum.
- Bosker, R.J. (1990). *Extra kansen dankzij de school? Het differentieel effect van schoolkenmerken op loopbanen in het voortgezet onderwijs voor lager versus hoger milieu leerlingen en jongens versus meisjes*. Proefschrift RU Groningen. Nijmegen: ITS (OOMO-reeks).
- Brandsma, T.F., Nijhof, W.J. & Kamphorst, J.C. (1990). *Kwalificatie en curriculum. Een internationaal vergelijkende studie naar methoden voor bepaling van kwalificaties*. (Forum 7). Amsterdam/Lisse: Swets & Zeitlinger.
- Brouwer, N. (1989). *Geïntegreerde lerarenopleiding, principes en effecten. Een longitudinaal onderzoek naar organisatie, didactiek en leereffecten van de Utrechtse universitaire lerarenopleiding*. Proefschrift RU Utrecht.
- Bijwaard, D. (1990). *Teaching as acting. A reconstructive study of an action theoretical approach to research and development in the domain of teaching*. Proefschrift Landbouwuniversiteit Wageningen.
- Cito (1990). *Meerjarenplan 1991-1994 en Jaarprogramma 1991 van het Instituut voor Toetsontwikkeling*. Arnhem: Cito.
- Cramer, G. & Kamp, M. van der (1990). *Leren op latere leeftijd*. (Forum 5). Amsterdam/Lisse: Swets & Zeitlinger.
- Driel, J.H. van (1990). *Betrokken bij evenwicht. Een chemiedidactisch onderzoek naar ontwikkeling van een chemisch evenwichtsbegrip*. Proefschrift RU Utrecht. Utrecht: CDB Press.
- Droogleeve Fortuijn, E. (1989). *De docerende staat: onderwijs en macht in de ontwikkeling van Venezuela*. Proefschrift Universiteit van Amsterdam. 's-Gravenhage: CESO.
- Du Bois-Reymond, M. & Eldering, L. (Red.) (1990). *Nieuwe oriëntaties op school en beroep. De rol van sexe en ethniciteit*. (Bijdragen aan de Onderwijsresearch, No. 31). Amsterdam/Lisse: Swets & Zeitlinger.
- Elshout-Mohr, M., Van Daalen, Kapteijns, M.M. & Baltzer, J.E. (1990). *Zelfstandig studeren in het hoger beroepsonderwijs*. Amsterdam: SCO.
- Geerligts, T. (1990). *Scriptiebeoordeling*. Maastricht: RU Limburg (Onderzoek van Onderwijs nr. 42).
- Giesbers, J.H.G.I. & Krogt, F.J. van der (Red.) (1990). *Werken met modellen in organisaties. De modellen van Ernst Marx: ontwikkeling, onderzoek en gebruik*. Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Herfs, P.G.P., Mertens, E.H.M., Perrenet, J.Chr. & Terwel, J. (1990). *Adaptief groepsonderwijs: implementatie en effecten van een curriculum in het voortgezet onderwijs* (Eindrapport Project AGO 12-16). Utrecht: ISOR, Afdeling Onderwijsresearch.
- Hertog, P.C. den (1990). *Persoonlijkheidskenmerken en causale attributies*. Proefschrift KU Nijmegen. Amsterdam: Thesis Publishers.
- Houtveen, A.A.M. (1990). *Begeleiding van vernieuwingen*. Proefschrift Rijksuniversiteit Utrecht.
- Hulshof, M.J.F., Hout, J.F.M.J. van & Prins, J.B.A. (1990). *Onderwijsvernieuwing in het HBO. Een onderzoek naar de richting en aanpak van het vernieuwingsproces*. Nijmegen: IOWO (Beleidsgerichte Studies Hoge Onderwijs en Wetenschappelijk Onderzoek 22).
- Imants, J. & Weijzen, W. (Red.) (1990). *Organisatie van onderwijsinstellingen*. (Bijdragen aan de Onderwijsresearch, No. 30). Amsterdam/Lisse: Swets & Zeitlinger.
- Ippel, M.J. & Elshout, J.J. (Red.) (1990). *Training van hogere-orde denkprocessen*. (Bijdragen aan de Onderwijsresearch, No. 32). Amsterdam/Lisse: Swets & Zeitlinger.
- Kemme, S.L. (1990). *Uitleggen van wiskunde*. Proefschrift RU Groningen. Utrecht: Vakgroep OW & OC.
- Kooij, R. van der, Been, P., Noordman, A. & Sluis, J. van der (1990). *Diagnostische expertmodellen. Richtlijnen voor de praktijk van de orthopedagogische diagnostiek*. Assen: Dekker & Van de Vegt.
- Liere, C. van (1990). *Lastige Leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen*. Proefschrift RU Groningen. Amsterdam: Thesis.
- Looij, T. & Ooijens, J. (Red.) (1989). *Alfabetisering in Noord en Zuid*. (CESO Paperback No. 5). 's-Gravenhage: CESO.
- Mutsaers, M. & Boendermaker, L. (1990). *Criminaliteitspreventie in het onderwijs. Eerste deelexperiment: spijbelcontrole*. Ministerie van Justitie/WODC.

- Ploeg, D. van der (1990). *Integratie van schipperskinderen in het basisonderwijs*. Utrecht: ISOR.
- Ruijs, A. (1990). *Vrouwen en schoolmanagement*. Amsterdam/Lisse: Swets & Zeitlinger.
- Vedder, P. (Ed.) (1990). *Fundamental studies in educational research*. Amsterdam/Lisse: Swets & Zeitlinger.
- Wijnstra, J.M. (Red.) (1990). *Verantwoording van de rekenpeiling medio en einde basisonderwijs 1987*. Arnhem: Cito (PPON-rapport nr. 5).
- IJzendoorn, W.J.E. van (1990). *Prognoses voor probleemleerlingen*. Proefschrift RU Groningen, Groningen: RION (Monografieën onderwijsonderzoek 7).
- Zoelen, L. van & M. Boekennoogen (1990). *Ondersteuning OVB*. Eindrapport van het onderzoek 'Ondersteuning OVB'. Utrecht: ISOR.
- Zoelen, L. van, Mertens, N., Stokhuyzen, Y. & Stokking, K. (1990). *Informatieverwerking en differentiatie op Nederlandse basisscholen*. Utrecht: ISOR.