

Tijdschrift voor onderwijsresearch 1987

<https://hdl.handle.net/1874/208862>

VW 8620

~~BIJL. A.3~~

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591X

Jaargang 12
1987

Swets & Zeitlinger B.V.
Lisse

Redactie:

Bert P.M. Creemers (secretaris)
G. d'Ydewalle
Bernadette van Hout Wolters
Leo J.Th. van der Kamp
Nijs A.J. Lagerweij
G. Wim Meijnen
P. Robert Jan Simons
Marinus J.M. Voeten
Peter Weeda
Wynand H.F.W. Wijnen

BIBLIOTHEEK DER
RIJKSUNIVERSITEIT
UTRECHT



VERENIGING VOOR ONDERWIJSRESEARCH

Artikelen

De toepassing van het weddenschapsmodel in een curriculum-evaluatie-onderzoek. <i>G. van den Berg</i>	196
Open en meerkeuzevragen Tekstbegrip: een onderzoek naar de relatie tussen prestatie op open en meerkeuze-vragen. <i>Huub van den Bergh</i>	304
De toepassing van een nieuw leerprocessen-model op het memoriseren van L.O.M.- en basisschoolleerlingen. <i>K.P. van den Bos, J. Kingma en H. Scheper</i>	207
Blending Behavioristic and Cognitive Approaches to the Study of Education. <i>Hans F.M. Crombag</i>	17
Rapportage van Parameters uit het Raschmodel. <i>Theo J.H.M. Eggen en Henk Kelderman</i>	121
On the Robustness of the "Minimum-Chi-Square" Method for the Rasch Model. <i>Dato N.M. de Gruijter</i>	225
Een verklaringsmodel voor onderwijsproblemen van promotie-assistenten. <i>J.F.M.J. van Hout en J.J.M. Schmeets</i>	257
Vocaalsubstitutiefouten bij bekende en onbekende eenlettergrepige woorden. <i>Margo G.H. Jansen en Anja W.M. Knuver</i>	185
De multiple-choice cloze-toets als algemene taalvaardigheidstoets. <i>Wim Jochems en Frans Montens</i>	133
Het leren van technische handvaardigheid. <i>L.F.W. de Klerk en P.G.F. Spit</i>	63
Het schatten van schooleffecten. Problemen bij de keuze van het model en de techniek. <i>G.G. Kreft en J. de Leeuw</i>	75
Richting van de school en schoolloopbanen in het voortgezet onderwijs. <i>P. van Laarhoven, B. Bakker, J. Dronkers en H. Schijf</i>	23
Group Effects on the Development of School Subject Motivation and Achievement. <i>Ton Mooij</i>	86
Beoordeling van de Predictieve Validiteit van een Schoolvorderingentoets door Gebruikers. <i>M. Nijse</i>	237
De Betekenis van Leren: een exporatief onderzoek naar Leerconcepties met de Semantische Differentiaal. <i>M.G.E. Pierar</i>	144
Het Rasch model: naar een alternatieve wijze van itemselectie. <i>Diana M.P. Smedts</i>	355
Interpretatie en anticipatie in the fysiotherapeutisch probleemoplossen. <i>A.A.W. Smits en P.C. van der Sijde</i>	273
Effectief gebruik van rekenmethoden: evaluatieonderzoek in het laatste jaar van de basisschool. <i>C. Suhre</i>	344
Convergentie en determinanten van beoordelingen van onderzoeksaanvragen voor en na de commissiebespreking. <i>T.H.A. van der Voort, B. Pangemanann en M.W. Vooijs</i>	285
Bekroond ORD-paper 1987: Overleving in de bovenstroom van het voortgezet onderwijs: Veranderde milieu-effecten voor jongens en meisjes. <i>Paula Willemse</i>	329
De perceptie van de leraar-leerlingrelatie; constructie en kenmerken van een instrument. <i>Th. Wubbels, H.A. Créton, J.M.G. Brekelmans en H.P. Hooymayers</i>	3

Notities en Commentaren

Beoordeling van het eindrapport "Vergelijking van rekenmethoden in het basisonderwijs" uitgevoerd door het RION (Projectleider Drs. E.G. Harskamp)	239
Beoordeling van het eindrapport "Vergelijking van taalmethoden in het basisonderwijs" uitgevoerd door het RION (Projectleider Drs. G. van den Berg)	243
Taalmethoden onderzocht: reactie op het SVO-beoordelingsrapport. <i>G. van den Berg en Th. A. van Batenburg</i>	247
De betrouwbaarheid van drie discriminatie-indices. <i>Henk J.M. van Berkel</i>	365
Commentaar op Pierar "De betekenis van Leren". <i>Pieter M. Kroonenberg</i>	160
Rekenmethoden: verwacht geen wonderen. Reactie op de SVO-beoordeling van het rekenonderzoek. <i>C. Suhre</i>	249
Semantische structuurkenmerken, cognitieve schemata en aanschouwelijke modellen bij aanvankelijke redactie-opgaven. <i>L. Verschaffel</i>	165
De samenstelling van de schoolbevolking in het basisonderwijs. <i>Johan M. Wijnstra</i>	99

Boekbesprekingen

H. Blok. <i>Taal voor alledag. Feiten en meningen over het taalgebruik van Ibo- en mavo-leerlingen in alledaagse situaties.</i> (Ton Mooij)	372
H. Brutsaert. <i>Gelijke kansen en leerlinggerichtheid in het secundair onderwijs.</i> (Roel Basker en Peter van Laarhoven)	369
H.A.M. Franssen. <i>Plannen van onderwijsleersituaties als Teamaktiviteit.</i> (G. Bergenhenegouwen)	47
F.J.G. Janssens. <i>De evaluatiepraktijken van leerkrachten.</i> (S.J. Pijl)	41
M.J. de Jong (red.). <i>Allochtone kinderen op Nederlandse scholen.</i> (H.G. Dors)	43
Ontvangen publikaties	173, 251

Redactionele medewerkers

In 1987 werd redactionele medewerking verleend door:

J.J. Beishuizen	VU Amsterdam
M. Beishuizen	RU Leiden
G. Beukhof	Universiteit Twente
A.G. Bus	RU Groningen
Th.J. ten Cate	UvAmsterdam
R.J.H. Engelen	Universiteit Twente
H.A.M. Franssen	RU Utrecht
W. van de Grift	SCO Amsterdam
A.D. de Groot	
J. Hartog	UvAmsterdam
W.Th.J.G. Hoebe	RION Groningen
W.K.B. Hofstee	RU Groningen
E. van Imhoff	EU Rotterdam
G.G.H. Jansen	RU Groningen
F.J.G. Janssens	CITO Arnhem
M. van der Kamp	RU Groningen
G. Kanselaar	RU Utrecht
K.B. Koster	RU Groningen
E.J.J. Kremers	CITO Arnhem
A. van der Leij	VU Amsterdam
J.M.G. Leune	EU Rotterdam
F. Luteijn	RU Groningen
G.J. Mellenbergh	UvAmsterdam
W.J. Nijhof	Universiteit Twente
S. Oosterloo	Universiteit Twente
G.R.E. Ouweneel	RU Groningen
J.L. Peschar	RU Groningen
V. Peters	KU Nijmegen
Y.J. Pijl	RION Groningen
R. Popping	RU Groningen
P. Reitsma	VU Amsterdam
G. Rijlaarsdam	SCO Amsterdam
E. Roede	SCO Amsterdam
P. Roeders	RION Groningen
J. Scheerens	Universiteit Twente
J.H. Slavenburg	SAD Rotterdam
K.M. Stokking	RU Utrecht
P.G. Swanborn	RU Utrecht
J. Terwel	RU Utrecht
P. Tesser	ITS Nijmegen
H.H. Tillema	RION Groningen
N. Veldhuijzen	CITO Arnhem
L. Verschaffel	KU Leuven
H. Verstralen	CITO Arnhem

VOR:	
Verslag VOR-ledenvergadering d.d. 13 november 1986	46
Van de bestuurstafel	106
Convocatie ledenvergadering VOR	106
Beknopt verslag Verenigingsjaar 1986	110
Van de bestuurstafel	176
VOR nieuws	314
Verslag VOR ledenvergadering 26 mei 1987	315
De VOR-prijs 1988	379
Artikelen:	
Over de kwaliteit van de arbeid in de wetenschap <i>Floris van den Berg van Saparoea</i>	317
De kwaliteit van de arbeid van assistenten-in-opleiding <i>Ellen van Bommen en Ellen Hasselerharm</i>	381
Problemen verkennen of problemen maken? De rol van de probleemverkenner in het programmeringsproces van de S.V.O. <i>P. van den Dool</i>	48
Ook bij programmering van onderzoek zijn korte communicatielijnen effectiever <i>D. de Jong</i>	115
De OTG's en de programmering van onderzoek <i>E. Roede</i>	52
Congressen, Symposia, Cursussen, Trainingen	57, 117, 177, 254, 323, 380
Mededelingen	56, 178, 255, 314
ORD:	
Themaplanning voor de Onderwijsresearch dagen	112
Mededelingen organisatie ORD 1987	113
De ORD in het buitenland	376
Thema's op de Onderwijsresearchdagen 1988	376
Personalia:	
In memoriam Co van Calcar	104
In memoriam Hildo Wesdorp	316
Proefschriften	180, 322, 384
Redactioneel	46, 176, 314
Nieuwe publicaties	183, 256, 324, 387

Van de redactie

Redactie en uitgever kijken op 1986 wat het Tijdschrift voor Onderwijsresearch betreft met tevredenheid terug. We zijn er in geslaagd een evenwichtige verspreiding van de uitgaven van het tijdschrift te realiseren, ook al is de overeenkomst met verschijningsdatum en voorgenomen planning meestal toevallig. Nog belangrijker is echter de inhoud van het tijdschrift. Door verandering in het kopij-aanbod is een beter evenwicht ontstaan tussen artikelen die handelen over statistisch-technische en methodologische problemen enerzijds en verslagen en reviews van onderwijsonderzoek anderzijds.

Bijdragen over psychometrische problemen en oplossingen zijn overigens van belang voor de verhoging van de kwaliteit van het onderwijsonderzoek. Als wetenschappelijk tijdschrift van de Vereniging voor Onderwijsresearch (VOR) is het tijdschrift er echter ook op gericht wetenschappelijke verslagen van onderwijsonderzoek in Nederland te publiceren. Bij de beoordeling van manuscripten hanteert de redactie strikte kwaliteitsnormen, daarbij geadviseerd door beoordelaars die anonieme beoordelingen geven. De redactie is dan ook veel dank verschuldigd aan de beoordelaars die het afgelopen jaar hun medewerking hebben verleend. Hun oordeel is van belang voor het bewaren van de kwaliteit van het tijdschrift, maar misschien nog meer voor de auteur, omdat het bewerken, herschrijven of verbeteren van het oorspronkelijke manuscript aan de hand van de veelal heel uitvoerige aanwijzingen van de beoordelaars tot een succesvolle herbeoordeling leidt.

Graag ontvangt de redactie in de toekomst manuscripten van nieuwe auteurs op het terrein van onderwijsonderzoek, die we op de hierboven aangegeven wijze willen begeleiden naar publikatie.

Zoals uit de gegevens, vermeld bij gepubliceerde manuscripten, blijkt, is het in principe mogelijk snel na indiening van manuscripten, zeker wanneer ze niet een al te grote omvang hebben, tot publikatie ervan te komen. Dit ondanks een herschrijfronde die meestal nodig is op basis van de redactionele beslissingen. De auteurs zelf kunnen publikatie bespoedigen door na ontvangst van de aanwijzingen voor herschrijving deze op korte termijn aan te brengen. In het verleden is gebleken dat daarmee vaak te lang wordt gewacht. Zo liggen er nu nog manuscripten uit 1984 die voor herschrijving teruggestuurd zijn, waarvan de auteurs aangeven dat ze nog steeds hiermee bezig zijn. Niet alleen wordt herschrijven dan een lijdensweg, maar ook de uitslag van een herbeoordeling wordt onzekerder.

De verminderde publication lag is ook te danken aan de uitbreiding van het aantal pagina's in het afgelopen jaar. De VOR en de redactie gaan verder in hun pogingen het volume van het tijdschrift te vergroten. Dit is echter sterk afhankelijk van zowel het aantal leden van de VOR c.q. abonnementen op het tijdschrift, als van de overige uitgaven van de VOR. Dat is dan ook de aanleiding geweest voor de VOR om tot een samenvoeging van het Tijdschrift voor Onderwijsresearch en het VOR-bulletin over te gaan. Vanaf dit nummer is het VOR-bulletin opgenomen in het Tijdschrift voor Onderwijsresearch, echter strikt daarvan onderscheiden en onder eigen redactionele verantwoordelijkheid. De rubriek Mededelingen zal in het vervolg niet meer onderdeel van het tijdschrift uitmaken, mededelingen zullen uitsluitend in het VOR-bulletin worden opgenomen. Ook hierdoor wordt het aantal pagina's voor wetenschappelijke publikaties groter.

Door de gebondenheid van het VOR-bulletin aan actualiteit is een extra verzekering ingebouwd voor het regelmatig en tijdig verschijnen van het tijdschrift.

Tenslotte is door deze samenvoeging een, door een aantal redactieleden, lang gekoesterde wens in vervulling gegaan, namelijk de vermelding van de naam van het tijdschrift, jaargang en aflevering op de rug ervan.

Ook op het personele vlak doen zich een aantal veranderingen voor.

Hans Crombag en Johan Hoogstraten zijn uit de redactie vertrokken. Hans Crombag was vanaf het begin voorzitter van de redactie. Hij vervulde die functie op een karakteristieke wijze, snel en efficiënt, onderbroken met afhankelijk van het onderwerp, sarcastische, cynische en humoristische opmerkingen en verhandelingen over onderwerpen binnen onderwijsonderzoek, universiteit, Nederland en de wereld. Door zijn benoeming bij de Rijksuniversiteit Limburg komt hij buiten de onderwijsresearch te staan. Op zijn eigen wijze spreekt hij in dit nummer nog 'het volk' toe over de richting die onderwijsonderzoek zou moeten gaan.

Johan Hoogstraten was vele jaren lid van de redactie en voerde gedurende enige tijd met Rinus Voeten het secretariaat. Hij was, mede wellicht op grond van zijn eigen onderzoek naar het publicatieproces, overigens terecht, kritisch in keuze van beoordelaars en de beoordeling zelf.

De redactie hoopt dat door de nieuw aangetrokken redacteuren de diversiteit en kwaliteit in de redactie wordt behouden.

De perceptie van de leraar-leerlingrelatie; constructie en kenmerken van een instrument

Th. Wubbels, H.A. Créton, J.M.G. Brekelmans en H.P. Hooymayers

Vakgroep Natuurkunde Didactiek, Ped. Did. Instituut voor de leraarsopleiding R.U. te Utrecht

ABSTRACT

In this paper we describe the construction of the Questionnaire for Interactional Teacher Behaviour (QUIT) and some of its characteristics. The questionnaire is based on the Leary model for interpersonal behaviour. This model maps interpersonal behaviour with the help of two dimensions: a proximity dimension and an influence dimension. These dimensions are then represented in a coordinate system divided into eight equal sections. In this manner eight types of interactional teacher behaviour are distinguished (e.g., leading behaviour, admonishing behaviour). The instrument consists of eight scales corresponding to the eight types of behaviour. It appears to be easy to administer the questionnaire to pupils. The reliability of the scales (test-retest and internal consistency) is good, whereas no item redundancy occurs. The pattern of correlations between scales is congruent with the Leary theory. Data that were gathered with the instrument confirm hypotheses about teacher behaviour. Confirmatory factor analyses show that both two dimensions and three dimensions may underlie the eight scales.

1. INLEIDING

In dit artikel rapporteren we over constructie en kenmerken van een instrument waarmee de leerlingen- en leraarsperceptie van de leraar-leerlingrelatie kan worden gemeten. Het belang van onderzoek naar de leraar-leerlingrelatie blijkt onder meer uit onderzoek naar problemen van beginnende zowel als ervaren leraren (b.v. Veenman 1984, Vonk 1982, Peters 1985). Docenten blijken veelvuldig problemen te ervaren in het omgaan met leerlingen hetgeen in de kwaliteit van de leraar-leerlingrelatie tot uiting komt.

Met behulp van het ontwikkelde instrument wordt het relationele aspect van het leraarsgedrag in kaart gebracht. Het relationele gedragsaspect kan op verschillende manieren worden bestudeerd, bijvoorbeeld via een intrapsychische benadering vanuit de persoon van de leraar, of ontwikkelingspsychologisch vanuit de leerling. In dit onderzoek kiezen we een interpersoonlijke benadering op basis van de systeemtheorie betreffende menselijke communicatie van Watzlawick, Beavin en Jackson (1970). Daarbij beschouwen we de klas als een communicatief systeem, waarin het gedrag van de leraar enerzijds het interactiepatroon in de klas (dat is de wijze waarop leraar en leerlingen zowel verbaal als non-verbaal op elkaar reageren) beïnvloedt, maar daardoor anderzijds ook wordt beïnvloed. We bestuderen daarmee wat we noemen 'het interactionele aspect van het leraarsgedrag'. Daaronder verstaan we die gedragsaspecten, die verwijzen naar de relatie tussen de leraar en zijn leerlingen en die tot uiting komen in de interacties tussen de communicerende personen in de classesituatie.

In de literatuur werd geen geschikt instrument gevonden om het interactionele aspect van het leraarsgedrag in kaart te brengen (Hooymayers, Créton en Wubbels 1982). Daarom is door ons een leraren- en leerlingenvragenlijst ontwikkeld op basis van het door Hooymayers e.a. (1982) ontwikkelde model voor interactioneel leraarsgedrag. Het model voor interactioneel leraarsgedrag is een bewerking van het model van Leary (1957) voor interpersoonlijk gedrag. Dit model ordent leraarsgedragingen met behulp van een machtsdimensie (Boven-Onder) en een nabijheidsdimensie (Samen-Tegen). Elders hebben we laten zien dat deze keuze voor twee dimensies empirisch en theoretisch goed onderbouwd is (Hooymayers e.a. 1982, Créton en Wubbels 1984).

blz. 71). Weliswaar wordt soms een derde dimensie verondersteld (b.v. Bales 1970) maar deze blijkt ook te worden aangetroffen in instrumenten die op het model van Leary zijn gebaseerd zodat deze instrumenten ten aanzien van de dimensies van interpersoonlijk gedrag waarschijnlijk wel een adequate beschrijving geven (zie Beekers en Halfens 1981).

De dimensies worden in een tweedimensionaal assenstelsel weergegeven. Door in het assenstelsel met behulp van twee lijnen de vier kwadranten te halveren ontstaan acht sectoren, die net als in een windroos worden benoemd met BS, SB etc. (fig. 1). Tot de BS-sector van het model bijvoorbeeld behoort gedrag met een sterk Boven-karakter (B) en een licht Samen-karakter (S), aangeduid als leidend gedrag. Daartoe wordt gedrag gerekend als: zien wat er gebeurt, zelfverzekerd optreden, structureren, boeiend en enthousiast vertellen. Het naastliggende gebied (SB) omvat gedrag dat een minder sterk Boven- en een sterker Samen-karakter heeft, aangeduid als vriendelijk/helpend gedrag. Daartoe wordt gedrag gerekend als belangstelling tonen, vertrouwen bieden, helpen, zich redelijk opstellen en tegen een grapje kunnen.

Teneinde een instrument te verkrijgen om interactioneel leraarsgedrag zoals dat door leerlingen en leraren wordt waargenomen in kaart te brengen hebben we de *Interpersonal Check-List* (ICL, Laforge en Suczek 1955), die bij het model van Leary behoort, bewerkt voor gebruik in de classesituatie. De ICL zelf was daarvoor niet geschikt omdat hij vele items bevat die niet relevant zijn om leraarsgedrag in de klas te beschrijven of die door leerlingen niet als zodanig herkend worden (b.v. wil geleid worden). Bovendien is de gerapporteerde interne consistentie van de Nederlandstalige versie niet altijd hoog (Knippenberg 1978, Beekers en Halfens 1981), terwijl het aantal items te groot is voor afname in één lesuur.

2. ONTWIKKELING VAN DE VRAGENLIJST VOOR INTERACTIONEEL LERAARSGEDRAG (VIL)

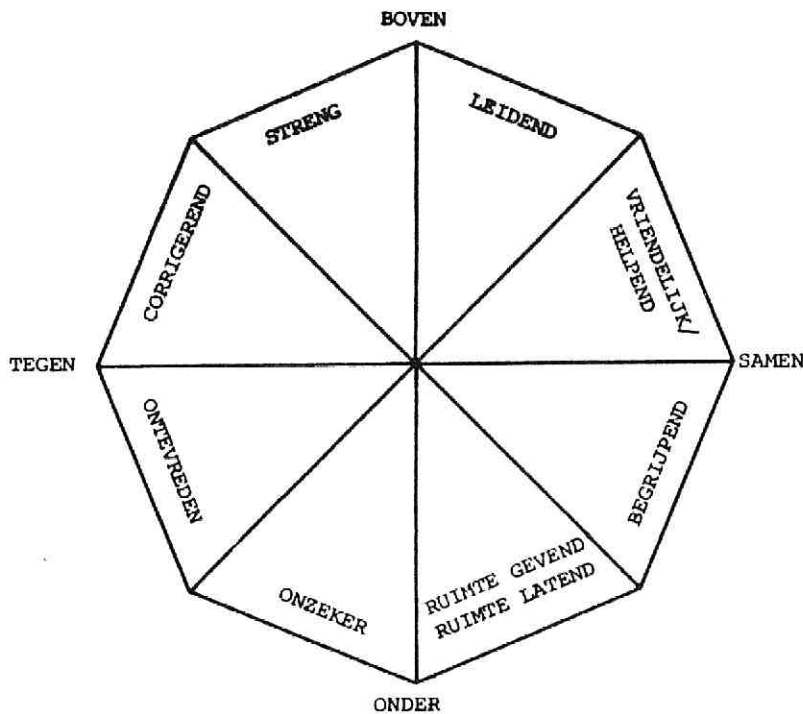
2.1 Werkwijze

Als basis voor de bewerking werd een pool van ongeveer 200 items gevormd uit de items van de door Diekstra (1972) vertaalde ICL, van het instrument voor beschrijvingsfeedback van het NIVOR (van Wijk 1979, blz. 103 e.v.) en uit nieuw geformuleerde items. Gezocht werd naar ongeveer 25 items per subschaal van het instrument waarbij elke subschaal overeenkwam met één van de sectoren van het model. Al deze items werden zodanig herschreven of door synoniemen vervangen dat ze ons begrijpelijk voor leerlingen leken en dat ze van toepassing waren op leraarsgedrag in de klas. Deze items werden vervolgens door twee lerarenopleiders beoordeeld op de mate waarin ze leraarspecifiek gedrag beschreven en op de mate waarin ze gesteld waren in taal die voor leerlingen begrijpelijk en gebruikelijk is. Op grond van deze beoordeling werden 160 items, soms in gewijzigde vorm, geselecteerd. Deze items werden in interviews voorgelegd aan 6 individuele leerlingen. Hun werd gevraagd een leraar in gedachten te nemen en voor elk item aan te geven of ze het betreffende gedrag op de leraar van toepassing achtten. Verder werd hun van elk item gevraagd of ze het begrepen en of ze in eigen woorden wilden weergeven wat ze zich bij het item voorstelden. Op deze wijze werden nog eens ongeveer 40 nieuwe of gewijzigde items verkregen.

Uit de zo ontstane pool werd een selectie gemaakt van 120 items, die beantwoord moesten worden op een vierpuntsschaal variërend van 'niet van toepassing' tot 'sterk van toepassing'.

De vragenlijst die zo ontstond werd voorgelegd aan 601 leerlingen van 29 leraren, waarbij hun gevraagd werd de items te beantwoorden voor de leraar waarvan zij op het moment van afname les hadden. De 8 subschalen overeenkomend met de 8 sectoren van het model waren door ons op inhoudelijke gronden samengesteld en bestonden elk uit 15 items. Er werden subschaalscores berekend als som van itemscores. Verder werden de itemrestcorrelaties, de correlaties tussen elk item en elke subschaal, de frequentieverdeling van de antwoordscores bij elk item en de interne consistentie voor de subschalen berekend.

Op basis van deze gegevens werden de subschalen met de beschikbare items geoptimaliseerd, door sommige items te verwijderen uit de vragenlijst en andere te verplaatsen naar een andere



Figuur 1: Het model voor interactioneel leraarsgedrag met in elke sector een globale karakteristiek van het betreffende gedrag.

subschaal. De daarbij gehanteerde criteria worden in paragraaf 2.2 besproken. Teneinde de vragenlijst verder te verbeteren werden enkele items opnieuw geformuleerd zodat een nieuwe proefversie ontstond. Deze versie werd op analoge wijze uitgetoetst, geanalyseerd en geoptimaliseerd. Aldus werden in totaal vier proefversies gemaakt voordat de definitieve versie van de VIL ontstond.

2.2 Criteria voor de beoordeling van items en subschalen

De kwaliteit van de ruwe en geoptimaliseerde versies van de subschalen werd aan de hand van 4 criteria beoordeeld.

Overeenstemming tussen inhoudelijke betekenis en psychometrische gegevens

Items zijn *à priori* ingedeeld in subschalen op inhoudelijke gronden. Dat wil zeggen dat het Boven-Onder karakter en het Samen-Tegen karakter van een item volgens de onderzoekers overeenkwam met het karakter van de subschaal. Aan de hand van de gegevens over item-subschaalcorrelaties werd vastgesteld in hoeverre de items ook op psychometrische gronden in de betreffende subschaal pasten. Daarvoor is nodig dat de item-restcorrelatie hoger is dan de correlatie met andere subschalen en dat de correlatie met de subschaal die er in het model tegenover ligt het laagst (dus het sterkst negatief) is.

Bij de uitgetoetste items kwamen drie typen afwijkingen voor van dit gewenste patroon. In de eerste plaats correleerden items soms hoger met een andere subschaal dan de door ons *à priori* gekozen schaal. Wanneer daartegen geen inhoudelijke bezwaren waren werd zo'n item verplaatst. Dit kwam alleen voor bij verplaatsing naar een naastliggende sector. Een voorbeeld is het item 'treedt zelfverzekerd op' dat werd verplaatst van de BT-schaal naar de BS-schaal. Voor sommige andere items was verplaatsing niet mogelijk. Zo gold dat een item als 'overdreven vriendelijk' niet in een andere dan die subschaal geplaatst kon worden waar ook 'vriendelijk' toe behoort (SB). Dit item correleerde echter het hoogst met de OT-subschaal en werd daarom uit de vragenlijst verwijderd. In de tweede plaats bleken enkele items weinig specifiek

tot een bepaalde subschaal te behoren omdat de correlaties van het item met alle subschalen laag was. Tenslotte waren er items die weliswaar hoger met sommige subschalen correleerden dan met andere, maar waarbij er twee of meer subschalen waren waartoe het item op grond van de correlaties vrijwel even sterk kon worden gerekend. Teneinde het patroon van item-subschaalcorrelaties te verbeteren werden voor elke proefversie wijzigingen in de formulering van sommige items aangebracht. Zo werd het item 'geeft je vrijheid' dat bedoeld was voor de OS-subschaal gewijzigd in 'je hebt bij hem vrijheid'.

Aansluiting bij de structuur van de ICL

Net als bij de ICL zou er in elke subschaal minstens één item moeten zijn dat aangepast gedrag beschreef en minstens één item dat onaangepast gedrag beschreef. Achtergrond van dit criterium is de gedachte achter het model van Leary dat elk type interpersoonlijk gedrag onaangepast en aangepast kan zijn onder meer afhankelijk van de intensiteit waarmee het vertoond wordt. Het bleek voor verscheidene sectoren erg moeilijk om aan dit criterium te voldoen. Items die gedrag beschrijven dat velen voor leraren wenselijk achten en bedoeld waren voor Tegen-schalen, bleken in het algemeen positief te correleren met subschalen aan de Samenkant van het model (speciaal de BS-, SB- en SO-schaal). Anderzijds correleren items bedoeld voor de Samen-schalen die voor leraren minder sociaal wenselijk gedrag beschrijven positief met subschalen aan de Tegenkant van het model. Voorbeelden zijn het al genoemde item 'is overdreven vriendelijk' bedoeld voor de SB-subschaal en het voor de OT-subschaal bedoelde item 'geeft toe als hij ongelijk heeft', dat het hoogst met de SO-subschaal correleerde.

Hoge interne consistentie van subschalen

De interne consistentie van de subschalen (Cronbachs α) zou zo hoog mogelijk moeten zijn (boven de .75) zonder dat dit ten koste zou gaan van de rijkheid aan inhoud van een subschaal. Deze rijkheid kan verloren gaan doordat er veel synoniemen in een subschaal staan. Er is dan sprake van redundantie aan items. Er is bij de revisies geprobeerd redundantie te beperken door geen correlaties tussen items van dezelfde subschaal hoger dan .70 toe te laten (verg. Pettegrew en Wolf, 1982). Verder is er naar gestreefd in subschalen zoveel mogelijk de nuances die in de ICL voorkwamen in de vragenlijst op te nemen. Zo werd er naar gestreefd in de SB-sector zowel items te plaatsen die het vriendelijke aspect verwoordden, als items die het helpende aspect benadrukken.

Discriminerend vermogen

De items en subschalen moeten discrimineren tussen leraren. Dit werd gezien door itemscores en scores op subschalen en de standaarddeviaties te berekenen voor alle leraren. Gestreefd werd naar items waarbij tussen leraren minstens een verschil van 1.0 in score voorkwam en waarbij subschaalscores minstens ($0.7 \times$ het aantal items) tussen leraren verschilden. De betreffende standaarddeviaties moesten voor items 0.9 en voor subschaalscores ($0.6 \times$ het aantal items) zijn.

2.3 Antwoordschaal

Bij het uitproberen van items voor de eerste proefversie bleek, dat de beantwoording met een vierpuntsschaal niet goed voldeed. Deze antwoordwijze, die aansluit bij de bij de ICL gebruikelijke ja/nee schaal bleek bij veel respondenten weerstanden op te roepen, omdat men zich gedwongen voelde tot een keuze tussen ja of nee, de rechter of linker helft van de schaal. Hierdoor raakten respondenten geïrriteerd. De behoefte aan een meer neutrale antwoordmogelijkheid werd uitgesproken. Er werd besloten een verdere nuancering in de beantwoording mogelijk te maken door een vijfpuntsschaal te gebruiken. Met het oog op de leeftijd van de respondenten (13-16 jaar) werd geen zevenpuntsschaal gebruikt. Een tweede bezwaar bleek bij de teruggave van resultaten van de afname van de vragenlijst aan leraren over wie de vragenlijst was ingevuld. Zij vonden de formulering in de vorm 'legt duidelijk uit, sterk van toepassing' niet helder. Zij wilden liever dat de vragenlijst zodanig werd afgenomen dat het resultaat in de vorm 'legt erg duidelijk uit' teruggegeven kon worden.

Besloten werd afhankelijk van de formulering van de stam aan elk item één van de volgende drie antwoordschalen toe te voegen:

	1	2	3	4	5
schaal A	geen	een beetje		veel	zeer veel
schaal B	nooit	soms		vaak	altijd
schaal C	niet	enigszins		erg	heel erg

3. KENMERKEN VAN DE RESULTERENDE VRAGENLIJST

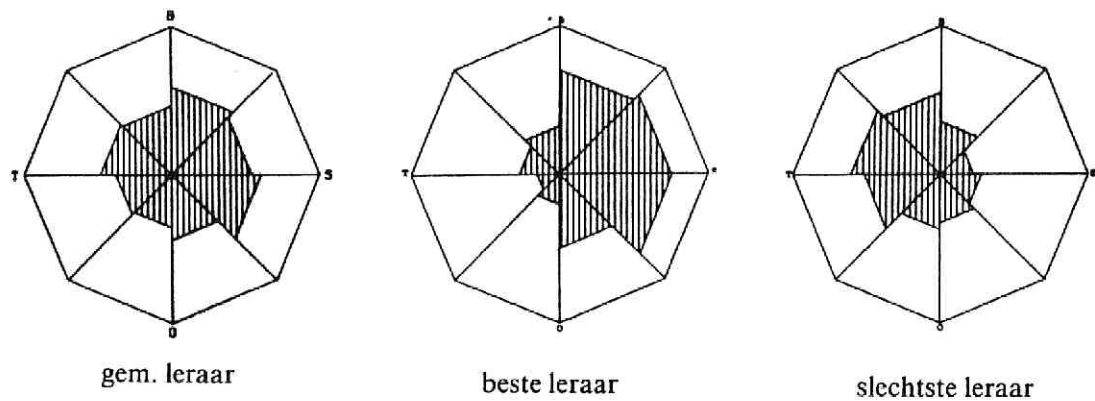
3.1 Inleiding

Nadat vier proefversies waren afgenomen bleken toch niet alle items en subschalen geheel positief beoordeeld te kunnen worden. Tien items bleken (iets) lager te correleren met de eigen subschaal dan met een naastliggende subschaal. Bij 8 van deze items was het verschil tussen de itemrestcorrelatie en de correlatie met de naastliggende subschaal niet hoger dan .03 en dus verwaarloosbaar¹. Voor de overige twee waren de verschillen groter. Het item 'verontschuldigt zich' werd niet verplaatst naar de SO-sector omdat dit volgens ons een OT-item is, terwijl het bovendien het enige in termen van aangepast gedrag geformuleerde item was dat in de OT-sector was overgebleven. Verder werd in de BS-sector het item 'is zelf aan het woord' gehandhaafd terwijl het hoger correleerde met de BT-subschaal. Dit item was echter het enige dat door combinatie met de antwoordschaal nog een 'onaangepast' item kon opleveren. Uiteindelijk ontstond dus de definitieve versie van de vragenlijst voor interactioneel leraarsgedrag, die bestaat uit 77 items verdeeld over 8 subschalen².

De ontwikkelde vragenlijst is in twee onderzoeken voorgelegd aan leerlingen en leraren. Het eerste onderzoek betrof 2407 leerlingen en 91 leraren van zes scholen. De gegevensverzameling vond plaats bij leraren van scholen waarmee al contacten bestonden en bij leraren die bereid waren daarvoor lestijd beschikbaar te stellen. Het betreft dus geen random steekproef. De leraren zijn afkomstig van MAVO-, HAVO- en VWO-scholen zowel van de openbare, rooms-katholieke als protestants-christelijke denominatie. Eén van de scholen was een montessorischool. De diversiteit van de scholen was dus redelijk groot. Aan leerlingen werd gevraagd de lijst te beantwoorden voor de leraar waarvan ze op het tijdstip van invullen les hadden en aan een aantal leerlingen ook om deze in te vullen voor de beste en voor de slechtste leraar waarvan ze les hadden. Aan leraren werd de vragenlijst voorgelegd met de vraag deze te beantwoorden voor het eigen gedrag in een bepaalde klas.

Het tweede onderzoek vond plaats in het kader van een Nederlandse optie van de Second International Science Study³. Gegevens werden verzameld bij 44 random gekozen natuurkundeleraren en leerlingen uit één derde klas van die leraren. In dit onderzoek werd alleen gevraagd de VIL in te vullen voor respectievelijk het eigen gedrag (leraren) en het gedrag van de leraar (leerlingen).

Bij de verwerking van de vragenlijsten is naast de itemscores steeds gebruik gemaakt van daaruit berekende subschaalscores. De som van de itemscores per subschaal werd omgezet in subschaalscores variërend van 0.0 tot 1.0 met behulp van een lineaire transformatie. De resultaten van een vragenlijst-afname kunnen behalve in item- en subschaalscores ook worden weergegeven door een figuur. In een in 8 sectoren verdeeld assenstelsel wordt dan per sector een gedeelte gearceerd, waarvan de straal een lineaire maat is voor de berekende score. Figuur 2 geeft als voorbeeld de gemiddelde scores van leerlingen over de 44 natuurkundeleraren uit de random steekproef. Ter vergelijking geven we ook de resultaten betreffende de gemiddelde beste en slechtste leraar in leerlingenogen. Uit de metingen met het instrument kunnen enkele conclusies over de hanteerbaarheid, de betrouwbaarheid en de validiteit worden getrokken die we achtereenvolgens zullen bespreken.



Figuur 2: Gemiddeld beeld van leerlingen van het interactioneel gedrag van 44 natuurkunde leraren en van de beste en slechtste leraar.

3.2 Praktische hanteerbaarheid

De hanteerbaarheid van de vragenlijst blijkt bevredigend. De vragenlijst kan door een leraar zelf zonder specifieke instructies in een klas worden afgenomen. Leerlingen hebben 20 (5e klas Atheneum) à 35 minuten (2e klas Mavo) nodig om de lijst te beantwoorden. Bij de afname in tweede klassen is gebleken, dat MAVO-, HAVO- en VWO-leerlingen (in die volgorde in afnemende mate) vragen stellen over de betekenis van enkele items. Een korte uitleg blijkt deze onduidelijkheden op te kunnen lossen.

3.3 Betrouwbaarheid

Over de betrouwbaarheid zijn gegevens verzameld betreffende de test-hertest betrouwbaarheid, de interne consistentie en de mate waarin de VIL onderscheid tussen leraren maakt.

We hebben de VIL met een tussentijd van 4 weken voorgelegd aan 80 leerlingen over 4 ervaren leraren, waarvan verondersteld mag worden dat ze een redelijk stabiel patroon van interactioneel gedrag vertonen. Tabel 1 geeft in kolom 3 de resultaten. Voor de meeste subschalen van de VIL is de test-hertest betrouwbaarheid redelijk. Voor de TO-schaal en in mindere mate de OT- en OS-schaal is deze het laagst.

Tabel 1 geeft in kolom 4 t/m 8 en kolom 10 en 11 ook de interne consistentie van de subschalen van de VIL bij afname aan leraren over zichzelf en aan leerlingen over een leraar waarvan ze op dat moment toevallig les hadden en over de in hun ogen beste en slechtste leraar. Aangezien de VIL vaak gebruikt zal worden om een beeld van het leraarsgedrag te krijgen door een gemiddelde over een gehele klas te nemen is ook de interne consistentie op klassenniveau (gebaseerd op scores gemiddeld over een klas) vermeld. De resultaten uit het eerste onderzoek (kolom 4 en 5) en uit het tweede (kolom 10 en 11) blijken vooral voor de leerlinggegevens goed vergelijkbaar. Dit geldt ook voor alle verdere analyses. Daarom worden alleen die van het eerste onderzoek verder vermeld. We kunnen constateren dat de betrouwbaarheid van de VIL, afgemeten aan de interne consistentie van subschalen, gunstig is. De waarden voor gebruik bij leraren over zichzelf en bij leerlingen over hun leraren in het algemeen en over de slechtste leraar liggen in de meeste gevallen niet ver onder de grens van .80, die door Nunnally (1967, blz. 226) wordt genoemd als grens waarboven voor researchdoeleinden geen verhoging van de interne consistentie meer nodig is. Voor de gegevens betreffende de beste leraar zijn de interne consistenties iets minder gunstig. Ze liggen echter ruim boven de waarde van .60 die Nunnally voor researchdoeleinden al vaak bruikbaar acht. Wat betreft de interne consistentie op klassenniveau kan worden opgemerkt dat deze voor alle subschalen ligt boven de minimumgrens van .90, die Nunnally noemt voor situaties waarin belangrijke beslissingen worden genomen op grond van testcores.

Tabel 1: Aantal items, test-hertest betrouwbaarheid, interne consistentie en betrouwbaarheid van klassegemiddelden van de definitieve versie van de VIL bij afname aan leerlingen en leraren.

sector	aantal items	test-hertest betrouwbaarheid	VIL (1e onderzoek)					VIL (2e onderzoek)	
			interne consistentie			betrouwbaarheid klassegemiddelden	leraren	interne consistentie	leerlingen
			leraren	leerlingen	beste				
			11.niveau n=91(lr) n=2407(11n)	klassenniveau n=109(klassen)	slechtste n=341(11n)	beste n=357(11n)			
BS	10	.80	.89	.85	.78	.72	.96	.80	.84
SB	10	.84	.76	.87	.87	.81	.95	.84	.89
SO	10	.80	.77	.86	.87	.79	.95	.83	.88
OS	9	.70	.87	.79	.84	.75	.96	.72	.74
OT	9	.69	.90	.87	.88	.72	.97	.82	.82
TO	11	.65	.86	.84	.86	.78	.94	.84	.82
TB	9	.76	.78	.80	.85	.79	.94	.78	.79
BT	9	.74	.80	.77	.81	.78	.94	.69	.79

De keerzijde van een hoge interne consistentie kan zijn dat de opgenomen items slechts een geringe variatie in inhoud laten zien. De interne consistentie kan men dus verhogen door alleen items met hoge onderlinge correlaties toe te laten. Om na te gaan in hoeverre dit het geval is hebben we de correlaties tussen items binnen een subschaal geanalyseerd. Deze correlaties variëren voor de leerlingengegevens van .01 tot .74 en liggen op 3 na onder de .70, zodat er op enkele uitzonderingen na geen sprake is van item-redundantie.

De vragenlijst kan gebruikt worden om met behulp van klassegemiddelden uitspraken over het leraarsgedrag in de ogen van leerlingen te doen. Daarom is ook de betrouwbaarheid van klassegemiddelden op de 8 subschalen berekend met behulp van een formule van Horst (1949). Deze betrouwbaarheden geven een indicatie van de mate waarin subschalen discrimineren tussen klassen in verhouding tot de verschillen binnen klassen. Uit tabel 1 (kolom 9) blijkt dat deze betrouwbaarheden zeer gunstig zijn. Ook uit variantieanalyse, uitgevoerd op subschaalscores, blijkt dat het leraarseffect vergeleken met het leerlingeffect zeer aanzienlijk is. Beschouwd werden 109 groepen leerlingen ($n=2407$) die als klas over een leraar de vragenlijst hadden ingevuld⁴. Per subschaal wordt 48 à 62% van de variantie verklaard door de factor leraar.

3.4 Validiteit

Aanwijzingen over de validiteit van de VIL zijn verkregen uit analyse van de correlaties tussen subschalen, uit de gemiddelde subschaalscores betreffende verschillende categorieën leraren en uit resultaten van confirmatieve factoranalyses. We bespreken deze aanwijzingen achtereenvolgens.

Correlaties tussen subschalen

Volgens het model van Leary vertonen gedragingen naarmate ze in het assenstelsel dichter bij elkaar liggen meer gelijkenis. Op grond van de circulaire ordening mag verwacht worden dat correlaties tussen subschaalscores lager zijn naarmate de subschalen in het model verder van elkaar liggen. Tabel 2 geeft de correlaties van de subschaalscores onderling bij afname aan 2407

Tabel 2: Correlaties tussen subschalen bij leerlingen ($n = 2407$, boven de diagonaal) en lerarengegevens ($n = 91$, onder de diagonaal).

	BS	SB	SO	OS	OT	TO	TB	BT
BS	–	.61	.50	–.12	–.72	–.48	–.33	.02
SB	.48	–	.86	.38	–.34	–.68	–.60	–.42
SO	.35	.76	–	.44	–.23	–.69	–.63	–.49
OS	–.41	.09	.30	–	.34	–.24	–.33	–.48
OT	–.72	–.37	–.15	.52	–	.44	.29	–.03
TO	–.40	–.47	–.45	–.08	.49	–	.76	.53
TB	–.17	–.44	–.57	–.40	.15	.60	–	.58
BT	.34	–.19	–.29	–.64	–.19	.44	.54	–

Tabel 3: Gemiddelde correlaties tussen sectoren bij verschillende sectorafstanden.

Aantal tussenliggende sectoren	11n. $n = 2407$	1rn. $n = 91$
0	.51	.50
1	.06	.05
2	–.29	–.37
3	–.63	–.60

leerlingen respectievelijk 91 leraren. Tabel 3 geeft de gemiddelde correlatie tussen 2 sectoren voor verschillende sectorafstanden.

In de tabellen valt in de eerste plaats de over het algemeen redelijke overeenkomst op tussen correlatiecoëfficiënten verkregen uit leerlingengegevens en uit lerarengegevens. Uit tabel 3 kan worden afgelezen dat de afname van correlaties met de sectorafstand redelijk gelijkmatig is, hetgeen de circulaire ordening van subschalen bevestigt. Ook in tabel 2 wordt bevestiging van deze ordening gevonden, hoewel er enkele ongerechtigheden zijn: de BS- en BT-sector correleren onderling niet hoog genoeg en niet altijd is de correlatie van een sector het laagst met de recht tegenoverliggende. Bij circulaire ordening zouden voor elke subschaal de correlaties met andere subschalen afhankelijk van de afstand tot die subschaal eerst afnemen en dan weer toenemen. Dit blijkt in ons onderzoek voor de gegevens verkregen met de VIL bij zowel leerlingen als bij leraren zo te zijn. Dit resultaat vormt een empirische ondersteuning van de theorie rond het circulaire model van Leary en voor de adequaatheid van de VIL als representant van dit model.

Subschaalscores bij verschillende categorieën leraren

De verzamelde gegevens over het beeld dat leerlingen gemiddeld hebben van leraren bevestigen naar onze mening in het algemeen bestaande veronderstellingen over de wijze waarop leraren zich gedragen. Wij geven daarvan een voorbeeld en verwijzen voor verdere gegevens naar Créton en Wubbels (1984). Gezien de aard van het leraarsberoep mag verwacht worden, dat leraren meer gedrag vertonen uit de BS-, SB- en SO-sector van het model van Leary dan gedrag uit de OT- en TO-sector. Bovendien mag worden verwacht dat degenen die volgens hun leerlingen zeer goede leraren zijn aanzienlijk meer gedrag uit de BS-, SB- en SO-sector vertonen dan leraren die door hun leerlingen als slechte leraren worden beschouwd. Deze verwachtingen worden ondersteund door bijvoorbeeld de resultaten betreffende effectief onderwijsgedrag van Rosenshine (1971, zie ook Brophy en Good 1986). De resultaten (figuur 2) bevestigen duidelijk de uitgesproken verwachtingen, wat kan worden opgevat als een aanwijzing voor de validiteit van de VIL.

Confirmatieve factoranalyses⁵

Aan de theorie van Leary ligt de veronderstelling ten grondslag dat de 8 sectoren van het model worden bepaald door twee factoren: macht en nabijheid. Deze veronderstelling hebben we getoetst voor onze operationalisatie van Leary's theorie met behulp van confirmatieve factoranalyse volgens de maximum likelihood methode (LISREL). De analyses werden uitgevoerd op de variantie-covariantiematrix van leerlingengegevens ($n=2407$). Voor leraren zijn onvoldoende gegevens voor een analyse beschikbaar (Boomsma, 1983).

De meest aannemelijke schattingsmethode gaat ondermeer uit van twee vooronderstellingen: de omvang van de steekproef moet zeer groot zijn en de geobserveerde variabelen moeten een multivariaat normale verdeling hebben. Blijkens het werk van Boomsma (1983) is bij ons aan de eerste voorwaarde voldaan ($n \geq 200$). Voor de tweede voorwaarde raadt hij aan (blz. 195) LISREL alleen te gebruiken wanneer de gemiddelde absolute waarde van de scheefheid van de geobserveerde variabelen kleiner is dan ongeveer 1.0. In ons geval bleek de gemiddelde scheefheid .63, zodat ook aan deze voorwaarde is voldaan. Tenslotte werd gecontroleerd in hoeverre er sprake was van lineaire verbanden tussen de gemeten variabelen zodat het berekenen van Pearson productmoment correlaties geoorloofd zou zijn. Inspectie van de 'normal probability' diagrammen voor de residuen en predictoren bracht bij drie van de 28 verbanden tussen 2 subschalen zwakke indicaties voor een afwijking van een lineair verband aan het licht. Gezien dit geringe aantal werd niet besloten tot transformatie van variabelen.

Bij de analyses werd voor verschillende modellen (veronderstellingen over relaties tussen de gemeten variabelen) nagegaan of de verzamelde gegevens in overeenstemming waren met die modellen. Gebruikelijk is om daarbij de 'chi-square goodness of fit test' te hanteren als maat voor overeenstemming met een algemeen model van gecorreleerde variabelen. Als aanvulling op die maat hebben we voor elk model ook twee 'fit indices' berekend ontleend aan Bentler en

Bonett (1980). Wanneer namelijk, zoals in ons geval, gegevens bij veel respondenten zijn verzameld zal op basis van de 'chi-square goodness of fit test' vrijwel elk model als inadequaat van de hand moeten worden gewezen: wanneer gegevens van veel respondenten zijn verzameld ben je immers in staat zelfs zeer kleine afwijkingen van een model waar te nemen in die gegevens. Bentler en Bonett stellen dat een dergelijke statistische inadequatheid van het model nog geen praktische inadequatheid hoeft te betekenen. Daarom raden zij aan om bij grote aantallen respondenten na te gaan in hoeverre een gepostuleerd model beter in overeenstemming is met de verzamelde gegevens dan een algemeen nul-model: een model waarin verondersteld wordt dat de gemeten variabelen onafhankelijk zijn. De 'fit indices' zijn een genormeerde (Δ_{kl}) en een niet-genormeerde (ρ_{kl}) index, die beide twee modellen k en l vergelijken. Hoe groter deze indices zijn hoe groter de verbetering in fit is die wordt verkregen door in plaats van model k model l te postuleren.

De indices Δ_{0l} en ρ_{0l} geven aan in hoeverre model l beter past op de gegevens dan het algemene nulmodel. Voordeel van de niet-genormeerde index boven de genormeerde is dat daarin de berekende χ^2 vergeleken wordt met het aantal vrijheidsgraden. De genormeerde index daarentegen heeft als voordeel dat zij altijd ligt tussen 0 en 1 en dat daardoor Δ_{0l} aangeeft in hoeverre praktische verbetering van het model l nog mogelijk is.

Tabel 4 geeft een overzicht van de resultaten van LISREL analyses die we in het volgende zullen bespreken.

Het model dat het meest met Leary's theorie in overeenstemming is wordt gevormd door twee onderling loodrechte factoren, waarbij twee sectoren die een gelijke hoek met de BO- of ST-as maken gelijke ladingen op de machts- of nabijheidsfactor hebben. De factorladingen voldoen dan aan het patroon van gelijkheidsrestricties uit tabel 5a. Dit model (1) past slecht op de gegevens maar betekent wel een zeer aanzienlijke verbetering ten opzichte van het algemene nulmodel. Een minder restrictief model, waarbij wordt aangenomen dat de factoren niet onderling loodrecht zijn en waarbij de grootte van de factorladingen op de twee factoren onafhankelijk van elkaar zijn gemaakt (tabel 5b) past slecht (model 2) maar de verbetering t.o.v. het algemene nulmodel is zeer groot. Een verdere verbetering is slechts mogelijk met $(1 - .84) = 0.16$. Van model 2 kan gezegd worden dat het nog behoorlijk in overeenstemming is met Leary's

Tabel 4: Overzicht van resultaten van confirmatieve factoranalyses

no. (1)	modelkenmerken	χ^2	df	p	ρ_{0l}	Δ_{0l}
0	8 onafhankelijke variabelen (algemeen nul-model)	13598	28	.000		
1	2 orthogonale factoren met gelijkheidsrestricties volgens tabel 5a	4209	24	.000	.640	.690
2	2 oblique factoren met gelijkheidsrestricties volgens tabel 5b	2124	19	.000	.771	.844
3	2 oblique factoren met 2 gelijkheidsrestricties op factorladingen	1730	13	.000	.727	.873
4	2 oblique factoren, 3 gelijkheidsrestricties, 2 correlaties residuen	951	12	.000	.839	.930
5	2 oblique factoren, 3 gelijkheidsrestricties, 13 correlaties residuen	.05	1	.83	.002	1.000
6	2 orthogonale factoren, 3 gelijkheidsrestricties, 13 correlaties residuen	24.8	2	.000	.976	.998
7	2 orthogonale factoren met gelijkheidsrestricties volgens tabel 5a, 23 correlaties residuen	1190	8	.000	.695	.912
8	3 orthogonale factoren, 2 factoren volgens tabel 5a	1206	16	.000	.847	.911
9	3 orthogonale factoren, 4 gelijkheidsrestricties	77.8	8	.000	.982	.994
10	3 orthogonale factoren, 3 gelijkheidsrestricties	77.1	7	.000	.979	.994

Tabel 5: Overzicht van gelijkheidsrestricties op factorloadingen in een tweefactormodel volgens Leary (kolom a) en met toestaan van verschillen in factorloadingen tussen de twee factoren (kolom b). De LISREL-schattingen van factorloadingen in het best passende tweefactormodel (kolom c) en in het driefactormodel (kolom d).

subschaal	(a)			(b)			(c)			(d)		
	factor 1	factor 2		factor 1	factor 2		factor 1	factor 2		factor 1	factor 2	factor 3
BS	p	q		p	t		.10	.07		-.08	.09	.13
SB	q	p		q	u		-.01	.14		-.04	.18	.03
SO	r	p		r	u		-.03	.14		-.01	.17	.01
OS	s	q		s	t		-.11	.09		.03	.09	-.08
OT	s	r		s	v		-.20	-.07		.03	-.04	-.19
TO	r	s		r	w		-.03	-.17		-.06	-.13	-.07
TB	q	s		q	w		.04	-.17		-.08	-.13	-.05
BT	p	r		p	v		.07	-.11		-.08	-.09	.02

theorie: verondersteld is in dit model dat de twee dimensies in verschillende eenheden gemeten worden en dat de dimensies niet geheel onafhankelijk zijn. De correlatie tussen de factoren blijkt .06 te zijn, hetgeen slechts een zeer kleine afwijking is van de orthogonaliteit. De resultaten betreffende de modellen 1 en 2 bevestigen toch slechts voor een deel de adequaatheid van Leary's theorie voor onze gegevens. Daarom hebben we ook andere modellen geprobeerd.

Het verder loslaten van de restricties op de factorladingen op basis van inspectie van de matrix met verschillen tussen de gemeten en geschatte variantie-covariantiematrix (verschilmatrix) leidt slechts tot marginale verbeteringen in de fit van het model. Wanneer op één van de best passende identificeerbare 2-factor modellen (slechts 3 gelijkheidsrestricties op de factorladingen) correlaties tussen de residuen worden toegelaten wordt een aanzienlijke daling in de χ^2 van de 'goodness of fit test' gevonden (model 3, 4). Het toelaten van gecorreleerde residuen is redelijk, wanneer verondersteld kan worden dat er correlaties tussen meetfouten optreden. Jöreskog en Sörbom (1979, blz. 172) achten correlaties bijvoorbeeld aannemelijk wanneer, zoals in het onderhavige geval, de instrumenten (subschalen) verwante begrippen meten. Correlaties tussen meetfouten kunnen dan optreden door toevallige kenmerken van de schaalconstructie en door geheugeneffecten. Stapsgewijs zijn steeds meer correlaties toegelaten op basis van inspectie van de verschilmatrix en de eerste-orde afgeleiden van parameterschattingen. Zodoende bleek het mogelijk te komen tot een model dat past op de data(5), waarbij bij elke stap een significante verbetering van de 'fit' optrad (afgemeten aan de χ^2 -afname vergeleken met de afname van df). Ook het oblique zijn van de factoren (correlatie .05) betekent een verbetering van de fit ten opzichte van een model met orthogonale factoren(6). Een model waarbij aan de twee factoren de voorwaarden van tabel 5a worden opgelegd blijkt, ook wanneer zoveel mogelijk correlaties tussen residuen worden toegelaten, niet te passen(7).

In het passende tweefactormodel zijn er 13 correlaties tussen residuen toegelaten van de in totaal mogelijke 28. In overeenstemming met de aanname betreffende de gecorreleerde meetfouten traden de correlaties over het algemeen op tussen residuen van subschalen die sterk laden op dezelfde factor. Tabel 5c geeft de factorlading in dit passende model. De eerste factor lijkt op een Boven-Onderdimensie en de tweede op een Samen-Tegendimensie. Hoewel het passende model enige gelijkenis vertoont met hetgeen op basis van Leary's theorie verwacht mocht worden, wordt deze theorie toch maar ten dele bevestigd. Het grote aantal gecorreleerde residuen geeft aanleiding te veronderstellen dat een model met meer factoren beter op de gegevens zal passen. Ook in exploratieve factoranalyses betreffende andere vragenlijsten, die op Leary's theorie zijn gebaseerd worden altijd 3 of meer factoren gevonden (bijv. Briar en Bieri 1963, Beckers en Halfens 1981). Daarom werd besloten ook een model met drie factoren te toetsen.

Als uitgangspunt werd een model genomen waarbij voor twee factoren het ladingenpatroon uit tabel 5a werd verondersteld, terwijl op de factorladingen van de derde factor geen restricties werden opgelegd. Dit model(8) past niet op de gegevens, maar gezien de fitindex wel aanzienlijk beter dan een tweefactormodel met ongecorreleerde residuen. Op basis van inspectie van de verschilmatrix en de eerste-orde afgeleiden van de parameterschattingen werden stapsgewijs restricties op de eerste twee factoren losgelaten hetgeen leidt tot de modellen 9 en 10. Gezien de verhouding tussen χ^2 en het aantal vrijheidsgraden en de niet-genormeerde fit index blijkt model 9 dan het beste te passen. De schattingen van de factorladingen van model 9 worden gegeven in tabel 5d. Vanuit praktisch oogpunt is de fit van dit model uitstekend te noemen: van de totaal mogelijke verbetering t.o.v. het nulmodel is 99% gerealiseerd. De resulterende driefactorstructuur kan beschreven worden als een Onder-Boven dimensie (factor 1) een Samen-Tegen dimensie (factor 2) en een factor langs de as BS-OT (factor 3). De verklaarde variantie behorende bij de factoren is respectievelijk 8,8%, 44,4% en 23,5%.

Samenvattend kunnen we stellen dat onze gegevens in overeenstemming zijn met de gedachte achter Leary's model dat aan de 8 subschalen twee dimensies ten grondslag liggen, mits correlaties tussen de meetfouten in de subschaalscores worden verondersteld. Het optreden van dergelijke correlaties is denkbaar door geheugeneffecten: response set.

Een wat minder goed, maar toch altijd nog zeer bevredigend passend model veronderstelt dat

er behalve de twee door Leary vermelde dimensies nog een derde dimensie ten grondslag ligt aan de acht subschalen van de VIL. Een dergelijk drie-factormodel is ook geopperd voor andere instrumenten gebaseerd op het model van Leary (Briar en Bieri 1963, Beekers en Halfens 1981, Knippenberg 1978). Deze auteurs analyseerden hun gegevens echter alleen met behulp van exploratieve factoranalyses. Wanneer we hun gegevens analoog aan het in het voorgaande beschrevene onderwerpen aan confirmatieve factoranalyses blijken ook daarop redelijk passende twee-factormodellen te vinden.

Teneinde meer definitief voor een twee-, dan wel drie-factormodel te kiezen zou speciaal daarop gericht onderzoek nodig zijn.

4. CONCLUSIES

Uit de voorgaande paragrafen blijkt, dat de VIL een vragenlijst is die gemakkelijk door leraren zelf in de klas aan leerlingen kan worden afgenomen. De vragenlijst kan goed gebruikt worden om leraren feedback te geven over hun functioneren in de ogen van leerlingen. Zij krijgen dan per item en per sector de gemiddelde scores van hun leerlingen over hun gedrag en tevens de gemiddelde scores van leerlingen over een groot aantal leraren. In een figuur krijgen ze verder de resultaten van de afname bij hun leerlingen per sector. Tevens kunnen in een figuur ter vergelijking de gemiddelde schaalscores van leerlingen over een groot aantal leraren worden weergegeven en de resultaten van twee door de leraar zelf ingevulde vragenlijsten: een lijst met betrekking tot zijn eigen gedrag en één over het gedrag dat hij graag zou willen vertonen in de betreffende klas. Voor onderzoeksdoeleinden kan de lijst gebruikt worden om een belangrijk aspect van de leeromgeving, de leraar-leerling relatie, gezien door de ogen van leraar en leerlingen in kaart te brengen. Voor resultaten die met behulp van de VIL zijn verkregen verwijzen wij naar Créton en Wubbels (1984) en Brekelmans, Wubbels en Hooymayers (1986).

De betrouwbaarheid afgemeten aan de test-hertest betrouwbaarheid is redelijk en afgemeten aan de interne consistentie goed, terwijl er vrijwel geen items in de lijst voorkomen die als synoniem moeten worden beschouwd.

Ten aanzien van de validiteit zijn er enerzijds bemoedigende resultaten, maar moet anderzijds een kanttekening worden gemaakt bij de veronderstelde tweedimensionaliteit van Leary's model. De resultaten die vertrouwen wekken in de validiteit zijn:

- Gegevens betreffende de correlaties tussen subschalen bevestigen de circulaire ordening van de subschalen.
- Vermoedens over het gedrag van leraren worden bevestigd door gegevens die met de VIL zijn verzameld.

De gegevens uit de confirmatieve factoranalyses geven zowel ruimte om te veronderstellen dat aan de operationalisatie van interactioneel leraarsgedrag door middel van de VIL twee dimensies ten grondslag liggen als om te veronderstellen dat nog een derde dimensie nodig is. In het laatste geval zouden de acht subschalen beter op een bol weergegeven kunnen worden dan in een plat vlak. Wanneer een derde dimensie nodig is blijft voorlopig onduidelijk welke dit is. Aan een derde dimensie worden volgens Beekers en Halfens (1981) verschillende betekenissen toegekend zoals inzet, onafhankelijkheid, minderwaardigheid, manifeste intelligentie of passiviteit. Cuvelier (1976) heeft de derde dimensie het verst uitgewerkt door aan de Boven-Onder-dimensie een activiteitsdimensie toe te voegen. Ook uit ons onderzoek volgt een dimensie die dicht bij de Boven-Onder-as ligt. Voorlopig lijkt het niet aan te raden de acht subschaalscores samen te vatten tot één punt in een tweedimensionaal vlak zoals Leary soms doet (b.v. Leary 1957, blz. 218). Dit leidt tot verlies van informatie.

Noten

1. Verplaatsing van deze items naar andere schalen leidde ertoe dat weer andere items niet het hoogst met hun eigen subschaal correleerden.
2. De vragenlijst is bij de auteurs verkrijgbaar.

3. Voor dit onderzoek werd subsidie verleend door de Stichting voor Onderzoek van het Onderwijs (SVO-project 1073).
4. We vatten dezelfde leraar in twee verschillende klassen dus op als verschillende leraren, hetgeen leidt tot een onderschatting van het leraarseffect.
5. Wij bedanken A. Dijkhuizen voor zijn hulp bij de LISREL-analyses.

LITERATUUR

- Bales, R.F. (1970). *Personality and interpersonal behavior*. New York: Holt, Rinehart and Winston.
- Beekers, M., & Halfens, R. (1981). Met het oog op vaardigheidstherapieën. Cliënten over hun interpersoonlijk gedrag. *Tijdschrift voor Psychotherapie*, 7, 318-327.
- Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality*. Groningen: dissertatie.
- Brekelmans, J.M.G., Wubbels, Th., & Hooymayers, H.P. (1986). Interactioneel leraarsgedrag, curriculum en onderwijsopbrengsten. In: W.J. Nijhof en E. Warries (red.), *De opbrengst van onderwijs en opleiding*, bijdragen tot de onderwijsresearchdagen 1985, Lisse: Swets en Zeitlinger, 41-55.
- Briar, S., & Bieri, J. (1963). A factor analytic trait inference study of the Leary Interpersonal Checklist. *Journal of Clin. Psychology*, 19, 193-198.
- Brophy, J., & Good, Th.L. (1986). Teacher behaviour and student achievement. In: M.C. Wittrock. *Handbook of research on teaching*, third edition. New York: McMillan.
- Créton, H.A., & Wubbels, Th. (1984). *Ordeproblemen bij beginnende leraren*, Utrecht: W.C.C. (dissertatie).
- Cuvelier, F.J. (1976). *De stad van axen*. Kapellen: Ned. Boekhandel.
- Dijkstra, R.F.W. (1972). *Interpersonal Checklist*: Nederlandse bewerking (niet gepubliceerd), Nijmegen.
- Hooymayers, H.P., Créton, H.A., & Wubbels, Th. (1982). Een model voor interactioneel leraarsgedrag in het bijzonder gedrag gericht op de sfeer in de klas. *Pedagogisch Tijdschrift, Forum voor Opvoedkunde*, 7, 108-121.
- Horst, P. (1949). A generalized expression for the reliability of measures. *Psychometrika*, 14, 21-31.
- Jöreskog, G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge: Abt Books.
- Knippenberg, F.C.A. (1978). *Exploratief onderzoek naar het interpersoonlijk gedrag van thuislozen*. Nijmegen: (doctoraalscriptie).
- Laforge, R., & Suczek, R.F. (1955). The interpersonal dimension of personality: III. An interpersonal checklist. *Journal of Personality*, 24, 94-112.
- Leary, T. (1957). *An interpersonal diagnosis of personality*. New York: Ronald Press Company.
- Nunnally, J.C. (1967). *Psychometric theory*. New York: McGrawHill.
- Peters, V. (1985). *Docenten en hun probleemsituaties*. Nijmegen: dissertatie.
- Pettegrew, L.S., & Wolf, G.E. (1982). Validating measures of Teacher Stress. *American Educational Research Journal*, 19, 3, 373-396.
- Rosenshine, B., (1971). *Teaching behaviours and student achievement*. London: National Foundation for Educational Research.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54, 143-178.
- Vonk, J.H.C. (1982). *Opleiding en praktijk*. Amsterdam: VU boekhandel/uitgeverij (dissertatie).
- Watzlawick, P., Beavin, H.J., & Jackson, D. (1970). *De pragmatische aspecten van de menselijke communicatie*. Deventer: Van Loghum Slaterus.
- Wijk, L.J. van (1979). *Verandering van lesgedrag van leraren*. Nijmegen: Nivor.

Manuscript ontvangen 24-6-1986

Definitieve versie ontvangen 11-11-1986

Blending Behavioristic and Cognitive Approaches to the Study of Education¹

Hans F.M. Crombag

University of Leiden, The Netherlands²

SUMMARY

In this paper it is argued that specifying precisely which cognitive repertoires we want our students to be able to perform is not enough as a basis for an educational technology. In order to learn and routinize such repertoires students will have to perform a variety of study behaviors, which are in part probably under operant control. An effective educational technology should therefore be built on a mixture of cognitive and behavioral insights.

1. A COMMENT ON THE COGNITIVE REVOLUTION

Sometime during the sixties a revolution is said to have occurred in psychology. Like the French, American, and Russian revolutions it had a name: 'the cognitive revolution'. And like those other revolutions it was directed against some 'ancien régime', which in this case bore the name of 'behaviorism'. 'Behaviorism is dead. Long live cognitive psychology,' many psychologists cried as they handed their copy of *Hilgard and Marquis' Conditioning and Learning* (Kimble, 1961) to the garbage man. Of course, there still remains good old B.F. Skinner, who mistakes Harvard for Elba instead of Saint Helen (see e.g., Skinner, 1984).

This is supposed to be the brief history of modern psychology as some would like it to be (see e.g., Dennett, 1979, p. 53 ff). But, of course, it is not. Classical or radical behaviorism may not be very popular any more, but various brands of behavioristic psychology are still very much with us, e.g., in the form of behavioral instruction in education (Johnson & Ruskin, 1977) and behavior modification in psychotherapy (Bandura, 1969). Cognitive psychology is, no doubt, the mainstream of present-day psychology and behaviorism is at best 'coexisting, however uneasily' with 'mentalism', which is the nickname that behaviorists preserve for cognitive psychology (see Schnaitter, 1986, p. 292).

To the extent that it is true that cognitive psychology superseded behaviorism as the mainstream of psychology, one may assume that many psychologists above a certain age started out as behaviorist before changing sides at some point in their careers. But many European cognitive psychologists cannot have made this transition, because they never were trained in a behavioristic tradition. When I was a student in the late fifties, behaviorism was taught as one among many 'schools' of psychology, on a par with e.g. Gestalt psychology and psychoanalysis. On leaving the university, I did not consider myself to belong to a 'school'. My early work in educational research did not reflect an allegiance to anyone 'school' in particular. My (applied) research was empirical and its methodology more or less Popperian (De Groot, 1969), but I had no qualms whatsoever using mentalistic terms such as 'intelligence', 'attitude', or 'motivation'.

I tell all this not because I assume that my personal history is of interest to anyone but me, but

¹ Address of the incoming president of the Division for Educational, Instructional and School Psychology of the International Association of Applied Psychology, Jerusalem, 16 July 1986.

² Presently with the Faculty of Law, State University of Limburg, P.O. Box 616, 6200 MD Maastricht.

because my biography so far seems typical for my generation of European psychologists. Because most of us never were real behaviorists, the cognitive revolution of the sixties wasn't much of a revolution for us and using mentalistic concepts was old hat.

Allow me to pursue my personal history somewhat further. Not because I continued to be prototypical of my generation, however. On the contrary, I ended up on the wrong side of the fence. But not immediately. Like many of my colleagues, I turned cognitive in the early seventies. I did so, because as an educational researcher (of higher education in particular) I had been studying teaching methods and their differential effectiveness and I was dissatisfied with the results. I wanted to get closer to the real thing in education, which is not the teaching of teachers, but the learning of students.

2. THE MIND OF STUDENTS

Learning takes place in the minds of students and education is about changing minds – I had not yet read Skinner at the time (Skinner, 1972, p. 92). Teachers are expected to put new things into the minds of students, which they call 'knowledge', and occasionally take old things out, which they call 'misunderstandings'. They are expected to instill some understanding in their students, which has something to do with the organization of knowledge in the students' memories. And they are expected to teach students certain skills, mostly mental, which means telling them ways to use their knowledge and making them go through such repertoires repeatedly in order to routinize their execution.

Apart from sounding naive, all this is rather commonplace. Some twenty five years ago it wasn't revolutionary either. But studying organization in memory or problem-solving repertoires, i.e., algorithms or heuristics, was not the big thing which it is nowadays. In turning cognitive, I did the right thing at the right time, not out of any strong theoretical conviction, but for what I thought was a purely practical reason. I wanted to get at the heart of education: shaping minds of students.

The School of Law in my university asked for my help with a fresh attempt at teaching its students something more than mere knowledge of the law. During exams the faculty used to put hypothetical cases to the students. On these cases the students had been doing poorly. They should be trained explicitly in handling such cases. Could I help the Law School devising a training procedure to that purpose?

It soon turned out that even among the faculty nobody really knew how you go about solving a hypothetical legal case. Experienced lawyers are able to do it allright, and they can distinguish with reasonable consensus between good and bad solutions. But to tell us precisely how they went about it, was something else.

We selected some of these hypothetical cases and presented them to experienced lawyers. We asked them to think aloud while solving each case. Each time we did not understand what was in their protocols, we asked for further explanations. Building some sort of a methodology as we went along (see Chang, 1974), we ended up with what we called 'a rational reconstruction' of the skill of solving hypothetical legal cases (Crombag, De Wijkerslooth, & Van Tuyll van Serooskerken, 1972 and 1975; Crombag, De Wijkerslooth, & Cohen, 1977). It took us several years to develop this model of judicial decision making and a book to describe and explain it.

By contrast we took comparatively little time to devise a teaching procedure based on this model. Early in the course we gave the students a simplified version of the model we had developed and made them go through it with simplified cases. Gradually we added details to the model, while increasing the complexity of the cases presented to the students, in order to make the students' knowledge of the model grow together with their skill in using it.

We did not spend much deep thought in devising this procedure. As an after-thought we added some rules. Students were obliged to attend all classes in the course -unusual in universities in my country. Moreover, at the beginning of each class the teacher should check whether each student was prepared. These rules seemed practical things, without much theoretic-

tical significance. The course worked: we were able to demonstrate that the students actually learned what they were supposed to learn (Crombag, 1974, p. 265).

Having this and a few similar things done, I proudly considered myself a cognitive educational psychologist. Henceforth, I considered the task of educational psychologists to be: to study and specify the mental operations we wish to instill in the minds of students in order to make them experts in their chosen fields.

3. MAKING A U-TURN

We were overlooking a few things, however. This only became apparent when, in the late seventies, my colleagues and I rather accidentally became involved in a quite different line of research. Around that time the length of university curricula, which in my country used to be between five and eight years, was shortened to four years by rule of law. As a result all university curricula had to be revised. To a large extent this was done by cutting courses from the curricula, but naturally part of the solution was sought in packing the four years left with more courses than before. This raised the question of study load, and we were asked to find out in various curricula how much time it took to complete them.

We took up this task with some dismay, because this was not cognitive, but behavioral psychology. We asked students in various departments to report to us daily how they spent their time and we got pictures of how students divided their time over courses, over days of the week and over weeks of the year. These 'behavioral' data turned out to be much more intriguing than we had anticipated.

With respect to the outcomes I'll restrict myself to a few generalities. If you know the organization of the curriculum, the work behavior of university students is highly predictable. Students put in time wherever there is a short-term demand. Whenever there are no immediate curricular demands, then little or no work is done. Curricular demands that are further away than at most two weeks do not elicit study behavior. Whenever there are conflicting curricular demands, such as two exams following each other closely, then the one whose deadline comes first takes up virtually all of the students' time, and the second is neglected. Students, as far as their work behavior is concerned, are externally motivated and shortsighted to the point of stupidity (Vos, 1985; Crombag, Van der Drift, & Vos, 1984/85).

By manipulating curricular demands, it is possible for teachers to manipulate the work behavior of students in such a way, that each course in the curriculum gets adequate attention from the students, keeping them busy all weeks of the academic year and most days of the week. Advocating this kind of behavioral engineering by rearranging environmental contingencies, I found myself wondering if I had become a behaviorist. A disturbing thought at a time when behaviorism had gone out of fashion.

4. LOOKING BACKWARD

Let me compare the two lines of research which I happened to have been engaged in. There must be a lesson in their conflicting implications.

I still think that it is the task of education to shape minds. If we are going to do this or at least try to do this, we better be explicit about the shapes we are going to give to various minds, such as the mind of the generally educated person, the mind of the lawyer, the mind of the mathematician. Educators have been in the business of shaping minds for a long time, but they have remained remarkably vague about the precise shape of the educated mind. If cognitive psychology is going to provide educators with precise models of the shapes of educated minds, that would mean a remarkable progress. And this is indeed what is going on. But is it enough?

I do not adhere to the belief that minds shape themselves through some sort of 'natural' development, Lauren Resnick's 'biological assumption' (Resnick, 1981). Nor to the idea that

this shaping is a unique process, different for every student, Resnick's 'individualist assumption'. I therefore think that we also need a technology of teaching which contains general rules about do's and don'ts in teaching. Here too cognitive psychology can be helpful, because it teaches us about the capabilities and limitations of the 'human information processing system' and their implications for teaching (Calfee, 1981). Do not put too many points into a single argument while teaching, for short-time memory has a limited capacity and you may therefore end up with information overload. Do not go too fast while teaching, for transfer of information from short-term to long-term memory takes time. Very sensible rules and based on solid research of cognitive psychologists.

When we know what we want the students to learn and what we want the teachers to do – of course we do not know it all yet, but we are working on it and making progress – all we need more is that everybody does 'the sensible thing'. For students the sensible thing is, that they always come to class, that they come well prepared, that they always pay attention in class and plan their work well ahead. For teachers the sensible thing is, that they choose the assigned literature with care and carefully design the exercises, that they always keep the students busy, but never too busy, that they prepare thoroughly for class and then behave impeccably: clear, thoughtful, and patient, always keeping in mind what it is that they are supposed to do and how. As cognitive educational psychologists are becoming more and more precise about what is the sensible thing, education is going to get better and better. Isn't that what we all hope for? And don't we all know that this is never going to happen? Why not? Why do people so often and so stubbornly refuse to do the sensible thing, in education as well as in other areas of human endeavor? Let me suggest a reason.

5. OPERANT BEHAVIOR

A good deal of human behavior is not *intelligent behavior*, controlled by conscious considerations or at least learned in that way and automatized afterwards (e.g., as described by Schneider & Shiffrin, 1977, and Shiffrin & Schneider, 1977), but *operant behavior*, 'shaped and maintained by its consequences' (Skinner, 1972, p. 18). It is difficult and in many instances impossible to decide whether and to what extent a particular behavior is either intelligent or operant. The fact that people will explain why they do certain things, does not make these behaviors controlled by intelligence. We know such explanations to be wrong frequently. The fact that people often cannot explain how they perform certain tasks, does not make these performances operant behavior for the same reason: people are unable to observe directly their cognitive processes, but they often act or, rather, talk as if they can (Nisbett & Wilson, 1977; Herrmann, 1982).

It also is not the case that particular kinds of behavior are always intelligent, while others are always operant. Language production in our native tongue is almost certainly acquired and maintained under operant control (Skinner, 1957, Bruner 1983). Language production in a second language, learned in school, is almost certainly intelligent behavior, which may become routinized over time. (Certain aspect like pronunciation may be under operant control even in a second language.)

Finally, in many behavior repertoires intelligent and operant components go together. When the intelligence-controlled components are routinized, intelligent and operant components become virtually indistinguishable. Routinization of intelligent repertoires minimizes the demands on short-term memory, which has a limited capacity. Robert Calfee (1981, p. 11) says: 'The person selects a routine that should produce a desired outcome, and then the routine runs off more or less on its own. The individual checks to see what is happening every once in a while to ensure that things are still on the right track.'

Operant (components of) behavior repertoires are even more economical on the capacity of short-term memory, because they are acquired and executed 'without any attention to the underlying cognitive mechanism,' Calfee (1981, p. 11) says. This statement is misleading, as it is

uncertain whether there is an 'underlying cognitive mechanism' in operant behavior at all. Classical behaviorists deny this and since we have no way of testing it, its presence is at best a conjecture.

6. MORALE

Which brings me to the point I wish to make. Behavior repertoires that are under operant control are hardly, if at all, accessible to intelligent control. There is abundant evidence that people often know perfectly well what is the sensible thing to do or not to do. But even so, they keep doing the insensible thing and cannot bring themselves to do the sensible thing. Doctors instruct their patients, parents their children, teachers their students to do the sensible thing, but the patients, children, and students are insensitive to their exhortations. The only way to make them mend their ways, is letting them find out for themselves. But finding out by yourself is, of course, the operant way.

The fact that operant behavior is (often) inaccessible to intelligent control does not make operant behavior irrational. Operant learning brings behavior under the control of its immediate consequences. It adapts behavior to its environment. There is a certain amount of rationality in this adaptive mechanism, but its rationality is shortsighted or, to use Herbert Simon's word, 'myopic'. Following the myopic rationality of operant learning, a person 'may find (him)self on top of all sorts of local hills with nowhere to go but down' (Simon, 1983, p. 66ff.). And that, in the long run, is stupid.

In our research of work behavior of university students, we found that students are very shortsighted in their work habits. Telling them so, pointing out the deferred consequences of their shortsightedness or explaining them the wisdom of time management, does not alter their behavior. In fact, they do not have to be told, because they already know. They are not stupid! Their work behavior, however, is almost entirely under curricular, i.e. operant control. Only operant techniques in the form of curricular reorganisations will change their behavior effectively.

Let me generalize from this. Telling teachers and students what is the sensible thing to do in order to achieve the goals of education, is not enough as a technology of teaching. Now that the cognitivists have helped us delineate what constitutes the sensible thing, i.e. precisely how we want to shape the minds of students, we must bring what is necessary to achieve this, under operant control. Thus we may bring under myopic, but effective operant control what our intelligence tells us to be the sensible thing.

REFERENCES

- Bandura, A. (1969). *Principles of behavior modification*. New York: Holt, Rinehart & Winston.
- Bruner, J. (1983). *Child's talk: Learning to use language*. New York: Norton.
- Calfee, R.C. (1981). Cognitive psychology and educational practice. In: D.C. Berliner (ed.), *Review of research in education, Vol 9*, 3-73. Washington, D.C.: American Educational Research Association.
- Chang, T.M. (1974). A method for analysing cognitive skills. In: W.A. Verreck (ed.), *Methodological problems in research and development in higher education* (pp. 383-396). Amsterdam: Swets & Zeitlinger.
- Crombag, H.F.M. (1974). Product and process in teaching and testing. In: H.F. Crombag & D.N. de Gruijter (eds.), *Contemporary issues in educational testing* (pp. 256-271). The Hague: Mouton, 1974.
- Crombag, H.F.M., Drift, K.D.J.M. van der, & Vos, P. (1984/85). De inrichting van curricula en het werkgedrag van studenten. *Universiteit en Hogeschool*, 31, 234-247.
- Crombag, H.F.M., Wijkerslooth, J.L. de, & Cohen, M.J. (1977). *Een theorie over rechterlijke beslissingen*. Groningen: H.D. Tjeenk Willink.
- Crombag, H.F.M., Wijkerslooth, J.L. de, & Tull van Serooskerken, E.H. van (1972). *Over het oplossen van casusposities*. Groningen: H.D. Tjeenk Willink.

- Crombag, H.F.M., Wijkerslooth, J.L. de, & Tuyll van Serooskerken, E.H. van (1975). On solving legal problems. *Journal of Legal Education*, 27, 68-202.
- Groot, A.D. (1969). *Methodology. Foundations of inference and research in the behavioral sciences*. The Hague: Mouton (Dutch original from 1961).
- Dennett, D.C. (1978). *Brainstorms. Philosophical essays on mind and psychology*. Hassocks, Sussex: Harvester Press.
- Herrmann, D.J. (1982). Know thy memory: The use of questionnaires to assess and study memory, *Psychological Bulletin*, 92, 434-452.
- Johnson, K.R., & Ruskin, R.S. (1977). *Behavioral instruction: An evaluative review*. Washington, D.C.: American Psychological Association.
- Kimble, G.A. (ed.) (1961). *Hilgard and Marquis' Conditioning and learning*. New York: Appleton-Century-Crofts.
- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes, *Psychological Review*, 84, 231-259.
- Resnick, L.B. (1981). Social assumptions as a context for science: Some reflections on psychology and education, *Educational Psychologist*, 16, 1-10.
- Schnaitter, R. (1986). A coordination of differences: Behaviorism, mentalism, and the foundations of psychology. In: T.J. Knapp & L.C. Robertson (eds.), *Approaches to cognition: Contrasts and controversies* (pp. 291-315). Hillsdale, N.J.: Erlbaum.
- Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention, *Psychological Review*, 84, 1-66.
- Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127-190.
- Simon, H.A. (1983). *Reason in human affairs*. Oxford: Basil Blackwell.
- Skinner, B.F. (1957). *Verbal behavior*. Englewood Cliffs, N.J.: Prentice-Hall.
- Skinner, B.F. (1972). *Beyond freedom and dignity*. New York: Knopf.
- Skinner, B.F. (1984). The shame of American education. *American Psychologist*, 39, 947-954.
- Vos, P. (1985). Zelfstudie als functie van onderwijsdeelname: Hypothesetoetsing. *Tijdschrift voor Onderwijsresearch*, 10, 228-238.

Manuscript ontvangen 3-10-1986

Definitieve versie ontvangen 12-11-1986

Richting van de school en schoolloopbanen in het voortgezet onderwijs*

P. van Laarhoven, B. Bakker, J. Dronkers en H. Schijf
RION, Groningen

ABSTRACT

School success in public, Catholic and Protestant secondary education.

The educational system in the Netherlands has a large private sector which is state-supported in the same manner as the public sector. It consists mainly of denominational schools. In this article, a sequel to an article on primary education, the effects of school sector (public, Catholic, Protestant) on students' scholastic achievement are studied within two levels of secondary education. At both levels sector effects on key aspects of school success are found. These are generally small, but tend to cumulate in favour of one of the three sectors. Protestant school students perform best at the top level (VHMO), Catholic school students at the level of ULO. In neither case do public school students manage to do better than their counterparts in the private sector. These results suggest that differences in school sector effectiveness may partly explain the persistence of denominational education in the Netherlands.

1. INLEIDING

Een sterke verzuimdheid kenmerkt het Nederlandse onderwijs. De verzuiling heeft het karakter gehad van een groei van het confessionele onderwijs ten koste van de openbare en algemeen-bijzondere scholen, wat niet zonder ideologische conflicten en machtsstrijd verlopen is. De verdeling naar richtingen is in het lager onderwijs al vóór 1940 voltooid, in het voortgezet onderwijs pas tegen het einde van de jaren zestig. Sindsdien vertoont het patroon weinig verschuivingen van betekenis.

Het aandeel van confessionele scholen in de onderwijsvoorzieningen is nog altijd hoog. In 1983 bezocht 65% van de leerlingen in het lager onderwijs en 62% van de leerlingen in het voortgezet onderwijs een school van confessionele signatuur. Dat is opmerkelijk gezien de secularisering en ontzuiling op een aantal andere terreinen van het maatschappelijk leven. Het werpt de vraag op waarom ouders hun kinderen nog steeds naar scholen sturen die een godsdienstige levensbeschouwing vertegenwoordigen.

* Dit artikel is een ingekorte versie van een paper, gepresenteerd tijdens de Vlaams-Nederlandse Studiedagen voor Sociologen en Antropologen in april 1986. De paper-versie, met uitgebreidere analyses en toelichtingen, kan bij de auteurs opgevraagd worden.

Het onderzoek is gefinancierd door het Ministerie van Onderwijs en Wetenschappen, via de Stichting voor Onderzoek van het Onderwijs en het Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. Het is uitgevoerd binnen de vakgroep Grondslagen en Methoden van de FSW-B van de Universiteit van Amsterdam.

De auteurs zijn werkzaam bij resp. RION (Groningen), CBS (Voorburg), Kath. Univ. (Tilburg) en de Universiteit van Amsterdam.

De gegevens van het Van Jaar Tot Jaar-onderzoek, verzameld door het Instituut voor Toegepaste Sociologie te Nijmegen, zijn beschikbaar gesteld via het Steinmetz Archief.

Correspondentie aan: P. van Laarhoven, RION, Postbus 1286, 9701 BG Groningen.

Daarvoor worden verschillende verklaringen aangedragen. In de eerste plaats zou het onderwijs, meer dan andere sectoren, aan de kern van de levensbeschouwing raken. Thurlings (1972) stelt, dat naarmate de affiniteit van een sector tot de levensbeschouwing groter is, de verzuiling eerder en de ontzuiling later zal inzetten. Deze hypothese wordt ondersteund door de resultaten van verschillende onderzoeken. Zo stelt Van Kemenade (1968) vast, dat veel katholieke ouders belang hechten aan de levensbeschouwelijke oriëntatie van de school. Dat is ook het geval bij ouders die weinig behoefte hebben zich te isoleren van de andersdenkende buitenwereld of die persoonlijk weinig binding met het katholieke onderwijs voelen. Van de katholieke scholen wordt een religieuze vorming in meer algemene zin en onderwijs in een sfeer van waarden- en normenbesef verwacht. Onderzoek onder protestants-christelijke ouders levert een soortgelijk beeld op (De Jonge 1978). Binnen deze kring moet onderscheid gemaakt worden tussen de vrijzinnig hervormden enerzijds en de gereformeerden en orthodox-hervormden anderzijds. De eerste groep levert een groot deel van de leerlingen van de openbare scholen. Onder de laatsten zijn traditionele godsdienstige opvattingen en een verzuilingsmentaliteit vrij sterk aanwezig. Het aantal lagere scholen van orthodox-protestantse signatuur is zelfs licht stijgende (Pelkmans 1984 : 153).

Een bezwaar tegen deze verklaring is, dat veel meer kinderen naar confessionele scholen gaan dan op grond van de voorkeuren van ouders verwacht mag worden.¹ Daarom is als tweede verklaring gewezen op het ontbreken van alternatieven. Het bestand aan scholen verandert maar langzaam, omdat door het teruglopende aantal leerlingen het aantal nieuw te stichten scholen gering is. Wanneer er nieuwe scholen worden gevestigd, wordt de bestaande verdeling als uitgangspunt genomen. Bij het vestigingsbeleid is de invloed van de gevestigde schoolbesturen en hun centrale zuilenorganisaties groot. Deze invloed wordt aangewend ter handhaving van de bestaande verdeling van onderwijsvoorzieningen (Boef-van der Meulen 1982; Randsdorp en Dronkers 1982). Ouders zien zich hierdoor bij hun schoolkeuze veelal beperkt tot het bestaande aanbod van scholen van het gewenste onderwijstype.

Deze tweede benadering verklaart evenwel niet waarom er zo weinig verzet bij de ouders bestaat tegen deze beperking in het aanbod. Een derde factor die dat zou kunnen verklaren, is het verschil in de 'kwaliteit' van het onderwijs op scholen van verschillende richting, zoals dat door ouders wordt waargenomen. Pelkmans spreekt in dit verband van een secundaire functie van het confessionele onderwijs (1984 : 25). Deze derde verklaring wordt in dit artikel empirisch onderzocht.

Onderzoek uit de Verenigde Staten en Israël laat verschillende onderwijskansen per richting zien. Coleman et al. (1982) en: Hoffer et al. (1985) maken een vergelijking tussen openbare en katholieke 'high schools' in de Verenigde Staten. De resultaten gaan in tegen twee naar hun zeggen gangbare vooroordelen over het katholieke onderwijs in de Verenigde Staten. De particuliere onderwijssector (met een aandeel van $\pm 15\%$) zou de scheiding van kansarme en kansrijke leerlingen versterken en afgestemd zijn op de kansrijke leerlingen. Het blijkt nu, dat katholieke scholen wel minder door kinderen uit kansarme groepen bezocht worden, maar dat de spreiding van deze leerlingen over de scholen binnen de katholieke sector beter is dan binnen de openbare sector. Bovendien komen kansarme leerlingen in het katholieke onderwijs bij een gelijke uitgangspositie tot betere leerprestaties dan op openbare scholen. De betere resultaten kunnen grotendeels verklaard worden uit de grotere motivatie en studie-inzet die de katholieke scholen bij hun leerlingen blijken te kunnen afdwingen. Ook Yogev en Chen (1985) vinden bij hun vergelijking tussen 'staats-' en 'staats-religieuze' scholen in Israël, dat met de richting van de school verschillende patronen van prestatie en selectie samenhangen.

In het Nederlandse onderwijs kan dit verschijnsel eveneens een rol spelen. Ouders met een religieuze achtergrond hebben alle reden om vast te houden aan hun voorkeur voor scholen van de eigen richting, wanneer zij vinden dat die het 'beter' doen juist met het oog op de leerprestaties. En omgekeerd, nu het religieus-vormende karakter van confessionele scholen op de achtergrond is geraakt kan voor niet-godsdienstige ouders de drempel verlaagd worden. De

kwaliteit van het onderwijs op confessionele scholen kan meer gaan meespelen bij de schoolkeuze.

Wát voor 'kwaliteit' precies van het onderwijs verlangd wordt, hoeft overigens niet voor alle ouders gelijk te zijn. De Jonge (1978) laat zien, dat onder protestant-christelijke ouders een flink aantal waarde hecht aan onderwijs van de eigen richting omdat het 'meer gericht is op het geluk van het kind en niet op een zo groot mogelijk maatschappelijk succes' en omdat 'de prestatiedwang er niet zo centraal staat'. Pelkmans (1984 : 25) vindt daarentegen aanwijzingen dat 'sommige ouders los van religieuze overwegingen de voorkeur geven aan de veronderstelde 'stevigheid' van de pedagogisch-didactische aanpak in het bijzonder onderwijs ten opzichte van de lossere of vrijere aanpak in openbare scholen'. Het is voor ouders meestal moeilijk vast te stellen of een school de kwaliteiten bezit die zij verlangen. Daarom kan de richting van de school gaan fungeren als een 'screening device' voor onderwijskwaliteit. Zijn 'inschaling' kan een school aan meerdere zaken ontleen: aan de reputatie die de school in zijn omgeving heeft, aan de pedagogisch-didactische aanpak die traditioneel met het onderwijs van de betreffende richting verbonden wordt of aan de gunstige randvoorwaarden waaronder het onderwijs er plaats vindt. Bijzondere scholen kunnen bijvoorbeeld extra schoolgelden heffen en daarmee buitenschoolse activiteiten organiseren. Ze kunnen wellicht ook op effectievere wijze schooldiscipline en studie-inzet bij hun leerlingen afdwingen – en wellicht ook bij hun leerkrachten.

Tegen deze achtergrond hebben wij eerder het effect van de richting van de school op de individuele schoolloopbaan in het lager onderwijs nagegaan (Van Laarhoven et al. 1986). We vinden een richting-effect bij verschillende aspecten van de lagere schoolloopbaan. De resultaten geven geen aanwijzingen, dat de lagere scholen van een bepaalde richting in alle opzichten beter zijn dan die van de andere richtingen. Evenmin doet het bijzonder lager onderwijs het beter dan het openbare. Zo doubleren leerlingen van katholieke scholen anderhalf tot twee keer zo vaak als leerlingen van protestantse scholen, maar komen zij op een hoger prestatieniveau uit. De leerlingen van openbare scholen nemen in beide opzichten een middenpositie in; bij deze groep valt op dat de jongens veel vaker doubleren en iets lager presteren dan de meisjes. Ook de verschillen bij de overgang naar het voortgezet onderwijs zijn nauwelijks in termen van beter of slechter op te vatten. Er lijken daarbij eerder godsdienst- en seksegebonden voorkeuren voor bepaalde typen voortgezet onderwijs in het spel.

De individuele schoolloopbanen zijn met de intrede in het voortgezet onderwijs nog niet vastgelegd. De aanvangsverschillen kunnen daar versterkt, gehandhaafd of ten dele gecompenseerd worden. Daarom trekken we in dit paper de analyse door in het voortgezet onderwijs. De *probleemstelling* luidt:

Heeft de richting van de school in het voortgezet onderwijs effect op het individuele schoolsucces van zijn leerlingen, ook na controle voor achtergrondkenmerken van de leerling (sekse, ouderlijk milieu, regio) en voor het verloop van de voorafgaande schoolloopbaan?

2. HET DATAMATERIAAL

Voor het antwoord op deze vraag is gebruik gemaakt van de longitudinale 'Van Jaar Tot Jaar'-gegevens. In 1965 is door het Centraal Bureau voor de Statistiek een steekproef ($n = 405$) getrokken uit de Nederlandse lagere scholen. Uit de zesdeklassers van deze scholen ($n = 11\,170$) is vervolgens een drietal steekproeven getrokken van in totaal 3042 leerlingen. Deze zijn door het Instituut voor Toegepaste Sociologie gevolgd in hun school- en beroepsloopbaan (Collaris en Kropman 1978). Voor de analyses in dit paper wordt het hele Van Jaar Tot Jaar-bestand gebruikt.²

De schoolloopbaan in het voortgezet onderwijs begint voor alle leerlingen uit dit cohort in

1965 en loopt tot hooguit 1974. In de meeste gevallen bezoeken zij de schooltypen van vóór de Mammoetwet, zoals ULO, MMS en HBS. Alleen de HAVO wordt al door vrij veel leerlingen bezocht, als tweede of derde schooltype, na afstroom of opstroom uit een van de oudere typen. Dat legt een historische beperking op aan de in dit artikel te trekken conclusies. Voor recentere landelijk representatieve gegevens is de variabele 'richting van de bezochte school' evenwel niet voor secundaire analyse beschikbaar. Mogelijke historische ontwikkelingen kunnen daarom niet worden onderzocht.

3. ANALYSE-OPZET EN ONDERZOCHE KENMERKEN

De analyse van schoolloopbanen in het voortgezet onderwijs in relatie tot de richting van de school kan langs twee lijnen worden uitgevoerd: 1) een analyse van het effect op het bereikte *eindniveau* bij het verlaten van het voortgezet onderwijs, en 2) een analyse per bezochte *schoolsoort*.

De eerste benadering stuit op bezwaren, omdat ruim een derde van de leerlingen in het Van Jaar Tot Jaar-cohort meer dan één soort voortgezet onderwijs volgt. Ook bezoekt, zoals uit *tabel 1* naar voren komt, in ca. 20% van de gevallen een leerling in het voortgezet onderwijs een school van een andere richting dan de lagere school. Het is daarom niet goed mogelijk om voor alle leerlingen de schoolloopbaan te relateren aan één richting.

Tabel 1. Richting van de school in VHMO, ULO en LBO naar richting van de lagere school

Richting van de lagere school	VHMO			ULO			LBO		
	Open- baar	Katho- liek	Chris- telijk	Open- baar	Katho- liek	Chris- telijk	Open- baar	Katho- liek	Chris- telijk
Openbaar	59%	2%	9%	78%	3%	5%	43%	4%	11%
Katholiek	24%	97%	6%	11%	96%	1%	25%	91%	3%
Christelijk	11%	1%	83%	10%	1%	94%	30%	4%	87%
Algemeen-bijzonder	6%	–	2%	2%	–	1%	2%	1%	–
(100%=)	407	447	262	442	650	489	447	465	231

Bovendien zijn de verschillende richtingen niet op alle niveaus van het voortgezet onderwijs even goed vertegenwoordigd. Een aantal leerlingen ziet zich waarschijnlijk genoodzaakt voor het schooltype van hun keuze naar een school van een 'andere' richting te gaan, of van een andere richting dan de lagere school. Dat speelt vooral in het LBO. Wanneer de schoolloopbaan in het voortgezet onderwijs toch wordt gerelateerd aan de eerstbezochte of 'meest bezochte' richting van de school, blijft een aantal belangrijke aspecten buiten beeld.

In dit artikel beperken we ons daarom tot de tweede benadering, waarin het gaat om de schoolloopbanen binnen elke soort voortgezet onderwijs afzonderlijk. Van een leerling wordt het schoolsucces voor ieder bezocht schooltype afzonderlijk onderzocht, wat ondubbelzinnig aan een bepaalde richting gerelateerd kan worden. De uiteenlopende schooltypen van het voortgezet onderwijs zijn gegroepeerd in vier 'schoolsoorten': VGLO, LBO, ULO en VHMO. Leerlingen die meer dan één schoolsoort bezoeken zijn bij elk van de bezochte schoolsoorten in de analyse betrokken, waarbij de schoolsoort van herkomst als controle-variabele is opgenomen. De 2982 leerlingen in het Van Jaar Tot Jaar-cohort die van de lagere school naar het voortgezet dagonderwijs zijn gegaan leveren zo 4171 te analyseren 'cases' op.

Het datamateriaal legt ons echter nog een tweede beperking op, namelijk dat de schoolsoor-

ten VGLO en LBO niet goed geanalyseerd kunnen worden. VGLO-onderwijs is door een te gering aantal leerlingen in het bestand gevolgd ($n = 156$). Bij het LBO ($n = 1143$) is de richting van de bezochte school hoogstwaarschijnlijk onbetrouwbaar gemeten³. Aldus blijven, na eliminatie van niet-volledige cases, voor analyse over: 1581 cases in het ULO en 1116 cases in het VHMO.

In het onderzoek zijn de volgende *kenmerken* betrokken (tussen haakjes de variabele-namen, waaraan in tabellen wordt gerefereerd):

A. Achtergrond	1. Geslacht	(SEKSE)
	2. Beroepsniveau vader	(BERVA)
	3. Opleidingsniveau vader	(OPVA)
	4. Opleidingsniveau moeder	(OPMOE)
	5. Regio woongemeente	(REGIO)
	6. Urbanisatiegraad woongemeente	(URB)
B. Lagere school	7. Richting van de school	(RILO)
	8. Doublures	(DOUBLO)
	9. VHMO-prestatiescore	(PREST)
	10. Advies voor voortgezet onderwijs	(ADVIES)
C. Voortgezet onderwijs (per schoolsoort:) - onafhankelijken	11. Voorafgaande schoolsoort	(WEG)
	12. Type binnen schoolsoort	(TYPE)
	13. Klas van schoolverlaten	(UITKLAS)
	14. Richting van de school	(RISEC)
	15. Meer opleidingen op school	(ANDOP)
	16. Extra-curriculaire activiteiten	(EXCUR)
	17. Participatie daaraan	(EXPART)
	18. Doublures	(DOUB)
	19. Gekozen afdeling/pakket	(PAKKET)
	20. Diploma behalen	(DIPLOMA)
	21. Doorleren na deze school	(DOOR)
- afhankelijk		

De onderwijsloopbanen in ULO en VHMO zijn door ons op twee manieren onderzocht. Eerst zijn deze uiteengelegd in een viertal kenmerken: doubleren, keuze van afdeling of vakkenpakket, diploma behalen, volgen van verder dagonderwijs na het verlaten van deze schoolsoort. Het zijn 'breukpunten' waarop de onderwijsloopbaan van een leerling op de school in kwestie zich gaat onderscheiden van die van andere leerlingen. Het al dan niet behalen van het diploma is het belangrijkste van deze breukpunten. Vervolgens zijn de wegen langs de breukpunten weer samengevoegd tot een typologie van doorlopen 'trajecten'.

4. STATISTISCHE METHODEN

Voor de analyse is gezocht naar een statistische procedure waarmee kan worden vastgesteld of de richting van de school effect heeft op de schoolloopbaan *na uitschakeling* van de invloed van zoveel mogelijk andere effecten. Het leerlingenbestand van scholen in het voortgezet onderwijs kan immers verschillen naar sociale achtergrond en voorafgaande schoolloopbaan. Bovendien kan het scholenbestand van de diverse richtingen een andere verdeling over schooltypen (binnen een schoolsoort) of regio's vertonen, meer categoriale scholen omvatten, etc. Statistische controles zijn nodig om na te gaan of verschillen in schoolsucces een gevolg zijn van de

richting van de school, of van deze andere kenmerken.⁴

Gekozen is voor covariantie-analyse, volgens ANOVA uit het SPSS-pakket. Dit is een techniek waarmee nominaal-metrische aanhangen onderzocht kunnen worden (zie voor een inleiding Wildt/Ahtola 1978). Het is een combinatie van variantie-analyse en regressie-analyse. De afhankelijke is een metrische variabele en als onafhankelijken kunnen zowel variabelen van nominaal meetniveau (aangeduid als 'factoren') als metrische variabelen (de 'covariaten') ingevoerd worden. De resultaten van ANOVA zijn tweeledig:

- (1) De eigenlijke covariantie-analyse, waarbij de bijdrage die elk van de onafhankelijke variabelen aan de variantie levert wordt berekend en getoetst op significantie.
- (2) Een 'multiple classification analysis' (MCA). Het programma berekent, voor elke categorie van een factor, een gecorrigeerde score op de afhankelijke variabele. Dat gebeurt door te controleren voor alle overige factoren en de covariaten. De gecorrigeerde score kan geïnterpreteerd worden als het gemiddelde dat de betreffende categorie zou realiseren – ceteris paribus – indien zij qua samenstelling exact aan de gemiddelde samenstelling van het hele analysebestand zou beantwoorden.

De covariantie-analyse is door ons toegepast op elk afzonderlijk breekpunt in de schoolloopbaan. De controlevariabelen zijn zoveel mogelijk als factoren opgenomen. Een variabele is alleen als covariaat behandeld indien deze geen interactie vertoont met een van de overige controle variabelen én indien het effect op de afhankelijke variabele niet significant afwijkt van lineariteit. In de uiteindelijke analyses zijn naast de richting van de school alleen de controlevariabelen opgenomen die een significante bijdrage (op .05-niveau) aan de verklaarde variantie leveren.

Interactie-effecten vragen in dit type analyse om een zorgvuldige behandeling. Hier is daar als volgt mee omgegaan:

1. Interacties van hogere dan de tweede orde zijn buiten beschouwing gelaten.
2. Interacties tussen factoren waarbij 'richting' niet betrokken is zijn verdisconteerd door een nieuwe variabele te construeren. Deze nieuwe factor is een combinatie van de oorspronkelijke factoren.
3. Interacties van de richting van de school met een andere factor raken de kern van onze probleemstelling en worden afzonderlijk gepresenteerd.

5. RESULTATEN

5.1. Richting en schoolloopbaan in het VHMO

Als eerste van de afzonderlijk onderzochte schoolsoorten bespreken we het VHMO. Deze schoolsoort is nogal heterogeen en omvat de schooltypen MMS, HAVO, HBS en Gymnasium, die als twee niveau's van voortgezet onderwijs kunnen worden opgevat, maar hier tot één schoolsoort gerekend worden. Ongeveer een derde van de leerlingen in het analysebestand volgt schooltypen die met de Mammoetwet in het leven zijn geroepen, voor het overgrote deel HAVO (28%). De meeste HAVO-leerlingen zijn opgestroomd uit het ULO (70%), een klein deel komt rechtstreeks van de lagere school (20%) of van een ander VHMO-type (10%). De schoolloopbanen in het VHMO zijn uiteen te leggen in vijf breukpunten, waarvoor de resultaten in *tabel 2a* worden gepresenteerd.

Op de meeste breukpunten komen leerlingen van christelijke VHMO-scholen tot betere resultaten dan leerlingen in de openbare of de katholieke sector. Zij kiezen vaker de B-kant, behalen vaker het diploma en leren vaker door, zowel met als zonder diploma.⁵ Alleen doubleren zij iets vaker dan katholieke leerlingen. De verschillen tussen de katholieke en de openbare sector zijn klein, maar – althans vóór controles – in het voordeel van katholieke leerlingen. Die doubleren minder, halen vaker het diploma en leren daarna in meer gevallen verder.

Tabel 2a. Richting-effecten op de *breukpunten* in de schoolloopbaan in het *VHMO*:

(a) vóór controles

(b) na controle voor de in tabel 2b opgenomen variabelen.

Richting van de school	Breukpunt in de schoolloopbaan									
	doubleren		keuze 'B'-afdeling of vakkenpakket		diploma behalen		doorleren met diploma		doorleren na schoolverlaten zonder diploma	
	a	b	a	b	a	b	a	b	a	b
openbaar	43%	44%	42%	40%	57%	60%	66%	67%	81%	81%
n =	407		244		407		234		173	
katholiek	38%	38%	37%	39%	61%	59%	73%	74%	79%	80%
n =	447		271		447		271		176	
christelijk	41%	40%	45%	45%	66%	65%	83%	81%	83%	82%
n =	262		175		262		173		89	
totaal	41%		41%		61%		73%		81%	
n =	1116		690		1116		678		438	

Tabel 2b. Resultaten van de covariantie-analyse.

onafhanke- lijke variabele	afhankelijke variabele				
	DOUB	PAKKET	DIPLOMA	DOOR met	DOOR zonder
R ² = waarvan voor rekening van	.15	.27	.31	.10	.39
TYPE	].11**	].15**	].16**	.02**	].14**
WEG				-	
DOUB	-	-	].06**	.04**	-
ADVIES	-			-	
PREST	-	-	.00**	-	-
DOUBLO	-	-		-	-
PAKKET	-	-	-	.01*	-
ANDOP	-	.01**	.04**	-	-
SEKSE	.01**	.04**	-	-	-
BERVA	.02**	-	-	.01**	.02**
OPVA	-	.02**	-	-	-
UITKLAS	-	-	-	-	.12**
EXCUR	.00 ⁺	-	.00 ⁺	-	-
EXPART	.00 ⁺	-	.03 ⁺⁺	-	-
Richting	.00	.00	.00	.02**	.00

Interacties van
richting en:

-

ANDOP*

ANDOP*
BERVA *
SEKSE *

BERVA*

** : (factor) } significant op .01 - niveau
 ++ : (covariaat) }
 * : (factor) } significant op .05 - niveau
 + : (covariaat) }

Door de invoering van controlevariabelen (gepresenteerd in *tabel 2b*) worden de verschillen in de meeste gevallen kleiner. Dat is vooral toe te schrijven aan verschillen in de verdeling over schooltypen binnen het VHMO. Het gevolgde schooltype heeft effect op alle breukpunten. MMS en HAVO komen daarbij naar voren als de wat 'gemakkelijkere' opleidingen (minder kans op doubleren, meer B-keuzes, meer kans op het diploma); HBS en Gymnasium zijn 'moeilijker', maar leiden vaker tot doorleren. Op katholieke scholen volgen leerlingen vaker de MMS- of HAVO-opleiding dan op de overige scholen (44% versus 36%). Dat gegeven heeft een belangrijk aandeel in de correcties bij het centrale breukpunt 'diploma behalen', waardoor het verschil tussen de katholieke en de openbare sector geheel verdwijnt. De gunstige positie van het christelijke VHMO wordt door de correcties nauwelijks geraakt.

Met de opgenomen variabelen wordt 10% (bij het doorleren met diploma) tot 39% (bij het doorleren na voortijdig schoolverlaten) van de variantie verklaard.⁶ Het richting-effect is slechts in één geval significant: bij het doorleren na het behalen van het diploma, waar het 2% van de variantie voor zijn rekening neemt (20% van de verklaarde variantie). Het verschil tussen leerlingen van openbare en christelijke VHMO-scholen is op dat punt dan ook groot: van de diploma-behalers in de openbare sector leert 67% door, in de protestants-christelijke sector 81%.

Interactie-effecten vinden we in het VHMO niet veel. Slechts drie controle-variabelen blijken bij de interacties betrokken. In *tabel 3* zijn de interactie-effecten op 'diploma behalen' weergegeven. Na controles behalen op confessionele scholen jongens vaker het VHMO-diploma dan meisjes, terwijl dit verschil op openbare scholen ontbreekt.

Het resultaat van leerlingen uit de verschillende milieugroepen verschilt naar de richting van de school. Op openbare scholen heeft het milieu van herkomst geen enkele invloed op de kans het diploma te behalen. Op confessionele scholen is dat wel het geval, maar op een verschillende manier. Kinderen uit de middengroepen halen in beide gevallen minder vaak het diploma. Als groep met de beste kansen komt in de katholieke sector de hoge milieugroep naar voren, terwijl

Tabel 3. Interactie-effecten in het VHMO:
percentage diploma-behalers*
naar richting van de school en
- sekse
- beroepsniveau vader
- aanwezigheid van meerdere opleidingen.

Controle variabele		Richting van de school							
		Openbaar		Katholiek		Christelijk		Totaal	
			n		n		n		n
SEKSE	jongens	59%	242	62%	283	71%	153	63%	678
	meisjes	59%	165	56%	164	59%	109	58%	438
BERVA	arbeider	59%	75	60%	80	72%	32	62%	187
	zelfstandige/	60%	163	54%	209	61%	121	58%	493
	lagere employee	60%	169	66%	158	67%	109	64%	436
	middelb. employee/ hogere beroepen								
ANDOP	categoriale school	51%	135	50%	124	48%	70	50%	329
	scholen- gemeenschap	63%	272	63%	327	71%	192	65%	787
(zonder interacties)		60%	407	59%	447	65%	262	61%	1116

* na controle voor de in tabel 2b opgenomen variabelen

in de christelijke sector kinderen van arbeiders het vaakst hun diploma halen (deze groep is in de christelijke sector kleiner dan in de beide andere sectoren, namelijk 12% versus 17% à 18% van het leerlingenbestand). De interactie met ANDOP tenslotte laat zien, dat het gunstige diplomaresultaat van het christelijke VHMO alleen op scholengemeenschappen wordt gerealiseerd. Leerlingen van categoriale christelijke scholen halen zelfs minder vaak hun diploma dan die van openbare of katholieke categoriale scholen.

Tabel 4. Schoolloopbanen in het VHMO naar richting van de school; *trajecten* langs de voornaamste breukpunten*

Doorlopen trajecten		Richting van de school						
		Openbaar	Katholiek	Christelijk				
zonder diploma	uitstroom	8%	8%	6%				
	afstroom	32%	33%	30%				
met diploma	A	40%	41%	41%				
					uitstroom	14%	11%	8%
					na doubl.	9%	9%	10%
					door-leren	17%	21%	23%
	B	20%	19%	23%				
					uitstroom	6%	4%	5%
					na doubl.	5%	6%	6%
					door-leren	9%	9%	12%
(100% =)		407	447	262				

* Na controle op elk breukpunt voor de in tabel 2b opgenomen variabelen.

In tabel 4 zijn de resultaten op de vijf breukpunten weer samengevoegd tot een typologie van schoolloopbanen in het VHMO. Drie gunstige effecten versterken elkaar in de christelijke sector: leerlingen in het christelijk VHMO kiezen vaker een B-richting, halen vaker hun diploma en leren vaker door dan de overige leerlingen. Daardoor stroomt van deze groep 51% met een diploma door naar het tertiair onderwijs, waarvan 33% met een A-diploma en 18% met een B-diploma. Dat is aanmerkelijk meer dan leerlingen van het openbare en katholieke VHMO, waarvan resp. 40% en 45% deze weg volgt. De resultaten in het openbare en het katholieke VHMO verschillen weinig, maar zijn licht in het voordeel van de katholieke sector. Op katholieke scholen wordt vaker het diploma gehaald, zij het in meer gevallen op MMS- en HAVO-niveau. Op openbare scholen halen iets meer leerlingen hun diploma met doubleervertraging en zetten minder leerlingen hun opleiding voort in het tertiair onderwijs.

Interessant tenslotte zijn twee interactie-effecten. Het eerste betreft jongens en meisjes (tabel 5). Bij de meisjes zijn de verschillen vrijwel geheel terug te voeren op verschillen in het gevolgde VHMO-type: op openbare scholen kiezen meisjes veelal de moeilijkere opleidingen (HBS en Gymnasium), terwijl zij op christelijke en katholieke scholen vaker MMS en HAVO kiezen.

Als daarvoor wordt gecontroleerd verdwijnen de verschillen in diploma behalen en in keuze voor de B-richting. Bij jongens komt het christelijk VHMO tot de beste resultaten en verschillen de openbare en katholieke sector weinig. Opmerkelijk is dat jongens op christelijke scholen veel

Tabel 5. Schoolloopbaan in het *VHMO* van *jongens en meisjes*, naar richting van de school*

Doorlopen trajecten		Richting van de school						
		Openbaar	Katholiek	Christelijk				
zonder diploma	uitstroom	38%	40%	32%				
	afstroom							
met diploma	A	62%	60%	68%				
					uitstroom na doubl.	13%	10%	8%
	doorleren					9%	8%	8%
					B	uitstroom na doubl.	14%	20%
	doorleren						9%	5%
						26%	22%	30%
	7%	8%	8%					

* Na controle per breukpunt voor de in tabel 2b opgenomen variabelen.

vaker een B-diploma halen, zo'n 10% meer dan op katholieke scholen. Op dat punt doet ook de interactie tussen richting en 'breedte' van de school zich gelden (tabel 6). Op categoriale scholen wordt in het christelijk VHMO veel vaker de B-kant gekozen dan in de beide andere sectoren, maar er wordt ook relatief weinig het diploma gehaald. Op protestants-christelijke scholengemeenschappen is niet het percentage B-keuzen hoger dan bij de rest, maar juist het diploma-rendement. Op beide soorten scholen realiseert het christelijk VHMO uiteindelijk meer B-diploma's.

5.2. Richting en schoolloopbaan in het ULO

De meest bezochte schoolsoort in het Van Jaar Tot Jaar-bestand is het ULO. Onder 'ULO' is hier ook het MAVO opgenomen, dat er in 1968 voor in de plaats kwam: bijna 10% van de leerlingen in ons ULO-bestand volgt het MAVO. In de schoolloopbanen op het ULO zijn zes

Tabel 6. Schoolloopbaan in het *scholengemeenschappen en categoriale scholen in het VHMO*, naar richting van de school*

Doorlopen trajecten		Richting van de school					
		Openbaar	Katholiek	Christelijk			
zonder diploma	uitstroom afstroom	37% 7% 30%	37% 8% 29%	29% 5% 24%			
met diploma	A	63% 41% 22%	63% 44% 19%	71% 49% 22%			
	uitstroom na doubl.				15%	11%	8%
					doorleren	8%	10%
	B				uitstroom na doubl.	7%	3%
doorleren		6%	6%	5%			
		9%	10%	12%			
<i>scholengemeenschap</i>							
(100% =)		272	323	192			
(100% =)		135	124	70			
<i>categoriale school</i>							
zonder diploma	uitstroom afstroom	48% 10% 38%	50% 8% 42%	52% 9% 43%			
met diploma	A	52% 35% 17%	50% 33% 17%	48% 21% 27%			
	uitstroom na doubl.				12%	14%	10%
					doorleren	9%	5%
	B				uitstroom na doubl.	3%	4%
doorleren		4%	5%	9%			
		10%	7%	15%			

* Na controle per breukpunt voor de in tabel 2b opgenomen variabelen.

breukpunten onderscheiden, die in *tabel 7* bijeen gebracht zijn.

De richting van de school heeft in het ULO, net als in het VHMO, invloed op het schoolsucces. Leerlingen van katholieke scholen behalen een beter resultaat op alle vier de elementen van een ononderbroken ULO-schoolloopbaan: zij doubleren minder, kiezen in meer gevallen een 'B'-examenpakket, halen vaker het diploma en blijven daarna vaker doorleren dan leerlingen van de beide andere richtingen. Zonder diploma opstroom naar het VHMO, een onderbroken maar succesvolle weg, komt op katholieke scholen minder vaak voor. De verschillen tussen de openbare en de christelijke sector liggen per breukpunt anders. Leerlingen van openbare ULO-scholen doubleren vaker, kiezen minder vaak een B-pakket en leren minder vaak door na het behalen van hun diploma; zij slagen er echter wel vaker in hun diploma te behalen of tussentijds op te stroom naar het VHMO.

Tabel 7a. Richting-effecten op de *breukpunten* in de schoolloopbaan in het ULO:

(a) vóór controles

(b) na controle voor de in tabel 7b opgenomen variabelen.

Richting van de school	Breukpunt in de ULO schoolloopbaan											
	doubleren		keuze 'B'- afdeling of vakkenpakket		diploma behalen		doorleren met diploma		doorleren na voortijdig schoolverlaten		opstroomen zonder diploma	
	a	b	a	b	a	b	a	b	a	b	a	b
openbaar n =	39%	38%	18%	17%	61%	58%	56%	57%	64%	69%	26%	28%
	442		286		442		268		174		112	
katholiek n =	32%	33%	23%	22%	66%	65%	67%	67%	69%	67%	19%	18%
	650		395		650		427		223		154	
christelijk n =	35%	34%	21%	22%	55%	58%	61%	61%	72%	70%	21%	22%
	489		290		489		271		218		157	
totaal n =	35%		21%		61%		62%		69%		22%	
	1581		971		1581		966		615		423	

Tabel 7b. Resultaten van de covariantie-analyse.

onafhanke- lijke variabele	afhankelijke variabele					
	DOUB	PAKKET	DIPLOMA	DOOR met	DOOR zonder	OP zonder
R ² =	.10	.07	.25	.13	.40	.35
waarvan voor rekening van						
TYPE	-	-	.00*	-	-	-
DOUB	-	-		.01 ⁺⁺		.02**
WEG	.01*	.01**	.02**	-	.03**	.01*
ADVIES		-	.01**	-	.02**	
PREST	.04**	-	.01**	.01**	-	.12**
PAKKET	-	-	-	.00*	-	-
SEKSE	.03**	.03**	-	.03**	.04**	-
BERVA	.01**	-	-	.04**	.03**	.05**
OPVA	-	.01*	-	.01 ⁺	.01 ⁺	-
OPMOE	-	-	-	-	.01 ⁺⁺	.03 ⁺⁺
DOUBLO	-	-	.01 ⁺⁺	-	-	.01 ⁺⁺
ANDOP	-	.01**	-	-	-	.01 ⁺⁺
EXPART	.01 ⁺⁺	-	.12 ⁺⁺	-	-	-
UITKLAS	-	-	-	.00 ⁺	.21 ⁺⁺	-
Richting	.00	.00	.01*	.01*	.00	.01*
Interacties van richting met:	SEKSE*	OPVA*	PREST* DOUB* WEG *	DOUB*	WEG*	ADVIES*

** : (factor) } significant op .01 - niveau
 ++ : (covariaat)

* : (factor) } significant op .05 - niveau
 + : (covariaat)

De gevonden verschillen kunnen ten dele worden toegeschreven aan verschillen in de samenstelling van de drie leerlinggroepen, want de statistische correcties leiden tot een verkleining van de verschillen op de meeste breukpunten. Vooral de groep leerlingen van protestants-christelijke scholen blijkt een relatief ongunstige uitgangspositie te hebben en wordt daarom in positieve zin gecorrigeerd. Toch veranderen de correcties weinig aan de positie van de richtingen ten opzichte van elkaar, met uitzondering van het breukpunt 'diploma behalen'. Daar verdwijnen de in eerste instantie vrij grote verschillen tussen leerlingen van openbare en christelijke ULO-scholen. Katholieke ULO-leerlingen halen nog steeds de beste resultaten.

De uitkomsten van de variantie-analyses zijn opgenomen in *tabel 7b*. De verklaarde variantie loopt van 7% bij de keuze van een B-examenpakket tot 40% bij het al dan niet doorleren na voortijdig schoolverlaten. Bij drie van de zes ULO-breukpunten kan van een significant richting-effect worden gesproken: diploma behalen, doorleren na behalen van het diploma en opstromen zonder diploma.

Op elk breukpunt is een of meer interactie-effecten van richting van de school en een controlevariabele significant. De omvang ervan is over het algemeen klein in vergelijking met het hoofdeffect van de controlevariabele, maar minstens even groot als het hoofdeffect van richting van de school. Hier bespreken we de meest interessante interactie-effecten, die op de afhankelijke variabele 'diploma behalen' (*tabel 8*).

Tabel 8. *Interactie-effecten in het ULO:*
percentage diploma-behalers*
naar richting van de school en
– prestatieniveau in de 6e klas lagere school
– schoolsoort van herkomst
– doubleerstatus.

Controle variabele		Richting van de school							
		Openbaar		Katholiek		Christelijk		Totaal	
			n		n		n		n
Prestatiescore in 6e klas glo	laag	50%	135	64%	157	47%	178	54%	470
	middel	56%	185	65%	277	61%	191	61%	653
	hoog	70%	122	69%	216	67%	120	68%	458
Schoolsoort van herkomst	lager ondw.	55%	351	65%	531	55%	428	59%	1310
	VHMO	79%	71	70%	98	81%	54	75%	223
	LBO/VGLO	44%	20	73%	21	34%	7	55%	48
Doubleer-status	niet	62%	269	67%	439	61%	320	64%	1028
	1 keer	50%	151	63%	189	53%	141	56%	481
	2 of meer keer	56%	22	68%	22	52%	28	59%	72
(zonder interacties)		58%	442	65%	650	58%	489	61%	1581

* na controle voor de in tabel 7b opgenomen variabelen

De drie interacties laten een overeenkomstig patroon zien: grote verschillen op openbare en christelijke ULO-scholen, geringe verschillen op katholieke scholen. Verschillen in prestatieniveau aan het eind van de lagere school hebben weinig effect op de kans dat een katholieke ULO-leerling het diploma haalt: het percentage leerlingen dat een diploma behaalt loopt op van 64 in de laagste tot 69 in de hoogste prestatie-categorie. Op openbare en christelijke scholen zijn de verschillen veel groter en liggen tussen ca. 50% en 70%. Bovendien scoren leerlingen in de middencategorie op openbare scholen dicht bij de lage categorie, op christelijke scholen dicht bij de hoogste categorie. Het katholieke ULO dankt zijn goede totaalresultaat dus niet

aan uitschieters naar boven, maar aan het goede resultaat bij leerlingen met een minder gunstige uitgangspositie. Ook verschillen in doubleerstatus en in voorafgaande schoolsoort hebben op katholieke scholen veel minder effect op het diploma behalen dan bij de andere richtingen. Doubleurs halen bijna net zo vaak of vaker het diploma als niet-doubleurs en omgekeerd zijn afstromers uit het VHMO veel minder in het voordeel ten opzichte van degenen die rechtstreeks van de lagere school komen.

Een samenvattend beeld van de schoolloopbanen in het ULO is gegeven in *tabel 9*, in de vorm van een typologie van trajecten langs de breukpunten. Bij katholieke leerlingen zijn er twee gunstige effecten die elkaar versterken: zij halen het vaakst hun diploma en van de gediplo-meerden leert het grootste deel verder. Ongeveer 45% belandt zo mét diploma in vervolgonderwijs. Dat is ruim 10% méér dan de leerlingen van openbare en christelijke ULO's (beide 34%). Leerlingen in de christelijke sector halen even vaak een B-diploma als in de katholieke sector, 11% à 12%; het grote verschil wordt gevonden bij de A-diploma's, 54% tegen 47%. De openbare en de christelijke sector houden elkaar ook in dit totaalbeeld in evenwicht. In het christelijke ULO worden meer B-diploma's gehaald en wordt vaker doorgeleerd, in het openbare onderwijs is er een iets grotere doorstroming zonder diploma naar het VHMO.

Tabel 9. Schoolloopbanen in het ULO naar richting van de school: *trajecten* langs de voornaamste breukpunten*

Doorlopen trajecten		Richting van de school		
		Openbaar	Katholiek	Christelijk
zonder diploma	uitstroom	13%	11%	12%
	afstroom	42% 20%	35% 19%	42% 24%
	opstroom	8%	4%	6%
met diploma	A	58%	65%	58%
	uitstroom	22%	19%	19%
	na	8%	10%	8%
	doubl.	18%	26%	21%
	door- leren	9%	12%	11%
	B	9%	12%	11%
	uitstroom	3%	3%	5%
	na	3%	2%	2%
	doubl.	3%	6%	6%
	door- leren			
(100% =)		442	650	489

* Na controle op elk breukpunt voor de in tabel 7b opgenomen variabelen.

6. CONCLUSIES

In dit artikel is nagegaan of de blijvende belangstelling voor confessionele scholen verklaard kan worden uit de kwaliteit van het confessionele onderwijs. Toegespitst op het voortgezet onderwijs luidt de onderzoeksvraag: heeft de richting van de school zelfstandig effect op de individuele schoolloopbaan in het secundair onderwijs? Het antwoord hebben we gezocht in afzonderlijke analyses van de schoolsoorten VHMO en ULO. De richting-effecten zijn uitvoerig gecontroleerd voor verschillen in achtergrond en eerdere onderwijsresultaten die mogelijk samenhangen met de richting van de school. Correctie voor deze compositie-verschillen leidt over het algemeen tot enige verkleining van de effecten, zonder evenwel de rangorde

van de richtingen ten opzichte van elkaar te veranderen. De gevonden effecten kunnen niet toegeschreven worden aan andere in het onderzoek betrokken variabelen.

De resultaten wijzen erop, dat de richting van de school effect heeft op belangrijke breukpunten in de individuele schoolloopbaan: zittenblijven, keuze van examenpakket, diploma behalen en doorstroming naar verder dagonderwijs. De verschillen zijn in het algemeen groter dan de verschillen die we eerder in het lager onderwijs vonden. De richting-effecten zijn per schoolsoort verschillend en vertonen op diverse punten interacties met de effecten van andere kenmerken.

In het VHMO komen leerlingen van protestants-christelijke scholen tot de beste resultaten en die van openbare scholen tot de minst goede. De gunstige effecten versterken elkaar bij leerlingen van het christelijke VHMO: zij kiezen vaker een B-examenpakket, behalen vaker het diploma en leren vaker door. Dat resulteert in een verschil in doorstroming mét diploma naar het tertiair onderwijs van ruim 10% met de openbare sector. Twee interacties spitsen de richting-effecten in het VHMO nader toe. Ten eerste levert de katholieke sector meer gediplomeerden af dan de openbare sector, maar dat blijkt grotendeels toe te schrijven aan de voorkeur voor MMS- en HAVO-opleidingen. In de openbare sector kiezen meisjes vrij vaak de HBS, een relatief moeilijke VHMO-opleiding; na controle daarvoor verdwijnt het richting-effect bij meisjes bijna geheel. Op categoriale scholen wordt in de christelijke sector relatief vaak de B-kant gekozen.

Op ULO-niveau zijn de verschillen aanzienlijk, deze keer in het voordeel van leerlingen van katholieke scholen. Op vier van de zes breukpunten komen zij tot betere resultaten: zij doubleren minder, kiezen vaker de B-kant, halen vaker het diploma en blijven daarna vaker doorleren. Ook hier versterken de effecten elkaar, zodat van de leerlingen ruim 10% méér met een diploma doorstromen naar vervolgonderwijs dan op scholen van andere richtingen. De openbare en christelijke ULO-sector houden elkaar over het geheel genomen in evenwicht. Ook de interactie-effecten lijken in het voordeel van het katholieke ULO. Leerlingen met een ongunstige uitgangspositie ondervinden daarvan in het katholieke ULO minder last dan in de beide andere sectoren.

De richting-effecten verschillen per schoolsoort en wijzen zeker niet op systematisch en eenduidig betere resultaten van een der richtingen op beide niveaus van voortgezet onderwijs. Toch kunnen voor elke richting enkele kenmerkende verschijnselen naar voren worden gebracht.

In de katholieke sector lijken middelpuntzoekende tendenzen werkzaam te zijn. Het duidelijkst is dat in het katholieke ULO, waar leerlingen met een ongunstige uitgangspositie tot vrij goede resultaten komen. Ook over de hele lijn komen leerlingen van katholieke scholen tot gemiddeld iets betere, maar minder gespreide resultaten dan de rest. Het goede eindresultaat komt tot stand in de vorm van relatief minder goede resultaten bij de meest kansrijke leerlingen en betere resultaten bij de overige. Bovendien heeft doubleren in het katholieke ULO én VHMO minder negatieve effecten dan bij de andere twee richtingen. Ook in het katholieke lager onderwijs wordt, zoals we eerder zagen, het meest gedoubleerd en toch het hoogste prestatieniveau gehaald. Het lijkt er dus op, dat in de hele katholieke sector het zittenblijven wordt gehanteerd als middel om leerlingen nog een extra kans te geven.

In het protestants-christelijke onderwijs treedt bijna het omgekeerde op: sterkere scheiding van laag- en hoogpresterenden naast betere prestaties bij leerlingen met een gunstige uitgangspositie. Relatief weinig leerlingen halen op de christelijke lagere school het hoogste prestatieniveau en gaan naar het VHMO. Toch blijkt dat op christelijke VHMO-scholen meer leerlingen van het hoogste prestatieniveau zitten dan in de rest van het VHMO en deze selectere groep komt tot iets betere resultaten. Bij de overgang van lager naar voortgezet onderwijs constateerden we een hoge instroom in het ULO van christelijke leerlingen met een LBO- of VHMO-advies. Die heterogene samenstelling van het ULO heeft tot gevolg dat leerlingen op het laagste aanvangsniveau matige resultaten behalen en veelvuldig weer afstromen naar het LBO. Leerlingen met een hoog aanvangsniveau behalen juist goede resultaten en stromen relatief vaak door naar het VHMO. In zowel ULO als VHMO worden op christelijke scholen relatief veel

B-diploma's gehaald. Al met al lijkt de christelijke sector tot goede resultaten voor een vrij selecte groep te komen.

Het openbare voortgezet onderwijs komt tot minder goede resultaten dan de beide confessionele sectoren. Het valt op door de relatief lage doorstroming van zijn leerlingen: Zowel leerlingen met een VHMO-diploma als die met ULO-diploma stromen minder door naar vervolgonderwijs dan in het confessionele onderwijs. Bovendien is op beide niveaus het aantal leerlingen van openbare scholen dat na schoolverlaten zonder diploma het dagonderwijs verlaat opmerkelijk groot. Dat is vooral bij jongens het geval en resulteert in een vrij groot aantal jongens in de openbare sector die zonder enig diploma het voortgezet onderwijs verlaten.

De uitkomsten van dit onderzoek moeten met enige voorzichtigheid worden gehanteerd. Het is nog niet duidelijk hoe de gevonden richting-effecten tot stand komen. Met name de vraag of er zoiets als een specifieke pedagogisch-didactische aanpak of een eigen schoolklimaat bestaat op openbare, katholieke en christelijke scholen, dient nader onderzocht te worden. Wellicht zijn de verschillen eerder terug te voeren op verschillen in de mate waarin bepaalde bevolkingsgroepen hun aspiraties middels het onderwijs proberen te realiseren. De verschillen in onderwijsresultaat kunnen dan samenhang vertonen met godsdienst en schoolrichting, zonder het resultaat te zijn van een karakteristieke onderwijspraktijk op scholen van een bepaalde signatuur.

De gebruikte gegevens bieden geen mogelijkheden richting-effecten op het niveau van de school te onderzoeken. In het voorgaande is gesproken van drie 'sectoren' en van verschillen tussen leerlingen die een school binnen een van deze sectoren bezoeken. Het is evenwel nog maar de vraag of de drie sectoren wel even homogeen zijn. De ene sector kan een veel heterogener scholenbestand hebben dan de andere (grote versus kleine scholen, categoriale versus brede scholen, standsscholen versus streekscholen, etc). Mogelijk worden in onze analyses verschillen binnen sectoren op een hoop gegooid en blijken de effecten bij nadere analyse groter te zijn.

De hier onderzochte leerlingen hebben ruim tien jaar geleden het voortgezet onderwijs verlaten. Het is niet onaannemelijk dat juist in de tussenliggende periode de veronderstelde kwaliteit van het onderwijs sterker in de schoolkeuze is gaan meespelen dan levensbeschouwelijke overwegingen. Wij constateren 'neven-effecten' van katholiek, protestants-christelijk en openbaar onderwijs op het schoolsucces van hun leerlingen. Dit wordt mogelijk door ouders als zodanig waargenomen en bij de keuze voor een concrete school betrokken. Vooral onder katholieken lijkt religieuze binding als schoolkeuze-motief aan betekenis te hebben verloren; het is ook deze groep die in de kwaliteit van het katholieke onderwijs voldoende reden heeft om aan onderwijs van de eigen richting vast te houden. Op VHMO-niveau komt ook het christelijk onderwijs tot goede resultaten.

Juist nu ook in het onderwijs de ontzuiling verwacht mag worden is wellicht een nieuw mechanisme in werking getreden dat ouders binding geeft met het onderwijs op confessionele scholen: de veronderstelde kwaliteit van het gegeven onderwijs. Dat kan ten dele verklaren waarom een uit de verzuiling overgeleverde bestuursstructuur gehandhaafd kan blijven, ook na de secularisering en ontzuiling van grote delen van de Nederlandse samenleving.

Noten

- 1 Al rond 1965 verklaarde een derde van de katholieke ouders zich bereid hun kinderen naar een 'algemene gezindten'-school te sturen (Van Kemenade 1968: 145 ev; een algemene gezindten-school omschrijft Van Kemenade als een school voor alle gezindten, waarin wel iedere groep zijn eigen godsdienst-onderwijs krijgt.). De instemming met dit alternatief was overigens lager waar het de lagere school betreft.

In 1979 had naar schatting twee derde van de ouders die zich tot een godsdienstige gezindte rekenen voorkeur voor confessioneel onderwijs (Sociaal en Cultureel Rapport 1980). Dat komt nagenoeg overeen met resultaten uit het onderzoek '*Opnieuw: God in Nederland*' (Goddijn e.a. 1979, aangehaald

bij Pelkmans 1984). 45% van de ondervraagden (waarvan $\pm 70\%$ zich tenminste zwak gelovig achtte) sprak een behoefte uit aan confessioneel onderwijs; in 1966 was dat nog 56% (de afname in het onderwijs was overigens kleiner dan die in andere sectoren).

- 2 De drie *subsamples* die voor het Van Jaar Tot Jaar-onderzoek zijn getrokken uit de Nationale Steekproef van het CBS, zijn:

a. een proportionele steekproef	: 1845 lln.
b. een oversample van ULO- en VHMO-leerlingen	: 851 lln.
c. een extra groep bestaande uit alle (nog niet onder (a) of (b) opgenomen) leerlingen die vóór oktober 1968 (na drie jaar voortgezet onderwijs dus) van schooltype gewisseld hebben	: 346 lln.
	3042 lln.

Voor de analyse van het lager onderwijs hebben we – om daar aangevoerde redenen – alleen de proportionele steekproef gebruikt. Gebruik van het hele bestand leidt tot een oververtegenwoordiging van leerlingen die ULO of VHMO bezoeken t.o.v. LBO-leerlingen en van jongens t.o.v. meisjes. In de analyse van het voortgezet onderwijs beperken we ons tot de schoolloopbanen binnen een bepaalde schoolsoort. Bovendien vertoont de steekproefcategorie waartoe de leerling behoort geen samenhang met de richting van de bezochte school voor voortgezet onderwijs. Daarom verwachten we van gebruik van de hele dataset geen invloed op eventuele richting-effecten binnen een schoolsoort.

- 3 Gezien het onwaarschijnlijk grote aandeel van de 'openbare' sector in vergelijking met de landelijke verdeling lijkt hier sprake van een verwarring van de richtingen 'openbaar' en 'algemeen-bijzonder'. Dat is waarschijnlijk veroorzaakt door het ontbreken van inzicht bij de onderzoekspersonen in de 'bijzondere' beheersvorm van hun school. We hebben enige analyses uitgevoerd met de aanname, dat de onbetrouwbaarheid zich beperkt tot de niet-confessionele sector en dat de richting van een katholieke of protestants-christelijke school wel goed gemeten is. Uit die analyses komen zo goed als geen verschillen tussen richtingen naar voren.
- 4 Een impliciet uitgangspunt bij de wijze waarop deze controleprocedure door ons is gehanteerd is, dat *alle* verschillen in samenstelling van de drie richting-groepen 'toevallig' zo gelden voor onze steekproef, maar niet noodzakelijkerwijs voor de populatie. Zowel verschillen in de individuele achtergrond van de leerlingen als verschillen in kenmerken van de scholen die zij bezoeken worden zoveel mogelijk geëlimineerd. Deze aanpak is streng te noemen; de kans is groot dat op het eerste oog aanwezige verschillen tussen richtingen door de statistische controles kleiner worden. Wat overblijft als mogelijk richting-effect is een hoeveelheid nog niet geïnterpreteerde, maar met de richting van de school samenhangende variantie.

Het is evenwel goed mogelijk – maar alleen met daarop toegesneden onderzoek op schoolniveau na te gaan – dat een aantal van de hier uitgeschakelde verschillen systematisch met de richting van de school samenhangt. Als bijvoorbeeld de VHMO-scholen van een bepaalde richting systematisch bezocht worden door leerlingen met betere schoolcapaciteiten of als deze scholen meer leerlingen aan een B-diploma helpen (desnoods door hen een jaar langer erover te laten doen), dan is evenzeer sprake van een richting-effect. Dergelijke 'ruwe' verschillen tussen richtingen komen er zelfs eerder voor in aanmerking, door ouders als richting-effect te worden waargenomen en in hun schoolkeuze mee te spelen; ouders voeren immers meestal geen statistische controles uit. Om deze reden verdient het aanbeveling ook aan de hier gepresenteerde verschillen 'vóór controles' aandacht te besteden.

- 5 De in §5.1 en §5.2 gepresenteerde schoolloopbaanresultaten kunnen niet representatief worden geacht voor de schoolloopbanen op VHMO en ULO in de eerste jaren na 1965. Door de opnemings van de twee extra subsamples uit het Van Jaar Tot Jaar-bestand gaan per schoolsoort specifieke afwijkingen optreden.
 - a. Het ULO-bestand groeit van ca. 700 tot bijna 1600 leerlingen. Dat leidt tot een oververtegenwoordiging van niet-succesvolle leerlingen: in de proportionele steekproef (PS) haalt 68% het ULO-diploma, in het hele bestand 61%.
 - b. Het VHMO-bestand groeit van ca. 450 tot ruim 1100 leerlingen. Ook hier vinden we een verschil in schoolsucces: in de PS verlaat 75% het VHMO met diploma, in ons bestand 61%. Deze afwijkingen vertonen geen samenhang met de richting van de school. Vergelijking van onze resultaten met die van Collaris en Kropman wijst uit, dat in de PS de verschillen tussen de richtingen groter zijn dan in onze bestanden. De rangorde van de richtingen verandert echter in geen van beide gevallen:

Percentage leerlingen in de proportionele steekproef dat het *de school met diploma* verlaat, naar *richting van de school*:

Richting	VHMO	N	ULO	N
openbaar	66%	157	70%	193
RK	77%	175	72%	290
PC	85%	100	60%	200
totaal	75%	432	68%	683

bron: Collaris en Kropman 1978, tabellen 5.42 en 6.35.

- 6 Aan de hoogte van de R^2 moet in dit verband niet te veel belang worden gehecht. Die is grotendeels afhankelijk van het verklarend vermogen van de controle-variabelen, wat hier niet ter discussie staat. De hoge R^2 die bijvoorbeeld bij 'doorleren zonder diploma' wordt bereikt komt voor een aanzienlijk deel voor rekening van de variabele UITKLAS, die het leerjaar aangeeft waarin de leerling de school voortijdig verliet.

Door onderlinge samenhang van de afzonderlijke variabelen sommeren hun bijdragen niet steeds tot de totale R^2 .

LITERATUUR

- Boef-van der Meulen, S. (1982). Wat is verlangd onderwijs? in: J.M.G. Leune en A.M.L. van Wieringen (red.), *Onderwijsplanning als middel van onderwijsbeleid*. Harlingen, 35-50.
- Coleman, J.S., Hoffer, Th. & Kilgore, S. (1982): *High School Achievement*. Public, Catholic, and Private Schools Compared. New York.
- Collaris, J.W.M., & Kropman, J.A. (1978), *Van Jaar Tot Jaar*. Tweede fase. Nijmegen.
- Jonge, J. de (1978). De motivatie voor protestants-christelijk onderwijs, in: L. Box et al., *Vrijheid van onderwijs*. Nijmegen.
- Hoffer, Th., Greeley, A.M., & Coleman, J.S. (1985). Achievement Growth in Public and Catholic Schools. *Sociology of Education* 58, 74-97.
- Kemenade, J.A. van (1968). *De katholieken en hun onderwijs*, Meppel.
- Laarhoven, P. van, Bakker, B., Dronkers, J., & Schijf, H. (1986). Some aspects of school careers in public and non-public primary schools. *Tijdschrift voor Onderwijsresearch*, 11, 83-96.
- Pelkmans, A.H.W.M. (1984), *Samenwerkingsscholen in ontwikkeling*. Nijmegen.
- Randsdorp, L., & Dronkers, J. (1982), *Verzuiling in het onderwijs en het beleid van een bijzonder schoolbestuur*. Amsterdam (SISWO).
- Thurlings, J.M.G. (1978). *De wankelende zuil*. Nederlandse katholieken tussen assimilatie en pluralisme. Deventer.
- Wildt, A.R., & Ahtola, O. (1978), *Analysis of covariance*. Beverly Hills.
- Yogev, A., & Chen, M. (1985). Sponsorship as school charter: educational mobility in religious versus secular schools in Israel. *International Review of Modern Sociology* 12.

Manuscript ontvangen 26-5-1986.

Definitieve versie ontvangen 17-11-1986.

Boekbesprekingen

Franssen, H. A. M., *Plannen van onderwijsleersituaties als teamactiviteit*. Een curriculumtheoretische verkenning. Dissertatie, vakgroep Onderwijskunde Rijksuniversiteit Utrecht, 1985.

Het proefschrift van H. Franssen heeft betrekking op zowel het beschrijven alsook het waarderen op kwaliteit van ontwikkelingsactiviteiten van schoolteams in het basisonderwijs aangaande voorbereiding, inhoudsbepaling en vormgeving vooraf van onderwijsleersituaties.

De theoretische grondslag van het onderzoek ligt enerzijds op het terrein van de curriculumtheorieën en vraagstukken van curriculumontwikkeling, anderzijds op het terrein van planningstheorieën. Het onderzoek en de onderzoekssetting is gelegen in het kader van het Curvo-project. Het onderzoek is verricht aan de hand van een beschrijvingsmodel, dat is ontleend aan Dror, dat op zijn betrouwbaarheid en validiteit wordt getoetst. De beschrijving van de ontwikkelingsactiviteiten wordt gevolgd door een kwaliteitsbeoordeling ervan met het doel de kwaliteit van ontwikkelingsprocessen te verbeteren.

Voor deze beoordeling ontwikkelt Franssen een 16-tal normen, die in overeenstemming zijn met het beschrijvingsmodel.

In deze kwaliteitsbeoordeling ligt de sterkte maar ook de zwakte van het proefschrift. Het is uiterst noodzakelijk om met het oog op de verbetering van kwaliteit en effectiviteit van ontwikkelingsprocessen op school criteria op te stellen en deze processen dien aangaande te evalueren.

De normen die in dit proefschrift worden gebruikt, zijn in hoofdzaak ontleend aan een 'ideaal' procesverloop en niet aan kwaliteitseisen van het 'produkt' van het proces, of de sociaal psychologische dan wel situationele aspecten van de organisatiecontext. Criteria waarom alleen eerstgenoemde normen zijn uitgewerkt en toegepast worden niet genoemd. Ook ontbreekt een toetsing van de 16 opgestelde normen, waardoor de kwaliteitsbeoordeling op losse (subjectieve) schroeven komt te staan.

Naast het zwakke gedeelte van de kwaliteitsbeoordeling bevat het proefschrift een helder overzicht van het curriculum theoretisch denken, een nauwkeurig uitgewerkt beschrijvingsmodel, illustratieve toepassingen op een aantal konkrete projecten en interessante perspectieven voor verder onderzoek.

G. Bergenhenegouwen,
Universiteit van Amsterdam

Janssens, Frans J. G. *De evaluatiepraktijken van leerkrachten*. Een beschrijvend onderzoek naar het evalueren tijdens het rekenen in het primair onderwijs. Proefschrift R.U. Groningen, Arnhem, 1986.

In Janssens's proefschrift worden de evaluatiepraktijken van leerkrachten in het basis- en speciaal onderwijs beschreven en wordt van een aantal factoren onderzocht of ze van invloed zijn op die evaluatiepraktijken. Janssens maakt direct een inperking naar het evalueren van schriftelijk werk van leerlingen: dat wil zeggen werkbladen, schriftelijke opgaven, toetsen.

Dat is een enorme inperking, beter: beperking van het onderzoek. Uit de literatuur wordt steeds duidelijker dat leerkrachten het onderwijs evalueren door na te gaan of de leerling geconcentreerd werkt, of de instructie snel begrepen is, of de dagtaak afgerond is en of de

leerling en leerkracht een plezierige werkdag gehad hebben (Huberman, 1983; Shavelson, 1983). Deze wijze van evalueren wordt gebaseerd op de dagelijkse voortgangscntrole met behulp van vooral informele technieken. Dat maakt onmiddellijk duidelijk dat veel van de op de evaluatie van leerlingvorderingen gerichte waarnemingen van de leerkracht, van andere dan schriftelijke bronnen komen. Janssens heeft daarmee een kleiner en wat formeler deel van de evaluatiepraktijk te pakken. Dat deel is daarom niet minder belangrijk. Onderzoek van onder andere Mirkin (1981) toont aan dat leerkrachten nogal eens een onjuist beeld van hun leerlingen hebben en op grond daarvan tot niet optimale onderwijshandelingen komen. Het wat formelere deel van de evaluatiepraktijk is dé kans voor leerkrachten om de beeldvorming op grond van informele evaluaties regelmatig te controleren. Het is bijna de evaluatie van de evaluatie. Uit Mirkins onderzoek blijkt overigens dat enige formalisering van de evaluatiepraktijk tot betere oordelen leidt. Uit Janssens' proefschrift wordt niet echt duidelijk welke rol Janssens de formele technieken in het geheel van evaluatieactiviteiten geeft; wel worden in een van de laatste paragrafen van het boek (6.2.3.) de professioneel ontwikkelde toetsen als controle op de bevindingen van de methodegebonden toetsen opgevoerd.

Janssens ordent de beschrijving van de evaluatiepraktijk van leerkrachten op een originele wijze met behulp van de configuratie-analyse, geïntroduceerd door Hall en Loucks. In de configuratie worden alle redelijkerwijs denkbare werkwijzen van leerkrachten met betrekking tot de componenten van het evalueren, bijvoorbeeld instrumentgebruik, interpretatie, maatregelen of registratie, opgenomen. Die werkwijzen worden per component geordend van niet tot zeer acceptabel, waarbij ook omslagpunten gegeven worden tussen de beslist niet en de wel te accepteren werkwijzen. Die ordening, de omslagpunten en de beschrijving van de meest acceptabele werkwijze zijn gebaseerd op een analyse van de prescriptieve literatuur. Deze zorgvuldig uitgevoerde werkwijze van Janssens bij de beschrijving van de evaluatiepraktijk lijkt me zeer acceptabel (minstens niveau 5) en ook in ander onderzoek goed toepasbaar. Een probleem in Janssens' studie is echter wel, dat de stap van de interviewgegevens naar de data in de configuratie groot is. Een aanzienlijk deel van het interview bestaat uit open vragen: het risico bij deze werkwijze is, dat respondenten te weinig (eenduidige) informatie geven om de keuze voor het juiste niveau in de configuratie te maken. Janssens heeft het probleem bestreden met meerdere beoordelaars en het zonodig opvragen van aanvullende gegevens. Het is de vraag of dat voldoende geweest is. In tabel 5, bijvoorbeeld, kloppen de gegevens onderling niet helemaal. Uit de tabel blijkt dat 79 leerkrachten extra toetsen voor de resultaatbeoordeling gebruiken (instrumentgebruik, niveau 3), terwijl maar 52 leerkrachten hun resultaatbeoordeling mede baseren op deze toetsen (resultaatbeoordeling, niveau 2 en 3). Als afgezien wordt van de verklaring dat 27 leerkrachten tests afnemen zonder ze te gebruiken, dan lijkt hier in de vertaling van de antwoorden op de open interviewvragen (21 en 31) naar de niveaus in de configuratie iets mis te zijn gegaan. Het lijkt raadzaam de dataverzameling meer door de inhoud van de configuraties te laten sturen. Dat kan onder andere door alle verschillen tussen de niveaus om te zetten in gesloten vragen en van te voren een combinatieregels voor de classificatie op te stellen.

Janssens heeft drie configuraties opgesteld voor de beschrijving van het evalueren met behulp van schriftelijk materiaal; te weten: één voor het gebruik van oefenwerk en methodegebonden toetsen; één voor de algemene niveautoetsen van het CITO en één voor de leerdoelgerichte toetsen van het CITO. De eerste configuratie beslaat het brede terrein van het evalueren door leerkrachten, terwijl de twee andere betrekkelijk geïsoleerd ingaan op het toetsgebruik. Als het gebruik van algemene niveautoetsen en leerdoelgerichte toetsen inherent aan evalueren is, dan zouden deze configuraties eigenlijk onderdeel van de eerste, meer omvattende moeten zijn, bijvoorbeeld als uitbreiding van de componenten 'interpretatie en feedback' en 'didactische maatregelen'. Nu wordt de indruk gewekt dat evalueren beschreven wordt in de eerste configuratie en dat correct toetsgebruik, beschreven in de andere configuraties, iets is wat daar los van staat.

De resultaten van Janssens' studie zijn volstrekt duidelijk: er schort nog heel wat aan de wijze waarop leerkrachten het werk van hun leerlingen evalueren: geen goede bepaling van de beginsituatie en resultaatbeoordeling, een inefficiënt registratiesysteem, onvoldoende rapportage en feedback. Het percentage leerkrachten dat op een aantal belangrijke componenten op acceptabele wijze evalueert is erg laag: 1% (blz. 111).

Er is niet veel onderzoek naar het effect van evalueren op de kwaliteit van het onderwijs. Empirisch gefundeerde uitspraken over de mate waarin leerkrachten een onjuist beeld van hun leerlingen hebben en daardoor suboptimale beslissingen nemen zijn niet te doen, maar – met Janssens' studie in de hand – is enige zorg gerechtvaardigd.

Janssens wijkt niet af van de traditie te besluiten met een aantal voorstellen voor verder onderzoek. Dat onderzoek is zeker nodig. Janssens heeft daaraan met z'n goed leesbare studie over een belangrijk deel van de evaluatiepraktijk van leerkrachten een waardevolle bijdrage geleverd.

LITERATUUR

- HUBERMAN, M. (1983) Recipes for busy kitchens: a situational analysis of routine knowledge use in schools. *Knowledge: creation, diffusion, utilization*, 4, 478-510.
- MIRKIN, P., FUCHS, L., TINDAL, G., CHRISTENSON, S. and DENO, S. (1984) The effect of IEP monitoring strategies on teacher behavior. *Research Report No. 62*, IRLD, University of Minnesota, Minneapolis.
- SHAVELSON, R.J. (1983) Review of research on teachers' pedagogical judgments, plans and decisions. *The elementary school journal*, 4, 392-413.

S.J. Pijl
Rion, Groningen

De Jong, M. J. (Red.), *Allochtone kinderen op Nederlandse scholen*. Onderwijskundige studies, deel 5. Swets & Zeitlinger b.v., Lisse, 1986.

Over de positie van kinderen uit etnische minderheidsgrepen in het onderwijs is er de laatste tien tot vijftien jaar nogal tamelijk veel geschreven. Men denke maar aan de talrijke publicaties in onderwijsbladen en de tientallen doctoraalscripties die telkens weer het etnocentristische begrip 'allochtoon' in de titel opnemen.

Een bezwaar dat men zou kunnen inbrengen tegen hetgeen tot nu toe is geschreven, is het feit dat de meeste van bedoelde publicaties een speculatief karakter hebben, al of niet gebaseerd op directe ervaring of oppervlakkige waarneming van de auteurs.

Het boek 'Allochtone kinderen op Nederlandse scholen; prestaties, problemen en houdingen' (Swets & Zeitlinger/Lisse 1985; 157 blz.) hoort bij de trend van de jaren tachtig, waarbij het thema meer empirisch wordt benaderd. Het is een onmiskenbaar feit dat beleidsmaatregelen in deze pas succesvol kunnen worden afgedwongen als er 'harde' cijfers m.b.t. de positie van onze 'visible minorities' op tafel kunnen worden gelegd. En die cijfers liegen er niet om: leerlingen uit etnische minderheidsgroepen, vooral die van Turkse en Marokkaanse afkomst, vertonen een grote 'achterstand' t.o.v. de blanke Nederlandse leerlingen (de auteurs spreken van 'autochtone leerlingen'). Zelfs op zgn. concentratiescholen, waarvan men mede op grond van sociaal-economische factoren zou mogen vermoeden dat het alle leerlingen slecht vergaat, blijken blanke Nederlandse leerlingen het best te presteren. Ook de door leerlingen uit etnische minderheidsgroepen in het voortgezet onderwijs bereikte resultaten blijken zeer mager: onge-

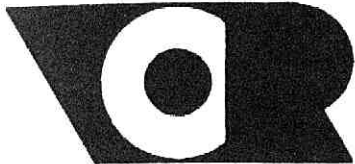
veer de helft van deze leerlingen behaalde een diploma, slechts 11% een diploma op avo-niveau. Voegt men hieraan toe de geconstateerde 'Nederlandse taalachterstand' en de daaraan gekoppelde 'onderschatting van de verbale intelligentie', dan lijkt het sombere beeld aardig compleet.

Het tweede deel van het door M.J. de Jong geredigeerde boek doet minder somber aan: op het sociaal-emotionele vlak blijken onze Surinaamse, Turkse, Marokkaanse en andere leerlingen uit etnische minderheidsgroepen minder problematisch te zijn. Dit geldt met name voor hun 'identiteitsproblemen' (een typering die overigens niet gevrijwaard is van 'ethnic-bias'). Van een identiteitsprobleem is er sprake 'als een persoon door zijn significante omgeving (aan wiens oordeel hij zich niet kan onttrekken) een identiteitsstempel krijgt opgelegd (vanwege een kenmerk b.v. huidskleur) dat hem dwingt zichzelf volgens dat stempel te definiëren terwijl hij tegelijkertijd omwille van die categorale indeling negatief gewaardeerd, geminacht of gediskrimineerd wordt' (p. 84/85). De gekozen definitie die in feite de 'etnische identiteit' tracht te verduidelijken, staat haaks op de sociale werkelijkheid, waarbij de persoon in het kader van een 'ruil' probleemloos de 'labels' van zijn significante omgeving kan overnemen en zich ook daarnaar gedragen ('self-fulfilling prophecy').

Bovendien wordt in de gekozen definitie teveel het accent gelegd op externe factoren (in dit geval door de dominante categorie bepaald). De zelf-definitie van de etnische identiteit en het al of niet kunnen stabiliseren van een consistente organisatiestijl van de eigen ervaringen blijven in de gekozen definitie voor 'identiteitsproblemen' buiten beschouwing. Jongeren uit raciaal-gemengde huwelijken kunnen in etnische zin 'identiteitsproblemen' ('ethnic-identity confusion') hebben, puur al omdat de significante omgeving de door die jongeren zelf gekozen categorisering destabiliseert! En dat hoeft niet perse door minachting of discriminatie. Dat kan al gebeuren door zo'n jongere 'allochtoon' of 'buitenlander' te blijven noemen, hoewel hij/zij zichzelf etnisch definieert in een richting die meer in overeenstemming is met die van zijn/haar Nederlandse moeder (of vader).

De bouwstenen voor beleidsvorming van de in het boek gepresenteerde onderzoeken zijn over het algemeen relevant, consistent afgeleid en helder gepresenteerd. En dat is al heel wat. Hierbij denk ik niet in de laatste plaats aan het onderzoek van Willem Fase en Gerda van den Berg over beleid en praktijk van intercultureel onderwijs. Dit alles neemt niet weg dat door het hele boek heen de deficiet-theoretische opvatting, waarbij kinderen uit etnische minderheidsgroepen en hun ouders in een pathologisch aandoend referentiekader worden geplaatst, zich voortdurend aan de lezer opdringt. De manier waarop de machtsverdeling en de principes van sociale controle regulerend optreden ten aanzien van de verdeling, de receptie, de participatiegraad en de verandering van dominante categorieën, blijven geheel buiten beschouwing. In Engeland waar ook termen als 'allochtonen' ('aliens') en daarvan afgeleide varianten zonder meer worden afgewezen, heeft de titel van een onderzoeksrapport 'West-Indian children in our schools' vanwege het 'our' tot felle protesten geleid. Het lijkt mij in dit verband erg zinnig dat de redacteur van 'Allochtone kinderen op Nederlandse scholen' ten behoeve van alle andere waardevolle bijdragen die we zeer zeker van hem kunnen verwachten, kritisch-ideologisch op de titel van het onderhavige boek reflecteert. Dit advies geldt uiteraard niet alleen voor de redacteur, maar ook voor al die auteurs die een bijdrage hebben geleverd aan de totstandkoming van het boek.

Henry G. Dors,
Amsterdam



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie

Kees Beintema (RUL), Peter van den Dool (SCO), Dirk de Jong (LICOR), Ewoud Roede (SCO) en Peter van der Sijde (UT, TO, secretaris).

Redactie-adres

Peter van der Sijde, Faculteit der Toegepaste Onderwijskunde, Vakgroep Instructietechnologie, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893560.

Jrg. 11, nr. 1

Februari 1987

Inhoudsopgave

Van de redactie	46
Verslag VOR-ledenvergadering dd. 13 november 1986	46
Problemen verkennen of problemen maken? De rol van de probleemverkenner in het programmeringsproces van de S.V.O. P. van den Dool	48
De OTG's en de programmering van onderwijsonderzoek E. Roede	52
Mededelingen	56
Congressen, Symposia, Cursussen, Trainingen	57

Van de redactie

Dit is het eerste nummer van de 11de jaargang van het VOR Bulletin. In het samengaan met het Tijdschrift voor Onderwijsresearch blijft het Bulletin redactioneel hiervan onafhankelijk: Het Bulletin blijft onder een eigen redactie bestaan.

Een aantal wijzigingen zijn echter wel het gevolg van het samengaan. De belangrijkste hiervan is dat nu ook de uitgever Swets & Zeitlinger het Bulletin gaat uitgeven. Gevolg: het Bulletin gaat vanaf heden officiëel 6 maal per jaar verschijnen en wel in de maanden: februari, april, juni, augustus, oktober en december.

Iedereen die copij wil aanleveren dient hiermee rekening te houden, de uiterste data om copij in te leveren zijn:

februari nummer	1 november
april nummer	1 februari
juni nummer	15 maart
augustus nummer	15 mei
oktober nummer	15 juli
december nummer	15 september

De redactie hoopt dat iedereen met deze data rekening wil houden en haar tijdig de bijdragen voor de verschillende rubrieken wil toesturen. Met name de rubriek 'Scripties' willen wij onder uw aandacht brengen (zie jrg. 10, nr. 5 voor informatie).

Verslag VOR-ledenvergadering d.d. 13 november 1986

Aanwezig: Hoebe, Slavenburg, Claessen, Roede, De Jong, Span, Gerritsen van der Hoop, v.d. Werf, Janssens, v.d. Sijde.

1. Opening

2. Verslag VOR-ledenvergadering 15 mei 1986

Naar aanleiding van:

punt 4 ten aanzien van themapanning ORD:

V.d. Sijde merkt op dat het artikel in het VOR bulletin binnenkort moet verschijnen, ook naar aanleiding van afwezigheid van het thema Onderwijsleerprocessen op het ORD-programma.

Geantwoord wordt dat het VOR-bestuur heeft besloten de vier vaste thema's niet uit te breiden en wel om twee redenen:

- a. eerst moet ervaring worden opgedaan met de vaste thema's;
- b. er moeten mogelijkheden open blijven voor plaatselijke bijdragen.

Het thema Onderwijsleerprocessen komt op de ORD 1987 aan de hand van het onderwerp Diagnostiek aan de orde. Ook in de vrije paper sessies is ruimte voor presentatie van onderzoek naar onderwijsleerprocessen. De voorstellen voor deze invulling van de thema's is voorgelegd aan de commissie ORD-themapanning. Deze ging accoord en vervolgens kreeg het programma de goedkeuring van het VOR-bestuur.

Naar aanleiding van dit punt wordt voorgesteld voortaan de thema's zoals vastgesteld ook zo te blijven noemen en ook op de bundels te vermelden. Deze suggestie zal worden doorgegeven aan de ORD organisatiecommissie.

punt 6 Op de vraag waarom bij de ORD thema's geen aandacht wordt geschonken aan de universitaire Lerarenopleidingen wordt vermeld dat aan Warries is geschreven dat dit thema aan de orde is bij het thema Scholing en Opleiding.

Hierna wordt het verslag ongewijzigd goedgekeurd.

3. Bestuursverkiezing

Naar aanleiding van de criteria voor selectie van bestuursleden zijn kandidaten gezocht.

Voorgesteld wordt in functie te benoemen:

Claessen: voorzitter

Roede: penningmeester

v.d. Werf: secretaris.

Er zijn geen tegenkandidaten. Met het voorstel de nieuwe leden te aanvaarden wordt accoord gegaan. Aldus besloten.

4. Begroting 1987

Er is geen behoefte aan een toelichting. Bij de begroting van de uitgaven moet één voorbehoud worden gemaakt. Afhankelijk van de onderhandelingen zal aan Swets en Zeitlinger 77,50 of meer van de ledencontributie moeten worden afgedragen. De begroting wordt ongewijzigd vastgesteld.

5. Samenstelling kascontrole commissie 1987

Dit agendapunt is abusievelijk opgevoerd. Pas bij de volgende ledenvergadering is sprake van kascontrole. Kandidaten voor 1988 worden dringend verzocht zich kenbaar te maken bij de penningmeester.

6. Rondvraag

- verzoek v.d. Sijde om namens het bestuur een brief te sturen naar de drukker van het VOR bulletin om deze op de hoogte te stellen van opname van VOR-bulletin in TOR en te bedanken voor bewezen diensten. De drukker is al wel op de hoogte gesteld.
- naar aanleiding van vraag Span: de volgende ledenvergadering is op de eerste dag van de ORD.
- naar aanleiding van vraag Span: de ORD organisatiecommissie vergadert zo weinig mogelijk. Er zijn enkele vergaderingen geweest over de thema's en begroting. Momenteel worden papervoorstellen ingewacht. Vanaf januari zal organisatiecommissie weer enkele keren moeten vergaderen.
- naar aanleiding van vraag Span: OTG's zijn benaderd voor het organiseren van symposia. Aan de themacoördinatoren is meegedeeld dat ze mensen kunnen uitnodigen voor thema's, onder voorbehoud dat genodigden geen voorkeursbehandeling krijgen bij beoordeling van papervoorstellen en opname van papers in bundels.

7. Dankwoord aan vertrekkende bestuursleden en sluiting van de vergadering.

Greetje v.d. Werf
Secretaris

Problemen verkennen of problemen maken? De rol van de probleemverkenner in het programmeringsproces van de S.V.O.

P.C. van den Dool
S.C.O.

INLEIDING

In deze bijdrage in de reeks van artikelen over de programmering van onderwijsonderzoek gaan we in op de rol van de probleemverkenner zoals die jaarlijks voor de S.V.O. op haar programmalijnen aan het werk worden gezet. We gaan in op een aantal van de problemen of dilemma's waarvoor een probleemverkenner in z'n werk komt te staan. Daarbij zal blijken dat een groot aantal van dat soort problemen mede veroorzaakt wordt door de specifieke positie van S.V.O. en door haar relatie met zowel opdrachtgevers als uitvoerders van onderzoek.

WAT DOEN ZE NU EIGENLIJK?

De werkzaamheden van een probleemverkenner en programmamaker zijn zeer divers. Er bestaan wel een aantal beschrijvingen van het type van activiteiten dat ondernomen moet worden (van Hoessel 1985, Van Ooijen 1985 en Van den Dool en De Jong 1984 en S.V.O. 1983). Ondanks het bestaan van deze, meestal in fraaie bewoordingen gestelde, beschrijvingen is er over de concrete activiteiten die een probleemverkenner en programmamaker onderneemt nog niet zeer veel duidelijkheid. In de diverse nota's (rapporten) van S.V.O.-probleemverkenner op diverse programmalijnen worden wel steeds een aantal typen van activiteiten genoemd. Wel is een bepaald patroon herkenbaar, waarbij er tussen de diverse probleemverkenningen en de werkwijze die daarbij is gevolgd nogal wat differentiatie bestaat. Zeer waarschijnlijk is een dergelijke differentiatie wel functioneel, gezien de verschillen tussen de probleemgebieden en met het oog op de kwalificaties van de betrokken probleemverkenner. Het gaat er in deze tekst niet in eerste instantie om vast te stellen wat de activiteiten van een probleemverkenner precies zouden moeten zijn, maar om een aantal van de dilemma's in dat werk aan te duiden.

Een aantal activiteiten blijken binnen de diverse probleemverkenningen steeds aan bod te komen.

Ten eerste is daar het verkennen van de probleemformuleringen en vragen om onderzoek van de betrokken beleidsmakers, overwegen die vanuit het ministerie van O&W en voor een aantal programmalijnen ook vanuit een aantal andere ministeries en de onderzoeksvragen zoals die geformuleerd zijn door vertegenwoordigers van de verzorgings- en de adviesstructuur of vanuit de onderwijskoepels. Dit verkennen van problemen kan sterk variëren in de mate van intensiteit, soms wordt er gewerkt met vrij uitvoerige en uitputtende vragenlijsten, soms wordt volstaan met het napluizen van beleidsdocumenten van de diverse instellingen waarin veelal, al dan niet expliciet topics voor onderzoek zijn geformuleerd. Die inventarisatie kan ook bestaan uit het gericht benaderen van een aantal personen waarvan reeds bekend is of verwacht mag worden dat ze ten aanzien van de mogelijkheden van onderzoek duidelijke opvattingen/of een goed overzicht hebben.

Een tweede type van activiteiten dat steeds opduikt is het wegen, beoordelen en analyseren

Adres: S.C.O., Grote Bickersstraat 72, 1013 KS Amsterdam

van de geïnventariseerde vragen op hun onderzoekbaarheid en op de mate waarin onderzoek ook werkelijk zou kunnen bijdragen aan het oplossen van de problemen waar de vragers van onderzoek voor gesteld staan. In de formuleringen van S.V.O., met een verwijzing naar Swanborn wordt gesproken van reconstructie van problemen en van beleidsopties, -maatregelen en -effecten. Waarschijnlijk zijn we hier aangeland op het belangrijkste maar tegelijkertijd ook het moeilijkste activiteitengebied van een probleemverkenner en programmamaker. Zoals bij het naslaan van Swanborn en ook andere literatuur blijkt zijn er weinig aanknopingspunten voor dit laatste type van activiteiten. Het onderzoekbaar maken en analyseren van de bijdragen van onderzoek aan de problemen van beleid en veld zijn in de sociale wetenschappen nog weinig methodisch uitgewerkt. Sowieso kan worden vastgesteld dat de methodologie van onderzoeksontwerp en zeker de problematiek van de keuze van een onderzoekstopic en de definiëring van het onderzoeksobject, in vergelijking met andere onderdelen van de methodologie nog steeds weinig sterk is uitgewerkt. Hoogstraten stelde in dit verband dat het vaak veel vruchtbaarder is om onderzoeksplannen in plaats van onderzoeksverslagen in het forum te bespreken. Niet alleen zou veel onderzoek dan niet uitgevoerd behoeven te worden ook de beeldvorming over optimale designs en objectformuleringen zou versterkt kunnen worden.

Het derde type van activiteiten dat steeds terugkeert is het prioriteren en selecteren op grond van de zojuist genoemde analyses van een beperkt aantal onderzoeksonderwerpen en vervolgens het uitwerken van een aantal schetsen voor onderzoek.

De opbrengst van dit laatste type van activiteit is meestal dat wat andere onderzoekers van de probleemverkenningen terugvinden in het onderzoeksprogramma van de S.V.O. Overigens is in de fase tussen het opleveren van de probleemverkenning- en programmeringsrapporten aan de S.V.O. en het verschijnen van een al dan niet groot deel van de voorgestelde onderzoeken in het onderzoeksprogramma en in de aanbestedingen van de S.V.O. al een heleboel gepasseerd, dat zich tot nu toe aan de waarneming van zowel probleemverkenners als andere onderzoekers onttrekt. Daarmee doelen we op

- a) Het hele proces van afwegen van programmalijnen ten opzichte van elkaar en het proces van wegen van de prioriteringen zoals door probleemverkenners aangegeven.
- b) De kwestie van het adopteren van de onderzoeksschetsen door enerzijds de vertegenwoordigers van MO&W en anderzijds de vertegenwoordigers van de koepelorganisaties in het bestuur van de S.V.O..
- c) De rol van het S.V.O.-bureau bij de voorbereiding van deze besluitvoering.

We willen niet op voorhand de indruk wekken dat de afwegingen niet voldoende degelijk zouden plaatsvinden, we signaleren slechts dat het toch bevreemdend is dat er over de diverse stappen in deze afwegingen zo weinig expliciet gerapporteerd wordt.

OM WIENS PROBLEMEN GAAT HET NU EIGENLIJK?

Uiteraard baseren alle probleemverkenners zich op de inventarisatie van problemen zoals ze die tot stand hebben gebracht. Dit neemt niet weg dat daarbij vaak ook sprake is van het aandragen van problemen bij de betrokkenen. Andersgezegd het kan zo zijn dat daar waar betrokkenen problemen/vragen voor onderzoek zien, een probleemverkenner (veelal terecht) kan aangeven dat een andere wijze van informatieverzameling beter op zijn plaats is. Andersom kan het zo zijn dat daar waar betrokkenen geen problemen zien, juist vanuit het gezichtspunt van de onderzoeker, bijv. op grond van zijn of haar eigen specialiteit, wel een aantal problemen en onderzoeksvragen gedefinieerd kunnen worden. In feite is in de structuur rondom de probleemverkenningen onvoldoende rekening gehouden met de vruchtbare wisselwerking die er kan ontstaan, indien onderzoeksvrager en probleemverkenner iets uitvoeriger met elkaar onderzoeksvragen en onderzoeksmogelijkheden, in een proces van loven en bieden en bijstellen van vraag en aanbod, tot stand kan komen. Dit wijst erop dat een wat sterkere zij het tijdelijke binding tussen vrager en aanbieder tot stand zou moeten komen. Misschien is het zo dat de

constructie van de programma-adviescommissie zoals die voor de eerste fase voor het v.o. wordt gevolgd aan dit probleem enigszins tegemoet kan komen. Anderzijds speelt hier waarschijnlijk een probleem op het vlak van de deskundigheid van zowel bij de personen aan de vraagkant die verantwoordelijk zijn voor de formulering en generering van onderzoeksvragen als aan de kant van probleemverkenner die (tot op zekere hoogte en alleen in een bepaalde fase van het programmeringsproces) zich in het gezichtspunt van de vrager moet kunnen verplaatsen. Het is verwonderlijk dat de S.V.O. aan deskundigheidsbevordering op dit vlak geen structurele aandacht geeft. Met name een zekere training van de probleemverkenner kan waarschijnlijk snel tot verbeterde resultaten leiden.

Er is uiteraard niks op tegen wanneer een probleemverkenner probeert om z'n eigen problematiek geadopteerd te krijgen door een van de potentiële klanten. Het is zelfs de vraag of het niet mogelijk zou zijn dit soort parachuteringen niet sterker systematisch in het hele programmeringsproces in te brengen. Wanneer de S.V.O. haar eigen positie definieert als makelaar op de onderzoeksmarkt dan is het in feite verwonderlijk dat wel de vraagkant wordt gestructureerd en ook wordt benoemd in termen van klanten en opdrachtgevers maar dat anderzijds de aanbodkant in termen van onderzoeksinstituten, groepen van onderzoekers en individuele onderzoekers in het hele proces zo weinig structureel wordt bewerkt. Anders gezegd wat zou er kunnen gebeuren op het moment dat van onderzoekers, instituten, lopende programmagroepen of clusters van onderzoekingen bij die instituten (en dus van méér dan alleen een x-tal probleemverkenner) gevraagd zou worden om in concrete termen in de zin van onderzoeksschetsen een onderzoeksaanbod te formuleren (daarbij ook denkend aan klanten uiteraard). Pas dan kan, om de vergelijking door te trekken, een reële makelaarssituatie ontstaan door voor de zoekende op de onderzoeksmarkt die naar een voor hun probleem geschikt onderzoeksontwerp op zoek zijn, ook een reële keuze te laten maken. Tegelijkertijd betekent het dat onderzoekers duidelijker hun eigen problematiseringen en hun eigen methodische en instrumentele oplossingen, die ze voor bepaalde problematieken zouden willen inzetten en uitwerken, ook duidelijk kunnen presenteren. Uiteraard is nog niet eenvoudig voor te stellen hoe een dergelijk marketing proces dan georganiseerd zou moeten worden. Wel valt er aan te denken, wat nu door veel probleemverkenner ook al wordt gedaan, dat de probleemverkenner niet alleen de opdracht krijgen om de vraagkant in kaart te brengen, te exploreren en te analyseren en daarin te prioriteren. Probleemverkenner zouden ook de opdracht moeten krijgen om de aanbodkant te exploreren, te inventariseren, te analyseren, te prioriteren. Vervolgens kan op grond van de beide apart uitgevoerde analyses een voorstel voor koppeling van vraag en aanbod gedaan worden. Eveneens is denkbaar dat een dergelijk proces in de vorm van een beperkt aantal werkconferenties in de vorm van eendaagse workshops tussen vragers en aanbieders tot stand zou kunnen komen.

WAT ZIJN NU EIGENLIJK INTERESSANTE PROBLEMEN?

De programmalijnen zoals die nu in het onderzoeksprogramma van de S.V.O. worden onderscheiden zijn buitengewoon divers, enerzijds zijn er de sectorale programmalijnen en anderzijds zijn er een aantal meer thematisch gekozen programmalijnen zoals emancipatie, relatie tussen onderwijs en arbeidssituatie en de informatietechnologie. Uit de probleemverkenningrapporten van de sectorale programmalijnen blijkt al direct het grote probleem dat er ontstaat wanneer je binnen zo'n sectorale opzet ook tot de aanduiding van een aantal thematieken of problematieken moet komen. Elke probleemverkenner bedenkt daartoe een eigen indelingsgrondslag die uiteindelijk al dan niet goed blijkt te werken.

In het onderzoeksprogramma van de S.V.O. en de daaruit voortkomende aanbestedingen zien we dan – even als voorbeeld – een drietal onderzoekingen op het terrein van schoolorganisatie en -management worden uitgezet. Voor een flink deel met een overlappende vraagstelling, evenwel in het programma worden ze als geheel losse flodders gepresenteerd. Dit lijkt een gemiste kans voor interessante en vruchtbare afstemming van projecten binnen het totale

S.V.O. programma. Een oplossing voor dit probleem zou eruit kunnen bestaan dat de programmalijnen van S.V.O. volgens een wat nadrukkelijker thematische of probleemgeoriënteerde definiëring zouden worden opgezet.

Dat biedt dan ook de mogelijkheid om waar dat zinvol is een verbinding te leggen met expertise- en onderzoeks aanbod zoals die zich nu binnen de instituten en de OTG's heeft geconcentreerd. Op dit vlak zijn uiteraard allerlei consequenties denkbaar: vanuit de OTG's wordt een aanbod geformuleerd of OTG's adopteren bepaalde problemen of onderzoeksvragen. Het lijkt ook niet ondenkbaar om op grond van een review of een schifting van alle projecten in het S.V.O. programma tot de afbakening van een aantal steeds weer terugkerende problematieken en thematieken te komen. Vervolgens kunnen op grond daarvan een aantal groepen van onderzoekers aan het werk op een vergelijkbare wijze als de OTG's.

ENKELE CONCLUSIES

Concluderend uit de opsomming van vraagpunten en de genoemde problemen hoeft niet geconcludeerd te worden dat het instituut van probleemverkenner en programmamaker niet werkbaar zou zijn. Wel valt op dat in de huidige situatie een aantal mogelijkheden blijven liggen om het werk van probleemverkenners en programmamakers beter te structureren en beter van een constante en kwaliteitsvolle informatiestroom te voorzien.

We vatten een aantal suggesties samen:

- A) Versterking van de inventarisatie en structurering van de aanbodkant.
- B) Verbetering van de wisselwerking tussen vraag en aanbodkant.
- C) Explicitering van het afwegingsproces zoals dat door de verantwoordelijke besturen en door de besluitvoorbereiders wordt uitgevoerd.
- D) Een reshuffling van de S.V.O. programmalijnen tot thematische programmaclusters.
- E) Een versterking van de wisselwerking tussen OTG programmering en overige programmering van de S.V.O.

Uiteraard betekent een eventuele realisatie van een aantal van dit soort mogelijkheden het een en het ander voor zowel de organisatie van de S.V.O. als voor de organisatie van het werk op de onderzoeksinstituten. Het zal in elk geval betekenen dat ook van de kant van de onderzoeksinstituten een aantal programmerings- en profilerings activiteiten moet worden verlangd. Daartoe zullen die instituten ook hun werkstructuur moeten inrichten en anderzijds zal de statutair aan de S.V.O. toegedeelde programmeringsopdracht zo moeten worden geherformuleerd dat het een gedeelde opdracht met de onderzoeksinstituten gaat worden. Tevens zal moeten worden nagegaan hoe de onderzoeksprogrammering zoals die door de grootste klant van S.V.O. namelijk MO&W wordt uitgevoerd gestalte zou moeten krijgen. Uiteraard zijn de afdelingen van het ministerie wel sectoraal ingedeeld maar mogelijkerwijze is de structuur, in de vorm van de coördinatie van de onderwijsresearch, op het ministerie al zodanig sterk dat het ook mogelijk is de vragen voor onderzoek al op dat niveau wat sterker in thematieken onder te brengen.

Al deze voorstellen komen voort uit de overweging dat bij een voortduren van de huidige opzet met losse projecten, zonder reële inhoudelijke kaders in het onderzoeksprogramma van de S.V.O., er van een groei van het kennisbestand van de onderwijsresearch, door het opdracht-onderzoek, slechts in geringe mate sprake zal zijn. Ook op het niveau van de individuele onderzoekers zal een voortgaande, zowel inhoudelijke als methodische, competentie ontwikkeling alleen maar door toevalligheden tot stand komen. Beide punten betekenen dat potentiële mogelijkheden tot kwaliteitsverbetering van het onderwijsonderzoek en expertise vergroting van onderwijsonderzoekers onnodig onbenut blijven.

De OTG's en de programmering van onderwijs-onderzoek

E. Roede

S.C.O.

In de eerste bijdrage over de programmering van onderwijsonderzoek werden twee overwegingen genoemd die ten grondslag liggen aan initiatieven tot programmering van onderzoek, namelijk:

1. de bijdrage aan de cumulatie van kennis en
2. de bijdrage aan het oplossen van problemen in het onderwijs.

Als belangrijke problemen werden in die bijdrage o.a. de onvoldoende cumulatie van kennis en het onvoldoende gebruik van kennis voor beleidsontwikkeling genoemd.

Deze wat pessimistische waardering van de opbrengsten van de programmering van het onderwijsonderzoek tot zover wordt bevestigd in de bevindingen van de door het Instituut voor Sociaal Beleidsonderzoek (ISB) uitgevoerde inventarisatie van beleidsgericht onderzoek in het onderwijsveld. (Nijssen & Mulders, 1985). Deze inventarisatie betrof zo'n tachtig via de SVO aanbestede projecten binnen de 40% ruimte voor overheidsvragen. Het meest opvallende is volgens Hetebrij en Nijssen (1986) de niet-theoretische en empiristische aanpak in vrijwel alle onderzochte projecten.

In de beschouwingen over de programmering van het onderwijsonderzoek tot nu toe is nauwelijks aandacht besteed aan de programmering door de op initiatief van de SVO ingestelde Onderzoeksthmagroepen (OTG's). Dat is niet zo verwonderlijk, omdat in het onderzoeksprogramma van de SVO zelf de OTG-programmering niet wordt genoemd. Hoewel de OTG's aanvankelijk waren gesitueerd binnen het kader van de SVO-programmering is in de loop van de jaren het OTG-onderzoek (en het ANGO-onderzoek) steeds meer een afzonderlijke lijn geworden. Het doel van de OTG's is daarbij vooral theorie-constructie en -toepassing.

Dit doel van de OTG's en de eerder gesignaleerde problemen met het SVO-onderzoek vormen de aanleiding om in te gaan op de vraag hoe de OTG-programmering kan bijdragen aan een verbetering van de SVO-programmering.

Eerst zal ik kort de bedoeling en realisering van de OTG-constructie bespreken. Daarna wordt een aantal mogelijkheden gegeven om de OTG's een grotere rol te geven in de programmering voor SVO-onderzoek en zo te bereiken dat er meer sprake is van cumulatie van kennis en van gebruik van kennis.

De OTG's

In de, in eerdere afleveringen al genoemde SVO nota 'programmering voor onderwijsonderzoek' werden de OTG's gezien als een van de elementen in de organisatie-structuur waarin de SVO-programmering zich zou moeten gaan afspelen. Andere elementen waren de sektoronderzoeksgroepen, de onderzoeksinstituten, wetenschappelijke adviseurs, het SVO bureau en het SVO bestuur. Omdat we nu te maken hebben met een wat andere organisatiestructuur zijn de verschillende elementen op zich niet zo relevant meer. Wel echter de benadering die werd voorgestaan. De rol van de OTG's zou daarbij moeten zijn dat de activiteiten van onderzoekers worden gecoördineerd. Dit omdat ervan uitgegaan werd dat programmering alleen zin heeft als het resultaat door de betrokkenen wordt aanvaard en uitgevoerd. De OTG's zouden moeten

zorgen voor het overleg over samenwerking, afstemming en prioritering op het werkniveau van de onderzoekers.

Daarbij zou ook aandacht moeten worden besteed aan de discussie over de kwaliteit en bruikbaarheid van onderzoeken (de forumgedachte). Omdat de mogelijkheden voor onderzoek beperkt zijn, zouden op belangrijke inhoudelijke thema's OTG's tot ontwikkeling moeten worden gebracht. Het instellen van OTG's hield daarbij dus in het stellen van prioriteiten in de aanpak van bepaalde probleemgebieden. Binnen de OTG's zouden verder prioriteiten worden gesteld wat betreft de meest vruchtbare aanpak binnen een probleemgebied.

Van die gedachte is vrij weinig terecht gekomen. Weliswaar zijn er OTG's ingesteld op grond van een inventarisatie van door onderzoekers belangrijk geachte probleemvelden. Ook is er onderzoek geprogrammeerd, beoordeeld en geprioriteerd door de OTG's. Ook wat betreft de realisering van overleg op het werkniveau van de onderzoekers lijken de OTG's een geslaagd initiatief. De opbrengsten liggen echter vooral binnen het terrein van de OTG's. Binnen de totale SVO-programmering is de OTG-programmering een apart gebeuren en dat was niet de bedoeling.

In de afgelopen 10 jaar zijn allerlei randvoorwaarden in het onderwijsonderzoek veranderd waardoor de oorspronkelijke ideeën wat betreft de realisering moeten worden bijgesteld. De oorspronkelijke bedoeling lijkt echter ook nu nog waardevol, vooral om de cumulatie van kennis en het gebruik van die kennis te vergroten.

Mogelijke bijdragen van de OTG's

Op een aantal belangrijke thema's heeft een groot aantal onderzoekers van het onderwijs de afgelopen jaren op verschillende manieren samengewerkt. Behalve tot een verbetering van de kwaliteit van het uitgevoerde onderzoek heeft dat geleid tot kennis over en ervaring met:

- de programmering van onderzoek
- de beoordeling van onderzoek.

Op verschillende manieren kan deze kennis en ervaring te zamen met de inhoudelijke deskundigheid op een aantal belangrijke thema's worden gebruikt voor de verbetering van de SVO-programmering. De manieren die ik hier wil noemen verschillen met name wat betreft de mate waarin de bestaande situatie van aparte geldstromen blijft bestaan. Gemeenschappelijk is steeds dat 'gebundelde' kennis en ervaring van groepen onderzoekers systematischer wordt gebruikt dan nu het geval is.

1. Hoewel indertijd de vijf probleemvelden waarop OTG's werden geïnitieerd zijn gekozen na een afweging tussen allerlei door onderzoekers geuite voorkeuren is het niet waarschijnlijk dat de huidige vijf thema's de enige belangrijke zijn op het terrein van onderwijsonderzoek. Tegen de gevolgde procedure zijn bezwaren ingebracht, maar belangrijker is waarschijnlijk de gehandhaafde beperking tot de oorspronkelijke vijf thema's. Overleg op het werkniveau van de onderzoeker is nog steeds onmisbaar voor programmering die gericht is op cumulatie van kennis. Alleen als meerdere onderzoekers hun inspanningen op een bepaald thema weten te coördineren en op elkaar af te stemmen is er een kans dat de opbrengsten niet alleen bij voortgaand onderzoek, maar ook bij elkaars onderzoek aansluiten.

Ervaringen binnen de huidige OTG's wijzen erop dat zo'n afstemming niet van de ene dag op de andere wordt gerealiseerd. Opeenvolgende programma's van bijvoorbeeld de OTG's Onderwijsleerprocessen en Motivatie in het Onderwijs laten zien dat pas na een aantal jaren sprake kan zijn van geaccepteerde normatieve keuzen in de aanpak van onderzoeksproblemen op een bepaald thema.

Dat neemt niet weg dat ook op andere thema's dan de huidige vijf dergelijke samenwerking en afstemming zou moeten worden gestimuleerd. Het meest effectief is daarbij het beschikbaar stellen van financiële ruimte voor bijvoorbeeld programmeringswerk. Een consequentie van het stimuleren of initiëren van meer OTG's betekent dat het budget voor OTG-onderzoek zou moeten worden verruimd. Als dat om statutaire redenen niet kan buiten de 20% vrij onderzoek, dan zou daarvoor de voor ANGO-onderzoek bestemde ruimte moeten worden gebruikt.

Een andere consequentie van het stimuleren of initiëren van (meer) OTG's is dat de OTG's

regelmatig worden geëvalueerd op cumulatie van kennis en de toepassing van kennis voor het oplossen van problemen in het onderwijs. Elke OTG zou daartoe elk 3-4 jaar meer moeten doen dan het beschrijven van wat is gedaan. Het uitputtend bespreken en bediscussiëren (tijdens een ORD) met OTG-deelnemers, maar vooral ook andere wetenschappers en gebruikers van onderzoekskennis zou regel moeten worden.

2. In het voorgaande werd ervan uitgegaan dat het gebruik van de kennis en ervaring in de OTG's beperkt blijft tot de (20%) zogenaamde vrije onderzoeksruimte. Dat is niet per se nodig zonder de SVO statuten te moeten veranderen. Van bestaande en nieuwe OTG's zou systematischer gebruik kunnen worden gemaakt door, onder het behoud van de bestaande financieringstromen, de nu geheel onderscheiden programmeringen voor een deel te integreren. Drie mogelijkheden daartoe zijn de volgende:

- a. De SVO-programmering (aanbesteding- en OTG-) kan zo worden ingericht dat onderzoeksvragen of problemen van de overheid en veld kunnen worden opgenomen in de OTG-programmering. Dat betekent dat die vragen en problemen soms meer en soms minder zullen worden aangepast op basis van beschikbare en gewenste kennis op het terrein van de betreffende OTG. Ook betekent het, dat onderzoeksvoorstellen naar aanleiding van aanbesteding van door de OTG's geadapteerde onderzoeken (ook) door de OTG worden beoordeeld.
- b. Evenzo kunnen door de OTG-deelnemers ontwikkelde onderzoeksvoorstellen geadapteerd worden door overheid of veld.
- c. Een belangrijk deel van de SVO-programmering van aanbestedingsonderzoek komt nu tot stand via probleemverkenningen door programma-architecten. Deze worden geacht de mogelijkheden voor onderzoek en vragen om onderzoek op elkaar af te stemmen. In de praktijk komt dat veelal neer op het in een aanbesteedbare vorm beschrijven van vragen om onderzoek. Voortbouwen op eerder onderzoek (cumulatie van kennis en toepassing van kennis) is veelal minimaal. Dat is niet verwonderlijk omdat de opdracht voor de meeste, ook de heel ervaren en wijze onderzoekers te moeilijk is – behalve als het 'toevallig' precies zijn of haar specialiteit betreft en de invulling daarvan overeenkomt met de opvattingen en kennis van de vragensteller. De onder a en b genoemde mogelijkheden laten de huidige organisatie van de programmering grotendeels in takt. Dat is niet zo met de derde mogelijkheid.

Deze houdt in dat, en dat was ook de bedoeling van de oorspronkelijke OTG-gedachte, dergelijke probleem verkenningen te laten uitvoeren door OTG's. Alleen dan is er een gereede kans dat de aanbestedingen mogelijkheden bieden om niet alleen veel data op te leveren, maar ook om bij te dragen aan kennis en cumulatie van kennis.

3. De drie varianten van de zojuist beschreven mogelijkheid tot systematischer gebruik van de OTG's in de programmering passen nog binnen de huidige structuur van het onderwijsonderzoek. Alleen de organisatie van een aantal elementen zou wat moeten worden aangepast.

De derde en laatste te bespreken mogelijkheid gaat een stuk verder.

Het onderwijsonderzoek en ook de programmering ervan betreft afzonderlijke projecten. Soms sluiten projecten aan bij elkaar en kunnen dezelfde constructen, theorieën en meetinstrumenten worden gebruikt. Meestal is dat niet het geval. Door een projectenbeleid wordt onderwijsonderzoek gefragmentariseerd zowel wat betreft thema's en vraagstellingen als wat betreft de deskundigheid van onderzoekers. Als cumulatie van kennis en toepassing van kennis echt belangrijk worden gevonden, dan ligt het voor de hand van projectfinanciering over te gaan op programmafinanciering.

OTG's zouden daarbij zowel wat betreft de programmering, als ook wat betreft de uitvoering van een programma een belangrijkere rol kunnen krijgen dan nu het geval is. Meerderjarige programma's van verschillende omvang al naar gelang het belang in termen van problemen, vragen en mogelijkheden zouden daarbij moeten kunnen worden gecontinueerd of afgesloten op grond van de opbrengst aan kennis en toepassing van die kennis.

Zoals eerder al geopperd zou dat kunnen door twee- of driejaarlijks (op grond van de inhoud

van een programma) de opbrengsten in wetenschappelijk en maatschappelijk opzicht door wetenschappers en gebruikers te laten beoordelen.

Tenslotte

Er zijn niet alleen maar problemen en tekorten in het onderwijsonderzoek. Als ze er zijn, zijn ze niet te verhelpen met alleen een betere programmering tengevolge van een grotere rol van OTG's. In de OTG's is echter meer dan elders sprake van samenwerking op het werkniveau van onderzoekers in verschillende geldstromen. Het systematischer gebruiken van die samenwerking bij de programmering van onderzoek – opgevat als het totale proces van prioriteiten; tot ontwikkeling brengen en uitvoeren van onderzoek; alsmede de verspreiding van resultaten en producten en de evaluatie daarvan – kan ertoe bijdragen dat het beschikbare budget voor onderwijsonderzoek niet alleen leidt tot veel gegevens, maar ook tot kennis en kennistoename.

Het zijn immers de onderzoekers die vooral bepalend zijn voor de kwaliteit van het uitgevoerde onderzoek. Indien zij systematischer betrokken worden bij de programmering en met name als dat gebeurt vanuit een bundeling van kennis en ervaring op een bepaald terrein zoals in de OTG lijken kennis en kennistoename ook voor het toegepaste onderwijsonderzoek beter realiseerbare doelen.

Mededelingen

REVIEW ARTIKELEN

Ook in 1987 zullen weer premies worden toegekend aan schrijvers van zeer goede overzichtsartikelen (review articles) verschenen in tijdschriften op het terrein van de gedrags- en maatschappijwetenschappen. Voor dit doel is een bedrag van f 15.000,- beschikbaar.

Om voor beoordeling in aanmerking te komen dienen overzichtsartikelen vóór 1 augustus 1987 te worden ingezonden naar de Sociaal-Wetenschappelijke Raad, Herengracht 410, 1017 BX Amsterdam.

BEROEPSBEKWAAMHEIDS-REGISTRATIE VOOR PEDAGOGEN, ONDERWIJSKUNDIGEN EN ANDRAGOLOGEN

De NVO, de Nederlandse beroepsvereniging van pedagogen, onderwijskundigen en andragologen, is per 1 december jl. overgegaan tot het instellen van een Beroepsbekwaamheidsregistratie (BBR). Deze registratie, gericht op de kwaliteitsbewaking van de beroepsuitoefening binnen de verschillende differentiaties in de NVO (Klinische Pedagogiek, Orthopedagogiek, Onderwijskunde, Andragologie), is van belang enerzijds als waarborg voor cliënten, anderzijds voor de identiteit van de beroepsgroep.

De BBR vindt plaats per differentiatie. In aanmerking komen diegenen, die gewoon lid zijn van de NVO.

Universitair afgestudeerden volgens de nieuwe 1e fase dienen, alvorens opgenomen te kunnen worden in het register, enige tijd onder supervisie werkzaam te zijn. Ten behoeve van degenen, die een universitaire opleiding vóór de invoering van de 2-fasen structuur gevolgd hebben, is een tijdelijke overgangsregeling voor opname in het register vastgesteld.

Voor nadere informatie omtrent de BBR kunt u schriftelijk contact opnemen met het van de NVO, Korte Elisabethstraat 11, 3511 JG Utrecht.

Congressen, Symposia, Cursussen, Trainingen

THE COURSEWARE INDUSTRY WORLDWIDE

A one day international conference bringing together representatives of courseware industry providers and clients

Op vrijdag 27 februari 1987 organiseert Courseware Europe bv een internationale conferentie te Zaandam.

De conferentie richt zich tot opleiders en opleidersmanagers die COO gebruiken of het gebruik van COO overwegen, tot beleidsmakers die zich bezig houden met opleiding, training en informatietechnologie en tot hen die in de courseware industrie werkzaam zijn.

De conferentie wordt georganiseerd ter gelegenheid van het door Courseware Europe in gebruik te nemen nieuwe hoofdkantoor in Zaandam. Aansluitend op de conferentie kan men kennis maken met courseware toepassingen en produkten van de Europese Courseware Group, waar Courseware Europe bv deelnemer van is. Courseware Europe is gespecialiseerd in het ontwikkelen van courseware en expert systemen voor het bedrijfsleven.

Voor inlichtingen/inschrijvingen: dr. M.J.A. Mirande, Courseware Europe bv, Nieuwstraat 59, 1441 CL Purmerend, tel. 02990-39376. Per 1 jan. 1987: Ebbehout 1, 1507 EA Zaandam, tel. 075-172201.

CONFERENTIE 'BASISVORMING IN DE BASISCHOOL'

Het Centraal Instituut voor Toetsontwikkeling (Cito) en de Stichting voor Onderzoek van het Onderwijs (SVO) houden op 31 maart en 1 april 1987 een conferentie, getiteld 'Basisvorming in de basisschool'.

Deskundigen vanuit uiteenlopende disciplines zullen hun, op de empirie gebaseerde, opvattingen geven over de kerntaken van het basisonderwijs. Daarbij zal de vraag centraal staan wat het basisonderwijs kan bijdragen aan de basisvorming zoals die beschreven staat in het rapport van de Wetenschappelijke raad voor het regeringsbeleid (WRR) 'Basisvorming in het onderwijs'. Hoewel de basisvorming zich over het basisonderwijs en de eerste leerjaren van het voortgezet onderwijs uitstrekt, wordt in het WRR-rapport geen aandacht besteed aan de bijdrage die de basisschool – toch acht van de elf jaren – aan de basisvorming dient te leveren. Dat is één van de redenen waarom Cito en SVO deze conferentie organiseren.

De tweede reden is het starten van periodieke peilingen naar het niveau van het basisonderwijs in Nederland. Het hoe en waarom zal tijdens de conferentie uitgebreide aandacht krijgen.

De resultaten van een onderzoek onder 'de Nederlanders' over basisvorming en van eenzelfde onderzoek onder de deelnemers aan de conferentie zullen op 31 maart worden bekendgemaakt.

Onderwijsspecialisten van de fracties van de grote politieke partijen in de Tweede Kamer zullen hun visie geven op de basisvorming in de basisschool.

Inschrijfformulieren zijn te verkrijgen bij mevrouw E. van der Veldt-Leenheer van het Cito, postbus 1034, 6801 MG Arnhem, telefoon (085) 521451.

LUSTRUM CONGRES NVO: STILSTAAN BIJ VOORUITGANG

1962-1987 25 jaar NVO

Tijd om even stil te staan bij de ontwikkelingen op pedagogisch, onderwijskundig en andragogisch terrein in de afgelopen 25 jaar en tijd om feest te vieren. De NVO organiseert daarom op 27 maart 1987 een lustrum-congres in De Reehorst te Ede met als thema:

Beweging en tegenbeweging in opvoeding, vorming en onderwijs

In een plenair ochtendprogramma zullen enige ontwikkelingen op het gebied van de pedagogische, onderwijskundige en andragogische wetenschappen in de afgelopen 25 jaar in een cultuurhistorisch kader geplaatst worden.

In de middag worden vier parallel-sessies gehouden voor de differentiaties:

Klinische- en Orthopedagogiek

Onderwijskunde

Andragologie

Wijsgerige, Historische en Empirische Pedagogiek

waarin per subdiscipline steeds de volgende aspecten aan bod komen:

- een stand van zaken in de wetenschap
- illustratie(s) vanuit de praktijk
- maatschappelijke implicaties.

Lustrum-Boek

Ter ere van het 25-jarige bestaan van de NVO wordt aan iedere deelnemer een lustrum-boek uitgereikt, waarin alle lezingen die op deze dag worden gehouden zijn opgenomen.

Kosten bedragen voor NVO-leden f 50,-, aspirantleden en werkeloze leden f 20,-, en niet NVO-leden f 125,-. Inschrijven: NVO, Antwoordnummer 9045, 3511 JG Utrecht.

Abonnement International Journal of Educational Research

In VOR-bulletin nr. 4 van het afgelopen jaar werd verslag gedaan van de onderhandelingen tussen de VOR, de SVO en Pergamon Press betreffende het International Journal of Educational Research (IJER). Mede op grond van een garantiesubsidie door de SVO heeft de VOR met Pergamon een overeenkomst gesloten voor een collectief abonnement op het IJER. Leden van de VOR kunnen zich vanaf nu voor f 40,- per jaar (ingande 1987) opgeven voor een abonnement op het IJER. Daarvoor kan gebruik worden gemaakt van het onderstaande inschrijfformulier.

Meer informatie over het IJER kunt u tegemoet zien in een volgend VOR-bulletin. Tijdens de ORD'87 in Groningen zullen enige exemplaren van uitgebrachte afleveringen kunnen worden ingezien.

Vereniging voor onderwijsresearch
Association for Educational Research
Inschrijfformulier IJER (alleen voor leden van de VOR).

Opsturen naar: drs. E. Roede
Ravenwerf 14^I
1018 HV Amsterdam

1. Naam :

2. Privé-adres :

3. Postcode :

4. Woonplaats:

Ik wacht met betalen op een acceptgiro van de VOR.

VOR lidmaatschap en contributiebetaling

Er blijken zo nu en dan misverstanden te bestaan over het lidmaatschap van de VOR. Daarom zijn enkele hoofdzaken nog eens samengevat.

- * Het lidmaatschap betreft steeds een heel kalenderjaar.
- * Opzeggen van een lidmaatschap voor een volgend kalenderjaar kan alleen gebeuren voor de eerste oktober van het lopende kalenderjaar.
- * Contributie is verschuldigd voor de eerste maart van het lopende kalenderjaar. Als, na een herinnering van de penningmeester de contributie niet voor 1 april is betaald kan het lidmaatschap vervallen worden verklaard, met behoud van de financiële verplichtingen.

Tot zover de belangrijkste regels.

Omdat een belangrijk deel van de contributie-inkomsten gebruikt wordt om aan het begin van het jaar de aan Swets & Zeitlinger verschuldigde collectieve abonnementen op het Tijdschrift voor Onderwijsresearch te voldoen is het nodig dat de contributie tijdig wordt betaald. Daarom is in de afgelopen jaren de toezending van de acceptgiro's geleidelijk aan vervroegd.

Graag tijdige overmaking!

Maakt u voor het overmaken zoveel mogelijk gebruik van de toegestuurde acceptgiro's. Het komt steeds vaker voor dat contributies via bedrijven of universiteiten worden overgemaakt, waarbij alleen door telefonisch informeren te achterhalen is voor wie de contributie is bedoeld. Voor een enkeling is dat niet zo'n probleem, maar voor meerderen wordt het al gauw veel extra werk.

Belgische leden kunnen het beste hun lidmaatschap overmaken met een cheque – d.w.z. als zij niet over een Nederlandse giro- of bankrekening beschikken. Overmaken via een Belgische bank betekent dat er extra kosten moeten worden betaald of, wat meestal het geval is, dat de VOR te weinig contributie krijgt overgemaakt.

Indien u vragen of opmerkingen heeft over het lidmaatschap of de contributie, dan kunt u daarvoor schriftelijk of telefonisch (overdag) het beste terecht bij de penningmeester van de VOR. Formeel lopen deze contacten via de secretaris, maar de ledenadministratie wordt gedaan door de penningmeester:

E. Roede

Ravenwerf 14 I

1018 HV Amsterdam

Tel.: 020 - 264547 (overdag).

Het Leren van een Technische Handvaardigheid

L.F.W. de Klerk en P.G.F. Spit

Vakgroep Functieer, Onderwijspsychologie en Ergonomie, Katholieke Universiteit Brabant

ABSTRACT

In this study the learning of a technical skill has been investigated. The task involved making center-points on a small sheet of metal with the aid of a center and a hammer. The points should be located exactly on a thin line that was marked on the surface of the sheet. It was assumed that during the learning process an internal representation would develop. In this internal representation, information is supposed to be stored that may serve as a basis for the execution of the psycho-motor actions that are involved in the technical skill. Three experiments have been carried out. These experiments reveal that student learning can be promoted by presenting an adequate orienting-base, by providing feedback, and by means of verbal (self)instructions that guide the execution of these actions. An instructional method that combines these three components appeared to be the most effective in stimulating storage and retrieval of the essential task information. Furthermore, it has been found that prolonged training did not significantly improve performance on a retention test. For this particular task, an adequate internal representation is formed already during the first few learning sessions.

INLEIDING

In het lager technisch onderwijs wordt betrekkelijk veel aandacht besteed aan het leren uitvoeren van praktijkopdrachten (zie o.a. van der Sanden, 1986). Dit zijn complexe taken die zowel een beroep doen op kennis en intellectuele vaardigheden als op *technische handvaardigheden*. Deze laatste hebben betrekking op handelingen waarbij handgereedschap(pen) gebruikt moet(en) worden. Het hangt van de praktijkopdracht af over welke technische handvaardigheden de leerling in een concreet geval moet kunnen beschikken. Indien de taak bijvoorbeeld bestaat uit het vervaardigen van een plaatwerkstuk, dan zal een beroep worden gedaan op technische handvaardigheden zoals: aftekenen, knippen, zagen, vijlen e.d.

In deze studie ligt het accent op het leren van motorische handelingen die aan de betreffende technische handvaardigheid ten grondslag liggen. Het gaat er uiteindelijk om aanwijzingen te vinden over de wijze waarop het leren van technische handvaardigheden geoptimaliseerd kan worden.

De vraag is hoe motorische handelingen geleerd worden. Volgens de zogenaamde 'open loop' opvatting worden de bewegingen waaruit de handeling is samengesteld gecontroleerd vanuit een *motorisch programma* dat in het lange termijngeheugen van de leerling is opgeslagen. Dit programma bevat alle respons-specificaties die nodig zijn om de bewegingen correct uit te voeren. De nodige bewegingsinformatie verkrijgt de leerling tijdens het leerproces op basis van zowel interne als externe *feedback*. Echter, naarmate het leerproces vordert zal de rol van de (externe) feedback afnemen. De bewegingen zullen steeds automatischer plaats vinden, waarbij bewuste aandacht en externe controle gereduceerd zullen worden.

Het hangt af van de aard van de taak welk niveau van automatisering uiteindelijk bereikt zal worden. Van Cranenburgh en Mulder (1986) merken in dit verband op: 'dat weliswaar evidentie gevonden is voor het bestaan van een motorprogramma, doch dat dit motorprogramma geenszins het absolute (feedbackloze) karakter heeft als oorspronkelijk werd aangenomen. Er kan dus niet worden gesteld dat alle bewegingen onder motorprogramma-controle staan en dat

feedback geen rol speelt. Beide controle-principes spelen een rol, waarbij de rol van motorprogramma of feedback afhankelijk is van het type taak, de vereiste nauwkeurigheid, het niveau van de uitvoerder, enz. Grofweg zou men kunnen stellen dat snelle, goed geleerde bewegingen voor een belangrijk deel gestuurd worden door een opgeslagen motorprogramma, terwijl langzame, nauwkeurige bewegingen voor een belangrijk deel gereguleerd worden door middel van feedback'.

Schmidt (1976) is van mening dat tijdens de eerste fase van het leerproces bewegingen zelden op identieke wijze en onder exact dezelfde bewegingscondities herhaald zullen worden. Er wordt dus steeds andere bewegingsinformatie opgenomen. Het is niet waarschijnlijk dat al deze gegevens afzonderlijk in het geheugen worden opgeslagen. Volgens Schmidt zal generalisatie plaats vinden. Van een bepaald bewegingspatroon worden alleen enkele algemene, abstracte kenmerken opgeslagen die gelden voor een hele verzameling van onderling verwante (maar niet noodzakelijk identieke) bewegingen. Deze kenmerken worden in het geheugen opgeslagen in de vorm van een *schema*. Dit fungeert tevens als ophaalplan voor de informatie die dan vervolgens vertaald wordt in concrete uitvoerings-instructies. Indien tijdens de oefening feedback gegeven wordt dan leert de leerling de relaties tussen bewegingsparameters (zoals kracht, snelheid en richting) en het resultaat van de beweging. Zo zal een flexibele en continue aanpassing tot stand komen van de uitvoeringswijze aan de steeds wisselende uitvoeringsomstandigheden.

De bewegingsinformatie die in het geheugen wordt opgeslagen wordt niet alleen bepaald door de feedback die wordt verschaft tijdens de uitvoering van de taak. De opslag van deze informatie kan ook worden bevorderd door het verschaffen van een adequate *oriënterings-basis*, bijvoorbeeld via demonstraties of verbale instructies en aanwijzingen. Zo merkt Pijning (1985) in het kader van het leren schrijven van letters op dat deze leertaak: 'niet neerkomt op een herhaling van toevallig juist uitgevoerde bewegingen; een dergelijke opvatting ontkent de organiserende en bepalende rol van het bewustzijn bij het leren van motorische vaardigheden, ontkent – om het preciezer te zeggen – dat de oriëntering en controle van de leerling de uitvoering van de schrijfbeweging mede bepalen'.

Dat het aanbrengen van een adequate oriënterings-basis van belang is voor het leren schrijven van letters heeft het onderzoek van Pantina (1957) aangetoond. Pantina heeft drie groepen kinderen vergeleken, die elk op een andere wijze letters kregen aangeboden. Zij spreekt van drie verschillende oriënterings-typen. Volgens het eerste oriënterings-type werd elke letter door de proefleider voorgedaan op speciaal ruitjespapier. Daarbij werd uitgelegd welke bewegingen gemaakt moeten worden. Vervolgens moesten de leerlingen de letters op dezelfde wijze naast die van de proefleider schrijven. Eventuele fouten werden aangewezen en gecorrigeerd door aan te geven hoe de lijn had moeten lopen. Volgens het tweede oriënterings-type kregen de kinderen een voorbeeld-letter. De proefleider liet zien waar zich in de letter de zogenaamde steunpunten bevinden. Dit zijn de punten waar de lijn sterk van richting verandert. Deze punten moesten eerst op aanwijzingen van de proefleider op het ruitjespapier worden aangebracht. Vervolgens moesten deze worden verbonden. Als de leerlingen de letter op deze wijze met voldoende precisie hebben leren schrijven, werden de punten successievelijk weggelaten. Zodra de leerlingen de letter correct 'uit het hoofd' konden schrijven werd met de volgende letter begonnen.

Het verschil tussen het derde en tweede oriënterings-type is hierin gelegen dat bij het derde type de uitleg over steunpunten alleen bij de eerste letter werd gegeven. Bij de volgende letters moest het kind zelf de steunpunten bepalen, waarbij de proefleider alleen de fouten verbeterde. Tijdens een later stadium van het oefen-proces werden de steunpunten weggelaten en moest het kind de verschillende punten in gedachten uitzetten.

Gebleken is dat de methode volgens het derde oriënterings-type het meest effectief is. Blijkbaar is de combinatie van *verbale uitleg* (van het principe van de steunpunten) en het in gedachten uitzetten van deze punten van groot belang voor de vorming van een adequaat motorisch programma. Hierbij dient te worden aangetekend dat het zich voorstellen van de steunpunten – zeker in de beginfase van het leerproces – in belangrijke mate geleid wordt door een reeks van verbale instructies (zie Pijning, 1985). Naarmate het leerproces vordert verloopt het proces van ophalen van relevante bewegingsinformatie en vertalen daarvan in uitvoerings-instructies vlotter en automatischer.

Ook uit experimenteel psychologisch onderzoek bij ervaren schrijvers (zie Thomassen e.a., 1983, 1985) is gebleken dat steunpunten de kern vormen van het motorisch programma dat zich ontwikkeld heeft. Er is evidentie dat in het programma vooral de halenstructuur van de letters is opgeslagen. Elke schrijffletter wordt gekenmerkt door een reeks op- en neerhalen. Dit zijn de verbindingen tussen wat Pantina de steunpunten heeft genoemd. Bij ervaren schrijvers wordt elke haal uitgevoerd door een korte, snelle beweging, die geheel onder controle van het motorisch programma staat. Andere parameters van de schrijfbeweging, zoals grootte, snelheid en kracht, worden tijdens de uitvoeringsfase van het schrijven ingesteld, afhankelijk van de specifieke schrijfoomstandigheden.

EXPERIMENT 1

Vraagstelling.

Het beheersen van een technische handvaardigheid veronderstelt dat de leerling over relevante informatie beschikt die uit het geheugen kan worden opgehaald en vertaald in een voor de vaardigheid in kwestie karakteristiek bewegingspatroon. Deze vaardigheid ontwikkelt zich tijdens een leerproces. Dit proces kan worden bevorderd door van te voren informatie te verschaffen, bijvoorbeeld door de bewegingen voor te doen (visueel) of door verbale toelichting te geven over het doel van de bewegingen, het te verwachten resultaat, de juiste houding, het gebruik van gereedschappen, de volgorde waarin de deelbewegingen moeten worden uitgevoerd e.d. Op basis van deze voor-informatie ontstaat een *interne representatie* die bepalend is voor de uitvoeringswijze.

Vooraf tijdens de beginfase van het leerproces zal de leerling gebruik maken van verbale activiteiten. Adams (1971) spreekt van de *verbaal-motorische fase*. Hierin speelt de taal een belangrijke rol bij het begrijpen en analyseren van de bedoeling van de taak, alsook bij het bepalen van een strategie. Ook het gestelde doel van iedere beweging wordt vaak gegeven in termen van verbale uitvoerings-instructies zoals: 'beweeg je arm zodanig dat...'. Bovendien wordt in vele gevallen op verbale wijze (in de vorm van kennis van resultaten) aangegeven of het gestelde doel inderdaad bereikt is.

Aangenomen wordt dat de vooraf gegeven informatie en de informatie die de feedback verschaft met elkaar in verband gebracht worden en dat het resultaat van deze vergelijking mede de basis vormt voor de uitvoering van de volgende beweging. Het gehele proces van vorming van een motorisch programma en het ophalen en vertalen van dit programma kan – zo luidt de veronderstelling – worden bevorderd door de leerling te instrueren de verschillende uitvoerings-instructies telkens voor zichzelf te herhalen, waardoor deze het karakter van *zelf-instructies* krijgen. De verwachting luidt dat de combinatie van deze verbale (zelf)instructies en feedback veel effectiever is dan een procedure waarbij alleen feedback wordt verschaft.

Taak.

De technische handvaardigheid die in het kader van dit experiment geleerd moest worden was het met behulp van een hamer en center slaan van center-punten op een dunne lijn die was aangebracht op een metalen plaatje. Deze vaardigheid komt voor in het curriculum van het lager technisch onderwijs.

Procedure.

In dit onderzoek werden twee instructie-procedures met elkaar vergeleken, een meer traditionele instructie, waarbij de nadruk lag op oefenen en feedback, en één waarbij een combinatie van feedback en verbale (zelf)instructies werd gebruikt.

Bij de traditionele methode was de procedure als volgt. Eerst werd de taak door de leraar in de klas voorgedaan. De verschillende handelingen werden gedemonstreerd. Dit ging vergezeld van een mondelinge toelichting (verbale instructies). Vervolgens werd de leerlingen gevraagd de bewegingen uit te voeren. Dit gebeurde gedurende een viertal bijeenkomsten (lessen). Per les

werden drie series van 13 center-punten geslagen. Aan het eind van elke les werd door de leerling nagegaan hoeveel center-punten hij per serie goed had geslagen (feedback in de vorm van kennis van resultaten). Dit aantal werd door de leraar gecontroleerd. Aan het begin van de derde les werd de klassikale instructie (i.e., de demonstratie plus verbale toelichting) herhaald. Tijdens de uitvoering van de taak werkten de leerlingen in eigen tempo.

De procedure bij de gecombineerde instructie kan als volgt worden samengevat. Ook hier werd begonnen met een klassikale demonstratie plus toelichting. Echter, gedurende de eerste serie van dertien centerslagen werd er in een vast tempo gewerkt, waarbij een verkorte verbale instructie hardop in de klas werd herhaald tijdens de uitvoering. Elke volgende serie werd begonnen met de klassikale demonstratie plus toelichting. Daarna mochten de leerlingen in eigen tempo aan de taak werken. Zij werden verzocht de verkorte verbale instructie steeds zachtjes voor zichzelf te herhalen. Ook hier werd gedurende vier lessen aan de taak gewerkt, waarbij per les drie series van elk 13 center-punten geslagen moesten worden. Aan het eind van elke les werd elke leerling verzocht zijn resultaten te beoordelen. De leraar controleerde dit.

Dit experiment vond plaats in een praktijk-lokaal van de school. Verschillende vakdocenten fungeerden als proefleider.

Leerlingen.

Aan dit experiment namen leerlingen deel die afkomstig waren uit de eerste twee leerjaren van de ito-afdeling en uit de eerste klassen van de lts-afdeling van de Katholieke Technische School te Oosterhout. Dat twee verschillende schooltypen in het onderzoek betrokken zijn heeft te maken met het feit dat dit experiment een onderdeel is van een project over het leren van psychomotorische vaardigheden. In het kader van dit project zijn niet alleen verschillende technische handvaardigheden onderzocht maar ook de relaties tussen het leren van deze vaardigheden en individuele verschillen. Om deze laatste reden zijn ook verschillende schooltypen in het project met elkaar vergeleken.

Opzet.

Het experiment is opgezet volgens een 2x2x4 factorieel design. De factoren zijn: schooltype (lts vs ito); instructie (alleen feedback versus feedback plus verbalisaties); en oefening (de lessen 1 t/m 4).

De leerlingen zijn per klas toegewezen aan de experimentele condities. Echter, om gelijke aantallen per conditie te verkrijgen, zijn de gegevens van enkele leerlingen buiten beschouwing gebleven. Op aselechte wijze is bepaald om welke leerlingen het ging. Per conditie waren er 15 leerlingen.

Resultaten en discussie.

Per serie is voor elke leerling het aantal goed geslagen center-punten vastgesteld. Het gemiddeld resultaat, berekend over de drie series waaruit een les is samengesteld, wordt beschouwd als de score van de leerling.

Over deze score is een variantie-analyse uitgevoerd volgens een 2x2x4 design, met herhaalde metingen op de laatste factor (oefening). De uitkomsten van deze analyse zijn samengevat in tabel 1.

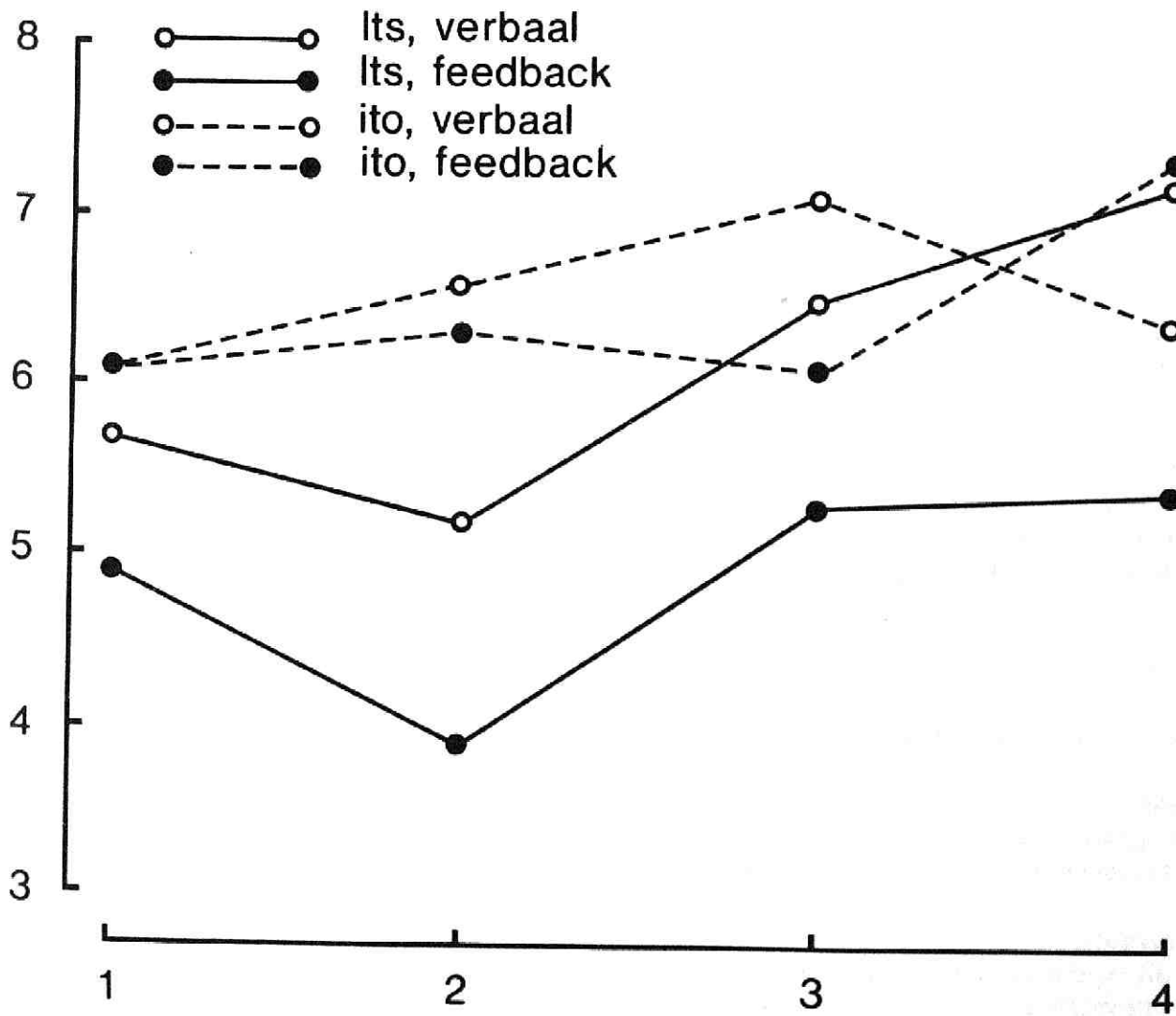
Uit de gegevens van tabel 1 blijkt dat de data de onderzoekshypothese bevestigen. De combinatie van verbale (zelf)instructies en feedback leidt tot betere prestaties dan wanneer alleen feedback wordt gegeven. Hierbij moet worden opgemerkt dat er sprake is van een significante interactie tussen instructie en schooltype. Nadere inspectie van de gegevens leert dat het verschil tussen beide instructies voor de lts-leerlingen groter is dan voor de ito-leerlingen.

Zowel uit figuur 1 als uit de gegevens van tabel 1 blijkt dat de ito-leerlingen gemiddeld betere prestaties behalen dan de lts-leerlingen. Dit kan een leraar-effect geweest zijn. Gebleken is dat de ito-leerkrachten het experiment enigszins in de competitieve sfeer getrokken hebben. Voorts moet worden gesteld dat de ito-leerlingen uit de eerste en tweede klassen gerecruteerd zijn

Tabel 1: De resultaten van een 2x2x4 variantie-analyse met herhaalde metingen.

bron	df	MS	F	P
A: schooltype	1	57.0	19.5	< .0001
B: instructie	1	27.3	9.4	< .001
AxB	1	22.2	7.6	< .001
error 1	56	2.9		
C: oefening	3	14.2	7.9	< .0001
CxA	3	5.6	3.1	< .05
CxB	3	1.3	.7	
CxAxB	3	3.6	2.0	
error 2	168	1.8		
totaal	239			

De gegevens zijn grafisch weergegeven in figuur 1.



Figuur 1: Gemiddelde score per les per conditie.

terwijl de lts-leerlingen alleen uit de eerste klassen afkomstig zijn. De ito-leerlingen zijn dus gemiddeld ouder en hebben ook meer ervaring, ofschoon deze specifieke taak voor geen van de leerlingen eerder is behandeld.

Uit de resultaten van tabel 1 blijkt dat er een significant leereffect is opgetreden. Bovendien is er een significante interactie tussen de factoren oefening en schooltype. Het blijkt dat de gemiddelde resultaten van de ito-leerlingen per les steeds verder verbeteren. Bij de lts-leerlingen ligt dit duidelijk anders. Bij deze leerlingen vindt na de eerste les eerst een daling plaats en vervolgens weer een stijging. Uit figuur 1 blijkt dat de daling vrijwel onafhankelijk is van de instructie; onder beide instructie-condities komt de daling voor. Dit is des te opmerkelijker omdat Spit (1983) een dergelijk verschijnsel ook heeft waargenomen bij een andere technische handvaardigheid, te weten vlak-vijlen. Kenmerkend voor de daarbij gebruikte methode is dat de leerlingen van tijd tot tijd zelf hun resultaten moesten controleren met behulp van een bepaald meetinstrument (een haar-lineaal). Na een aanvankelijke stijging van de resultaten trad een daling op. Dit ging gepaard met een afname van de frequentie waarmee de leerlingen hun resultaten controleerden. De daling zou dus een gevolg kunnen zijn van de verminderde feedback.

In de inleiding van dit artikel hebben we uiteengezet dat zowel voor-informatie als feedback van belang zijn voor het verloop van het leerproces. In experiment 1 is de voor-informatie verstrekt in de vorm van verbale uitleg en voordoen door de leerkracht in de klas. De demonstratie kan verbeterd worden door gebruik te maken van video-apparatuur. Wanneer er verschillende proefleiders zijn dan bestaat het gevaar dat de een het anders (en beter of slechter) doet dan de ander. Bovendien kan er tijdens een demonstratie in de klas van alles mis gaan. Met behulp van video kan de demonstratie gestandaardiseerd worden. Bovendien kunnen allerlei bewegingen herhaald, vertraagd en vergroot worden, hetgeen het instructieve karakter van de demonstraties ten goede kan komen.

EXPERIMENT 2

Vraagstelling.

In experiment 1 werd de feedback gegeven in de vorm van kennis van resultaten. Ook de informatie die de feedback verschaft kan worden vergroot door gebruik te maken van video-apparatuur. Niet alleen kan beter worden aangegeven wat goed en wat fout is, maar tevens kunnen veel nauwkeuriger 'gedragsgecentreerde' instructies (Van Parreren, 1983) worden gegeven, waardoor de gemaakte fouten beter gecorrigeerd kunnen worden.

In experiment 2 is er gebruik gemaakt van de mogelijkheden die video biedt voor het verschaffen van informatie. De verwachting is – evenals bij experiment 1 – dat de combinatie van feedback en verbale (zelf)instructies veel effectiever is dan wanneer alleen feedback wordt gegeven. Bovendien wordt aangenomen dat wanneer in de gecombineerde situatie meer informatie wordt verschaft, zowel vooraf (demonstratie) als achteraf (feedback), de eerder gesignaleerde daling tijdens het leerverloop zich niet zal voordoen.

Taak.

In experiment 2 is dezelfde taak gebruikt als in experiment 1: het met behulp van hamer en center slaan van center-punten op een dunne lijn in een metalen plaatje.

Procedure.

In dit experiment is eveneens een onderscheid gemaakt tussen twee instructie-condities. Volgens de ene methode werd begonnen met een klassikale uitleg door de leraar aan de hand van een demonstratie (voordoen). Deze instructie werd eenmaal herhaald, waarbij de leraar de stappen die achtereenvolgens doorlopen moesten worden hardop noemde. Daarna werd gedurende een viertal lessen geoefend, waarbij elke les weer uit drie series bestond. Na elke les werden de

resultaten ingenomen en door de leraar nagekeken. Bij het begin van de volgende les werden de resultaten medegedeeld (feedback). De leerlingen werkten in eigen tempo.

Bij de andere methode werd eveneens begonnen met een klassikale uitleg waarbij de leraar de handelingen voordeed. Daarna werd een demonstratie-film (video) gepresenteerd waarbij de correcte handelingen nog eens getoond werden en waarbij de verbale instructie stap voor stap herhaald werd.

Tijdens de eerste serie moesten de leerlingen de handelingen zelf uitvoeren, begeleid door de leraar die de verbale instructies in verkorte vorm herhaalde. Voorafgaande aan de tweede serie werd de demonstratie-film nogmaals gepresenteerd. Daarna volgde de tweede serie oefeningen, waarbij van enkele willekeurige leerlingen video-opnamen gemaakt werden. Deze werden aan het eind van de serie getoond en klassikaal besproken. Hierbij werd vooral geattendeerd op de onjuiste handelingen.

De volgende serie verliep op analoge wijze, zij het dat de gemaakte video-opnamen pas bij het begin van de volgende les getoond en besproken werden. Daarbij kregen de leerlingen tevens kennis van resultaten over de voorafgaande les.

De volgende lessen verliepen steeds volgens het stramien van de derde serie. Aan het begin van de les werden de resultaten van de vorige les verstrekt en werden de video-beelden in de klas besproken. Tijdens de vierde (laatste) les werden geen opnamen meer gemaakt.

Zes weken na de laatste les werd een retentie-toets afgenomen. Deze bestond uit de opdracht om een serie center-punten te slaan. Hierbij werden geen verdere aanwijzingen of instructies gegeven.

Het onderzoek vond plaats in het praktijklokaal van een technische school. Een vakdocent fungeerde als proefleider.

Leerlingen.

Ook experiment 2 maakte deel uit van het eerder genoemde project: het leren van psychomotorische vaardigheden. In het kader van dit project hebben verschillende scholen deelgenomen. Voor experiment 2 waren zowel gegevens beschikbaar van leerlingen die afkomstig waren uit de eerste klassen van de lts-afdeling van de Katholieke Technische School te Oosterhout, als van leerlingen uit de eerste klassen van het St. Oelbert Gymnasium, eveneens te Oosterhout.

Opzet.

Dit experiment is opgezet volgens een 2x2x4 factorieel design. De eerste factor was de instructie (feedback versus feedback plus video en verbalisaties); de tweede factor was het schooltype (lts versus gymnasium) en de derde factor was de oefening (les 1 t/m 4).

De leerlingen zijn per klas aan de condities toegewezen. Om gelijke aantallen per conditie te krijgen zijn de gegevens van enkele leerlingen buiten beschouwing gelaten. Op aselecte wijze is bepaald welke leerlingen hiervoor in aanmerking kwamen. Per conditie waren er 16 leerlingen.

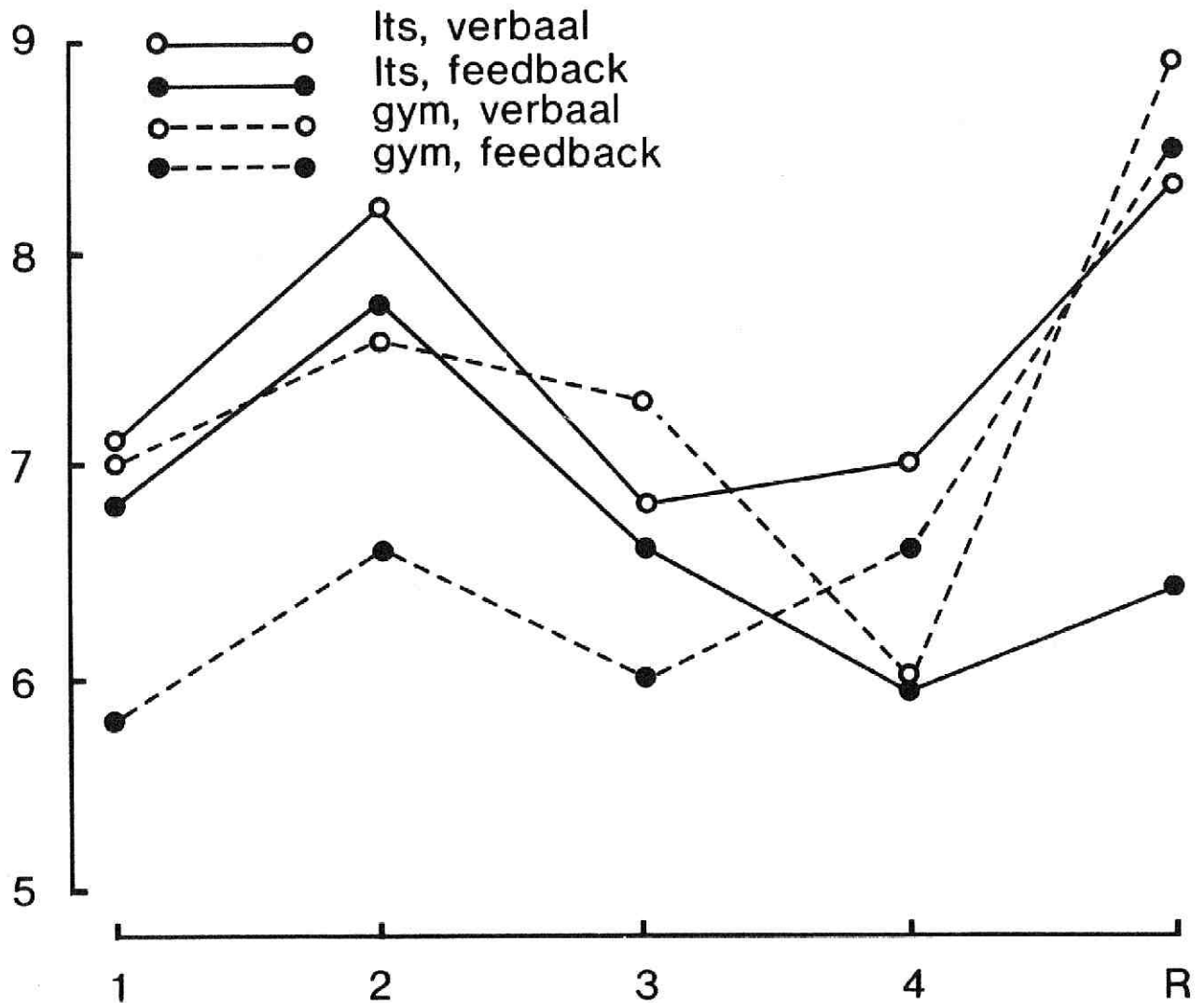
Resultaten en discussie.

Evenals in experiment 1 werd per les voor elke leerling de gemiddelde score berekend over de drie series waaruit een les is samengesteld. Op grond van deze gegevens is een variantie-analyse uitgevoerd volgens een 2x2x4 design, met herhaalde metingen op de laatste factor.

Gebleken is dat er slechts één significant hoofdeffect is. De uitgebreide instructie leidt gemiddeld tot significant betere resultaten dan de instructie waarbij alleen feedback wordt verschaft ($F(1,60) = 6.6; p < .05$). Bovendien is de interactie tussen alle drie de factoren significant ($F(3,180) = 2.9; p < .05$).

De resultaten zijn grafisch weergegeven in figuur 2.

Uit de gegevens van de variantie-analyse blijkt dus dat de factor instructie significant is. Opmerkelijk is dat het effect minder groot is dan in experiment 1, terwijl het verschil tussen de beide niveau's van deze variabele in experiment 2 groter is. De uitgebreide methode heeft in experiment 2 niet alleen betrekking gehad op verbale (zelf)instructies. Tevens vond een fouten-



Figuur 2: Gemiddelde score per les (en retentietoets = R) per conditie

analyse plaats aan de hand van video-opnamen van enkele willekeurige leerlingen. Het kan zijn dat de 'gedragsgecentreerde' aanwijzingen te specifiek waren, d.w.z., alleen van toepassing waren op enkele toevallig voorgekomen fouten. Het is de vraag in hoeverre deze aanwijzingen van belang waren voor andere leerlingen dan die waarvan de opnamen gemaakt zijn. Het is zelfs niet uit te sluiten dat hierdoor de aandacht is afgeleid van het eigen gedrag van de leerlingen waardoor de feedback minder effectief is geweest. Het is zelfs mogelijk dat correctieve aanwijzingen, die in feite voor anderen bestemd waren, geïnterfereerd hebben met het proces van opslaan, ophalen en vertalen van bewegingsinformatie na de uitvoering van de handelingen. Hierdoor zou het totale effect van de uitgebreide instructie-methode wel eens minder groot geweest kunnen zijn vergeleken met de in experiment 1 gebruikte 'verbalisatiemethode'.

Uit de gegevens van de variantie-analyse blijkt ook dat de gymnasium-leerlingen gemiddeld niet slechter of beter presteren dan de lts-leerlingen. Voor beide groepen geldt dat zij niet – althans vanuit de schoolsituatie gezien – bekend waren met de leertaak. Bovendien doet de taak geen beroep op hogere cognitieve vaardigheden, terwijl ook geen complexe bewegingspatronen geleerd moesten worden. Er is dan ook weinig reden om aan te nemen dat het ene type leerling hogere prestaties zou leveren dan het andere.

Tussen alle factoren die in dit experiment gemanipuleerd zijn bestaat een significante interactie. Deze houdt in dat het verloop van de verschillende prestatie-curven niet voor alle condities gelijk is. Dit blijkt ook uit figuur 2. Wel geldt voor alle condities dat eerst een stijging plaats

vindt en dan een daling. Echter, in sommige gevallen lijkt de daling zich voort te zetten, terwijl in andere gevallen de daling gevolgd wordt door een verbetering van de prestaties.

Eerder hebben wij opgemerkt dat een dergelijke terugval ook geconstateerd werd in experiment 1 bij dezelfde leertaak en door Spit (1983) bij het leren vlak-vijlen. Bezien we de resultaten van de retentie-toets (zie figuur 2) dan blijkt dat de prestaties na 6 weken nog betrekkelijk hoog zijn.

Met behulp van t-toetsen is nagegaan in hoeverre de resultaten van de retentietoets verschillen van die van de 2e les, waarin gemiddeld de beste resultaten waren behaald. Deze gegevens zijn per conditie berekend en gebaseerd op individuele verschil-scores. Gebleken is dat bij de gymnasium-leerlingen die alleen feedback hebben gekregen tijdens het leerproces, sprake is van een verbetering van de prestaties na een tijdsinterval van 6 weken ($t(15) = 3.39; p < .01$). Voor de gymnasium-leerlingen die de uitgebreide methode gevolgd hebben, is het verschil eveneens positief, maar niet significant. Voor de lts-leerlingen die de uitgebreide methode gebruikt hebben geldt dat hun prestaties vrijwel gelijk zijn aan die van de 2e les. Alleen bij de lts-leerlingen die alleen feedback gekregen hebben treffen we een verslechtering van de prestaties aan na het retentie-interval ($t(15) = -2.27; p < .05$). Blijkbaar zijn de meeste leerlingen – en dat geldt a fortiori voor de gymnasium leerlingen – erin geslaagd tijdens de eerste lessen een interne representatie van de handelingen te vormen waarin de relevante informatie op een zodanige wijze is opgeslagen dat deze na verloop van tijd op adequate wijze opgehaald en gebruikt kan worden. Aangenomen wordt dat het opslaan van relevante informatie een cognitieve activiteit is. Sommige leerlingen zijn hiertoe beter in staat dan andere. Het is niet onwaarschijnlijk dat gymnasium leerlingen hierbij in het voordeel zijn. (De prestaties op de retentietoets zijn voor de gymnasium leerling gemiddeld significant beter dan voor de lts-leerlingen ($F(1,62) = 6.47; p < .05$). Voorts kan worden gesteld dat deze cognitieve activiteit in belangrijke mate gereguleerd wordt door verbale (zelf)instructies en/of correctieve aanwijzingen tijdens het leerproces. (Ook de instructie-factor met betrekking tot de retentietoetsgegevens is significant ($F(1,62) = 4.57; p < .05$).

EXPERIMENT 3

Vraagstelling.

Naar aanleiding van de resultaten van zowel experiment 1 als experiment 2 kan worden geconcludeerd dat er tijdens het leerproces sprake is van een – al dan niet tijdelijke – terugval. De vraag is in hoeverre dit een gevolg is van gebrekkige (intrinsieke of extrinsieke) motivatie. Het feit dat in het onderzoek van Spit (1983) gevonden is dat leerlingen hun resultaten steeds minder vaak controleerden, zou daarop kunnen wijzen.

Een tweede vraag betreft het aantal oefen-series. Zowel in experiment 1 als in experiment 2 werd in alle condities voorafgaande aan de uitvoering informatie over de taak verschaft. Er vond een meer of minder uitvoerige oriëntering op de handelingen plaats. Deze vormde de basis voor een interne representatie die de uitvoering in het begin van het leerproces bepaalt. Door de combinatie van verbale (zelf)instructie en feedback worden eventuele fouten in de uitvoeringswijze gecorrigeerd. Na verloop van tijd wordt weinig extra informatie meer toegevoegd bij datgene wat inmiddels in het geheugen van de leerling is opgeslagen. Om deze reden zal – gelet ook op de aard van de taak – verder oefenen weinig zin meer hebben. De vraag is dan ook, of met minder oefening niet toch hetzelfde resultaat wordt bereikt.

Beide hierboven gestelde vragen hebben ten grondslag gelegen aan de opzet van experiment 3.

Taak.

Evenals in beide vorige experimenten bestond de taak uit het leren slaan van centerpunten op een dunne lijn op een metalen plaat.

Procedure.

Alle leerlingen kregen dezelfde instructie. Deze was gelijk aan de uitgebreide instructie-conditie van experiment 2.

Dit experiment is opgezet volgens een $2 \times 2 \times 3$ factorieel design. De eerste factor had betrekking op de motivatie. In het ene geval werd de leerlingen verteld dat de resultaten van het experiment meetellen voor het rapport; in het andere geval werd gezegd dat deze niet meetellen. De tweede factor betrof de hoeveelheid oefening. Een deel van de leerlingen oefende gedurende een tweetal lessen, terwijl de overige leerlingen gedurende vier lessen oefenden. De derde factor had betrekking op de herhaalde metingen. Het ging hierbij om de gegevens van de eerste twee lessen, alsmede om de gegevens van een retentietoets die na een interval van 6 weken is afgenomen.

Het experiment was zo ingericht dat alle leerlingen die werkten onder de conditie waarbij de resultaten wel meetellen voor het rapport uit dezelfde klas afkomstig waren. De leerlingen die werkten onder de conditie waarbij de resultaten niet meetellen kwamen allen uit een andere klas.

Leerlingen.

Aan het experiment namen 32 leerlingen deel uit twee eerste klassen van de Katholieke Technische School te Oosterhout. Via random selectie werd ervoor gezorgd dat er gelijke aantallen leerlingen per conditie waren.

Resultaten en discussie.

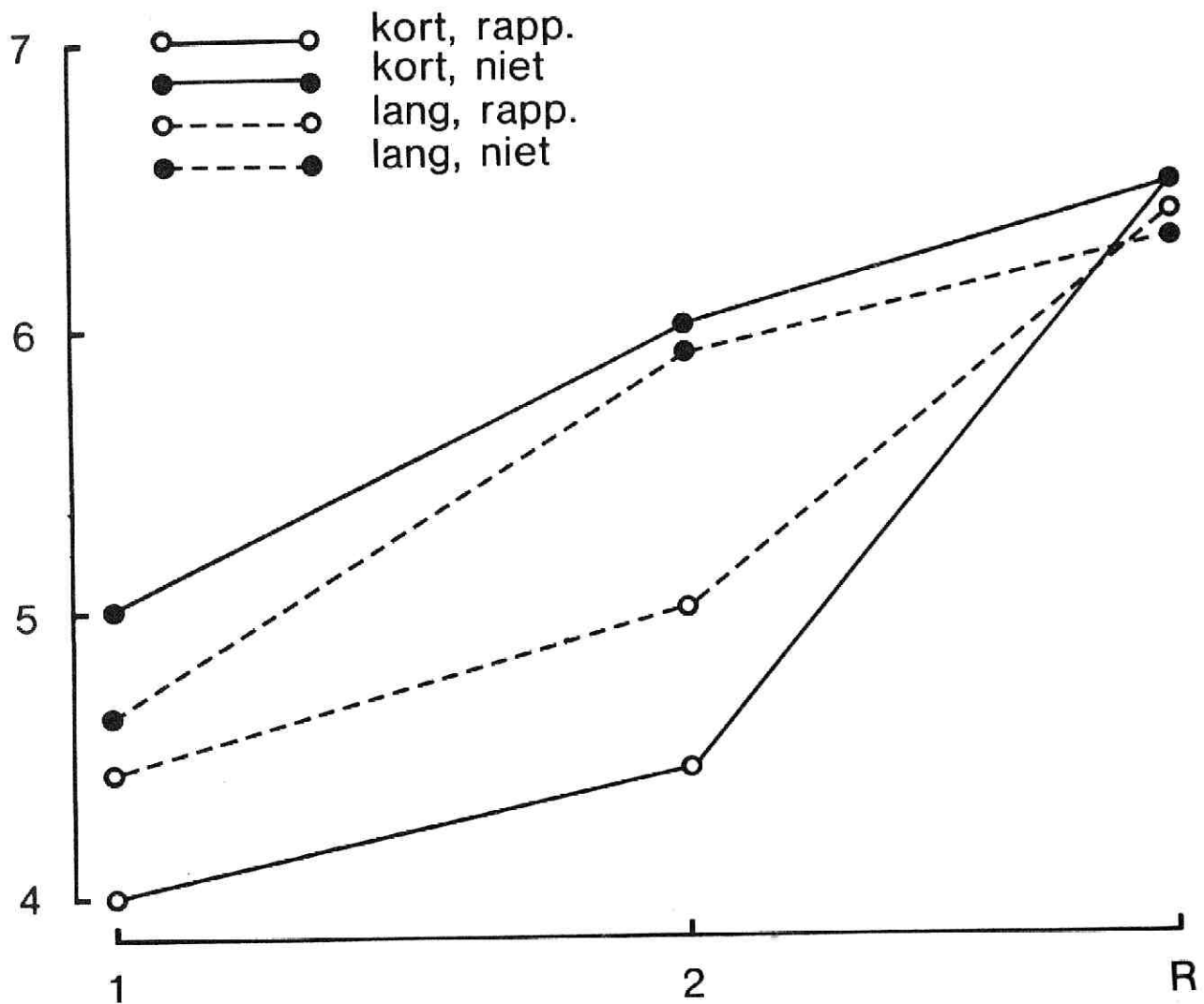
Voor de uitgevoerde analyses zijn alleen de individuele scores gebruikt die verzameld zijn tijdens de eerste twee lessen, als mede de retentietoets scores. Deze gegevens zijn geanalyseerd met behulp van variantie-analyse volgens een $2 \times 2 \times 3$ design, met herhaalde metingen op de derde factor. De factoren waren, respectievelijk, motivatie (wel versus niet meetellen van de resultaten voor het rapport), lengte van de oefening (twee versus vier lessen) en de drie herhaalde metingen.

Uit de analyse is gebleken dat alleen de eerste factor ($F(1,28) = 4.2; p < .05$) en de derde factor ($F(2,56) = 28.4; p < .0001$) significant zijn. Er waren geen significante interactie-termen. De resultaten van dit experiment zijn grafisch weergegeven in figuur 3. Hieruit blijkt in de eerste plaats dat de leerlingen die veronderstelden dat de resultaten meetellen voor het rapport, slechtere prestaties behaalden dan de leerlingen waaraan verteld was dat dit niet het geval is. Het is onduidelijk waaraan dit onverwachte effect moet worden toegeschreven. Mogelijkerwijs is hier sprake van een klas-effect (zie procedure).

Uit de variantie-analyse is ook gebleken dat de verschillen tussen de opeenvolgende metingen significant zijn. Opmerkelijk hierbij is dat de stijgende lijn die bij de eerste lessen te zien is, zich voortzet tijdens de retentie-meting. Dit effect is vrijwel onafhankelijk van de hoeveelheid oefening die de leerlingen hebben gehad. De interactie tussen hoeveelheid oefening en de herhaalde metingen is niet significant ($F(2,56) = .3; p > .05$).

SAMENVATTING EN CONCLUSIES

De hypothese dat de combinatie van feedback en verbale (zelf)instructies effectiever is dan een leersituatie waarin alleen feedback wordt gegeven, is in beide eerste experimenten bevestigd. Hierbij deed zich een opmerkelijk verschijnsel voor. In experiment 1 bleek dat de leerprestaties – althans bij de lts-leerlingen – eerst een stijging vertonen, dan een daling en vervolgens weer een stijging. Aanvankelijk hebben we gemeend dit verschijnsel te moeten verklaren op grond van het feit dat onvoldoende informatie aan de leerlingen verschaft werd bij de uitleg (vooraf) en de feedback (achteraf). Echter, in experiment 2 werd meer informatie verstrekt in die situatie waarin het leren plaats vond via een combinatie van verbale (zelf)sturing en feedback. Toch



Figuur 3: Gemiddelde score per les (en retentietoets = R) per conditie.

bleek ook daar hetzelfde verschijnsel zich ook weer voor te doen. In experiment 3 is toen nagegaan in hoeverre het (tijdelijk) afnemen van de prestaties toegeschreven zou kunnen worden aan (gebrekkige) motivatie van de leerlingen. Ook deze hypothese kon niet worden bevestigd.

De resultaten van de drie experimenten wijzen erop dat gedurende de eerste lessen op basis van zowel externe regulatie (via voorinformatie en feedback) als interne regulatie (via verbale (zelf)instructies) een adequate interne representatie wordt gevormd. Deze bevat blijkbaar voldoende informatie om een bepaald prestatie-niveau gedurende langere tijd te handhaven.

De gegevens van deze studie hebben slechts betrekking op het leren van één technische handvaardigheid: het leren slaan van center-punten in metaal op een dunne lijn. In feite gaat het hier om een successief-motorische taak, waarbij de nadruk ligt op de nauwkeurigheid waarmee de taak moet worden uitgevoerd. De taak vereist een zekere oog-hand coördinatie en het vermogen om zich te concentreren. Er worden geen complexe bewegingspatronen verlangd. Hierdoor behoeft betrekkelijk weinig informatie in het geheugen te worden opgeslagen. De interne representatie bevat in hoofdzaak gegevens over de volgorde waarin de verschillende bewegingen uitgevoerd moeten worden, het gebruik van bepaalde gereedschappen (hamer en center) en een aantal bewegingsparameters, zoals het correct plaatsen van de center en de kracht waarmee de hamer gehanteerd moet worden. Deze kenmerken van de bewegingen werden in deze taak onder betrekkelijk constante bewegings-condities uitgevoerd. Er werd steeds gewerkt

met dezelfde hamer, dezelfde center en hetzelfde materiaal waarop de center-punten geslagen moesten worden. Het is de vraag in hoeverre hierdoor een zodanige interne representatie ontstaat dat een flexibele en continue aanpassing tot stand kan komen aan wisselende uitvoerings-omstandigheden. Het is met name de vraag in hoeverre de hier gebruikte instructiemethode geschikt is voor het realiseren van een dergelijke flexibele opslag, vooral in het geval van taken die meer complexe bewegingen vereisen. Vervolgonderzoek is gewenst om hierover duidelijkheid te verkrijgen.

REFERENTIES

- Adams, J.A. (1971). A closed loop theory of motor control. *Journal of Motor Behavior*, 3, 111-150.
- Cranenburgh, B. van, & Mulder, Th. (1986). *Van contractie naar actie*. Utrecht/Antwerpen: Bohn, Scheltema & Holkema.
- Pantina, N.S. (1957). De wijze van oriëntering in de taak bij het aanvangsschrijfonderwijs. In C.F. van Parreren en J.A.M. Carpay (red.) (1972): *Sovjetpsychologen aan het woord*. Groningen: Wolters Noordhoff.
- Parreren, C.F. van (1983). *Leren door handelen*. Apeldoorn: Van Walraven.
- Pijning, H.F. (1985). Psychologische fundering van het aanvankelijk schrijfonderwijs. In: A.J.W.M. Thomassen, G.P. Van Galen, & L.F.W. de Klerk (red.), *Studies over de schrijfmotoriek: theorie en toepassing in het onderwijs*. Lisse: Swets & Zeitlinger.
- Sanden, J.M.M. van der (1986). *Het leren van technische vaardigheden: individuele verschillen bij het uitvoeren van praktijkopdrachten in het lager technisch onderwijs*. 's-Gravenhage: SVO.
- Schmidt, R.A. (1975). *Motor Skills*. New York: Harper & Row.
- Spit, P.G.F. (1983). *Het aanleren van basisvaardigheden in het lager technisch onderwijs. Verslag van een vooronderzoek*. Oosterhout: Katholieke Technische School.
- Thomassen, A.J.W.M., Keuss, P.J.G., & Galen, G.P. van (eds.) (1983). Motor aspects of handwriting. *Acta Psychologica*, 54, nos 1-3, 1-354.
- Thomassen, A.J.W.M., Galen, G.P. van, & Klerk, L.F.W. de (red.) (1985). *Studies over de schrijfmotoriek: theorie en toepassing in het onderwijs*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 29-7-1986.

Definitieve versie ontvangen 26-11-1986.

Het schatten van schooleffecten

Problemen bij de keuze van het model en de techniek.

G.G. Kreft

Vakgroep AVO-UvA, Amsterdam

J. de Leeuw

Vakgroep Datatheorie FSW-RUL, Leiden

ABSTRACT

There is some confusion in the recent literature on multilevel analysis between specification of statistical models and between the techniques used to fit these models. We try to clarify the distinction between 'one-step' and 'two-step' procedures, discussed by Boyd and Iversen, and by Van den Eeden and Saris. For this clarification it is necessary to emphasize the difference between modelling and between estimation, i.e., between models and techniques. We illustrate the importance of careful statistical modelling by comparing various versions of the linear model which have been used in multilevel analysis. In this we follow recent more technical papers by Aitkin and Longford, and De Leeuw and Kreft. Our comparisons are illustrated with the analysis of a subset of the CBS-SMVO cohort, using CITO-score as an individual level predictor. A model with fixed and constant slopes and with random intercepts described the data well.

1. THEORETISCH KADER

Inleiding.

In dit paper houden we ons bezig met de karakteristieken van onderzoek naar schooleffecten op schoolloopbanen, en de implicaties daarvan voor de keuze van statistische modellen. Dit betekent, dat we niet op de theoretische en methodologische redenen voor de keuze van een multi-niveau benadering zullen ingaan. Hiervoor verwijzen we naar Boyd and Iversen (1979) en Blalock (1984). We hebben het daarom alleen over de statistische aspecten van de verschillende modellen, de rol van de theorie, en van de juiste keuze van een model.

De definitie van het schooleffect door middel van een model.

Het vaststellen van schooleffecten op schoolloopbanen kunnen we beschrijven als: hoe is de relatie tussen de uitkomsten van onderwijs, schoolprocessen en eigenschappen van leerlingen. Een belangrijk onderdeel van dit proces is de verklaring van variabiliteit van genoemde relaties tussen scholen. Een veel gebruikt analysemodel hierbij is multi-pele regressie, waarbij de uitkomsten van het onderwijs (de 'output') gerelateerd worden aan aanvangskenmerken van de leerlingen en aan schoolvariabelen (de 'input'). Hoewel de meeste sociologen het er over eens zullen zijn, dat contexten het gedrag van individuen beïnvloeden, is er op contextuele data analyse veel kritiek geleverd. Zie Swanborn (1973, 1978) en Hauser (1974). De discussie wordt echter bemoeilijkt omdat de gebruikte termen slecht gedefinieerd worden (De Leeuw en Kreft, 1984), en ook omdat de discussie zich soms op verschillende niveaus beweegt, zonder dat de

discussianten zich hiervan bewust zijn. Met name het gebruik van een analysetechniek en het gebruik van een analysemodel worden door elkaar gehaald. In dit paper doen we een poging dit te voorkomen door in deze paragraaf te definiëren wat we bedoelen als we het hebben over een contextueel-of schooleffect.

Binnen het multiële regressie model zijn veel variaties mogelijk. Deze variaties worden in onderzoek ook door elkaar gebruikt, met als gevolg dat het belang van het schooleffect zeer verschillend is beoordeeld. Schooleffecten worden gedefinieerd met behulp van een gekozen model. De techniek wordt toegepast, **nadat** de veronderstellingen over het model gemaakt zijn. De keuze van een techniek is aan het model ondergeschikt, en in feite in gelijke mate irrelevant als de keuze of een correlatiecoëfficiënt berekend wordt met SPSS op een computer of met de hand op een rekenmachine. Een computer programma heeft binnen deze opzet geen andere functie dan het schatten van de vrije parameters en het toetsen van de veronderstellingen, die gemaakt zijn in het model. Hierna vindt een eventuele aanpassing van het model plaats op basis van de *goodness of fit* en kunnen eventueel weer nieuwe keuzen gemaakt, of nieuwe computer programma's aangesproken worden. Al deze keuzen hebben gevolgen voor de definitie van het geschatte schooleffect. Dus een ander model, met andere specificaties, zal een andere schooleffectmaat opleveren. We zien dit geïllustreerd in Aitkin en Longford (1986).

De techniek is hierbij het hulpmiddel voor de onderzoekers waarmee gekeken wordt, of het door hen aangenomen model op de gegevens past, waarna een volgende cyclus kan volgen van modelaanpassing en modeltoetsing. In feite vervangen deze cycli van modelopstelling en modelaanpassing het oorspronkelijke experiment uit de natuurwetenschappen. We kunnen (daarom) zeggen dat 'modellen fitten' het antwoord is van de sociale wetenschap op het ontbreken van mogelijkheden tot sociale experimenten. In de latere paragrafen zullen we de bovenstaande opvattingen in de praktijk toetsen, en aantonen welke invloed de verschillende modelspecificaties hebben op de grootte van het gemeten schooleffect.

We gaan hierbij in op de door Tacq (1986) gestarte discussie over de relatie tussen probleemstelling en analysetechniek. Onze discussie gaat echter verder, met name omdat we de keuze van de techniek zien als een keuze die afhankelijk is van meer zaken dan de probleemstelling alleen. Aan de hand van de probleemstelling, de kennis van het terrein waarop het onderzoek plaats vindt, en het domein waaruit de gegevens komen, stellen we een model op. De toetsing van de veronderstellingen van dat model gebeurt met behulp van een techniek. De analyse-techniek verschaft ons de parameters van het model op basis van de onderzoeksgegevens, alsmede een maat voor de juistheid van het vooraf vastgelegde model. De vraag of een model past op de gegevens wordt via statistische toetsing beantwoord. Hierbij geldt de regel dat, bij gelijke passing, een eenvoudige model met weinig parameters te verkiezen is boven een model met veel parameters. Eenvoudige modellen zijn informatiever en stabiel, ingewikkelde modellen daarentegen moeilijker interpreteerbaar.

Het analyse-niveau. Aanvankelijk werd het probleem van de analyse van meer-niveau gegevens gezien als het probleem van een keuze van het juiste analyse-niveau, in plaats van de keuze voor het juiste model. Dit niveau zou in principe bepaald moeten worden door de vraagstelling, maar werd in de praktijk vaak bepaald door omstandigheden, zoals het beschikbaar zijn van gegevens van een bepaald niveau. Dit laatste speelt een grote rol bij de analyse van economische gegevens, omdat men hier vaak alleen gegevens op groepsniveau ter beschikking heeft. Het is dan ook niet verwonderlijk dat vanuit deze traditie het meso-niveau (de school) als de belangrijkste analyse-eenheid wordt gezien.

Vanuit de Verenigde Staten komt een tweede traditie, die het micro niveau (de leerling) als de belangrijkste en soms als de enig juiste analyse-eenheid beschouwt. Hauser (1970, 1974) is daar de bekendste vertegenwoordiger van. Hij stelt dat groeps effecten artefacten zijn van falende onderzoekstheorieën. Verklaarde varianties vanuit context variabelen zouden meestal het gevolg zijn van onvoldoende doorgezette pogingen om de invloeden te verklaren vanuit het individu. In feite verwijt Hauser de onderzoekers hier, dat deze proberen het tekort schieten van een theorie op individueel niveau te maskeren door deze aan te vullen met "armzalige" maten

voor onderliggende sociale processen, zoals bijvoorbeeld een gemiddeld IQ. In principe komt de opvatting van Hauser er op neer, dat context-variabelen nooit gebruikt kunnen worden, omdat altijd de mogelijkheid blijft bestaan dat de verklaarde variantie toe te schrijven valt aan een niet in het model opgenomen individuele variabele, waardoor en waarlangs de werking van de context-variabele verklaard kan worden.

Beide tradities verenigen zich in de onderzoeksaanpak van de contextuele analyse, zoals deze nu in ontwikkeling is. Het is daarom dat contextuele analyse soms beschouwd wordt als een synthese van de Amerikaanse micro- en de Europese macrotraditie in onderzoek (Allardt, 1968).

De contextuele analyse. Micro- en macro-traditie hebben elkaars werkwijzen sterk bekritiseerd. Als we de bezwaren tegen analyses op één enkel niveau samenvatten, dan komen we tot de conclusie dat aggregatie en analyse op een hoger niveau, maar ook disaggregatie en analyse op een lager niveau, tot problemen kunnen leiden als we uitspraken willen doen die zowel met kenmerken van het individu als met kenmerken van de context rekening houden. Alle pogingen om variabelen van verschillend niveau te herleiden tot één enkel niveau hebben echter gevolgen voor de betrouwbaarheid van de resultaten van de analyse. Aggregatie naar het schoolniveau elimineert de binnen-scholen variantie van alle individueel gemeten variabelen, en dit leidt tot groot verlies van informatie. Als we disaggregeren naar het leerlingniveau zitten we weer met andere problemen. Dan krijgen leerlingen op eenzelfde school gelijke scores voor school- of leerkrachtkenmerken, waardoor deze laatste variabelen een binnen-scholen variantie van nul hebben.

In de contextuele analyse kunnen beide werkwijzen vermeden worden, omdat met meerdere niveaus tegelijk rekening wordt gehouden. Het door ons voorgestelde model is een uitbreiding van de algemeen gebruikte modellen voor lineaire regressie en variantie analyse, waarin verschillende niveaus kunnen worden opgenomen. Het fundamentele idee van deze familie van multilevel lineaire modellen is in principe eenvoudig. Er is sprake van een micromodel, dat apart gedefinieerd wordt voor elke macro- (of meso-) eenheid. In het macromodel worden de parameters van het micromodel gerelateerd aan schoolkenmerken. In ons onderzoek betekent dit, dat elke school zijn eigen model heeft. Schoolkenmerken kunnen globale structurele of geaggregeerde kenmerken zijn. Een voorbeeld van geaggregeerde kenmerken is de centrale tendentie van leerlingkenmerken. In ons onderzoek hebben we slechts de beschikking over het laatste soort kenmerken.

2. HIERARCHISCHE MODELLEN

Bij het meten van schooleffecten op schoolloopbanen stuiten we op enkele methodologische en statistische problemen, waar tot nu toe geen bevredigende oplossingen voor waren. Ze komen vooral voort uit de hiërarchische indeling van leerlingen in klassen, van klassen in scholen, en van scholen in regio's. Diverse selectie en recruteringsprocessen liggen aan de toewijzing van leerlingen aan groepen ten grondslag. Het zijn deze selectie processen, die in het verleden hebben geleid tot discussies over de "confounding effects" van leerling- en schoolkenmerken. Of in de woorden van Pedhazur (1975, p. 255): "The overlap between peer quality and school quality is so large that there seems hardly any point in referring to them as different factors." Hieruit zouden we een pleidooi kunnen afleiden, dat lijkt op het eerder aangehaalde standpunt van Hauser, om context-invloeden op te vatten als gevolgen van op individuele kenmerken gebaseerde selectie processen. Ons onderzoek heeft echter niet deze meer fundamentele theoretische discussie op het oog, als we ons als doel stellen na te gaan wat deze hiërarchische *nesting* van leerlingen binnen klassen en van klassen binnen scholen betekent voor de keuze van het lineaire model.

De laatste ontwikkelingen op het gebied van de statistiek in schooleffectstudies suggereren, dat variantie component modellen en random coëfficiënt versies van het gebruikelijke lineaire

model effectieve methoden bieden voor de geschetste multilevel problemen. We verwijzen naar Burstein e.a. (1978), Tate en Wongbundit (1983), De Leeuw en Kreft (1986), en Aitkin en Longford (1986). Met name maken deze nieuw ontwikkelde modellen het mogelijk om school-effecten van individuele effecten te scheiden, en daarbij te corrigeren voor aanvangsverschillen in de populatie. Daarnaast wordt het mogelijk de aanpassing van de verschillende modellen ook te toetsen. Dit betekent, dat enkele bezwaren tegen de wat oudere vormen van contextuele analyse, zoals we die vinden in Boyd and Iversen (1979), worden ondervangen.

3. DE TWEE-STAPSPROCEDURE

Voordat we beginnen met de bespreking van de door ons voorgestelde random coëfficiënt modellen, en de verbeteringen die dergelijke modellen betekenen, enkele opmerkingen over de *twee-stapsprocedure*. Deze techniek voor contextuele analyse wordt al in detail besproken in het boek van Boyd en Iversen (1979), en is in Nederland geïntroduceerd door Van den Eeden en Saris (1984). Om verwarring te voorkomen benadrukken we dat de twee-stapsprocedure niet gezien moet worden als een techniek op basis van een duidelijk gespecificeerd model. Deze indruk zou kunnen ontstaan naar aanleiding van de inhoudsopgave van het nuttige overzichtsartikel van Tacq (1986), waarin deze procedure als zodanig wordt voorgesteld. Ook uitspraken als: "aangezien de analyse technieken ... niet veel verder staan dan de tweestapsprocedure..." (l.c., p. 2), wijzen in deze richting.

Van den Eeden en Saris (1984) omschrijven zelf de procedure als volgt: "Wij suggereren een procedure bestaande uit twee stappen. In de eerste stap kiest men de elementen per context als eenheden en analyseert men de relaties tussen de variabelen voor deze eenheden per context. Dit betekent dat men evenveel analyses doet als er contexten zijn." (l.c., p. 166). Even verderop vervolgen de auteurs: "Hierbij krijgt men voor elke context waarden voor de coëfficiënten ... waarvoor men kan toetsen of de context invloed heeft door na te gaan of de coëfficiënten van context tot context verschillen." (l.c., p. 166). Maar er is een zekere spanning tussen het eerste en het tweede citaat. Het eerste citaat omschrijft een *techniek*, die uitgevoerd kan worden zonder dat een bepaald model verondersteld wordt. Daar staat tegenover dat het toetsen, waarover in het tweede citaat wordt gesproken, een *model* vereist, omdat de toets alleen afgeleid kan worden op basis van een model. Een model met vooronderstellingen over bijvoorbeeld ongecorrleerde fouten en onafhankelijke normaal verdeelde storingen. De volledige twee-stapsprocedure, inclusief toetsing, is dus niet toepasbaar zonder het opstellen van een model met de bijbehorende veronderstellingen over storingstermen. Dit betekent, dat binnen het recept van de twee-stapsprocedure er een aantal modelkeuzen overblijven.

Een van de belangrijkste is de keuze tussen een random- of een fixed-coëfficiënt model. Van den Eeden en Saris schrijven hun vergelijkingen alsof ze een random coëfficiënt model veronderstellen. Immers de coëfficiënten uit de eerste stap worden in de tweede specificatiestap voorzien van een storingsterm. Ze verbinden daar echter geen consequenties aan voor de verdeling van de storingstermen, zodat bij hen niet duidelijk wordt welk model verondersteld wordt.

Op basis hiervan is het ook verstandig de "twee-stapsprocedure" te onderscheiden van "een berekening in twee stappen". In het eerste wordt een model aangenomen, terwijl in het laatste alleen wordt gezegd hoe de procedure verloopt, nl. in twee stappen. Dit laatste is een handige en gemakkelijk te beschrijven methode om aan te geven hoe je iets uit kan rekenen als je te maken hebt met gegevens van verschillend niveau. Bij een "twee-stapsprocedure" wordt echter uitgegaan van een vooraf geëxpliciteerd model, waarin voor beide niveaus aannamen voorkomen over de manier waarop de gegevens tot stand zijn gekomen. Dit betekent dat men theorie nodig heeft omtrent de aard van het bestudeerde experiment.

Een dergelijke twee-stapsmodellering wordt gebruikt, als men er van overtuigd is, dat twee-stapsschatters beter zullen zijn dan één-stapsschatters. Of dit zo is zal in belangrijke mate afhangen van de nauwkeurigheid en de correctheid waarmee het model op ieder van de niveaus

is gespecificeerd. Het blijft echter bij het gebruik van de twee-stapsprocedure toch zaak, om direct, bij de eerste stap al een geschikt model met een daarbij behorende geschikte techniek te kiezen, omdat keuzen in de eerste stap in volgende stappen doorwerken. Ieder van de afzonderlijke stappen heeft daarbij betrekking op slechts een enkel niveau, zodat we bij de afzonderlijke stappen niet van een multilevel techniek kunnen spreken. Door combinatie van de stappen achteraf kunnen we wel multilevel uitspraken doen.

Een andere bron van verwarring is het onderscheid tussen de "single- and separate equation", termen die we bij Boyd and Iversen (1979) tegen komen. Vaak worden deze termen verward met de termen "één-staps- en twee-stapsprocedure". En dit geheel ten onrechte, omdat beide laatst genoemde procedures als technieken gebruikt kunnen worden om een single-, maar ook om een separate-equation te berekenen. Een volgende verwarring kan gemakkelijk ontstaan, omdat er ook een "single- and separate-equation procedure" bestaat (Tate and Wongbundit, 1983). Onder de single equation procedure wordt een enkele multi-pele regressie analyse verstaan, waarbij variabelen van verschillende niveaus in één enkele analyse worden berekend. Voor verdere discussie op dit punt verwijzen we hier naar De Leeuw en Kreft (1986).

4. BEPERKINGEN VAN COVARIANTIE ANALYSE MODELLEN MET VASTE EFFECTEN

Bij een keuze voor een van de vele mogelijke randomcoëfficiënt modellen spelen een groot aantal overwegingen een rol. Sommige zijn gebaseerd op kennis van het onderzoeksterrein, andere op a-priori overwegingen van statistische aard. We geven als voorbeeld de keuze voor een speciaal (restrictief) geval uit deze familie van modellen: het bekende covariantie analyse model. Dit model is "fixed" in de parameters, dwz. dat de parameters worden opgevat als constanten, en niet (zoals in het random model) als variabelen met een storingsterm.

De keuze voor dit vaste effectenmodel is gebaseerd op drie veronderstellingen. Ten eerste wordt verondersteld dat de scholen uit het onderzoek deelpopulaties zijn, en geen steekproef uit een populatie van scholen. De geschatte effecten hebben daardoor alleen betrekking op deze scholen en zijn in principe niet generaliseerbaar naar scholen die niet in het onderzoek zijn opgenomen. In experimenten betekent iets dergelijks dat alle mogelijke variaties van het experimenteel ontwerp waarin we geïnteresseerd zijn in het onderzoek voorkomen. We willen bijvoorbeeld uitsluitend uitspraken doen over deze drie meststoffen, of over deze vier leer machines. In veel onderzoek van onderwijs moet echter aangenomen worden, dat de in het onderzoek aanwezige scholen een steekproef zijn uit de mogelijke scholen, waarover je wel generaliserende uitspraken wilt doen. In een dergelijke situatie is het dan ook logisch uit te gaan van "random" effecten in plaats van "fixed" effecten (zie verder Hays, 1974, hoofdstuk 12 en 13). Een andere beperkende veronderstelling van het covariantie analyse model is de aanname, dat de regressiecoëfficiënten (de "slopes") binnen contexten niet van elkaar verschillen.

Het is duidelijk, dat een keuze voor dit bepaalde vaste effecten model in schooleffect onderzoek niet in alle situaties juist is. Ten eerste gaat het er wel om generalisaties te maken naar de Nederlandse situatie in het algemeen, terwijl er bovendien vanuit verschillende theorieën verondersteld wordt, dat een bepaald schoolbeleid van invloed is op de regressiecoëfficiënt. De aanname van gelijke "slopes" binnen contexten verhindert onderzoekers om de invloed van schoolbeleid op de regressiecoëfficiënt te schatten. Met name bij de evaluatie van het voor-rangsbeleid zal men geïnteresseerd zijn in schoolkenmerken, die invloed uitoefenen op bijvoorbeeld de relatie tussen het ouderlijk milieu en de schoolprestatie.

Een derde veronderstelling van het vaste effecten model, die al of niet impliciet is bij de keuze voor covariantie analyse, is een aanname over de storingstermen van het individuele niveau. De storingstermen worden (immers) geacht de invloed samen te vatten van een groot aantal variabelen, die niet gemeten zijn, maar die wel invloed uitoefenen op de afhankelijke variabele. Van storingstermen wordt aangenomen dat ze random zijn, dwz. niet onderling gecorreleerd. In onderzoek naar schooleffecten ligt het echter voor de hand te veronderstellen, dat een gedeelte

van deze ongemeten variabelen binnen een school een zelfde effect zullen hebben, nl. het schooleffect. Hieruit volgt dat veronderstelling van gecorreleerde storingstermen binnen scholen eveneens voor de hand ligt. Deze gecorreleerde storingstermen komen te voorschijn in de random coëfficiënt versies van contextuele analyse, die we hierna bespreken.

5. RANDOM COEFFICIENT MODELLEN

Random of fixed.

Het effect van de school wordt binnen een niet-experimentele situatie gedefinieerd door het gehanteerde model en de veronderstellingen ervan. Een van de veronderstellingen van de random coëfficiënt modellen is de aanname met betrekking tot de storingstermen van het tweede niveau. Een uitgebreide discussie over het principiële verschil tussen "fixed" en "random" modellen vinden we in het eerder aangehaalde hoofdstuk 12 uit Hays (1974). Het verschil tussen beide modellen ligt in de veronderstellingen van het onderzoek met betrekking tot de representativiteit van de verschillende contexten.

Het in dit paper gebruikte basismodel is

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + e_{ij}, \quad (1a)$$

met

$$\alpha_j = \alpha + u_j, \quad (1b)$$

$$\beta_j = \beta + t_j. \quad (1c)$$

Met onderstreping van de variabelen en coëfficiënten in deze vergelijkingen geven we hier en in het vervolg van dit paper aan dat deze random zijn. Model (1a) heeft slechts één onafhankelijke variabele op individueel niveau. De X_{ij} zijn de metingen op die variabele (van individu i op school j). In (1b) zien we een storingsterm u_j , die evenals de storingstermen van het individuele niveau e_{ij} de ongemeten variabelen vertegenwoordigt, maar nu van het contextuele niveau. In de tweede vergelijking hebben we het random intercept $\alpha_j = \alpha + u_j$ gedefinieerd. Op dezelfde manier is t_j de storingsterm van de helling, en definieert (1c) de random slope $\beta_j = \beta + t_j$. De structuur van het random-coëfficiënt model is hiermee aangegeven. Een structuur, die op een natuurlijke wijze voortkomt uit de manier waarop de gegevens in de werkelijkheid gegroepeerd zijn.

De behoefte aan random coëfficiënt modellen wordt het duidelijkst beschreven in Aitkin en Longford (1986), die de verschillende variantie analytische modellen met elkaar vergelijken. Zij concluderen, dat de eerst genoemde modellen onmisbaar zijn in situaties waar het context effect "appreciable" is. (l.c. p. 26). Juist in genoemde situaties geven andere modellen misleidende informatie. Als voorbeeld geven zij de "serious overstatement" van de significantie van de contextuele variabele in een model, waarbij beide niveaus in de regressie vergelijking de variantie van de afhankelijke variabele verklaren. Deze "overstatement" is het gevolg van de onderschatting van de standaardfout in de regressie op het individuele niveau. In Kreft en Van den Eeden (1985) vinden we een soortgelijke opmerking naar aanleiding van het proefschrift van Van Marwijk Kooy.

Het lineaire model, random of fixed.

Bij analyses van schooleffecten is het tot nu toe meest gebruikte model het lineaire model, en de meest gebruikte techniek multiple regressie analyse. Deze modellen worden in onderzoek van onderwijs ook wel aangeduid met "educational production function" of met de naam "input-output modellen". Het kenmerk van deze modellen is, dat eigenschappen van leerlingen en eigenschappen van scholen (de "inputs") gerelateerd worden aan de uitkomsten van het leerproces (de "outputs"). Een klassiek voorbeeld van een "single equation production function" is

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + e_{ij}, \quad (2)$$

Hierbij is Y_{ij} de score van de schoolprestatie van leerling i op school j , X_{ij} is de observatie van de onafhankelijke variabele op individueel of op schoolniveau van dezelfde leerling, α_j en β_j de onbekende parameters, die geschat moeten worden, en e_{ij} is de random errorterm, met de gebruikelijke veronderstellingen van een gemiddelde nul en een variantie σ^2 .

Ook in het random coëfficiënt model is het lineaire model het uitgangspunt van de analyse. Het bovenstaande model (2) kunnen we beschouwen als een speciaal geval van het algemene model, nl. een model, dat de schooleffecten als vaste coëfficiënten opneemt en geen homogeniteit veronderstelt van de individuele storingstermen binnen contexten. Dit model is daarom behoorlijk restrictief. Meer algemeen is het random coëfficiënt model, waarbij een grotere homogeniteit verondersteld wordt tussen populaties op eenzelfde school, dan binnen de gehele steekproef gesommeerd over alle scholen. Deze homogeniteit is in het model opgenomen in de context-storingsterm. Anders gezegd de covariantie van de storingstermen op het individuele niveau, binnen eenzelfde context, is een maat voor schooleffect, of voor homogeniteit, veroorzaakt door een gelijke onderwijs situatie. De vergelijking van het random coëfficiënt model op het individuele niveau, of zo men wilt als "single equation", is

$$Y_{ij} = \alpha + \beta X_{ij} + u_j + t_j X_{ij} + e_{ij} \quad (3)$$

Hierbij is α het intercept van de gemiddelde school (het vaste deel), en is u_j de bijbehorende storingsterm (het random deel). Beide termen behoren bij de stochastische variabele, die het intercept in deze modellen is. Hetzelfde geldt voor de helling, waarbij we een vast deel vinden in de parameter β van de gemiddelde school, en een random deel in de storingsterm t_j , die tevens gecorreleerd is met u_j . Beide contextuele storingstermen zijn ongecorrleerd met de storingsterm op het individuele niveau. Deze laatste heeft een variantie σ^2 . De gecorreleerdheid van de storingstermen binnen contexten wordt daardoor bepaald door zowel de variantie van de individuele storingsterm als door de gewogen contextuele variantie. Het is duidelijk dat zowel (2) als (3) speciale gevallen zijn van (1). Als we in (1) aannemen, dat u_j en t_j nul zijn, dan hebben we het fixed effect model (2). Wanneer we aannemen in (1) dat de vaste delen van het intercept en de helling gelijk zijn, dus $\alpha_j = \alpha$ en $\beta_j = \beta$ voor alle scholen, dan hebben we model (3).

Als we het meest algemene model (1) als uitgangspunt nemen op het individuele niveau, dan betekent dit, dat de drie parameters van dit model, het intercept, de regressie coëfficiënt en de variantie van de individuele storingsterm "random" zijn. Dit betekent, dat niet gezocht wordt naar de "best schatter" van intercept of regressiecoëfficiënt, maar naar de variatie van deze "constanten" over contexten, uitgedrukt in de variantie van de contextstoringstermen. De aannamen omtrent deze storingstermen zijn van belang, omdat in het lineaire model de eigenschappen van toets en schatters geheel bepaald wordt door de aannamen omtrent de storingstermen.

Slopes en intercepts "as outcomes".

We geven in deze paragraaf een aantal basisbegrippen van het regressiemodel en de manier waarop we deze gebruiken in de beide stappen van onze analyse. De regressiecoëfficiënten van de binnen-scholen regressieanalyses uit de eerste stap worden in de tweede stap gebruikt als de afhankelijke variabelen in een regressieanalyse op schoolniveau. Deze modelmatige vorm van aggregatie vinden we al eerder in de literatuur. De procedure staat bekend onder de naam "slopes as outcomes", naar aanleiding van een artikel van Burstein, Linn en Capell, uit 1978. Dit laatste betekent, dat de afhankelijke variabele in de tweede stap de regressiecoëfficiënten van de individuele analyses per school zijn. Dergelijke schoolspecifieke regressiecoëfficiënten kunnen van belang zijn voor het beleid bij evaluaties van de effectiviteit van scholen, bijvoorbeeld vanuit de vraag welke scholen in aanmerking komen voor innovatie-maatregelen. Om vergelijkbare reden kan je ook spreken over "intercepts as outcomes". In dat geval wordt de afhankelijke variabele in de tweede stap gevormd door de intercepten van de individuele analyses per school. Ook deze schoolspecifieke intercepten kunnen van belang zijn voor het beleid bij het vaststellen van verschillen in aanvangsposities van de leerlingen van scholen, bijvoorbeeld bij de uitvoering van het zgn. onderwijsvoorrrangsbeleid.

In beide gevallen betekenen dit, dat eerst per context een eenvoudige regressie berekend wordt. Er zijn n_j scores op de afhankelijke variabele op school j , en n_j scores op de onafhankelijke variabele (in ons voorbeeld resp. de individuele schoolloopbanen en de individuele CITO-scores). Voor iedere context apart geldt het volgende: het intercept (a_j) is het aanvangsniveau van de afhankelijke variabele. Het geeft de waarde aan bij $x = 0$. Ongelijkheid in de intercepten tussen scholen betekent, dat alle leerlingen van een zelfde school hoger of lager scoren op de afhankelijke variabele. Het intercept is een algemene maat voor het schooleffect, en het wordt daarom in covariantie analyse gebruikt om verschillen tussen scholen te toetsen. De helling (b_j) geeft per school aan hoeveel een punt meer op de onafhankelijke variabele waard is op de afhankelijke variabele. Als de hellingen, vergeleken tussen contexten, significante verschillen vertonen, wijst dit op interactie tussen de context en individuele variabelen uit het model. Concreet betekent dit, dat leerlingen met eenzelfde score op een onafhankelijke variabele in de ene school beter uitkomen dan op een andere school.

Modellen, parameters en toetsen.

Het voordeel van het random coëfficiënt model is duidelijk. In dit model worden de ongemeten variabelen van het individuele en de ongemeten variabelen van het context niveau, die samengevat zijn in de storingstermen, uit elkaar gehaald. Deze modellen hebben echter nog een ander, niet onaanzienlijk voordeel boven vergelijkbare vaste effecten-modellen. Dit voordeel zit in de reductie van het aantal te schatten parameters. In het door ons in de volgende paragraaf geanalyseerde databestand van 106 scholen werden voor het ANCOVA model (fixed effect covariantie analyse) 108 parameters geschat. Dit grote aantal komt doordat voor elk van de 106 scholen een intercept geschat moet worden. Daarnaast hebben we, in het geval van het ANCOVA-model, nog een parameter voor de regressie coëfficiënt en de variantie van de storingsterm van het individuele niveau te schatten. Wanneer we hiernaast ook de regressiecoëfficiënten en varianties per school vrij laten, dan hebben we zelfs 318 parameters. Bij toepassing van het random coëfficiënt model (3) wordt het aantal te schatten parameters beperkt tot de drie varianties van de individuele en contextuele storingstermen, het fixed intercept, de fixed regressiecoëfficiënt, en de covariantie van de de contextuele storingstermen. Deze laatste modellen zijn daardoor aanzienlijk eenvoudiger. Er is immers een groot verschil tussen het schatten van 318 of 6 parameters.

Over het toetsen van modellen nog het volgende. De toetsen volgen min of meer vanzelf uit de maximum likelihood methoden, waarmee de parameters uit onze modellen worden geschat. Voor details verwijzen we naar De Leeuw en Kreft (1986). Elk model wordt getoetst ten opzichte van een nulmodel, of ten opzichte van een ander, minder restrictief model. Alleen het nulmodel zelf kan niet getoetst worden. Het nulmodel is het meest algemene en minst restrictieve model van de familie van modellen.

6. MODELTOETSEN MET VOORBEELDEN

Voor het fitten van de verschillende modellen gebruiken we gegevens uit het CBS databestand, het SMVO cohort 1977. We trekken hieruit een deelsteekproef van 3590 leerlingen verdeeld over 106 scholengemeenschappen voor mavo, havo en vwo. Voor de selectie procedures verwijzen we naar Kreft en De Leeuw, 1986. De door ons geselecteerde individuele achtergrondvariabele is de CITO-score op een verkorte versie van de reken- en taaltoets. De afhankelijke variabele is de speciaal voor dit doel geconstrueerde schoolloopbaanvariabele (zie ook hiervoor Kreft en De Leeuw, 1986). In onderstaande Tabel 1 staan de verschillende modellen met het aantal te schatten parameters. In deze tabel zien we ook de modeltoetsen, met bijbehorende χ^2 -waarden en vrijheidsgraden.

Tabel 1. Samenvatting modeltoetsen.

Model	# vrije Parameters	Verschil Log- Likelihoods	Chikwa- draat	DFR
(1) nulmodel fixed	318	****	****	****
(2) covariantie analyse	108	(1)-(2)	280.16	210
(3) gelijke intercepten fixed	3	(2)-(3)	437.40	105
(4) contextueel fixed	4	(2)-(4)	418.95	104
		(5)-(4)	162.58	1
(5) contextueel random	5	****	****	****
(6) gelijke intercepten random	4	(5)-(6)	4.45	1

Onder het nulmodel verstaan we hier het model van formule (2), met $3 \times 106 = 318$ vrije parameters. Covariantieanalyse heeft $106 + 1 + 1 = 108$ vrije parameters, en als we veronderstellen dat de 106 intercepten gelijk zijn krijgen we het fixed model met gelijke intercepten, met nog $1 + 1 + 1 = 3$ parameters. In de contextuele modellen die in de tabel genoemd worden veronderstellen we dat de hellingen allemaal gelijk zijn, en dat het intercept een lineaire functie is van een contextuele variabele, in ons geval de gemiddelde CITO score per school. Dus

$$\alpha_j = p + qx_j + u_j, \quad (4)$$

waarbij in de vaste effecten versie de storingsterm u_j verdwijnt. In formule (4) is x_j de gemiddelde CITO-score op school j . Het contextuele vaste model heeft daardoor 4 parameters (p , q , β , en σ^2), het random model heeft er 5 (ook nog ω^2 , de variantie van de contextuele storingsterm u_j). Het gelijke intercepten random model tenslotte behoort bij formule (3) met t_j de random component van de helling, gelijk aan nul. De vrije parameters zijn α , β , ω^2 , en σ^2 . In Tabel 2 vinden we schattingen van de vrije parameters.

Tabel 2. Samenvatting modelschattingen.

nulmodel: intercepten, hellingen, en individuele varianties verschillen per school.

covariantie analyse: intercepten verschillen per school.

contextuele modellen: $p = -2.0820$ en $q = .0163$

Model	Intercept	Helling	Variantieschattingen	
			Individueel	School
nulmodel	****	****	****	.0
covariantie analyse	****	.0622	.5335	.0
gelijke intercepten fixed	-2.7410	.0599	.6026	.0
contextueel fixed	****	.0622	.5996	.0
contextueel random	****	.0622	.5496	1.6868
gelijke intercepten random	-2.8224	.0616	.5497	1.8012

De varianties en de covariantie van de beide individuele variabelen uitgesplitst naar tussen en binnen staan in Tabel 3. We kunnen in de gevallen van covariantie-analyse en gelijke intercepten fixed zien hoe de parameterschattingen in Tabel 2 uit deze statistieken afgeleid zijn.

Onderaan Tabel 3 staan drie verschillende schattingen van regressie-coëfficiënten van de regressie van CITO-toets op schoolloopbaan, met de daarbij behorende schatting van de

Tabel 3. Partitionering van de varianties en covarianties in tussen- en binnencomponenten.

	Totaal	Tussen	Binnen	T/B
$v(x)$	95.1289	13.6509	81.4780	0.1435
$v(y)$	0.9436	0.0948	0.8488	0.1005
$c(x, y)$	5.6957	0.6265	5.0692	0.1100
β	0.0599	0.0459	0.0622	
σ^2	0.6026	2.2355	0.5334	

variantie van de individuele storingsterm. In iedere kolom is de beta berekend door de covariantie van xy te delen door de variantie van x . Dit levert de schatting op van het model, waarvan aangenomen wordt, dat alle intercepten en ook alle hellingen voor alle scholen gelijk zijn (het gelijke intercepten fixed model uit Tabel 1). In de tweede kolom van Tabel 3 zien we schattingen voor een ander veel gebruikt model in schooleffectiviteits onderzoek. In Nederland bijvoorbeeld in het proefschrift van Van Marwijk Kooy. In dit model is de "unit of analysis" de school, en de regressie is berekend over 106 scholen, waarbij de binnen scholen variantie wordt verwaarloosd. We zien dat de bij het model horende coëfficiënt onbetrouwbaar is, vanwege de grote foutenvariantie. In de derde kolom vinden we de schatter van de helling die hoort bij het covariantie analyse model.

De laatstgenoemde twee modellen zijn modellen die schoolverschillen kunnen meten, maar niet verklaren (dat wil zeggen relateren aan andere variabelen die schoolverschillen beschrijven). Dit gebeurt wel in de contextuele modellen (Boyd en Iversen, 1979). Een dergelijk model, te beschouwen als de tweede stap na de covariantie analyse, paste slecht op onze gegevens. Terwijl de daarvoor uitgevoerde covariantie analyse wel goed paste, zoals we kunnen zien aan de chi-kwadraat en vrijheidsgraden in Tabel 1.

Een toepassing van een model onder de aanname van random coëfficiënten op dezelfde gegevens geeft ons de gelegenheid om binnen dit model te toetsen, of de aanname waaronder de vorige modellen getoetst zijn, juist is. We toetsen of ω^2 , de variantie van de context storingsterm, gelijk is aan nul. Dit levert een chi-kwadraat van 162.58 met een enkele vrijheidsgraad. Zie weer Tabel 1. De conclusie is hier gerechtvaardigd, dat voor onze gegevens de intercept storingstermen binnen contexten gecorreleerd zijn. Een traditioneel contextueel model past niet bij deze data.

Na het fitten van de verschillende modellen lijkt het erop dat we de variabiliteit van de intercepten kunnen 'verklaren' door binnen scholen gecorreleerde storingstermen aan te nemen. De correlatie tussen de storingstermen van leerlingen op eenzelfde school varieert van .05 (voor scholen met 60 leerlingen) tot .35 (voor scholen met 6 leerlingen). Anders gezegd: voor grote scholen kunnen we zeggen dat de variantie van de storingstermen 90% individueel 'bepaald' is, voor kleine scholen is wel 40% door de school 'bepaald'.

7. REFERENTIES

- Aitkin, M. A., & Longford, N. T. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, A* 149, 1-43.
- Allardt, E. (1968). The merger of american and european traditions of sociological research: contextual analysis. *Social Sciences Information*, 7, 151-168.
- Blalock, H. M. (1984). Contextual-effects models: Theoretical and methodological issues. *Annual Review of Sociology*, 10, 353-372.
- Boyd, L. H. & Iversen, G. R. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth.
- Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics*, 3, 347-383.

- Eeden, P. van den, & Saris, W. E. (1984). Empirisch onderzoek naar multilevel uitpraken, *Mens en Maatschappij*, 59, 165-178.
- Hauser, R. M. (1970). Context and consex: a cautionary tale. *American Journal of sociology*, 75, 645-663.
- Hauser, R. M. (1974). Contextual analysis revisited, *Sociological Methods and Research*, 2, 365-375.
- Hays, W. L. (1974) *Statistics for the social sciences*. Holt, Rinehart & Winston, London.
- Kreft, G. G. & Leeuw, J. de. (1986). *Schooleffecten op schoolloopbanen*. Paper gepresenteerd op de gezamenlijke bijeenkomst van de SISWO-werkgroepen "Longitudinaal" en MULOG.
- Kreft, G. G. & Eeden, P. van den (1985). Valkuilen bij het proefschrift van L. van Marwijk Kooy, en hoe deze uit de weg te gaan, In: D. P. Kroes (red), *Studiebijeenkomst over de dissertatie "Scholen verschillen"*. SISWO-Publicatie 306. Amsterdam: SISWO.
- Leeuw, J. de. & Kreft, G. G. (1984). *Over multilevel analyse*. Leiden, The Netherlands: Department of Data Theory FSW/RUL.
- Leeuw, J. de. & Kreft, G. G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11, 57-85.
- Marwijk Kooy, L. (1984). *Scholen verschillen. Een verkennend onderzoek naar het intern functioneren van vijftientig scholengemeenschappen vwo-havo-mavo*. Dissertatie Rotterdam.
- Pedhazur, E. J. (1975). Analytic methods in studies of educational effects. *Review of Research in Education*, 3, 243-286.
- Swanborn, P. G. (1973), *Variabelen en hun meting*. Boom, Meppel.
- Swanborn, P. C. (1978), Schoolkenmerken en leerlinggedrag. Enkele methodologische merkwaardigheden in het onderzoeksbedrijf. *Mens en Maatschappij*, 53, 209-216.
- Tacq, J. J. A. (1986), *Van multiniveau probleem naar multiniveau analyse*. Paper gepresenteerd op de sociologendagen april, 1986.
- Tate, R. L. & Wongbundit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics*, 8, 103-120.

Manuscript ontvangen 29 mei 1986.

Definitieve versie 12 januari 1987.

Group Effects on the Development of School Subject Motivation and Achievement

Ton Mooij*

Catholic University of Nijmegen, Institute for Social and Behavioural Studies, Graafseweg 274, Nijmegen.

ABSTRACT

Group or mean teaching situation characteristics during a school-year were expected to interact with and influence a pupil's orientation, competence, and achievement with respect to a school subject. Specific hypotheses for the effects of internal differentiation and individual evaluation of learning achievement were formulated. Other group characteristics were included for exploratory reasons. A longitudinal investigation was carried out with 39 first year groups of pupils in secondary education. The multi-level analysis used was a two step multiple regression procedure. The analysis revealed various group effects in the investigated school subjects (Dutch language, mathematics/arithmetic, and manual dexterity). However, the hypotheses were scarcely supported.

1. INTRODUCTION

A pupil's learning can be seen as processes in which the pupil's abilities and skills are stimulated by an interaction with the learning environment. In this respect Glaser (1976) referred to a theory on the acquisition of competence by a pupil: 'Such a theory would be concerned with how an individual acquires increasingly complex performances by assembling the present components of his repertoire, manipulating the surrounding conditions and events, and employing his knowledge of how he learns' (p. 310).

Attention to the individual side of such an interactional person-environment relationship has been given, among others, by White (1959). He used the concept of competence to indicate: '(...) an organism's capacity to interact effectively with the environment' (p. 297). This competence concept was further extended by Harter (1981, 1982a, 1982b). She is especially interested in developmental and socializational aspects of competence in children.

The educational system is the 'official' learning environment of a pupil. The general goal of this system is to develop each pupil's cognitive, social, and creative abilities and skills. To realize this goal a certain learning content is supplied to (usually) groups of pupil by means of different didactic procedures. The pupils can be organized in various ways, and parallel to this the teaching-learning processes can be differentiated internally to different degrees (De Corte, Geerligs, Lagerweij, Peters, & Vandenberghe, 1976; Nijhof, 1981). Other relevant educational system variables are the teacher's management behaviour (Rosenshine & Furst, 1973; Emmer & Evertson, 1980; and Kounin, 1970), the number of pupils in the group (Glass, Cahen, Smith, & Filby, 1982), and the kind of evaluation of the pupils' learning achievement (Rheinberg, 1980; Van Oudenhoven, 1983).

* An earlier version of this article was presented at the 'Onderwijsresearchdagen', Utrecht 1986. The author is indebted to the reviewers of TOR, and to dr. S. F. Foster, and dr. M. J. M. Voeten for their comments on this earlier version.

Address: Graafseweg 274, 6532 ZV Nijmegen

The above-mentioned educational system variables reside at the group or teaching situation level. During a school-year these variables interact with each pupil's abilities and skills, and also with his or her motivation and achievement within a school subject. This kind of interaction is the kernel of the aptitude-treatment interaction or ATI paradigm (Cronbach, 1976, 1983; Cronbach & Snow, 1977; Snow & Lohman, 1984). In ATI theorizing and research attention is usually restricted to one or two of the above group level variables and their effects on pupil variables. This article, however, focuses on the relative importance of group level variables to the way in which a pupil's motivation and achievement concerning some specific school subjects develops in the course of a school-year. The score on a motivational or achievement variable at the end of a school-year may vary according to its dependence on the score at the start of the school-year. Within a class or group of pupils this dependence on initial status may be either clearly different or about the same for each of the pupils. Group level variables may influence the within-group developmental process. The research question is then formulated as: *which educational system variables at the group level are most important in influencing the degree of dependence on initial status with respect to the pupils' motivation and achievement in school subjects?*

Because both pupil level variables and group level variables are included, the study reported here can be characterized as a multi-level study (cf. Burstein, 1980; Van den Eeden & Hüttner, 1982; Mooij, 1980, 1984). The data used are from a research project funded by the Dutch Foundation for Educational Research (SVO; SVO-project 0483).

2. THEORY

The lessons in a specific school subject usually function as a series of encounters between pupils and a particular teacher, including specific tasks, rather specific organizational and didactic characteristics, and more or less articulated norms on learning achievement (see Beckers, 1982; Kounin, 1970; Mooij, 1982; Speklé, 1982). According to the theory of Neisser (1976), a pupil has cognitive maps in which information about his or her skills, abilities, performance, and expectations has been integrated in a school subject specific way (Cattell & Child, 1975, chapter 8). The information in a cognitive map also contains appraisals or values concerning characteristics of the environment, for example learning task characteristics, the school subject teacher, the other pupils, and performing alternative actions during the lesson. Those characteristics or the pupil's expectations concerning them may function as a motive guiding the pupil's actions (cf. Heckhausen, 1980; Boekaerts, in press). A school subject specific cognitive map enables making choices out of the motives. *School subject motivation* is then defined as the process in which a pupil chooses among different possibilities to act with respect to the lessons in a specific school subject.

This school subject motivation expresses itself in the pupil's actual *task behaviour* during the lesson (cf. Emmer & Evertson, 1980; Evertson, Emmer, & Clements, 1980; Mooij, 1985). This motivation is also indicated by information available in the school subject specific cognitive map, that is in beliefs of *competence* in and *orientation* towards the school subject (Bowerman, 1978; Mooij, 1986). *School subject competence* is defined as a pupil's beliefs of effectiveness in getting along with a certain school subject. This competence will influence the pupil's choices with respect to alternative next actions regarding the school subject, so a higher school subject competence will positively stimulate the pupil's orientation towards the school subject and promote the occurrence of next learning actions (concerning more difficult tasks). *School subject orientation* is defined as a pupil's interest in and experienced intensity of learning actions concerning a specific school subject. The school subject orientation is hypothesized to indicate the degree to which a school subject functions as a motive for the pupil's task behaviour during the lesson.

The skills and abilities of a pupil during the lessons in a specific school subject are expressed in specific performances. These performances are used to measure the pupil's achievement in the

school subject. *School subject achievement* is defined as the degree to which a pupil's performances meet the learning requirements of the school subject.

According to the interactional theorizing of Bowerman (1978), Harter (1981, 1982a, 1982b), and Heckhausen (1980) an action-effect-affect sequence is hypothesized to exist: a pupil's task behaviour (action) leads towards school subject achievement, which is evaluated with a certain school subject mark (effect), which leads towards specific affects. These affects will influence the beliefs of effectiveness in and the orientation towards the particular school subject. At the start of a school-year a pupil within a certain group or class of pupils can be characterized by his or her scores on school subject orientation, school subject competence, and school subject achievement. During the school-year different variables within the educational system influence the pupil's standing on these orientation, competence, and achievement variables. At the group or teaching situation level the degree of internal differentiation within the teaching situation, but also the kind of evaluation of the pupils' learning achievement, may be important.

Internal differentiation refers to organizational and curriculum conditions promoting separate learning processes for the pupils within a class-room. A pupil within a group characterized by a higher degree of internal differentiation has more possibilities to learn according to his or her own tendencies and preferences than a pupil in a group with only low or no internal differentiation. In highly internally differentiated teaching situations a pupil's *initial* school subject competence, orientation, and achievement (at the start of the school-year) determines the school subject competence, orientation, and achievement (at the end of the school-year) to a higher degree than occurs in low internally differentiated teaching situations (cf. Bloom, 1984). In other terms: as the degree of internal differentiation in a teaching situation during a school-year becomes higher, the degree of dependence on the initial status in the pupils' school subject orientation, school subject competence, and school subject achievement is induced to become higher (*hypothesis 1*).

A comparable hypothesis can be set up with respect to the kind of evaluation of pupils' learning achievement in the pupil group. If this occurs *individually* by comparing each pupil's learning achievement with his or her own earlier achievement, or by evaluating the effort of the pupil, each pupil can influence the own school subject achievement and consequent school subject competence and school subject orientation. Other ways to evaluate the learning achievement are *group comparative* (a pupil's mark is based on the comparison with the results of the class-mates), and *normative* (a pupil's mark is based on the learning content's result (criterion referenced measurement)). With group comparative evaluation or normative evaluation a pupil can get an insufficient mark despite his or her effort, which may reduce the within-group dependence on initial status in school subject competence, school subject orientation, and school subject achievement in the course of a school-year. In groups that are evaluated individually a higher degree of dependence on initial status in school subject achievement, school subject competence, and school subject orientation during a school-year will therefore exist than in groups evaluated in a group comparative way or normative way (*hypothesis 2*; cf. Rheinberg, 1980; Van Oudenhoven, 1983).

Other group variables potentially relevant to the degree of dependence on initial status in the pupils' school subject orientation, competence, and achievement during a school-year will be explored in this article. These variables concern the organization of pupils, the kind of lesson task, specific didactic procedures in transmitting the learning content, teacher's management behaviour, the number of pupils, and the mean school subject marks. The relevance of those variables is based on the research literature and on participant observation within class-rooms (Beckers, 1982; Mooij, 1982; Speklé, 1982). In combination with the internal differentiation and evaluation variables, those educational system variables at the group level are meant to influence the development in the school subject motivation and school subject achievement of each pupil. From a methodological point of view these educational system variables function as treatment variables.

3. METHOD

3.1. Sample

A selection procedure has been used to find secondary educational schools in which teachers said they were internally differentiating in one or more school subjects, and schools in which teachers said they did not do so (see Mooij, 1981). Ten out of 11 selected schools agreed to participate. All 39 first-year classes were included in the research. In those classes different educational types for relatively low-achieving pupils were represented (junior vocational education (lbo), intermediate general education (mavo), and individual vocational education (ibo)). The classes have been followed during four school-years (including the final examination results in 1985). In this article attention will be restricted to the pupils' first school-year (1981-1982) and to three school subjects: Dutch language, mathematics/arithmetic, and manual dexterity. In Dutch language and mathematics/arithmetic the administrative classes were the same as the school subject specific pupil groups present during the lessons throughout the school-year ($n = 39$). In manual dexterity, however, the classes in one school were divided into sex-specific year-long pupil groups. In this school subject the number of pupil groups is 41.

3.2. Design

The research design is a nonmanipulative comparative design (Cronbach, 1983), in which effects of year-long group variables on the development of the pupils' school subject motivation and school subject achievement during the school-year are investigated. The dependent variables consist of the pupils' dependence on initial status in school subject orientation, school subject competence, and school subject achievement. The independent variables are the educational system variables at the group level mentioned in section 2. In order to determine the net effects of those treatment variables, they are corrected for background variables characterizing the pupil group at the start of the school-year. The relevant data were collected at three levels: the pupil level, lesson level, and group level.

3.3. Data collection

In August and September 1981 all 860 first-year pupils filled in a pupil questionnaire containing background items and school subject motivation items. They also completed tests on achievement in the school subjects Dutch language and mathematics/arithmetic. At the end of the school-year, the data collection with respect to the school subject motivation items and achievement tests was repeated. Moreover, the school administrations' information on each pupil's report marks was collected at the start of the second school-year.

During the first school-year the group data were gathered in three ways. First, information on the characteristics of the teaching situations of each group were obtained by observation of lessons with a structured observation system (see Mooij, 1985). A second way was to ask the school subject teacher to complete a teacher questionnaire with four items on his or her evaluation modes of the achievement results within each specific pupil group. Third, pupil data were aggregated to the group level (see below).

An overview of the variables, their measurement level, the time of measurement, and the abbreviations used henceforth is presented in Table 1.

3.4. Operationalization of variables

Table 1 starts with three background variables measured at the *pupil level* by a pupil questionnaire at the start of the first school-year: *father's occupation* has been asked by different questions (see Van Westerlaak, Kropman, & Collaris, 1975a, 1975b) and has been coded in a scale with six ordered categories:

higher occupation	(6);
secondary employee	(5);
small self-employed	(4);

Table 1. Pupil, lesson, and group level variables and their abbreviations

variable name	measurement level	measurement time first school-year	abbreviation
father's occupation	pupil	start	fathoc
sex	pupil	start	sex
held back	pupil	start	heldb
(initial) school subject orientation	pupil	start/end	(in)ssor
(initial) school subject competence	pupil	start/end	(in)ssc
(initial) school subject achievement	pupil	start/end	(in)ssach
organization of pupils	lesson	during	orgpupils
kind of task	lesson	during	lessontask
oral	lesson	during	oral
writing	lesson	during	writing
visual	lesson	during	visual
constructive	lesson	during	constructive
creative	lesson	during	creative
teacher's management	lesson	during	teachmanag
internal differentiation	lesson	during	intdif
number of pupils	lesson	during	numbpup
Kind of evaluation:			
comparative	group	during	evcomp
normative	group	during	evnorm
individual comparative	group	during	evinc
individual effort	group	during	eveff
school subject mark	pupil	start second year	ssmark

lower employee (3);
 skilled labourer (2);
 unskilled labourer (1);

sex has been indicated by putting a cross in a box behind the right sentence: I am a boy (coded 1); I am a girl (coded 0); *held back* has been operationalized in question form: How many times were you held back in elementary school? (coded: 0 times = 0, 1 or more times = 1).

The operationalization and procedures in creating motivation and achievement scores at the *pupil level* are:

school subject orientation: the questionnaire items indicating this concept are:

1. do you pay attention to the learning content during the lesson?
2. do you work hard for this school subject?
3. are you interested in this school subject?

Those items are coded as: (almost) never = 1, now and then = 2, most of the time = 3, always = 4. The alpha coefficients at the start and the end of the school-year are: in Dutch language .62 and .69, in mathematics/arithmetic .74 and .80; and in manual dexterity .81 and .85;

school subject competence: the next three items have been chosen:

1. do you sometimes have problems in this school subject?
2. do you have difficulty with this school subject?
3. is this school subject easy for you?

The coding of the items is the same as with respect to school subject orientation. The codes of

the items 1 and 2 have been reflected. The alpha coefficients are: in Dutch language .69 and .78, in mathematics/arithmetic .78 and .85, and in manual dexterity .72 and .78;

school subject achievement: a pupil's school subject achievement in Dutch language and mathematics/arithmetic has been measured by using standardized achievement tests used in a Dutch national cohort research (Smulders, 1979). Each test has been subjected to a Rasch scale analysis in order to construct a Rasch homogeneous scale (see Mooij, in preparation). In the analyses the sum of the items included in a Rasch scale is used (which correlate .99 with the Rasch scale score). Because of the lack of a standardized achievement test in manual dexterity, the achievement in this school subject has not been measured.

Lesson level data were obtained by using a structured observation system meant to measure lesson characteristics including the pupils' task behaviour. The development, interrater-reliability, and first analysis results with this system have been reported in Mooij (1985). The numbers of observed lessons in the respective school subjects are: Dutch language (96), mathematics/arithmetic (100), and manual dexterity (96). A summary of the lesson characteristics observed is given next.

The incorporation of *organizational lesson characteristics* has been inspired by the work of Evertson et al. (1980). Two dichotomous variables will be used here (coded: 0,1, respectively):

1. *organization of pupils* (whole-class or pupils treated as one group versus pupils working in subgroups or individually);
2. *kind of task* (the same for all pupils versus different for different pupils).

With respect to the ways to *transmit the learning content* five variants, which may overlap, have been distinguished. All of these were scored dichotomously:

1. *oral* (by human voice e.g. clarification);
2. *writing* (e.g. in workbook, drawing);
3. *visual* (e.g. movie, model in manual dexterity);
4. *constructive* (three-dimensional activity e.g. building a model of a house);
5. *creative* (own contribution with regard to form, colour, content; e.g. presentation on a self-chosen subject).

The operationalization of *teacher's management* has been based on the relevant literature (Rosenshine & Furst, 1973; Emmer & Evertson, 1980; and Kounin, 1970). The scale used consists of six items on teacher's clarity and discipline-preserving and discipline-correcting behaviour. Each item has five answer alternatives ranging from hardly = 1 to very (well) = 5. The alpha coefficient on 292 observed lessons in the school subjects Dutch language, mathematics/arithmetic, and manual dexterity is .94. The theoretically relevant *internal differentiation* throughout the lesson has been operationalized with the aid of some of the criteria of Nuy (1981). The items used refer to the teaching-learning process characteristics: learning time, learning content, learning actions, and learning media. Each item has five answer alternatives ranging from very (much) the same = 1 until hardly the same = 5. The alpha coefficient on 292 lessons observed in the school subjects Dutch language, mathematics/arithmetic, and manual dexterity is .95.

The *number of pupils* has been operationalized as the mean of the number of pupils present at three registration moments spread throughout the lesson.

At the *group level* the *kind of evaluation* of pupils' learning achievement has been distinguished into four different kinds of evaluation: group comparative, normative, individual comparative, and individual effort-based. The formulation of those alternatives in the teacher questionnaire was:

1. Do you give marks based on comparing the results on the pupils' tests?
2. Do you give marks based on a norm determined before giving a test?
3. Do you give a mark based on an individual comparison of a test's result with the pupil's earlier results?
4. Do you take a pupil's effort into account in giving marks for a test?

Each question had to be answered with yes or no.

Finally, the three *school subject marks* on a pupil's report during the school-year were

obtained from the school administrations. The mean of the school subject specific marks served as the pupil's school subject mark. Though these data were collected at the start of the second school-year, it has to be noted that the school subject marks were given to the pupils *during* the first school-year.

3.5. Analysis

Assuming random coefficients and fixed regressors, the data can be analyzed by unweighed least squares estimation using a two step procedure (see De Leeuw & Kreft, 1986; Van den Eeden & Saris, 1984).

In a first step the school subject specific orientation, competence, and achievement scores of the pupils within each group will be regressed on their respective initial orientation, competence, and achievement scores. This first step results in group specific b coefficients expressing the dependence on initial status in school subject orientation, school subject competence, and school subject achievement. In a second analysis step those group parameters are treated as *dependent variables*: between group differences among b coefficients may then be explained by effects of specific group level variables.

In the second analysis step all *independent variables* (including the covariates) have to reside at the group level. To realize this the relevant variables measured at the pupil level and lesson level (see Table 1) have to be aggregated to the group level. With respect to the variables fathoc, sex, heldb, inssor, inssc, inssach, and ssmark the mean of the pupils in the group is used to indicate the group variable. The standard deviation of inssach within a group can be seen as indicating the group's degree of heterogeneity in initial school subject achievement. Because of the relevance of this heterogeneity in Dutch educational policy, this standard deviation (sdinssach) will be included in the analysis.

The number of lessons observed in a group varies from one to five (see Mooij, 1985). The lesson variables are aggregated to the group level by using the mean of a lesson variable scored in the same group as the score of the group level variable. In aggregating, each pupil or lesson score has an equal weight in determining the group score (cf. Hüttner, 1985). To underline the nonequivalence of the aggregated group variables with their pupil or lesson level counterparts, the abbreviations of the aggregated group variables start with m or sd. The means of the variables morgpupils - mcreative vary between 0 and 1.00. With respect to these variables a variable with a mean higher than .90 or lower than .10 will be excluded from the school subject specific analysis.

School level variation (for example school achievement or school climate aspects) may contaminate the group level variation (cf. Finn, 1974, p. 78-83). It seems necessary to check this possibility by a correction for school. This is done by subtracting from each group's score on a variable the corresponding school mean on this variable (see also Hox & Willemse, 1985).

The analysis results from a two step procedure are less bothered by multicollinearity and easier to understand and interpret, but also somewhat less precise than results from the alternative single-equation method (De Leeuw & Kreft, 1986).

4. RESULTS

In the first analysis step the end-of-the-year orientation, competence, and achievement scores are regressed upon their respective initial scores, separately for each group. The correlations between the intercepts and the b coefficients for the motivation variables vary between -.91 and -.98; for the achievement variables the relationships are -.78 (Dutch language) and -.77 (mathematics/arithmetic). De Leeuw (1986) showed that this relationship is influenced by the zero of the scale on which the independent variable is measured.

Though the within-group residual variances might be interesting too, in this article only the b coefficients are included in the second analysis step. The means and standard deviations of those coefficients (bssor, bssc, and bssach) are, by school subject, given in Table 2. All means

are positive and the standard deviations suggest that the variation among *b* coefficients is worth further exploration.

Table 2. Means and sd's of the *b* coefficients, by school subject and kind of variables

<i>school subject</i>	<i>n</i> <i>groups</i>	<i>kind of variables</i>					
		<i>bssor</i>		<i>bssc</i>		<i>bssach</i>	
		<i>m</i>	<i>sd</i>	<i>m</i>	<i>sd</i>	<i>m</i>	<i>sd</i>
Dutch language	39	.46	.22	.34	.26	.36	.21
mathematics/arithmetic	39	.43	.24	.44	.25	.53	.23
manual dexterity	41	.40	.24	.39	.23	—	—

In the second step of the analysis attention is directed at the potential group level influences on *bssor*, *bssc*, and *bssach*. The groups' mean background variables (*mfathoc*, *msex*, *mhelddb*), mean initial variables (*minssor*, *minssc*, *minssach*), and heterogeneity in initial school subject achievement (*sdinssach*) will serve as covariates. The treatment variables are: the mean lesson variables, the evaluation modes, and the mean school subject marks. The second analysis step intends to select the most relevant treatment variables effecting the development in a pupil's school subject orientation, competence, and achievement during the school-year (while controlling for the group initial characteristics). The covariates are entered at once as independent variables, and then a first treatment variable is introduced. The variable selected is the one which aids most in forming a better prediction in terms of a higher R^2 , given the covariates. Then the best combination of the covariates and two treatment variables is sought, and so on. Inclusion of a next treatment variable may heighten or lower the *F*-value (the ratio of the mean square of the included independent variables to the error or residual mean square). In order to select the potentially most powerful treatment variables, the equation with the relatively highest or maximal *F*-value will be presented. The analyses were carried out with uncorrected and for school corrected independent variables (cf. section 3.5). An inspection of the analysis results showed that the correction for school increased the R^2 explained concerning *bssor* in Dutch language, and *bssor* and *bssc* in manual dexterity. It lowered the R^2 explained with respect to *bssc* in Dutch language, and *bssor*, *bssc*, and *bssach* in mathematics/arithmetic. Only the for school corrected results are presented here (see Mooij, in preparation, for both uncorrected and corrected results).

The results of this stepwise analysis in the three school subjects Dutch language, mathematics/arithmetic, and manual dexterity are given in Tables 3-5. The dependent variables are the *b* coefficients (*bssor*, *bssc*, *bssach*); in the Tables the second step *b* coefficients and their standard errors are presented. At the bottom of each Table the following information appears:

- $R^2(7)$ or $R^2(5)$: variance explained by the 7 or 5 covariates;
- n* pred. extra : number of predictors joining the covariates in order to reach the maximum *F*;
- R^2 : variance explained by covariates and selected treatment variables;
- F* : *F* with respect to R^2 ;
- $R^2(19)$ or $R^2(20)$: variance explained by covariates and all treatment variables.

The results in Table 3 suggest (by comparing the magnitude of the *b* coefficients to their standard errors) that a higher teacher's management and individual effort evaluation of learning achievement stimulate a pupil's dependence on initial status in orientation towards Dutch language. The result on the covariate *minssc* indicates that a group's higher initial competence in Dutch language lowers the dependence on initial orientation towards this school subject: pupils in groups high on initial competence show less consistency in their orientation scoring. A similar tendency seems present in the effect of the treatment variable mean school

Table 3. Second step multiple regression results in Dutch language (for school corrected results, 7 covariates, n groups = 39)

	bssor	se	bssc	se	bssach	se
intercept	.65		.37		.23	
mfathoc	-.20	.16	-.12	.18	.06	.16
msex	.44	.70	.43	.81	-.15	.64
mheldb	-.52	.31	.73	.37	.07	.29
minssor	-.24	.28	.24	.29	-.52	.24
minssc	-.73	.24	.41	.29	.71	.22
minssach	.12	.06	.12	.08	.08	.06
sdinssach	-.10	.12	-.02	.13	.06	.10
morgpupils						
moral						
mwriting						
mcreative						
mteachmanag	.25	.10				
mintdif			.28	.12		
mnumbpup						
evcomp			-.18	.12		
evnorm					-.20	.13
evinc					-.28	.09
eveff	.26	.11				
mssmark	-.20	.13				
R ² (7)	.25		.35		.35	
n pred. extra	3		2		2	
R ²	.49		.48		.50	
F	2.71		2.99		3.23	
R ² (19)	.64		.56		.61	

Note: se = standard error of b coefficient

subject marks. The three treatment variables account for 24% of the variance of bssor, after controlling for the covariates.

The dependence on initial competence in Dutch language is stimulated by higher internal differentiation in the lessons, but becomes less by comparative evaluation of learning achievement. The dependence on initial achievement in Dutch language is decreased by the individual comparison and the normative method in evaluating a pupil's learning achievement. A decreasing effect is also indicated by the covariate minssor, while minssc has a stimulating effect on this dependence. The covariates and the selected 2 or 3 treatment variables explain about 50% of the variance of bssor, bssc, and bssach; each treatment variable is responsible for about 7% of the variance.

In Table 4 the second step regression results in mathematics/arithmetic are given. With respect to bssor and bssc no clear treatment effects seem present. An interesting covariate effect is the influence of heterogeneity in initial achievement: a greater heterogeneity in the group stimulates the dependence on initial competence and achievement in mathematics/arithmetic during the school-year. In Dutch language these heterogeneity effects are not so high (see Table 3). In Table 4 the results on bssach make clear that organization of pupils more in subgroups or individually, a higher oral transmission of learning content, a lower degree of internal differentiation, and not evaluating in a normative way contribute to the dependence on initial achievement in mathematics/arithmetic. The four treatment variables account for 41% of the

Table 4. Second step multiple regression results in mathematics/arithmetic (for school corrected results, 7 covariates, n groups = 38)

	bssor	se	bssc	se	bssach	se
intercept	.13		— .06		.13	
mfathoc	— .02	.23	.23	.20	.21	.16
msex	— .58	.94	— .72	.86	— 1.02	.59
mhelddb	— .07	.47	.08	.43	.63	.36
minssor	— .15	.36	.44	.29	.37	.23
minssc	.10	.27	— .41	.24	— .30	.17
minssach	— .14	.07	— .03	.05	.06	.04
sdinssach	.12	.09	.19	.09	.15	.06
morgpupils	— .21	.29			.55	.22
mlessontask						
moral					.45	.25
mwriting	.75	.51				
mvisual	.53	.44				
mteachmanag						
mintdif					— .29	.12
mnumbpup						
evcomp	.40	.27				
evnorm					— .61	.17
evinc	1.52	.79				
eveff	— .95	.41				
mssmark						
R ² (7)	.11		.24		.26	
n pred. extra	6		0		4	
R ²	.32		.24		.67	
F	.85		1.32		4.71	
R ² (20)	.36		.43		.73	

Note: se = standard error of b coefficient

variance of the within group slopes in mathematics/arithmetic achievement.

The results in manual dexterity in Table 5 show that a higher degree of creativity, lower degrees in writing, constructive, and school subject marks, and not evaluating in a comparative way promote the dependence on initial orientation concerning manual dexterity. The five treatment variables are responsible for 38% of the within group slope variance (after controlling for five covariates). The dependence on initial competence in manual dexterity seems to be stimulated by higher internal differentiation and by a higher number of pupils in the lessons. The chosen five treatment variables add 30% of the within group slope variance in competence in manual dexterity, after eliminating the variance explained by the covariates. Within these covariates, a higher proportion of pupils being held back in elementary education and a lower initial competence in the group stimulate the dependence on the initial manual dexterity competence.

Table 5. Second step multiple regression results in manual dexterity (for school corrected results, 5 covariates, n groups = 40)

	bssor	se	bssc	se
intercept	.40		.39	
mfathoc	.14	.18	.25	.20
msex	.08	.22	-.27	.21
mheldb	-.28	.38	1.14	.39
minssor	.08	.26	.42	.25
minssc	.01	.25	-.66	.30
morgpupils				
mlessontask				
moral				
mwriting	-.74	.21		
mvisual			-.46	.26
mconstructive	-.68	.26		
mcreative	.96	.34		
mteachmanag				
mintdif			.35	.11
mnumbpup			.05	.02
evcomp	-.89	.30	.42	.28
evnorm			-.24	.19
evinc				
eveff				
mssmark	-.26	.21		
R ² (5)	.04		.08	
n pred. extra	5		5	
R ²	.42		.38	
F	2.08		1.79	
R ² (20/19)	.47		.45	

Note: se = standard error of b coefficient

5. DISCUSSION

In the Tables 3-5 it can be seen that internal differentiation in the lessons during a school-year positively influences the dependence on initial competence concerning Dutch language and manual dexterity, while it interferes with the dependence on the initial mathematics/arithmetic achievement. With respect to the individual evaluation, the effort evaluation promotes the dependence on the initial orientation towards Dutch language, but it interferes with the dependence on initial orientation towards mathematics/arithmetic. Individual comparative evaluation disturbs the dependence on the initial achievement in Dutch language. Those results hardly confirm the formulated hypotheses. Furthermore, the exploration of group effects on the pupil's motivation and achievement suggests that different group level variables seem relevant to the dependence on the initial orientation, competence, and achievement variables in Dutch language, mathematics/arithmetic, and manual dexterity. The research results suggest the existence of interactions between the individual development in motivation and achievement variables, and specific group or mean teaching situation variables throughout the school-year. Those interactions are school subject specific.

The importance of the study can be indicated by the following points. A first point is the explicit theoretical and methodological attention to the (potential) interactions between indi-

vidual and a number of environmental variables. In this respect an extension of the aptitude-treatment interaction paradigm has been presented. Second, it should be noted that the correction for school has an impact on the group level results: such a correction seems to be necessary (cf. Cronbach, 1976). Third, the potential relevance of 15 educational treatment variables concerning the development in a pupil's motivation and achievement in three different school subjects has been checked. In combination with other research results in SVO-project 0483 (see Mooij, in preparation) the results suggest *which* concrete processes might occur at and between different educational levels and *why* they might take place. For example, at the pupil level the pupil's school subject marks especially influence the development of his or her school subject competence, but also the school subject achievement, in the course of a school-year (see also Mooij, 1986). In this article the potential relevance of group level variables to the development in the same pupil level variables has been dealt with. The different kinds of results contribute to ecological based knowledge of multi-level educational processes and effects.

It should be kept in mind, however, that the number of groups in this study ($n = 39$) is not large, and that the findings should be cross-validated. The cross-validation is in particular relevant to a more valid construction of practice-based ATI-theorizing.

REFERENCES

- Beckers, T. (1982). *Participerende observatie in een ibo-school. Rapportage in het kader van het projekt voortijdige schoolverlaters* (SVO-project 0483). Nijmegen: Instituut voor Toegepaste Sociologie.
- Bloom, B. S. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational leadership*, May, 4-17.
- Boekaerts, M. (in press). Arousal, telic dominance and learning behaviour. In R. Gupta & P. Coxhead (Eds.), *Cultural diversity and learning efficiency: recent developments in assessment*. London: Macmillan.
- 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.
- Cattell, R. B., & Child, D. (1975). *Motivation and dynamic structure*. London: Holt, Rinehart, & Winston.
- Corte, E. de, Geerligs, C. T., Lagerweij, N. A. J., Peters, J. J., Vandenbergh, R. (1976). *Beknopte didaxologie*. Groningen: Wolters-Noordhoff.
- Cronbach, L. J. (1976). *Research on classrooms and schools: formulation of questions, design, and analysis*. Stanford: Stanford University, Stanford Evaluation Consortium.
- Cronbach, L. J. (1983). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington Publishers.
- Eeden, P. van den, & Hüttner, H. J. M. (1982). Trend report multi-level research. *Current Sociology*, 30, 3, 1-182.
- Eeden, P. van den, & Saris, W. E. (1984). Empirisch onderzoek van multilevel uitspraken. *Mens en Maatschappij*, 59, 165-178.
- 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.
- Finn, J. D. (1974). *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston.
- Glaser, R. (1976). Cognitive psychology and instructional design. In D. Klahr (Ed.), *Cognition and instruction* (pp. 303-315). New York: Wiley.
- Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size*. Beverley Hills: Sage.
- Harter, S. (1981). A model of mastery motivation in children: individual differences and developmental change. *Minnesota Symposium on Child Psychology* (pp. 215-255). Hillsdale: Lawrence Erlbaum.
- Harter, S. (1982a). The perceived competence scale for children. *Child Development*, 53, 87-97.

- Harter, S. (1982b). A developmental perspective on some parameters of self-regulation in children. In P. Karoly & F. H. Kanfer (Eds.), *Selfmanagement and behavior change: from theory to practice* (pp. 165-204). Oxford: Pergamon Press.
- Heckhausen, H. (1980). *Motivation und Handeln: Lehrbuch der Motivationspsychologie*. Berlin: Springer.
- Hox, J. J., & Willemse, J. M. (1985). De empirische samenhang van sociometrische scores: een analyse op twee niveau's. *Tijdschrift voor Onderwijsresearch*, 10, 82-85.
- Hüttner, H. J. M. (1985). *Multilevel onderzoek: de operationalisatie van omgevingskenmerken*. Nijmegen: Katholieke Universiteit, Sociologisch Instituut.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart and Winston.
- Leeuw, J. de (1986). *The use of two-step methods in multilevel analysis (concept)*. Leiden: University of Leiden, Department of Data Theory FSW.
- Leeuw, J. de, & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11, 57-85.
- Mooij, T. (1980). Multi-level onderzoek van samenwerking in het part-time onderwijs. *Mens en Maatschappij*, 55, 148-169.
- Mooij, T. (1981). *Voortgangsrapport in het kader van SVO-project 0483 over de periode augustus-december 1981*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Mooij, T. (1982). Onderwijsleersituatie en lesondergravend gedrag van Ito-leerlingen. In E. Diekerhof (Ed.), *Leren, wat moet je ermee?* (pp. 39-52). Muiderberg: Coutinho.
- Mooij, T. (1984). Multiniveau-aspecten in de verklaring van leerling (de) motivatie. In P. van den Eeden, & J. Hauer (Eds.), *Analyse van multiniveaudata: theorie en toepassing* (pp. 103-123). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- Mooij, T. (1985). Leskenmerken en taakgedrag van leerlingen. *Tijdschrift voor Onderwijsresearch*, 10, 249-262.
- Mooij, T. (1986). Evaluation of learning achievement and the interactive multilevel development of school subject motivation and achievement. In J. H. L. van den Bercken, E. E. J. de Bruijn, & Th. C. M. Bergen (Eds.), *Achievement and task motivation* (pp. 247-262). Lisse and Berwyn: Swets & Zeitlinger.
- Mooij, T. (in preparation). An interactional multi-level approach to pupil behaviour, orientation, competence, and achievement.
- Neisser, U. (1976). *Cognition and reality*. San Francisco: Freeman.
- Nuy, M. J. G. (1981). *Interne differentiatie*. 's-Hertogenbosch: Katholiek Pedagogisch Centrum.
- Nijhof, W. J. (1981). *Interne differentiatie als een innovatie*. 's-Gravenhage: Staatsuitgeverij, SVO-reeks 9.
- Oudenhoven, J. P. L. M. van (1983). *Onderwijsongelijkheid en evaluatieve feedback*. Apeldoorn: Van Walraven.
- Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation*. Göttingen: Verlag für Psychologie.
- Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 122-183). Chicago: Rand McNally.
- Smulders, R. H. M. (1979). *C. B. S.-onderzoek: 'Schoolloopbaan en herkomst van leerlingen bij het voortgezet onderwijs'*. Paper Onderwijsresearchdagen 1979. Voorburg: Centraal Bureau voor de Statistiek.
- Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology*, 76, 347-376.
- Speklé, C. (1982). *Voortijdig schoolverlaten en schoolbeleving in het l.h.o.* Nijmegen: Katholieke Universiteit, Subfaculteit Onderwijskunde.
- Westerlaak, J. M. van, Kropman, J. A., & Collaris, J. W. M. (1975a). *Beroepenklapper*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Westerlaak, J. M. van, Kropman, J. A., & Collaris, J. W. M. (1975b). *Beroepenklapper. Toelichting*. Nijmegen: Instituut voor Toegepaste Sociologie.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66, 297-333.

Manuscript ontvangen: 16 mei 1986

Definitieve versie ontvangen: 23 december 1986

Notities en Commentaren

DE SAMENSTELLING VAN DE SCHOOLBEVOLKING IN HET BASISONDERWIJS

Johan M. Wijnstra

Centraal Instituut voor Toetsontwikkeling, Arnhem

INLEIDING

Sinds de invoering van de Wet op het Basisonderwijs in 1985 geldt voor de basisscholen een nieuwe formatieregeling waarbij het aantal formatieplaatsen wordt bepaald aan de hand van het gemiddelde van het gewogen aantal leerlingen over het voorgaande schooljaar. De zg. schoolscore, de ratio van het gewogen en het nominale aantal leerlingen, weerspiegelt de samenstelling van de schoolbevolking naar de combinatie van opleidingsniveau, sociaal-economische status en etnische herkomst van de ouders. In onderwijskundig onderzoek kan deze schoolscore een bijzonder geschikte variabele zijn voor het controleren van de representativiteit van steekproeven van scholen, niet in de laatste plaats omdat er zonder al te veel problemen over volledige gegevens beschikt kan worden. Het Ministerie van Onderwijs en Wetenschappen verzamelt ieder schooljaar op drie teldata gegevens, terwijl de scholen de gegevens in hun administratie dienen te bewaren, althans wanneer ze in aanmerking willen komen voor extra formatie. In voorkomende gevallen kan de schoolscore ook gebruikt worden bij het trekken van gestratificeerde of systematische steekproeven uit de populatie van scholen.

In verband met de voorbereiding van de steekproeftrekking voor het project Periodieke Peiling van het Onderwijsniveau (PPON) in het basisonderwijs is de samenstelling van de schoolbevolking in het basisonderwijs per 16 september 1985 verkend. De gegevens uit dit (nagenoeg volledige) teldatumbestand zijn door het Ministerie van Onderwijs en Wetenschappen verstrekt via het Rijkscomputercentrum in Heerlen. Recentere bestanden, waarvan bij de feitelijke steekproeftrekking gebruik zal worden gemaakt, zijn op dit moment nog niet volledig genoeg.

In de analyses is gebruik gemaakt van de gegevens over de verdeling van de leerlinggewichten per school en het herkomstland van de ouders. Omdat deze gegevens ook van belang kunnen zijn voor andere onderzoekers, wordt in deze notitie de samenstelling van de schoolbevolking in het basisonderwijs in het kort beschreven. Daarbij wordt ook aandacht besteed aan de scholen die met ingang van het schooljaar 1986/87 in een onderwijsvoorrangsgebied vallen.

SCHOOLScores EN DE VERDELING VAN DE GEWICHTSCATEGORIEËN

Voor de bepaling van het gewogen aantal leerlingen worden vijf gewichten gehanteerd (zie voor de exacte omschrijvingen de betreffende circulaire):

- 1.25 voor Nederlandse arbeiderskinderen (in termen van opleidings- en/of beroepsniveau);
- 1.40 voor schipperskinderen in een internaat of pleeggezin;
- 1.70 voor kinderen uit de reizende en trekkende bevolking;
- 1.90 voor kinderen uit gezinnen waarvan tenminste een van de ouders van niet-Nederlandse herkomst is (met beperkingen wat betreft opleidings- en beroepsniveau);
- 1.00 voor alle andere kinderen.

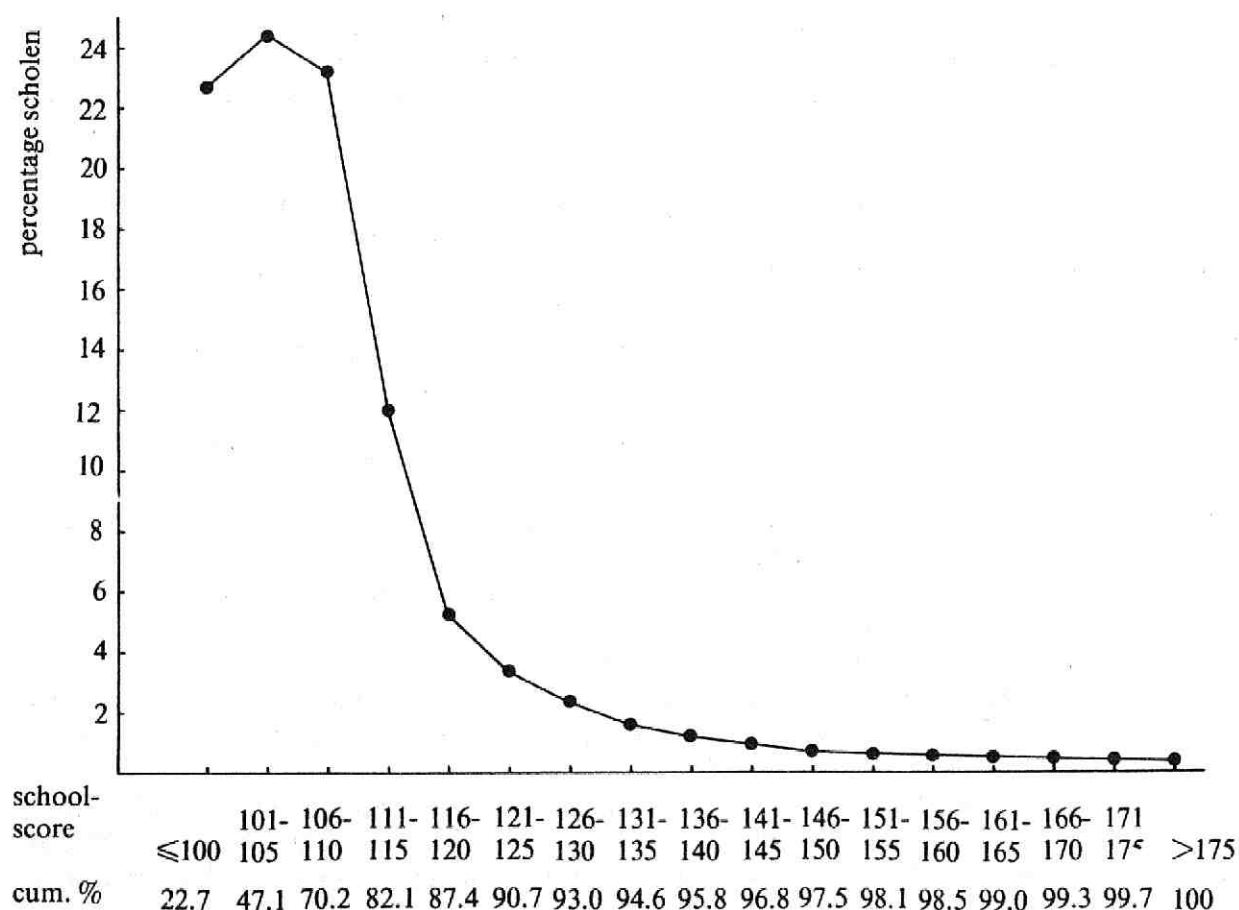
De schoolscore bestaat uit de ratio van het gewogen aantal leerlingen en het nominale aantal, na aftrek van een correctiefactor van het gewogen aantal leerlingen. In het schooljaar 1985/86 bedroeg de correctiefactor 5% van het nominale aantal leerlingen. Omdat voor het schooljaar

1986/87 de correctiefactor is verhoogd tot 9%, hebben we deze laatste gehanteerd, hoewel die op de gegevens feitelijk niet van toepassing was. Als door de correctieprocedure het gewogen aantal leerlingen lager uitkomt dan het nominale aantal, wordt bij de formatiebepaling uitgegaan van het nominale aantal leerlingen. Schoolscores <1.00 worden feitelijk dus op 1.00 gesteld.

Om te beginnen wordt in Figuur 1 de percentuele verdeling van de schoolscores ($\times 100$) grafisch weergegeven ($n = 8313$). Tevens worden de cumulatieve percentages vermeld. De scholen met een schoolscore ≤ 1.00 tellen nauwelijks kinderen van niet-Nederlandse herkomst, terwijl onder de Nederlandse kinderen slechts weinig arbeiderskinderen voorkomen. Bij een schoolscore van 1.01-1.15 neemt het percentage Nederlandse arbeiderskinderen sterk toe. Het aantal buitenlandse kinderen is ook in dit gebied nog betrekkelijk gering. Boven een schoolscore van 1.15 loopt het aantal buitenlandse kinderen snel op. In Tabel 1 wordt de gemiddelde samenstelling van de scholen naar de gewichtscategorieën van de leerlingen in ieder van de in Figuur 1 onderscheiden schoolscore-intervals in percentages vermeld.

Met ingang van het schooljaar 1986/87 zijn 70 onderwijsvoorrangsgebieden gevormd, overigens nog voordat de Onderwijsvoorrangswet volledig in het parlement is behandeld. Via de coördinator voor de evaluatie van het onderwijsvoorrangsbeleid hebben we de beschikking gekregen over de lijst van scholen met hun CISO-identificatienummers, zodat het mogelijk was om deze scholen afzonderlijk in het bestand te onderscheiden. Van de in totaal 908 basisscholen konden er 896 worden gelokaliseerd. In Tabel 1 wordt in de kolom 'waarvan OVB-scholen' per schoolscore-interval aangegeven hoeveel scholen in het schooljaar 1986/87 in een onderwijsvoorrangsgebied vallen.

Figuur 1. Percentuele en cumulatieve percentuele verdeling van de schoolscores per 16.09.85 volgens het formatiebesluit voor het schooljaar 1986/87



Tabel 1. Gemiddelde samenstelling van de basisscholen naar de gewichtscategorieën van de leerlingen per schoolscore-interval (16.09.85)

school-score	aantal scholen	waarvan OVB-scholen	gem. aantal leerl.	gemiddeld percentage per categorie				
				G1.00	G1.25	G1.40	G1.70	G1.90
≤100	1888	1	179.7	84.0	15.1	0.1	0.1	0.8
101-105	2028	1	178.1	56.8	41.0	0.1	0.1	2.0
106-110	1921	18	159.5	40.8	55.9	0.1	0.2	2.9
111-115	992	90	150.9	30.3	62.9	0.2	0.5	6.2
116-120	434	106	161.9	29.4	56.1	0.2	0.7	13.5
121-125	278	110	167.3	27.9	50.0	0.8	0.7	20.6
126-130	191	95	172.6	22.4	50.3	0.4	0.8	26.1
131-135	129	97	165.4	17.3	49.2	0.4	1.4	31.7
136-140	102	76	168.4	15.4	44.6	0.2	0.4	39.4
141-145	86	68	172.3	13.3	39.7	0.2	0.4	46.4
146-150	60	48	161.4	10.5	36.1	0.1	1.7	51.6
151-155	47	43	174.9	8.8	30.7	0.1	1.3	59.2
156-160	36	32	185.4	6.3	27.6	0.0	0.1	66.0
161-165	38	35	173.9	4.9	20.3	0.4	0.3	74.0
166-170	28	27	200.3	3.5	14.4	0.2	0.0	82.0
171-175	30	26	216.9	2.0	10.2	0.0	0.6	87.2
>175	25	23	203.8	0.8	3.5	0.0	0.1	95.6
totaal	8313	896	169.4	49.7	42.1	0.2	0.3	7.7

KINDEREN VAN NIET-NEDERLANDSE OORSPRONG

Het is bijzonder moeilijk een goed beeld te krijgen van het aantal leerlingen van niet-Nederlandse oorsprong in het basisonderwijs, doordat de teldatumbestanden twee soorten gegevens in elkaar niet uitsluitende categorieën bevatten. Allereerst is per school het aantal leerlingen met een gewicht van 1.90 bekend. Dit zijn de arbeiderskinderen van niet-Nederlandse oorsprong. Mits aan de voorwaarden met betrekking tot opleidings- en/of beroepsniveau wordt voldaan, vallen hieronder de leerlingen van wie tenminste een van de ouders:

- afkomstig is uit een van de zg. wervingslanden (Griekenland, Italië, Joegoslavië, Marokko, Portugal, Spanje, Tunesië, Turkije, Kaap-Verdië);
- afkomstig is uit Suriname of van de Nederlandse Antillen;
- afkomstig is van de Molukse eilanden;
- afkomstig is uit een ander niet-engelstalig land buiten Europa (m.u.v. Indonesië);
- de vluchtelingenstatus heeft.

Onder andere in verband met het onderwijs in eigen taal en cultuur worden daarnaast afzonderlijk de aantallen leerlingen uit de wervingslanden en van de Molukse eilanden vermeld. In dit geval worden echter geen voorwaarden gesteld aan het opleidings- of beroepsniveau van de ouders.

Het probleem van de elkaar gedeeltelijk overlappende gegevens wordt in Tabel 2 geïllustreerd. Het totale aantal leerlingen van niet-Nederlandse oorsprong en de aantallen per cel kunnen niet exact worden bepaald, omdat in het bestand alleen de som van het aantal leerlingen in cel A+B (onderverdeeld naar land van herkomst), respectievelijk cel C+D beschikbaar is. Aangezien het opleidings- en beroepsniveau van de ouders van kinderen uit de wervingslanden over het algemeen laag ligt, mag worden aangenomen dat cel B in de meeste gevallen weinig of geen leerlingen zal bevatten. Van dit gegeven is gebruik gemaakt om per school een schatting te maken van het aantal leerlingen in cel C. Cel D is door gebrek aan gegevens buiten beschouwing gebleven.

Tabel 2. Beschikbare gegevens over de aantallen leerlingen van niet-Nederlandse oorsprong

herkomst leerlingen	gewichtscategorie		
	1.90	1.00	
leerlingen van ouders uit wervingslanden + Molukse eilanden	A = ?	B = ?	A + B
overige leerlingen van niet-Nederlandse herkomst	C = ?	D = ?	?
	A + C	?	

Tabel 3. Onderverdeling per schoolscore-interval van het gemiddelde percentage leerlingen van niet-Nederlandse oorsprong (16.09.85)

school-score	gem. % totaal	Turkije	Marokko/ Tunesië	Griekenl./ Joegosl.	It./Sp. Port./K.V.	overige
≤100	0.9	0.1	0.1	0.0	0.1	0.6
101-105	2.0	0.2	0.3	0.1	0.1	1.3
106-110	3.0	0.4	0.5	0.1	0.2	1.7
111-115	6.2	1.1	1.2	0.2	0.4	3.3
116-120	13.5	3.1	2.6	0.3	1.0	6.5
121-125	20.6	5.1	4.9	0.4	1.6	8.5
126-130	26.1	7.5	6.6	0.7	1.7	9.6
131-135	31.7	9.4	8.5	0.7	1.8	11.3
136-140	39.4	13.3	9.7	1.1	2.5	12.9
141-145	46.4	13.2	13.1	1.2	2.9	15.9
146-150	51.6	19.4	13.9	0.8	2.0	15.4
151-155	59.3	18.1	15.5	1.4	3.7	20.5
156-160	66.0	23.4	17.4	1.2	3.0	20.9
161-165	74.4	24.8	21.1	2.3	3.2	23.0
166-170	82.1	28.8	23.0	1.7	4.4	24.2
171-175	87.5	32.0	19.6	2.6	6.6	26.8
>175	95.6	35.8	29.8	1.6	5.1	23.4
totaal	7.8	2.0	1.8	0.2	0.5	3.3

In Tabel 3 wordt per schoolscore-interval een overzicht gegeven van het gemiddelde procentuele aandeel van de leerlingen van niet-Nederlandse oorsprong, onderverdeeld in de volgende categorieën:

- Turkije;
- Marokko, Tunesië;
- Griekenland, Joegoslavië;
- Italië, Spanje, Portugal, Kaap-Verdië;
- overige.

Bij de kinderen in de eerste vier categorieën is geen onderscheid gemaakt naar gewichtscategorie. Dit geldt ook voor de Molukse kinderen, $\pm 0.3\%$ in de populatie, die in de categorie 'overige' zijn ondergebracht. Om ook het aandeel van de niet nader benoemde groep kinderen met een gewicht van 1.90 te schatten, is het aantal Molukse leerlingen vermeerderd met het verschil tussen het aantal leerlingen in cel A+C en cel A+B (vgl. Tabel 2). Als dit verschil negatief was, is het op 0 gesteld. Hierdoor kan het totaalpercentage in enkele gevallen ook iets hoger zijn dan het percentage kinderen met een gewicht van 1.90 in Tabel 1.

Op deze wijze gedefinieerd bedraagt het totale aantal leerlingen van niet-Nederlandse oor-

sprong 7.8%, zoals uit Tabel 3 blijkt. De percentages in de categorie 'overige' zijn echter schattingen, waarbij de leerlingen uit cel D van Tabel 2 niet zijn meegeteld. Hoe groot deze aantallen zijn, is onbekend. De leerlingen uit de categorie 'overige' vormen de grootste groep kinderen van niet-Nederlandse oorsprong. Dit betreft vermoedelijk in de meeste gevallen kinderen van Surinaamse herkomst. Tot en met schoolscore 120 vormt deze categorie ongeveer de helft van het totale aantal leerlingen van niet-Nederlandse oorsprong. In de groep scholen met de hoogste schoolscores is hun aandeel ongeveer een kwart. In deze scholen zijn de Marokkaanse en met name de Turkse kinderen sterker vertegenwoordigd dan het totaalpercentage aangeeft. De kinderen uit de noordelijke mediterrane landen en Kaap-Verdië vormen een betrekkelijk kleine groep.

Ontvangen 18-12-1986

In memoriam Co van Calcar

In hetzelfde jaar dat Co van Calcar aan een nieuwe periode van zijn loopbaan was begonnen is hij door een verkeersongeluk om het leven gekomen. Het is hem niet gegund geweest alle ideeën en plannen die hij voor deze periode bestemd had, ten uitvoer te brengen. Met zijn benoeming tot buitengewoon hoogleraar met als leeropdracht 'onderwijs en samenleving' zou het zoeklicht weer op theorievorming rond onderwijsachterstanden worden gericht. Voor zijn vele internationale contacten zou meer tijd beschikbaar zijn waarvan het onderzoek zou kunnen profiteren. In de SCO had Co de leiding over een aanvankelijk kleine, later groter wordende groep onderzoekers. Daar een instituut als de SCO afhankelijk is van de opdrachten van derden, kostte het extra energie in de gevarieerde orderportefeuille zowel de inhoudelijke betrokkenheid bij het onderwijs aan groepen in achterstandssituaties als het doen van onderzoek in een nauwe relatie met het veld, vast te houden en verder te ontwikkelen. Die energie ontbrak hem allerminst. Met veel doorzettingsvermogen werd een onderzoek op stapel gezet naar 'cultureel kapitaal' en de relatie daarvan met maatschappelijke loopbaan, met behulp van nieuwe gegevens van de Enschedese generatie leerlingen uit de zestiger jaren waarmee Co zijn eigen onderzoekersloopbaan was begonnen.

Tevens werden projecten op terreinen als innovatie basisschool, milieueducatie, creativiteitsontwikkeling, arbeidsoriëntatie en kort-mbo onder zijn inspirerende leiding uitgevoerd.

Co gebruikte met name de spreekbeurten en conferenties waarvoor hij vaak werd uitgenodigd om zijn gedachten opnieuw te ordenen en nieuwe ideeën te formuleren. Mede door de recent geopende mogelijkheid tot een follow-up op de Enschedese onderzoekingen zag 1987 er veelbelovend uit.

Co was mild de laatste jaren, minder fanatiek dan in de tijd van het IPA. Maar de gave anderen te inspireren, wakker te houden, bleef. Daarbij had hij inzicht en overzicht: hij zag de dingen tegelijkertijd in hun kleinste detail en in hun vergaandste consequenties.

Het is elders gezegd en geschreven: velen zullen hun best doen in zijn geest verder te werken. Maar zonder zijn inzet en solidariteit zal dat niet makkelijk vallen.

Marja van Erp
Jan Karel Koppen
Peter van den Dool

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie

Kees Beintema (RUL), Peter van den Dool (SCO), Dirk de Jong (LICOR), Ewoud Roede (SCO) en Peter van der Sijde (UT, TO, secretaris).

Redactie-adres

Peter van der Sijde, Faculteit der Toegepaste Onderwijskunde, Vakgroep Instructietechnologie, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893560.

Jrg. 11, nr. 2

April 1987

Inhoudsopgave

Van de bestuurstafel	106
Convocatie ledenvergadering VOR	106
Beknopt verslag Verenigingsjaar 1986	110
Themaplanning voor de Onderwijsresearch dagen	112
Mededelingen organisatie ORD 1987	113
Ook bij programmering van onderwijsonderzoek zijn korte communicatielijnen effectiever D. de Jong	115
Congressen, Symposia, Cursussen, Trainingen	117

Van de Bestuurstafel

Convocatie ledenvergadering VOR

- Tijdens de ORD 1986 belegt de Vereniging voor Onderwijsresearch (VOR) zijn jaarlijkse algemene ledenvergadering. Toegang: uitsluitend voor leden van de VOR.
- De vergadering vindt plaats aan het eind van de middag van de eerste ORD-dag 26 mei a.s., Tijd en plaats zullen tijdens de ORD nader bekend worden gemaakt.

Agenda

1. Opening, mededelingen, vaststelling agenda, ingekomen stukken.
2. Verslag ledenvergadering d.d. 13 november 1986 (zie VOR-Bulletin 1987 nr. 1)
3. Jaarrekening 1986 (zie stukken in dit VOR-Bulletin).
De kascontrolecommissie zal per vergadering verslag doen van de controle over 1986).
4. Beknopt verslag verenigingsjaar 1986 (zie dit VOR-Bulletin)
5. Samenstelling kascontrolecommissie 1987.
6. Frequentie en tijdstip algemene ledenvergadering (zie opmerking hieronder).
7. Wat verder ter tafel komt.
8. Sluiting.

Ad 6. De belangstelling van VOR-leden voor het bijwonen van de twee ledenvergaderingen per jaar blijkt niet voldoende voor een grote opkomst. Het bestuur wil naar aanleiding hiervan overleggen over de tot nu toe gevolgde opzet van de ledenvergadering.

Leden met suggesties hierover worden uitgenodigd om hun ideeën kenbaar te maken tijdens de vergadering of ervoor in contacten met bestuursleden.

Jaarrekening 1986

Het verslag bestaat zoals gewoonlijk uit drie delen:

- I Rekening en Verantwoording (R&V) over 1986;
- II Toelichting op de posten opgenomen in de R&V;
- III Balansoverzicht per 31-12-1986.

Hieronder volgt een overzicht van alle drie. De toelichting op de R&V is ingekort voor de opname in het VOR-bulletin. Geïnteresseerden kunnen bij de penningmeester indien gewenst een exemplaar van de uitvoerigere toelichting krijgen.

I. Rekening en Verantwoording over 1986

Samengevat zien de inkomsten en uitgaven over 1986 er uit zoals in overzicht I is aangegeven.

Overzicht 1: Inkomsten en uitgaven in 1986:

Inkomsten

100	Contributie				
	1985 :	nog ontvangen			
		38 gewone leden	38 x f 95,—	f 3.610,—	
		— student-leden	x f 65,—	f —	
		te weinig betaald		f —	
		totaal			f 3.610,—
	1986 :	375 gewone leden	375 x f 105,—	f 39.375,—	
		3 student-leden	3 x f 75,—	f 225,—	
		te weinig betaald		f 185,25	
		totaal			f 39.414,75
700	Inkomsten TOR				f 43.024,75
800	SVO-subsidie TOR				f —
850	SOR-bijdrage VOR-prijs				f 5.000,—
900	Diversen (rente)				f —
					f 684,56
				totaal	f 48.709,31

Uitgaven

400	Activiteiten				
	410 ledenwerving		f —		
	430 werkgroepen		f —		
	440 colloquia		f —		
500	Tijdschriften				f —
	510 Tijdschrift voor Onderwijsresearch				
	511 Abonnementen				
	11 leden ('85)	11 x f 45,—	f 495,—		
	450 leden	450 x f 60,—	f 27.000,—		
	9 student-leden	9 x f 40,—	f 360,—		
	afronding BTW berekening		f 1,01		
	512 Redactiekosten		f —		
	515 150 exempl. TOR Engelstalig		f 393,75		
	520 VOR-bulletin				f 28.249,76
700	VOR-prijs				f 10.134,74
600	Bestuur				f —
800	Ledenadministratie				f 918,76
1000	Diversen				f 279,63
1100	Aflossing lening				f 5.438,73
					f 6.500,—
					f 51.521,53

II. Toelichting op de R&V

1. Inkomsten

100 Contributie

In vergelijking met het aantal leden in 1986 – zie aantal leden waarvoor aan Swets & Zeitlinger abonnementsgelden is afgedragen – is er een relatief groot aantal wanbetalers. Afgezien van problemen met de betalingsdiscipline wordt dit “tekort” vooral veroorzaakt doordat zomer 1986 de acceptgirokaarten voor de inning van de contributie op waren. Vanwege de conversie bij de PCGD op nieuwe formulieren, een minimale afname van 1500 formulieren en onduidelijkheid over de afwerking (bedrukking) van de formulieren waren er pas eind januari 1987 acceptgiro's om te verzenden aan nieuwe leden en wanbetalers — wat inmiddels is gebeurd.

700 Inkomsten TOR

De VOR ontvangt van Swets & Zeitlinger 10% van de inkomsten van de niet VOR-abonnementen op het TOR over het voorafgaande jaar. Deze inkomsten zijn nog niet ontvangen.

2. Uitgaven

700 VOR-prijs

In het kader van de VOR-prijs zijn in 1986, net als in 1985, geen uitgaven gedaan. De VOR-prijs houdt in dat de onderscheiden onderzoeker tot een maximum van f 1000,— kosten in het kader van zijn verdere professionalisering bij de VOR kan declareren.

1000 Diversen

In 1986 is in totaal f 5.438,73 uitgegeven aan ‘diversen’. Het grootste deel betrof het overmaken van f 5.000,— naar de Stichting Werkgroep Onderwijsresearch (SOR). Dit bedrag was op 27/3-1985 ten onrechte overgemaakt naar de VOR als opbrengst van de ORD'84: de opbrengsten van de ORD zijn bestemd voor de SOR. Voor het overige bestaan de diverse onkosten uit reproductiekosten VOR-prijs, cadeau voor organisatiecommissie ORD, kosten Kamer van Koophandel.

1100 Aflossing lening

In 1986 is het nog resterende bedrag van de lening van de SOR aan de VOR overgemaakt. Er resteerde nog f 7.000,— van de oorspronkelijke lening van f 10.000,—. Uit de SOR-gegevens bleek er ook een lening van de VOR aan de SOR van f 500,— te zijn. In totaal is daarom f 6.500,— aan de SOR overgemaakt.

III. Balans-overzicht per 31-12-1986

activa

– saldo op rekeningen	
. gewone rekening	f 12.445,85
. sterrekening	f 7.243,29
. plusrekening	f 3,38
. renterekening	f 1.452,85
– tegoed van SVO	f 5.000,—
– royalties TOR	f 1.872,40
– SOR-bijdrage VOR-prijs	f 1.000,—
– tegoed aan contributies	f 4.725,—
	f 33.742,77

passive

– op korte termijn te betalen	
. TOR-abonnement	f 181,71
. Redactiekosten	f 4.000,—
. VOR-bulletin	f 1.600,—
. Bestuurskosten	f 115,—
. kosten ledenadm.	f 750,—
– VOR-prijs	f 1.000,—
– eigen vermogen	f 26.096,06
	f 33.742,77

Toelichting op bezit, tegoed en schuld per 1-1-1987

- Het tegoed van SVO betreft de bijdrage over 1986 aan de VOR betreffende TOR.
- De royalties van het TOR betreffen de 10% die de VOR van Swets & Zeitlinger ontvangt van niet VOR-abonnementen op het TOR.
- De SOR-bijdrage VOR-prijs: De VOR-prijs wordt betaald door de SOR, maar wordt pas aan de VOR overgemaakt indien er uitgaven hiervoor zijn gedaan.
- Tegoed aan contributies: omdat een aantal nieuwe leden en alle wanbetalers over 1986 nog geen acceptgiro voor 1986 hebben ontvangen, wordt verwacht dat in de loop van 1987 na het versturen van acceptgiro's (half februari) alsnog 40 gewone en 7 student-leden hun contributie over 1986 zullen voldoen.
- TOR-abonnement: betreffende 1986 moeten aan Swets & Zeitlinger nog 11 abonnementen worden betaald.
- Redactiekosten TOR: Over 1986 moeten aan het RION de redactiekosten ten behoeve van het TOR nog worden vergoed.
- VOR-bulletin: Aan Grafoplan b.v. moet nog f 1.566,18 worden betaald voor VOR-bulletin 86,6. Voorts wordt rekening gehouden met porto en reproductiekosten gemaakt door de TH Twente ten behoeve van het VOR-bulletin.
- Bestuurskosten: De kosten van twee vergaderingen (november en december) van het bestuur moeten nog worden betaald.
- Kosten ledenadministratie: Behalve de zaalhuur voor de ledenvergadering (f 165,—) wordt nog een aantal uitgaven voor 1986 verwacht: reproductie en portokosten.
- VOR-prijs: Er wordt rekening gehouden met een uitgave van f 1.000,— ten behoeve van de VOR-prijs.
- Eigen vermogen: De reserve van de vereniging.

Opmerking: Het grootste deel van de tegoeden en schulden is bekend, waardoor het balans-overzicht een goed beeld geeft van de financiële positie van de vereniging. Wat betreft de activa is alleen de precieze hoogte van het tegoed aan contributies onzeker. Wat betreft de passiva is er nog onzekerheid over f 618,82 (f 33,82 wat betreft VOR-bulletin en f 585,— wat betreft de ledenadministratie).

BEKNOPT VERSLAG VERENIGINGSJAAR 1986

1. Bestuurssamenstelling

Het bestuur was in 1986 als volgt samengesteld:

- a. voorzitter: Dr. W. Hoeben (tot 13-11-1986)
Prof. Dr. J. Claessen (vanaf 13-11-1986)
- b. vice-voorzitter: Mevr. Dr. B. van Hout Wolters
- c. secretaris: Dr. J. Slavenburg (tot 13-11-1986)
Mevr. drs. M.P.C. v.d. Werf (vanaf 13-11-1986)
- d. penningmeester: drs. E. Roede
- e. leden: Prof. Dr. P. Span
drs. D. de Jong (tot 13-11-1986)
Dr. W. Gerritsen van der Hoop
Prof. Dr. J. Scheerens (vanaf 13-11-1986)
drs. K. Beintema (vanaf 13-11-1986)

2. Internationale contacten

- Een groot aantal VOR-leden bezocht in 1986 de AERA-bijeenkomst te San Francisco. Voor de AERA 1987 zijn op instigatie van de VOR door de AERA-divisies Nederlandse reviewers ingeschakeld bij de beoordeling van papers.
- Met Pergamon Press is overeengekomen dat VOR-leden voor £ 10.— een abonnement kunnen nemen op het International Journal of Educational Research (IJER). De SVO zal de eerste jaren borg staan voor eventuele exploitatietekorten. Inmiddels zijn de VOR-leden van deze overeenkomst op de hoogte gesteld.

3. Onderwijs Research Dagen

De Onderwijs Research Dagen vonden plaats in Utrecht. Prof. Dr. N.A.J. Lagerweij was voorzitter van de organisatiecommissie. Het aantal deelnemers bedroeg 750. Rond eind 1986 waren reeds 3 van de geplande 6 ORD-bundels bij Swets en Zeitlinger verschenen.

De komende jaren zullen de Onderwijs Research Dagen georganiseerd worden in:

- 1987 Groningen
- 1988 Leuven
- 1989 Leiden

4. Tijdschrift voor Onderwijs Research

Ook in 1986 was Prof. Dr. B. Creemers redactie-secretaris van het tijdschrift. De bredere redactieformule is in de in 1986 verschenen nummers duidelijk zichtbaar geworden.

5. VOR-bulletin

Redactiesecretaris van het bulletin was drs. P. v.d. Sijde. Met Swets en Zeitlinger is overeengekomen dat het bulletin met ingang van 1-1-'87 wordt opgenomen in het Tijdschrift voor Onderwijs Research.

6. VOR-aanmoedigingsprijs

De jury onder leiding van Prof. Dr. W. van der Linden kende de prijs van f 1.000,— toe aan Mevr. drs. M. J. Crombach.

7. Professionalisering

Het overleg met de SVO over door de VOR te stimuleren professionaliseringsactiviteiten is definitief beëindigd. De instituten en vakgroepen zullen nu zelf zorg moeten dragen voor de professionalisering van hun medewerkers.

8. Financiën

Zie Kasverslag 1986 van de penningmeester VOR.

Greetje v.d. Werf,
Secretaris.

Programmering van onderzoek

Door de redactie van het VOR-bulletin zal op woensdagochtend 27 mei van 9.30 uur tot 12.30 uur tijdens de ORD in Groningen een bijeenkomst worden georganiseerd over de Programmering van Onderwijsonderzoek. Na de bijdragen in het VOR-bulletin zal in deze bijeenkomst vooral aandacht worden besteed aan mogelijkheden om de programmering te verbeteren. Nadere informatie over opzet en plaats zullen in het ORD-programma bekend worden gemaakt.

Themaplanning voor de Onderwijsresearch dagen

De presentatie van onderzoek op de ORD vindt al jarenlang plaats rondom verschillende thema's. In het verleden was dat een aangelegenheid die volledig aan de lokale organisatiecommissie werd overgelaten. Afhankelijk van de 'couleur locale' werd een aantal zwaartepunten gekozen voor de presentatie van het onderzoek.

Van het VOR-bestuur onder wiens auspiciën de jaarlijkse ORD wordt georganiseerd, was dit een onbevredigende situatie. Naar keuze van onderwerpen, naar aantal thema's en naar gehanteerde kwaliteitsmaatstaven, traden grote fluctuaties op. Het VOR-bestuur heeft daarom eind 1985 besloten te komen tot een zekere mate van themaplanning en tot het benoemen van een VOR-commissie die hierop zou toezien.

De Commissie Themaplanning ORD bestaat uit vijf leden: een VOR-bestuurslid en een viertal deskundigen op een bepaald themagebied. De commissie is als volgt samengesteld:

- prof. dr. P. Span (namens het VOR-bestuur)
- prof. dr. H. P. M. Creemers
- prof. dr. L. F. W. de Klerk
- prof. dr. J. M. G. Leune
- prof. dr. W. J. van der Linden

De Commissie Themaplanning ORD heeft een lange termijnplanning opgesteld voor de inhoud van de thema's die tijdens de ORD aan de orde dienen te komen. Het is de bedoeling dat deze vaste thema's jaarlijks, desnoods per twee jaar, aan bod komen waarbij al dan niet deelgebieden kunnen worden onderscheiden. Aan deze vaste thema's kunnen een of twee lokale thema's door de betreffende organisatiecommissie ORD worden toegevoegd. In de lokale thema's en in de invulling van de vaste thema's heeft de organisatiecommissie ORD het recht zich te profileren. Het VOR-bestuur stelt, zonodig na advies van de Commissie Themaplanning ORD, de lokale invulling van de thema's vast.

Voor de komende vijf jaren zijn de volgende vaste thema's vastgesteld:

- a. onderwijsleerprocessen
- b. curriculum (incl. schoolorganisatie en verzorgingsstrukturaspekten)
- c. methodologie/technologie (incl. toetsen, media e.d.)
- d. onderwijs en samenleving (incl. economie, sociologie, beleid, recht en opleidingen).

De verslagen van de ORD die door Swets & Zeitlinger B.V. op de markt worden gebracht, zullen bestaan uit een viertal bundels met daarin opgenomen de bijdragen aan de vier vaste thema's en uit maximaal twee bundels met daarin opgenomen de bijdragen aan de lokale thema's. De Commissie Themaplanning ORD heeft een redaktiestatuut opgesteld ten behoeve van de ORD-bundels waarin ook de regels m.b.t. de kwaliteit van de bijdragen vermeld staat. De betreffende thema-bundelcoördinatoren moeten dit statuut en de omschreven kwaliteitseisen in acht nemen. Voordat de bundels naar de uitgever worden gezonden zal de Commissie marginaal controle uitoefenen m.b.t. de vraag of de bijdragen aan de gestelde kwaliteitsregels voldoen. De Commissie Themaplanning ORD, die fungeert als een overkoepelende redactiecommissie voor de verslagen van de ORD, zal bevorderen dat de beste bijdragen aan de bundels in het Engels worden gepubliceerd.

Het is de verwachting van het VOR-bestuur dat door het instellen van de Commissie Themaplanning de continuïteit inzake de keuze van de thema's wordt bevorderd en dat de kwaliteit van de bijdragen wordt gehandhaafd.

prof. dr. J. Claessen
Voorzitter VOR-bestuur

Mededelingen organisatie ORD 1987

Onderwijsevaluatie

In de oproep voor papers is aangegeven dat op de Onderwijsresearchdagen 1987 het accent zal liggen op onderwijsevaluatie. Nu is dat betrekkelijk: ook andere papers, mits ze aan de kwaliteitscriteria voldoen, worden door de beoordelaars geaccepteerd en sommige thema's lenen zich zeker meer dan andere voor evaluatieonderzoek. De groeiende belangstelling voor effectief leerkrachtengedrag, effectiviteit van de school en het onderwijs in het algemeen, brengt echter een verhoogde belangstelling voor onderwijsevaluatie met zich mee. Daarbij wordt niet alleen gedacht aan evaluatie op het niveau van de klas en de school, maar ook evaluatie van programma's als het onderwijsvoorrrangsbeleid, vernieuwingen in het secundair onderwijs etc.

Openingsprogramma

Mag de accentuering van onderwijsevaluatie in het totaal van het programma enigszins betrekkelijk zijn, het openingsprogramma zal duidelijk in dit teken geplaatst worden. Gasten, zoals de rector magnificus van de Rijksuniversiteit Groningen en de Minister van Onderwijs en Wetenschappen, drs. Deetman, zullen in hun toespraken op onderwijsevaluatie ingaan. Ook de gastspreker, dr. Richard Shavelson, zal zijn bijdrage richten op onderwijsevaluatie, mede op basis van de resultaten van onderzoek dat op dit moment wordt uitgevoerd. Shavelson sloot onlangs een evaluatie-onderzoek af voor de U.S. Corporation for Public Broadcasting en is nu in opdracht van de National Science Foundation bezig een systeem te ontwerpen waarmee op nationaal niveau het wiskunde- en science onderwijs gescreend kan worden.

Shavelson ontving zijn opleiding in de experimentele psychologie en zijn Ph.D.-studie in Educational Psychology in Stanford. Na de Ph.D.-studie werkte hij in Stanford als Assistant Professor in Educational Psychology en daarna op de University of California, Los Angeles. Daarnaast is hij Research Scientist van de RAND-corporation. Hij is Consulting Editor van Educational Psychology en van diverse andere tijdschriften. In 1986 was hij programmavoorzitter van de AERA-conventie en in 1987 zal hij de AERA-voorzitter zijn. Shavelson heeft een groot aantal publicaties op zijn naam staan op het terrein van cognitieve psychologie, zowel wat leerprocessen van leerlingen als het planningsgedrag en de beslissingsstrategieën van leerkrachten betreft.

Wetenschappelijk programma

De organisatiecommissie heeft meer dan 200 papervoorstellen ontvangen. Het aanbod voor de diverse thema's is erg groot: voor praktisch ieder thema werden 20 tot 30 voorstellen ingediend. Er kunnen 11 presentaties per thema plaatsvinden; alle coördinatoren moeten dus streng selecteren zodat het samenstellen van een kwalitatief sterk programma mogelijk lijkt.

Daarnaast is een groot aantal voorstellen voor symposia op uiteenlopende terreinen ontvangen, wellicht door de specifieke keuzes die ten aanzien van de thema-invulling in Groningen zijn gedaan.

Omdat de beoordelingen nog niet zijn afgerond is nog niet definitief bekend welke symposia doorgang zullen vinden. Waarschijnlijk zullen er echter op de ORD symposia gehouden worden over:

- de ontwikkeling van leesvaardigheid in het basisonderwijs in longitudinaal perspectief,
- oorzaken en gevolgen van veranderingen in onderwijsdeelname,
- planmatig werken in het speciaal onderwijs,
- school- en klasklimaat,
- issues in bedrijfsopleidingen,
- meisjes, exacte vakken en techniek,
- motieven, motivatie en gemotiveerdheid van leerlingen,
- algemene redeneerfactoren en metacognitie bij rekenvaardigheden.

Congressecretariaat

Voor alle informatie over de ORD 1987 en inschrijvingen kunt u zich wenden tot het congressecretariaat ORD '87:

G. J. Reezigt/A. Y. Weersing

RION

Postbus 1286

9701 BG Groningen

tel. 050-636635/636660

Ook bij programmering van onderwijsonderzoek zijn korte communicatielijnen effectiever

D. de Jong
LICOR, RU Leiden

Tien jaar geleden werden in instituten voor onderwijsonderzoek dikke programmadocumenten geproduceerd, bol van de maatschappelijke uitgangspunten en wetenschapstheoretische overwegingen. Anno 1986/87 worden bescheidener stukken gemaakt, die flexibeler zijn: marktgevoelig als we zijn geworden, is er meer ruimte voor wisselende vragen van (potentiële) opdrachtgevers. De stijl is veranderd, de programmeringsdruk lijkt niet afgenomen. Mijn vermoeden is dat inperking van budgetten en de algemene hausse tot herschikking in en rond het hoger onderwijs (TVC, fuseringen in HBO, enz. enz.) die druk veroorzaakt. Pragmatische, productiegerichte vormgeving eerder dan idealistische onderbouwing van de documenten doet het nu beter.

Een verandering bij deze en dergelijke planningsactiviteiten is ook dat een kleiner aantal individuen de lijnen uitstippelen en daarover beslissingen nemen.

Van doorslaggevend belang voor de kwaliteit en relevantie van planningsdocumenten is uiteraard niet het aantal betrokkenen, maar de rationaliteit van het proces van programmering. Of op dit punt winst is geboekt, is moeilijk te schatten. Verdere rationalisering van het proces van probleemverkennen, isoleren van onderzoekthema's, de prioritering en ordening daarvan op een tijdsbasis blijft nodig. Van den Dool (VOR Bull. 11/'87, nr. 1, par. 3) doet een aantal suggesties. Roede gaat uitgebreid in op één daarvan i.c. de rol van de onderzoeksthemagroep in de programmering van het totale onderwijsonderzoek (VOR Bull. 11, nr. 1). De recente zeer ingrijpende inperking van de financiële ruimte voor de OTG's zal de uitvoering van Roede's voorstel tot integratie van programma's afkomstig van verschillende geldstromen en daarmee verschillende wetenschappelijke invalshoeken tenminste op de lange baan kunnen schuiven.

Des te meer reden om op het nivo van de onderzoeksinstituten te bevorderen dat vraagstellingen vanuit de meer theorie-georiënteerde programma's (1e, 2e geldstroomonderzoek) verweven worden met vragen uit opdrachtenonderzoek (3e). Van den Dool beargumenteert, ook om deze reden, dat onderwijsonderzoekers in het programmeringsproces een duidelijker rol spelen dan voorheen. Dan veronderstellen we dat onderzoekers uit de vak- en werkgroepen en die in instituten in regelmatige samenspraak zijn bij de uitvoering en planning van onderwijsonderzoek. Wanneer in de werkrelatie met opdrachtgevers wordt gesproken over toekomstig, gewenst onderzoek – en dat is zeer regelmatig op een informele basis het geval – dan kunnen dus mede vanuit meer theoretische overwegingen geformuleerde vraagstellingen doorklinken.

Uiteraard zal men de rollen van opdrachtgever en onderzoeker keurig dienen te scheiden. Maar intensivering van deze relatie, mede met het oog op probleemverkenning en programmering, zou wel eens tot meer onderzoekbare en door bestaande wetenschappelijke inzichten gedragen aanbestedingen kunnen leiden. Men hoeft niet bevreesd te zijn dat beleidsmakers of veldvertegenwoordigers zich snel laten afbrengen van hun eigen vragen voor onderzoek. De herformulering van deze vragen in de communicatie met onderzoekers is winst voor beiden. Omdat onderzoekers veelal over ruime "veldkennis" beschikken, kunnen ze bovendien veel bijdragen aan de inhoudelijke afgrenzing van door veldvertegenwoordigers (Koepels) en beleidsambtenaren te formuleren vragen.

¹ Mede gebaseerd op een leunstoelgesprek per telefoon met Dr. Wijnand Hoebe. Verantwoordelijk voor de inhoud van dit stuk is hij niet.

Krijgen onderzoekers langs deze weg niet teveel invloed? Zo zij al de fout maken op de stoel van de opdrachtgever te gaan zitten, is er nog de onafhankelijkheid van apart aangewezen probleemverkenner en programmamakers om dit in belangrijke mate te neutraliseren. SVO als statutair verantwoordelijke voor de programmering moet dit communicatieproces dan ook bevorderen en voorkomen dat communicatielijnen te lang worden.

Ook uit een oogpunt van continuïteit is het voor onderzoeksinstituten noodzaak om nauw met potentiële opdrachtgevers te communiceren en de informatie over te verwachten onderwerpen en opgebouwde deskundigheid in de instituutsplanning integreren. Hoewel er minder korte-termijn opdrachten lijken te worden verstrekt, blijft het inschatten van onderzoeksonderwerpen geen eenvoudige zaak en is continuïteit in het instituut daarmee een kwetsbare aangelegenheid. Maar ja: wil onderwijsonderzoek voor beleidsmakers interessant blijven dan zal er ruimte voor ad-hoc, politiek-geladen incidentele vragen moeten zijn. We kunnen dat ook beter appreciëren dan enige jaren terug, lijkt me.

Weiss heeft gesteld dat het effect van onderwijsonderzoek op beleid voornamelijk is gelegen in de andere wijze waarop de beleidsmakers tegen problemen aankijken. Men mag aannemen dat – als dit juist is – dit vooral een gevolg van een geregelde communicatie met onderzoekers over hun producten zal zijn. Een niet onbelangrijk neveneffect kan zijn dat er meer appreciatie voor de verschillende methoden van onderzoek ontstaat: een wat irrationele voorkeur voor een bepaalde opzet van onderzoek is soms aanwezig, wat een verarmd programma oplevert.

Men kan zich afvragen of de relevantie van te stellen onderzoeksvragen door beleidsmakers niet wordt verhoogd door explicitering van de door beleidsmakers gehanteerde impliciete beleidstheorie. Het verhelderen van de veronderstellingen – al of niet wetenschappelijk onderbouwd – van waaruit het beleid in een bepaald probleemgebied is opgebouwd, zou enig therapeutisch effect kunnen hebben, zoals dat omgekeerd het geval kan zijn voor de onderzoeker die zich geconfronteerd ziet met zijn impliciete beleidsoverwegingen.

Het voorhouden van die spiegels zou de kwaliteit van de communicatie kunnen verhogen.

Tenslotte:

Meer directe communicatie tussen onderzoekers en opdrachtgevers met het oog op programmering wordt in het voorgaande bepleit. De ergste fout die men kan maken als onderzoeker is om op de stoel van de beleidsmaker/veldvertegenwoordiger te gaan zitten. Zelfs de stoel aanschuiven is uit den boze. Het streven zij: maximale toenadering met behoud van distantie. Uit angst voor distantieverlies structuren bouwen die lange communicatielijnen betekenen, werkt averechts.

D. de Jong.

Licor.

Congressen, Symposia, Cursussen, Trainingen.

SYMPOSIUM COMPUTER ONDERSTEUND HOGER ONDERWIJS AAN DE R.U.G.: PC OF WP?

Onder deze titel zal op 25 mei a.s. aan de Rijksuniversiteit Groningen een eendaags symposium over computer ondersteund hoger onderwijs gehouden worden. De nadruk bij dit symposium zal liggen op ontwikkelmethoden, prototypes, en demonstraties. Deelnamekosten bedragen fl.50,-, inclusief symposiumbundel, lunch en borrel. Plaats van het symposium: Zernikegebouw Zaal 15, Landleven 12, Paddepoel, 9700 AV Groningen. Aanvang: 9.15 uur (inschrijving vanaf 8.30 uur). Reservering wordt aanbevolen gezien de beperkte capaciteit van de ruimtes (100 personen). Voor programma en reserveringsformulier, schrijf naar COWOG, Postbus 800, 9700 AV Groningen, of bel 050-633636.

ZEVENDE ONDERWIJSSOCIOLOGISCHE CONFERENTIE 'ONDERWIJS EN SOCIALE BEHEERSING'

Maatschappelijke oorzaken en effecten van schoolkenmerken

De Zevende Onderwijs sociologische Conferentie, georganiseerd onder auspiciën van de Stuurgroep Onderwijs sociologie, zal plaatsvinden op *12 en 13 november 1987* in het Conferentie-oord 'De Blije Werelt' te Lunteren.

Het conferentiethema is uitgewerkt naar vijf deelthema's:

1. sociale beheersing door socialisatie binnen scholen;
2. sociale beheersing van het onderwijs en van schoolkenmerken door het arbeidssysteem;
3. theoretische en methodologische multi-level problemen, toegespitst op sociale beheersing binnen scholen;
4. openbaar en bijzonder onderwijs als vormen van sociale beheersing;
5. etnische segregatie/integratie in het onderwijs en sociale beheersing.

Voor de vijf genoemde thema's zijn centrale inleiders uitgenodigd, resp.:

- Dr. P. Woods, Open University, Engeland
- Prof. dr. C. J. M. Schuyt, W.R.R.
- Prof. dr. S. Lindenberg, Sociologisch Instituut, RU Groningen
- Prof. dr. J. Billiet, KU Leuven
- Dr. M. Stone, Surrey, Engeland.

EARDHE CONFERENTIE

De Europese Associatie voor Onderzoek en Ontwikkeling van het Hoger Onderwijs (EARDHE) organiseert 22 t/m 25 april 1987 een internationaal congres te Utrecht. Het hoofdthema van het congres is '*Higher Education and New Technologies*'. Het congres is bestemd voor docenten, bestuurders en wetenschappers uit HBO en WO en zal worden geopend door de minister van onderwijs en wetenschappen Drs. W. J. Deetman. Het congres thema wordt op vier gebieden nader uitgewerkt:

- computerondersteund onderwijs (COO) in het HBO en WO

- automatisering in onderwijsbeleid en -beheer van de instellingen voor Hoger Onderwijs
- onderwijskundige ondersteuning
- onderwijstechnologie voor de derde wereld.

Tijdens het congres zal afzonderlijke aandacht besteed worden aan ontwikkelingen binnen verschillende sectoren van het onderwijs.

- 22 april '87 gamma-wetenschappen
- 23 april '87 beta- en medische wetenschappen
- 24 april '87 alpha-wetenschappen
- 25 april '87 technische wetenschappen

Het congres-secretariaat wordt verzorgd door de afdeling Onderzoek en Ontwikkeling van Onderwijs van de Rijksuniversiteit Utrecht, Heidelberglaan 8, 3584 CS Utrecht, tel. 030-534472.

INTERNATIONAL CONFERENCE LEARNING DISABILITIES 3 en 4 september 1987, AMSTERDAM

EERSTE AANKONDIGING EN OPROEP VOOR PAPERS

Op 3 en 4 september 1987 vindt te Amsterdam een internationale conferentie plaats over het thema **LEERSTOORNISSEN**, georganiseerd onder auspiciën van PAOS (Postacademisch Onderwijs Sociale Wetenschappen) en de IARLD (International Academy for Research in Learning Disabilities), een gezelschap van enkele honderden wetenschapsbeoefenaren uit 23 landen en diverse disciplines.

Circa 15 prominente sprekers uit Nederland, West-Duitsland, Engeland, de Verenigde Staten en Canada werken aan de conferentie mee.

De hoofdthema's zijn:

- definitie en predictie
- neuro-anatomische aspecten
- neuropsychologische aspecten
- cognitieve en informatieverwerkingsaspecten
- persoonlijkheids- en sociaal-psychologische aspecten
- uitkomsten
- behandeling/remediëring

De conferentie is bestemd voor hen die wetenschappelijk onderzoek doen, onderwijs geven, of in de beroepspraktijk werken op het gebied van leerstoornissen.

OPROEP VOOR PAPERS

Tijdens de conferentie is er voor deelnemers beperkte mogelijkheid tot het presenteren van een kort research-paper (15 minuten per paper), aansluitend bij één van de hoofdthema's. Abstracts in de engelse taal dienen ongeveer 500 woorden te bevatten, een beknopt beeld te geven van het doel van de studie, de gehanteerde methoden, de resultaten en de conclusies en vóór 1 mei 1987 ingezonden te worden aan onderstaand adres. Inzenders krijgen spoedig daarna bericht of hun paper al dan niet is geaccepteerd door de programmacommissie. De mogelijkheid bestaat dat enkele van de geaccepteerde papers worden gepubliceerd in het verslag van de conferentie.

INFORMATIE EN INSCHRIJVING

Uitgebreide informatie over het conferentie programma, alsmede inschrijfformulieren zijn verkrijgbaar bij:

PAOS, Postbus 325, 2300 AH Leiden, tel.: 071-278027.

Rapportage van Parameters uit het Raschmodel*

Theo J.H.M. Eggen en Henk Kelderman
Cito Arnhem *Universiteit Twente*

ABSTRACT

The article gives a survey of the scale properties of the parameters in the Rasch model. Scale properties of the model as such and of parameterizations of the model are distinguished. Desirable properties of scales for reporting individual abilities are given. Several possible choices for the scales are discussed. The article ends with the special possibilities of graphical representation of scores via so-called scale score reporting.

In de psychometrie staat het onderzoek naar latente-trek of itemresponsmodellen sinds de publikatie van Rasch (1960) enorm in de belangstelling. Geconstateerd kan worden dat dit onderwerp zowel theoretisch maar vooral ook naar de praktijk (Hambleton, 1982; Lord, 1980) in deze korte periode resultaten heeft opgeleverd. Te verwachten is dat de toepassingen van latente-trekmodellen door de technologische ontwikkelingen toe zullen nemen. De superioriteit op een aantal punten van itemresponsmodellen boven het voor de analyse van testresultaten ter beschikking staande klassieke testmodel (Fischer, 1974) kan zich alleen maar laten gelden, doordat computersystemen met veel meer mogelijkheden ter beschikking komen. Een voorbeeld van deze ontwikkeling zijn de projecten voor de opzet van centrale en schoolitembanken bij het Cito (Zwarts & Keizer, 1986).

Bij deze itembanken zal waar mogelijk gebruik gemaakt worden van itemresponsmodellen. Een model dat door zijn eenvoud en statistische eigenschappen hiervoor een belangrijke kandidaat is, is het Raschmodel. Dit model specificeert de kans op het goed maken van een dichotoom scorebaar item als functie van een op dezelfde latente schaal liggende waarde van een eigenschap van een persoon en van een eigenschap van een item. Doorgaans wordt het Raschmodel als volgt geparameteriseerd

$$P(+ | \theta_v, \sigma_i) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)} \quad (1)$$

Hierin is θ_v de persoonsparameter – de waarde van persoon v op de te meten latente eigenschap – en is σ_i te interpreteren als de moeilijkheid van item i met betrekking tot deze eigenschap. Het model wordt gecompleteerd door aan te nemen dat bij meerdere items de kansen op goed onafhankelijk van elkaar zijn en dat bij meerdere personen de kansen op goed van een item voor een vaste waarde van θ onafhankelijk van elkaar zijn.

* Deze studie werd uitgevoerd toen beide auteurs, werkzaam bij de vakgroep Onderwijskundige Meetmethoden en Data-analyse van de faculteit Toegepaste Onderwijskunde in opdracht van het Centraal Instituut voor Toetsontwikkeling te Arnhem, een onderzoek verrichtten naar 'Het gebruik van het Raschmodel voor een decentraal toetsservicesysteem'.

Adres: Cito, Postbus 1034, 6801 MG Arnhem;

Universiteit Twente, TO/OMD, Postbus 217, 7500 AE Enschede.

of $\prod_i \delta_i = 1$ daarvoor mogelijkheden. Dus bij deze parameterisering van het Raschmodel kunnen de parameters worden gemeten op een ratioschaal, hetgeen betekent dat absolute uitspraken over verhoudingen tussen schaalwaarden kunnen worden gedaan.

Bij de toepassing van het Raschmodel worden op basis van itemresponsen de itemparameters en persoonsparameters geschat. Relevant in dit verband is dan de vraag of de geschatte parameterwaarden dezelfde schaaieigenschappen hebben als de parameters in het model. Het blijkt dat, onder milde voorwaarden, zowel de niet conditionele als de conditionele aannemelijkheidschattingen van de parameters in het Raschmodel bij de parameterisering (1) en (7) precies onder de daarbij gegeven normeringen eenduidig zijn (Fischer, 1981). Dat wil zeggen uit de schattingsvergelijkingen volgen de schattingen voor de parameters uit (1) $\hat{\sigma}_i$ en $\hat{\theta}_v$ eenduidig op een additieve constante na; terwijl bij het multiplicatieve model (7) de schattingen $\hat{\xi}_v$ en $\hat{\delta}_i$ eenduidig zijn op een multiplicatieve constante na. De schaaieigenschappen van de parameters uit de modellen worden hierdoor dus duidelijk bevestigd voor de schattingen hiervan.

Uit het voorgaande kunnen we de conclusie trekken dat de schaal waarop de item- en persoonsparameters uit het Raschmodel gemeten kunnen worden, bepaald worden door de gekozen parameterisering. De vrijheid in de parameterisering geeft aan dat het minstens een schaal met onrinale eigenschappen is; echter door het vastleggen op een bepaalde structuur kan men meer eigenschappen aan de schaal geven. Duidelijk is echter dat het spreken over bijvoorbeeld differentieschaaieigenschappen van parameters alleen zinvol is als men het betreft op de bijbehorende parameterisering en niet als men het presenteert als eigenschappen van het Raschmodel in het algemeen.

GEWENSTE SCHAALEIGENSCHAPPEN

Alvorens in te gaan op mogelijke schaalkeuzes van de parameters uit het Raschmodel, kan opgemerkt worden dat het kiezen van een geschikte schaal voor de rapportage van testcores in het algemeen uitvoerig in standaard psychometrische literatuur wordt behandeld. Met name de moeilijkheden die er zijn met de interpretatie van ruwe scores op een test en mogelijke transformaties van de ruwe scores waardoor deze problemen mogelijk minder worden hebben daarbij de aandacht gekregen. Lyman (1978) geeft hiervan een aardig overzicht. De overwegingen die daarbij een rol spelen en de voor- en nadelen van een bepaalde transformatie zijn ook van toepassing als we over de schaalkeuze voor de Raschparameters spreken. Heel globaal kan gesteld worden dat een schaal beter is naarmate hij beter begrepen wordt door de gebruiker. De interpretatiemogelijkheden moeten zo duidelijk mogelijk zijn voor de gebruiker. Een aantal mogelijkheden om dit te bevorderen wordt nu besproken.

1. De schaal moet duidelijke referentiepunten hebben en een betekenisvolle eenheid. Het bezit van bijvoorbeeld een betekenisvol nulpunt of ander referentiepunt op de schaal en van een betekenisvolle eenheid zal het begrip van de schaal bevorderen. Enerzijds is het mogelijk een absoluut referentiepunt te kiezen (bijv. percentage correcte scores); dit noemt men criterium georiënteerde toetsen (Nitko, 1980). Vaak echter kiest men referentiepunten en eenheden op de schaal die gebaseerd zijn op interindividuele vergelijkingen van personen die de toets maakten (bijv. refereren aan groepsgemiddelden). In dat geval spreekt men van normgeoriënteerde toetsen.

2. Het werken met gehele positieve getallen op een schaal bevordert het begrip ervan. Decimalen en negatieve waarden moeten indien mogelijk vermeden worden, omdat deze waarden met name in relatie tot toetsscores door gebruikers moeilijk worden gevonden.

3. Als een schaal gekozen wordt moet er voor worden opgepast dat de gekozen waarden niet worden verward met andere gangbare schalen. Om de scores van de eindtoets basisonderwijs van het Cito niet te verwarren met intelligentiecoëfficiënten en schoolcijfers is daar gekozen voor een bereik op de schaal van 500 tot 550. Voor de goede interpretatie van scores moet verwarring met standaardtesten voorkomen worden; identificatie met een schaal van een

algemeen aanvaarde standaardtest zou echter het begrip wel eens kunnen bevorderen. Als elke testscore geïnterpreteerd zou kunnen worden als evenveel waard als een score op een bepaalde algemeen aanvaarde en overbekende standaardtest zou dit zeer welkom zijn. Echter zo'n algemeen aanvaarde standaardtest hebben we niet, zodat de mogelijke nadelen van verwarring voorshands groter lijken dan de voordelen van identificatie.

In relatie tot het Raschmodel kan men zich afvragen of gebruikers van de resultaten andere eisen of wensen hebben ten aanzien van de schaal voor persoons- of itemparameters. In het algemeen zullen de gebruikers meer waarde hechten aan een goed begrip van de persoonsparameters dan van de itemparameters. De gebruikers willen de prestaties van de personen kennen; dat die een samenhang hebben met de eigenschappen van de opgaven zullen ze inzien doch van ondergeschikt belang vinden. Zij zullen een duidelijk interpreteerbaar maatgetal voor de gemeten personen willen hebben. Daarom zullen we ons in het hierna volgende concentreren op mogelijke schaalkeuzes voor de persoonsparameter. Alle voor te stellen mogelijkheden en transformaties zijn in principe ook toepasbaar op itemparameters. Niet alle zijn echter relevant. Voor de praktijk valt het aan te raden de itemparameters op dezelfde schaal te rapporteren als de persoonsparameters. Daarnaast is de rapportage van de klassieke index p-waarde (percentage van de personen dat een bepaald item goed heeft) in een voor de concrete toepassing relevante populatie, vanwege de algemene bekendheid ermee en de modelonafhankelijkheid ervan, aanbevelenswaardig.

MOGELIJKE SCHAALKEUZEN

Aangezien, zoals eerder betoogd is, de vrijheid in de schaalkeuze afhangt van de gekozen parameterisering in het Raschmodel wordt in het hiernavolgende uitgegaan van de volgende modelstructuur:

$$P(+|\theta_v, \sigma_i) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)} \quad (9)$$

Hierin zijn θ_v en σ_i respectievelijk de persoonsparameter van persoon v en de itemparameter van item i . De parameters kunnen bij deze structuur op een differentieschaal gemeten worden. Daarnaast wordt ervan uitgegaan dat op een of andere wijze schattingen van de persoonsparameters θ (en ook de itemparameters) ter beschikking staan. De schaal waarop deze parameterwaarden worden gegeven is de zogenaamde logitschaal. Deze schaal dankt zijn naam aan het feit dat de logittransformatie – dat wil zeggen de natuurlijke logaritme van de verhouding van de kans op goed en de kans op fout – van (9) gelijk is aan het verschil van de parameters:

$$\ln(P(+|\theta_v, \sigma_i)/(1 - P(+|\theta_v, \sigma_i))) = \theta_v - \sigma_i \quad (10)$$

Uitgaande van schattingen $\hat{\theta}$ op deze logitschaal, die genormeerd is op $\sum_{i=1}^n \sigma_i = 0$, zullen hierna enkele mogelijke transformaties en dus schalen waarop deze $\hat{\theta}$ kunnen worden gerapporteerd worden besproken.

Positief Lineaire Transformaties

De schaalwaarden op de logitschaal behouden hun differentieschaaleigenschappen als we de transformatie van de schaal beperken tot de additieve transformatie in (2). Voor de praktijk is het instandhouden van de intervaleigenschappen van de schaal voldoende. Dit kunnen we bewerkstelligen door de transformaties van de schaal te beperken tot de positief lineaire transformaties:

$$\hat{k} = a + b \cdot \hat{\theta} \quad a, b > 0 \text{ constanten} \quad (11)$$

Door de keuze van a wordt de locatie – het referentiepunt – van de schaal vastgelegd, terwijl b de schaalfactor of eenheid op de schaal geeft; de oorspronkelijke vorm van de verdeling van de persoonsparameters blijft echter hetzelfde. Voor de keuze van het locatiepunt en de schaalfactor zijn zeer veel mogelijkheden. Afhankelijk van de soort interpretatie die men wenst kan men 3 subgroepen keuzes onderscheiden: (a) schalen met normatieve referentiepunten, (b) schalen met eenheden die onafhankelijk zijn van een groep en (c) schalen met eenheden die in de lijn van itemresponsmodellen te interpreteren zijn in termen van kansen. Achtereenvolgens zullen deze interpretaties worden toegelicht.

Normgroep Interpretatie

De schaal wordt in dit geval opgehangen aan een normgroep. Dat wil zeggen als schaalgemiddelde kiest men in eerste instantie het gemiddelde van de normgroep en als eenheid een standaarddeviatie in de verdeling van de normgroep; men standaardiseert de scores. Vervolgens kan men om praktische redenen besluiten het referentiepunt anders te kiezen (schaal opschuiven) en de eenheid van de standaarddeviatie van de normgroep te veranderen (schaal opblazen). De transformatie ziet er als volgt uit:

$$\hat{k} = a^* + b^* (\hat{\theta} - \bar{\theta}) / sd(\theta). \quad (12)$$

Hierin zijn θ en $sd(\theta)$ respectievelijk het gemiddelde en de standaarddeviatie van de normgroep, terwijl de keuze van a^* en b^* respectievelijk eventueel een handig referentiepunt (bijvoorbeeld 100 of 500) oplevert of een handzamer normgroepstandaarddeviatie (bijvoorbeeld 10 of 50) vastlegt. Transformaties van gestandaardiseerde ruwe scores met $a^* = 50$ en $b^* = 10$ (T-scores) en met $a^* = 500$ en $b^* = 100$ (Scholastic Aptitude Test Scores) uit de psychometrische literatuur (zie bijvoorbeeld Lyman, 1978) zijn alom bekend. De eerder genoemde score van de eindtoets basisonderwijs van het Cito werkt met $a^* = 535$ en $b^* = 10$.

Ook voor de rapportage van de persoonsparameter in het Raschmodel zijn enkele voorbeelden te geven: Wright en Stone (1979) nemen $a^* = 50$ en $b^* = 10$ voor een van hun NITS ('Normative Scaling Units') schalen, terwijl bij de Portland Public Schools RITS ('Rasch Scaling Units') in gebruik zijn die weer op een andere keuze voor de waarden van a^* en b^* berusten.

De op bovenstaande wijze te creëren schalen zijn te interpreteren met behulp van de normgroepprestaties. Voor de goede orde zij hierbij vermeld dat deze groep de beperkte groep zal kunnen zijn waarbij een test op een bepaald tijdstip is afgenomen, maar ook een standaardgroep of grotere populatie. Voorwaarde hiervoor is dat de gegevens van deze groep zijn vastgelegd. De bruikbaarheid van de schaal zal groter zijn naarmate de beschikbare normgroep bekender is en beter past bij de personen waarvan de resultaten geïnterpreteerd moeten worden. Voor de exacte keuze van de eenheid en locatie kan men zich laten leiden door de praktische overwegingen uit de vorige paragraaf. Echter ook het navolgen van eerder gemaakte keuzen kent geen enkel bezwaar.

Criteriuminterpretatie

Naast de mogelijkheid een normgroep voor referentie op een schaal te kiezen, is het ook mogelijk referentiepunten te kiezen die onafhankelijk zijn van een groep en bepaald worden door inhoudelijke criteria. Te denken valt hierbij aan de volgende situatie: op de logitschaal wijzen we twee punten aan die een in- en uitgangsniveau van een bepaalde cursus aangeven. Deze punten zouden we kunnen aanwijzen door bij de genoemde niveau's typische items aan te wijzen. De parameterwaarden van die items zijn dan deze twee inhoudelijke punten op de logitschaal. Vervolgens zouden we de schaal zodanig kunnen transformeren dat deze twee inhoudelijke punten eenvoudig te onthouden waarden (b.v. 0 en 10) krijgen. De overige waarden op deze schaal zijn dan te interpreteren in termen van deze inhoudelijke punten. Als we met $\hat{\sigma}_1$ en $\hat{\sigma}_2$ de punten op de logitschaal aanduiden en met $\hat{m}_1$ en $\hat{m}_2$ de referentiepunten op de nieuwe schaal, dan wordt de transformatie van de oorspronkelijke schaal gegeven door:

$$\hat{k} = a^* \cdot \frac{(\hat{m}_1 \hat{\sigma}_2 - \hat{m}_2 \hat{\sigma}_1)}{(\hat{\sigma}_2 - \hat{\sigma}_1)} + b^* \cdot \frac{(\hat{m}_2 - \hat{m}_1)}{(\hat{\sigma}_2 - \hat{\sigma}_1)} \hat{\theta} \quad (13)$$

De eenheid en de precieze referentiepunten worden in dit geval, evenals in het voorgaande, weer bepaald door de keuze van a^* en b^* . Wright en Stone (1979) geven voorbeelden – met SITS ('Substantive Scaling Units') eenheden – van deze inhoudelijk te interpreteren schalen.

Interpretatie in termen van kansen

Vanwege het probabilistisch karakter van het Raschmodel kan men de schaal zodanig kiezen dat bijvoorbeeld vergelijkingen van succeskansen van personen op items en van de succeskans van een persoon op verschillende items eenvoudig te maken zijn. Daarvoor kiest men een schaal voor de parameters met eenheden die geïnterpreteerd kunnen worden in termen van kansen. Omdat ook dit weer met behulp van een lineaire transformatie kan, is de keuze van handig te kiezen eenheden en referentiepunt weer aan de gebruikers. De schaal komt als volgt tot stand: men kiest de schaalpunten zodanig dat veranderingen in succeskansen op items van b.v. .10 naar .25 of van .25 naar .50 overeenkomen met eenheden van bijvoorbeeld 5, 10 of 20 op de nieuwe schaal. In tabel 1 is dit geïllustreerd.

Tabel 1. Voorbeeld schaaltransformatie met responskansen als basis

Succeskans p	$p/(1 - p)$	$\log(p/(1 - p))$	Nieuwe Eenheden ($c = 5$)
.10	1/9	-2.2	-10
.25	1/3	-1.1	- 5
.50	1/1	0.0	0
.75	3/1	1.1	5
.90	9/1	2.2	10

De bijbehorende transformatie van de logitschaal is:

$$\hat{k} = a^* + (c/1.1) \hat{\theta} \quad (14)$$

De verkregen schaal, waarvan de eenheden door Wright en Stone (1979) CHIPS ('Chicago Probability Units') worden genoemd, biedt de mogelijkheid om bijvoorbeeld een verschil van 10 eenheden tussen een persoon en een item direct te vertalen in de succeskans van deze persoon op dit item. Is men vertrouwd met het begrip odds – de verhouding van de kans op goed tot de kans op fout ($p/(1 - p)$) – dan zien we dat de nieuw gecreëerde eenheden zich eenvoudig laten vertalen in 'odds'-eenheden. Met name Wilmott en Fowles (1974) benadrukken en deze oddsinterpretatie; de door hen verkregen eenheden ($a^* = 50$ en $c = 5$) noemen zij WITS of ook wel BRYTES.

Monotoon Stijgende Transformaties

De in het voorgaande gegeven mogelijke schaalkeuzen voor de parameters van het Raschmodel hebben de gemeenschappelijke eigenschap dat ze positief lineaire transformaties van de logitschaal zijn. Aldus houden ze het intervalkarakter van de logitschaal in stand. Alternatieven hiervoor, die het intervalkarakter van de logitschaal verstoren maar andere prettige eigenschappen hebben, worden nu besproken.

Ware Score

Een mogelijkheid is de persoonsparameterschatting te transformeren naar de ware score op een bepaalde test of toets. Lord (1980) geeft aan dat de zogenaamde testkarakteristieke curve gelijk is aan de aantal goed ware score – de verwachting van het aantal goed op de lange duur – op de test:

$$\zeta_T(\theta) = E(X_T) = \sum_{i=1}^n P(+ | \theta_v, \sigma_i) \quad (15)$$

Het rechterlid van (15) is de zogenaamde testkarakteristieke curve van een test bestaande uit n

items, X_T is het aantal goed op de test, $\xi_T(\theta)$ de ware score op de test T . Uit (15) en (9) valt op te maken dat $\xi_T(\theta)$ een monotone transformatie van de persoonsparameter θ is. De rapportage van personen op de ware score schaal is mogelijk door in (15) de ter beschikking staande schattingen van θ en σ in te vullen:

$$\xi_T(\theta) = \sum_{i=1}^n P(+B\theta_v, \sigma_i) \quad (16)$$

Vanwege de relatie met het aantal goed op een test en de bekendheid met het begrip ware score kan rapportage op deze schaal van de persoonsparameters de voorkeur verdienen. Dit voordeel kan nog groter worden, doordat voor de keuze van de test waarop men de ware score rapporteert men zich niet gebonden hoeft te voelen aan de test die toevallig is afgenomen. Voor de test T in (15) kan iedere willekeurige test, waarvan de itemparameters bekend zijn worden gebruikt. Eventueel zou dat een nog te ontwikkelen standaardtest kunnen zijn. Met name het idee van een vaste standaardtest als referentie van scores van personen lijkt aantrekkelijk. Een andere aantrekkelijke mogelijkheid is alle items uit een itembank voor de scoring te gebruiken. Indien de items gecalibreerd zijn, vindt de scoring plaats op wat je zou kunnen noemen de bankkarakteristieke curve.

Percentielen

Een tweede, slechts monotone transformatie van de persoonsparameter, die zeker in aanmerking komt voor de rapportage van de persoonskenmerken is de transformatie naar percentielen – het percentage van personen in een gespecificeerde groep dat lager scoort dan een bepaalde waarde van de parameter:

$$p(\theta) = 100.F(\theta). \quad (17)$$

Hier is $F(\hat{\theta})$ de cumulatieve verdelingsfunctie van schattingen $\hat{\theta}$ in een referentiegroep. De groep die hiervoor gekozen wordt kan een willekeurig te kiezen referentiegroep zijn, mits men de persoonsparameters in de groep kent. Beperking tot de groep die de test op een bepaald moment maakt is niet noodzakelijk. Het voordeel van rapportage van persoonskenmerken met behulp van percentielen is de bekendheid ervan en de eenvoudige interpretatie; problemen met een goede keuze voor de referentiegroep en het verlies van intervaleigenschappen van de schaal staan daar echter tegenover.

VISUELE WEERGAVE VAN SCORES ('SCALE SCORING')

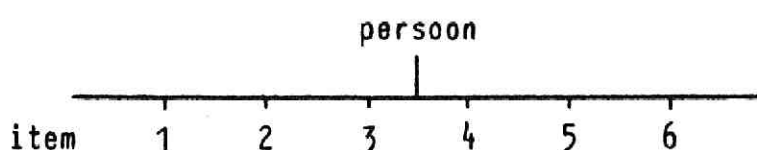
Alle voorgaande rapporteringswijzen van de parameters van het Raschmodel leveren een getal met een bepaalde interpretatie op. Deze getallen worden numeriek gerepresenteerd. Vaak is het wenselijk om de score met andere getallen te vergelijken. In dat geval kan het verstandiger zijn een visuele representatie te kiezen. Zowel een enkele score als een verzameling scores kan visueel worden weergegeven. In beide gevallen levert dit een grote toename in hoeveelheid informatie die in een keer kan worden overgebracht.

Visuele Weergave van een Enkele Score

Thurstone (1925) gaf reeds de moeilijkheid van opgaven weer als punten op een lijnstuk. Daardoor was hij in staat de inhoud van de individuele testitems te gebruiken voor de interpretatie van de door hen gemeten variabele. Hiervan kan dan gebruik gemaakt worden om de scores van de personen die door deze items op hetzelfde lijnstuk geplaatst worden te interpreteren. Dergelijke scoring staat bekend als 'Scale Scoring' (Bock, Mislevy en Woodson, 1982; Mislevy, Reiser en Zimowski, 1981). De in het voorafgaande besproken scoringen kunnen eveneens als punten op een lijnstuk worden weergegeven. Dit heeft als voordeel dat op hetzelfde

lijnstuk meerdere objecten kunnen worden weergegeven waarvan de onderlinge posities in een oogopslag kunnen worden vastgesteld. Op deze wijze kunnen vaardigheidsscores worden vergeleken met voor interpretatie relevante objecten. Die objecten kunnen zo worden gekozen dat de vaardigheidsscore van de leerling geïnterpreteerd kan worden in termen van gedrag. Voor de keuze van die objecten zijn een aantal mogelijkheden die ook in combinatie gehanteerd kunnen worden.

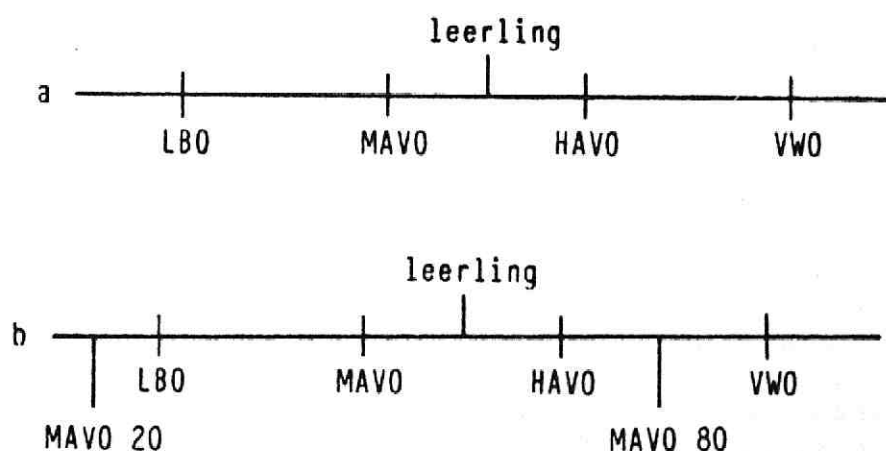
Ten eerste kan de vaardigheid van een leerling worden vergeleken met de moeilijkheid van een aantal zorgvuldig gekozen items. In Figuur 1 is de vaardigheid van een persoon afgebeeld op een θ -schaal waarop ook de opgavemoeilijkheden staan afgebeeld. Alle opgaven die links van de persoon staan afgebeeld worden door de persoon in minstens de helft van de gevallen goed gemaakt. Door de positie van de persoon in relatie tot die van de verschillende opgaven te beschouwen is de gebruiker beter in staat de score te interpreteren in termen van gedrag dan bij alleen rapportage van de score mogelijk zou zijn geweest.



Figuur 1. Itemmoeilijkheden en persoonsvaardigheid op een theta schaal.

Ten tweede kan men de gemiddelden van een aantal normgroepen op de schaal plaatsen. Bijvoorbeeld de gemiddelde prestaties bij verschillende schooltypen: LBO, MAVO, HAVO, enz. In dit geval kan de prestatie van de leerling worden vergeleken met die van de leerlingen in de betreffende schooltypen. In Figuur 2a is bijvoorbeeld naast de score van een leerling voor elk van de schooltypen MAVO, HAVO en VWO de gemiddelde scores afgebeeld. We zien dat het vaardigheidsniveau van de leerling hoger is dan het gemiddelde van de MAVO maar lager is dan dat van de HAVO.

Om een beeld te kunnen krijgen van de relatieve positie van de leerling binnen elk van de schooltypen kunnen enkele percentielen (of decielen) van de verdeling van elk van de schooltypen op de schaal worden aangegeven. In Figuur 2b zijn het twintigste en het tachtigste

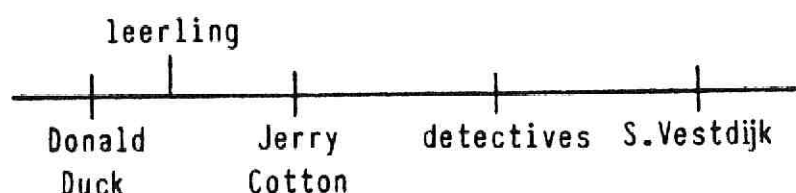


Figuur 2. Vaardigheidsniveau van leerling in relatie tot normgroepen.

percentiel ook in de figuur aangegeven. Merk op dat de schaal in principe in alle eerder besproken eenheden kan worden uitgedrukt.

Ten derde kan men het vaardigheidsniveau van de leerling in plaats van met normgroepen te vergelijken, vergelijken met het vaardigheidsniveau dat ten minste gehaald moet worden om aan bepaalde inhoudelijke eisen te voldoen. Men wil dan een criteriumgeoriënteerde in plaats van een normgeoriënteerde interpretatie van de testscore.

In Figuur 3 is een voorbeeld gegeven van een score voor een leesvaardigheidstoets. Naast het vaardigheidsniveau van de leerling is hier ook voor een aantal typen literatuur het vaardigheidsniveau aangegeven dat nodig is om dat type probleemloos te kunnen lezen.



Figuur 3. Leesvaardigheid van leerling in relatie tot teksttypen.

Visuele Weergave van Verscheidene Scores

Door gebruik te maken van informatie over de positie op de schaal van andere objecten kan men in een oogopslag een indruk krijgen van het vaardigheidsniveau van de leerling. Op een soortgelijke manier kan men zonodig ook een indruk krijgen van de positie van een leerling op meerdere schalen tegelijk. Sommige toetsprocedures leveren niet een score op een toets maar een profiel van scores op verschillende deeltolsten. Dergelijke profielen kunnen nu inzichtelijk visueel worden gerepresenteerd door de scores op elk van de deeltolsten weer te geven door een punt op een lijnstuk. De interpretatie van die lijnstukken kan nu worden vereenvoudigd door geschikte keuze van de oriëntatie van die lijnstukken ten opzichte van elkaar. Dat wil zeggen door een aantal punten die in de figuur boven elkaar liggen een op een bepaalde manier overeenkomstige interpretatie te geven. We kunnen dit opvatten als een vorm van equivalering met als doel een overeenkomstige interpretatie van qua inhoud verschillende schalen te verkrijgen.

De eerste manier om bovenliggende punten een overeenkomstige interpretatie te geven is om elk van de schalen in percentielscores voor een bepaalde groep uit te drukken. Boven elkaar liggende punten komen nu overeen met gelijke percentielen. Beschouwen we nu op elk van de lijnen de relatieve positie van de scores van de leerling op elk van de deeltolsten, dan kunnen we in één oogopslag zien op welke terreinen de leerling in die groep relatief achter ligt.

Ten tweede kan de oriëntatie van de lijnstukken ten opzichte van elkaar worden gebaseerd op een interpretatie met behulp van meerdere normgroepen. Vaak zal het zo zijn dat de normgroepen op elk van de deeltolsten een gelijke volgorde van vaardigheid zullen hebben, met name wanneer de normgroepen bestaan uit verschillende leerjaren van een bepaald schooltype. De kennis die leerlingen door de leerjaren heen opdoen zal door de bank genomen monotoon toenemen. De oriëntatie van de lijnstukken ten opzichte van elkaar wordt nu zo gekozen dat de normgroepen op elk van de lijnstukken recht boven elkaar komen te liggen. Merk op dat deze methode verschilt van de eerder beschreven percentielscore methode. In het eerste geval is er sprake van één normgroep, terwijl in het tweede geval er meerdere normgroepen zijn.

Ten derde kan de oriëntatie van de verschillende lijnstukken ten opzichte van elkaar worden ontleend aan externe gedragscriteria. De afteftgrens is bijvoorbeeld zo'n criterium. Verder kan

men bij elk van de lijnstukken een aantal punten van een kwalificatie voorzien, bijvoorbeeld in de vorm van een schoolcijfer, dat de kwaliteit van de met die punten overeenkomende gedragingen uitdrukt. Soms kan die kwalificatie worden ontleend aan de opgavenbank zelf (Nitko, 1980). bijvoorbeeld het percentage opgaven uit de bank die naar verwachting door de leerling goed gemaakt kunnen worden. Door de schaalpunten behorende bij overeenkomstige percentages goed bij de verschillende itembanken onder elkaar weer te geven kan men in een oogopslag zien welke deelbanken de leerling relatief slecht beheerst.

Voor welke representatie gekozen wordt, zal afhangen van de aard van het gebruik van de toets en van de aard van de toets zelf. Een visuele representatie van het vaardigheidsniveau kan echter extra mogelijkheden voor de interpretatie van de toetsscores opleveren.

DISCUSSIE

In het voorgaande is betoogd dat de schaaieigenschappen van de parameters in het Raschmodel afhangen van de gekozen parameterisering van het model. Daarnaast zijn de verschillende mogelijkheden voor rapportage en scoring van personen besproken. Duidelijk is dat visuele weergave van scores ('scale scoring') door zijn aard veel extra biedt. De vraag in hoeverre dit door gebruiker als positief wordt ervaren zal nog nader moeten worden onderzocht.

Voor scoring van leerlingen zal het noodzakelijk blijven een getalswaarde voor de vaardigheid van een leerling te presenteren. Ten aanzien van de gepresenteerde mogelijkheden kan geadviseerd worden met grote, positieve niet decimale getallen te werken, die niet verward kunnen worden met andere gangbare maten. Ten aanzien van de responskans interpretatie van de scores verdient het aanbeveling na te gaan of de in Engeland veel geprezen (o.a. Willmot en Fowles, 1974) vertaling naar log odds ook in Nederland zulke voordelen heeft. Transformatie van de vaardigheidsscores naar een ware scoreschaal is vooral interessant als men deze kan relateren aan de totale itembank.

Ten slotte is het vanwege de bekendheid en eenvoud aan te raden naast een andere maat te blijven werken met percentielscores in een relevante populatie.

SAMENVATTING

Het artikel geeft een overzicht van de schaaieigenschappen die de parameters in het Raschmodel kunnen hebben. Onderscheiden worden daarbij schaaieigenschappen van zowel het model als van parameterisering van het model. Ingegaan wordt op gewenste eigenschappen van schalen voor rapportage van persoonsvaardigheden. Diverse mogelijke schaalkeuzen worden besproken, waarna de extra mogelijkheden van visuele weergave van scores ('scale scoring') worden belicht.

REFERENTIES

- Bock, R.D., Mislevy, R., & Woodson, C. (1982). The next stage in educational assessment. *Educational Researcher*, 11, 4-11.
- Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Hans Huber Verlag.
- Fischer, G.H. (1981). On the existence and uniqueness of maximum likelihood estimator in the Rasch-model. *Psychometrika*, 46, 59-77.
- Hambleton, R.K. (Ed.) (1982). *Applications of Item Respons Theory*. Vancouver BC: Educational Research Institute of British Columbia.
- Kelderman, H., & Eggen, T.J.H.M. (1986). *Eenvoudige aanschouwelijke interpretaties van het Raschmodel*. Enschede: Universiteit Twente.
- Lord, F.M. (1981). *Applications of Item Respons Theory to Practical Testing Problems*. Hillsdale NJ: Lawrence Erlbaum Associates.

- Lyman, H.B. (1978). *Test scores and what they mean*. Englewood Cliffs NJ: Prentice Hall.
- Mislevy, R.J., Reiser, M.R., & Zinowski, M. (1981). *Scale reporting of National Assessment Data*. Final report to the Education Commission of the States under contract No. 02-81-20314. Chicago: International Educational Services.
- Nitko, A.J. (1980). Distinguishing the many varieties of criterionreferenced tests. *Review of Educational Research*, 50, 461-485.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Educational Tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Thurstone, L.L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451.
- Van der Linden, W.J. (1986). *Item banking met een dialoog gebaseerd op klassieke item- en testparameters*. Enschede: Universiteit Twente.
- Willmot, A.S., & Fowles, D.E. (1974). *The objective interpretation of test performance. The Rasch model applied*. Slough: NFER, Publishing Company Ltd.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.
- Zwarts, M.A., & Keijzer, H. (1986). Opzet en onderzoek van een toetservicesysteem. In W.J. van der Linden (red.), *Moderne methoden van toetsconstructie en -gebruik*; 8-14. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 11-12-1986.

Definitieve versie ontvangen 9-3-1987.

De multiple-choice cloze-toets als algemene taalvaardigheidstoets

Wim Jochems en Frans Montens
Technische Universiteit Delft

ABSTRACT

This article presents and discusses a number of empirical findings concerning the validity of multiple-choice cloze tests as tests of general language proficiency. Foreign students of Dutch at Delft University of Technology were given both a series of separate proficiency tests in listening, reading, speaking, writing and comprehension, and a series of multiple-choice cloze tests. Scores on the multiple-choice cloze tests were found to correlate significantly with each of the proficiency tests. In addition, scores on the multiple-choice cloze test appeared to form a solid basis for prediction of the total scores for listening, speaking, reading and writing taken together. Finally, a close, structural similarity was found to exist between curves of average scores on successive multiple-choice cloze tests and the presupposed growth of vocabulary during the language learning process. Together, these findings tend to show that multiple-choice cloze tests constitute a valid instrument for measuring general language proficiency.

Een multiple-choice (m.c.) cloze-toets¹ is een variant op de cloze-procedure. Een m.c. cloze-toets wordt geconstrueerd door in een (fragment van een) tekst regelmatig een woord te witten, bijvoorbeeld elk zevende of achtste woord. Voor elk 'gat' dat aldus ontstaat, worden in de rechter marge van de tekst enkele woorden als antwoordmogelijkheid gegeven, waarvan er steeds slechts één correct is. De opdracht is voor elk gat het juiste woord uit de aangegeven mogelijkheden aan te kruisen. Het aantrekkelijke van een m.c. cloze-toets is, dat niet alleen de scoring, maar ook – zoals nog zal blijken – de afname weinig tijd vergt, terwijl de constructie slechts enkele uren vereist. Het is dus een zeer efficiënt meetinstrument. Bovendien is de scoring geheel onafhankelijk van de beoordelaar, zodat het eveneens een volstrekt objectief instrument is.

Met name in situaties waar grote aantallen personen getoetst moeten worden, is de m.c. cloze-toets vanwege zijn efficiëntie interessant. Dat is bijvoorbeeld het geval aan de T.U. Delft waar zich jaarlijks 300 tot 400 buitenlandse studenten aanmelden. Overeenkomstig het bepaalde in het Academisch Statuut kunnen zij een studie pas aanvangen, nadat gebleken is, dat zij beschikken over 'voldoende kennis van de Nederlandse taal om het onderwijs met vrucht te kunnen volgen' (Artikel 66, lid 2). In de praktijk is de gang van zaken aan de T.U. Delft aldus. Elke buitenlandse student neemt bij aankomst deel aan een ingangsmeting, waarin zijn of haar kennis van het Nederlands wordt nagegaan. Afhankelijk van de geleverde prestaties wordt een student ofwel onmiddellijk toegelaten tot de eindmeting waarin getoetst wordt of er sprake is van voldoende kennis van het Nederlands, ofwel verplicht als beginner, halfgevoerde of gevorderde deel te nemen aan de cursus 'Nederlands voor buitenlanders' van de T.U. Delft. Zodra de cursus is doorlopen, kan de eindmeting worden ondergaan. Zowel bij de ingangsmeting als de eindmeting wordt gebruik gemaakt van een aantal m.c. cloze-toetsen.

Hoewel efficiëntie en objectiviteit niet onbelangrijk zijn, dienen bij de keuze van een toetsvorm uiteraard eisen gesteld te worden ten aanzien van betrouwbaarheid en validiteit. Met betrekking tot de validiteit van taalvaardigheidstoetsen raakt men onmiddellijk aan de discussie

Adres: Kanaalweg 2b, 2628 EB Delft

De auteurs zijn dank verschuldigd aan Maria Klijn-Wuisman voor haar bijdrage bij het verzamelen van de data.

rond 'integrative versus discrete-point testing' (Oller, 1979). 'Discrete-point testing' houdt kort gezegd in, dat de beheersing van specifieke elementen van de tweede taal gemeten wordt. Een vorm hiervan is de vier-vaardigheden-toets, waarbij luister-, spreek-, lees- en schrijfvaardigheid in die taal afzonderlijk getest worden. Bij 'integrative testing', ook wel 'global testing' genoemd, wordt gemeten in hoeverre een persoon in staat is die taal in een context te begrijpen en te hanteren. De m.c. cloze-toets is hiervan een voorbeeld. Deze discussie is verzand in een controverse over de zogenoemde 'global factor hypothesis', of anders gezegd: in de vraag of er aan taalvaardigheid iets al dan niet ondeelbaars ten grondslag ligt (Oller, 1982). Het resultaat is onder meer geweest, dat de m.c. cloze-toets als voor de hand liggende toetsvorm in discrediet is geraakt (zie voor het Nederlandstalige gebied bijvoorbeeld Beheydt, 1985).

Gelet op het voorgaande is het opmerkelijk, dat er weinig empirisch onderzoek naar de (in-)validiteit van m.c. cloze-toetsen is verricht. De enkele studies die er zijn, geven de indruk, dat een m.c. cloze-toets wel degelijk een betrouwbaar en valide instrument zou kunnen zijn voor het meten van taalvaardigheid (Cranney, 1973; Jonz, 1976; Wijnstra, 1977; Hinofotis & Snow, 1980). Helaas beperken zij zich tot het rapporteren van correlaties tussen prestaties op m.c. cloze-toetsen enerzijds en anderzijds (één of enkele van) de gebruikelijke discrete-point toetsen. Maar zij gaan volledig voorbij aan de vraag wat nu precies gemeten wordt door een m.c. cloze-toets.

In het navolgende wordt nagegaan in hoeverre een m.c. cloze-toets als algemene taalvaardigheidstoets kan functioneren. Met name zal getracht worden een aanzet te geven tot beantwoording van de eerder genoemde validiteitsvraag: wat meet een m.c. cloze-toets. Daartoe zullen we een drietal onderzoeken bespreken. In het eerste onderzoek worden van 85 buitenlandse studenten die de cursus 'Nederlands voor buitenlanders' hebben gevolgd, de prestaties op m.c. cloze-toetsen vergeleken met hun prestaties op een in Nederland veel gebruikte vier-vaardigheden-toets, waarin luisteren, spreken, lezen en schrijven afzonderlijk worden gemeten. In het tweede onderzoek worden van 76 buitenlandse studenten die de cursus gevolgd hebben, de prestaties op een gewone m.c. toets voor tekstbegrip vergeleken met hun prestaties op een m.c. cloze-toets. Aanleiding is de in de literatuur geopperde verwachting, dat door een m.c. cloze-toets vooral tekstbegrip gemeten zou worden (Porter, 1976, p. 152 en 153). In het derde onderzoek worden bij 125 buitenlandse studenten vorderingen tijdens de cursus gemeten, teneinde na te gaan in hoeverre zelfs een korte m.c. cloze-toets in staat is vooruitgang in de beheersing van het Nederlands zichtbaar te maken. Achtergrond van dit onderzoek is onze verwachting, dat er een structureel verband bestaat tussen de voortgang van het leerproces in de cursus enerzijds en anderzijds de vorderingen zoals gemeten door een m.c. cloze-toets. Zou een dergelijk, nog nader te specificeren verband gedemonstreerd kunnen worden, dan is dat een belangrijk element in het kader van de validering van m.c. cloze-toetsen. In de discussie ten slotte zullen we nagaan wat de beschreven bevindingen betekenen voor de beantwoording van de vraag naar de validiteit van m.c. cloze-toetsen. Tevens worden er kanttekeningen geplaatst bij het uitgevoerde onderzoek.

ONDERZOEK I: VIER-VAARDIGHEDEN-TOETSING

Methode

Onmiddellijk nadat zij de cursus 'Nederlands voor buitenlanders' aan de Technische Universiteit Delft hadden gevolgd, is aan 85 buitenlandse studenten² een zogenoemde vier-vaardigheden-toets afgenomen alsmede enkele multiple-choice cloze toetsen. In de vier-vaardigheden-toets, samengesteld door de landelijke werkgroep Centrale Toets Nederlands, worden vier taalvaardigheden afzonderlijk getest, te weten luisteren, spreken, lezen en schrijven. Luisteren wordt getest door een video-opname van drie fragmenten uit een populair-wetenschappelijk televisieprogramma aan te bieden, waarover 31 vragen worden gesteld, voornamelijk multiple-choice vragen. Spreken omvat twee onderdelen: een toets en een opdracht. In beide gevallen wordt een bandopname gemaakt van de mondelinge reacties van de kandidaat. Bij de toets

reageert hij op 21 opmerkingen uit dagelijkse communicatieve situaties van een student. Bij de opdracht moet de kandidaat een monoloog van twee minuten houden binnen een nader bepaald kader. Lezen wordt getoetst door het aanbieden van vijf korte teksten over verschillende onderwerpen waarbij per tekst 5 multiple-choice vragen gesteld worden. Schrijven omvat weer twee onderdelen: een toets en een opdracht. Bij de toets moet de kandidaat een ontbrekende zin toevoegen aan 14 korte tekstfragmenten. Bij de opdracht krijgt hij twee afbeeldingen waarover hij in een essay van 150 à 200 woorden enkele vragen moet beantwoorden.

Naast deze vier-vaardigheden-toets zijn een aantal multiple-choice cloze-toetsen afgenomen. Daarbij werden teksten aangeboden waarin, te beginnen bij een willekeurig woord, in de regel elk zevende woord is weggelaten. De kandidaat diende voor elk weggelaten woord er één te kiezen uit drie alternatieven, waarvan er telkens slechts één juist is. De onjuiste alternatieven werden gekozen uit de foute antwoorden van gevorderde studenten uit voorgaande cursussen, aan wie deze teksten met dezelfde 'gaten' als open cloze-toets waren voorgelegd. De toetsen waartoe we ons hier beperken, betroffen de teksten van twee uitzendingen van de radionieuwsdienst (elk met 100 items) en een fragment uit een leerboek natuurkunde voor het voortgezet onderwijs (46 items).

De scoring verliep als volgt. Bij alle m.c. cloze-toetsen alsmede bij de luister- en leestoets is per item 0 (fout) of 1 (goed) gescoord overeenkomstig de sleutel. Bij de spreektoets en de schrijftoets wordt per item de score 0, 1 of 2 gegeven: 2 indien een adequaat, begrijpelijk antwoord is gegeven waarin nog wel een enkele grammaticale fout mag voorkomen; 1 indien een antwoord is gegeven dat niet geheel adequaat is, grammaticale fouten bevat maar wel redelijk begrijpelijk is; 0 in alle overige gevallen. De prestaties bij de spreekopdracht en de schrijfoopdracht worden beide beoordeeld op een schaal die loopt van 0 tot en met 7. Voor deze twee opdrachten is er slechts een summier scoringsvoorschrift.

Resultaten

In Tabel 1 zijn de resultaten van de 85 personen op de onderdelen van de vier-vaardigheden-toets weergegeven. De betrouwbaarheid van luistertoets en leestoets is bepaald met behulp van Kuder-Richardson-formule 20; van spreektoets en schrijftoets waar elk item de score 0, 1 of 2 kan opleveren, is de betrouwbaarheid bepaald via de splitsingsmethode, waarbij per toets de items met een oneven nummer gesplitst zijn van die met een even nummer.

In Tabel 1 zijn geen gegevens opgenomen met betrekking tot de spreekopdracht en de schrijfoopdracht uit de vier-vaardigheden-toets. De reden is dat de psychometrische kwaliteit van deze twee te wensen overlaat. Teneinde een indruk te krijgen van de betrouwbaarheid van de beoordeling van beide opdrachten zijn van een steekproef van personen de antwoorden voorgelegd aan een tweede beoordelaar. De productmoment-correlatiecoëfficiënt tussen beide beoordelingen bleek voor de spreekopdracht 0.63 en voor de schrijfoopdracht 0.66. Opvallend was dat

Tabel 1. Gegevens over de vier onderdelen van de vier-vaardigheden-toets voor dezelfde 85 personen, exclusief de spreekopdracht en de schrijfoopdracht; voor spreken en schrijven met als score per item 0, 1 of 2 is p' in plaats van $\bar{p}$ en r_{xx} in plaats van KR20 aangegeven.

	Vier-vaardigheden-toets			
	Luisteren	Spreken	Lezen	Schrijven
k	31	21	25	14
$\bar{X}$	15.99	26.67	13.02	17.89
s_x	5.20	9.07	4.20	5.45
$\bar{p}$ of p'	0.52	0.64	0.52	0.64
KR20 of r_{xx}	0.79	0.88	0.72	0.83

beide beoordelaars de aangegeven normen kennelijk verschillend hanteerden. Bij de spreekopdracht gaf beoordelaar 1 in vergelijking met beoordelaar 2 gemiddeld 1.5 punt meer op een schaal die loopt van 0 tot en met 7. Bij de schrijfofdracht bedroeg het verschil gemiddeld bijna 1 punt. In absolute zin bestaat er dus een beduidend verschil tussen beide beoordelingen. Vanwege de gebrekkige kwaliteit laten we de spreekopdracht en de schrijfofdracht verder buiten beschouwing.

Vatten we de vier-vaardigheden-toets op als een samengestelde toets bestaande uit vier componenten, namelijk luisteren, spreken, lezen en schrijven, dan is de betrouwbaarheid van deze samengestelde toets geschat met behulp van de alpha-coëfficiënt voor $k=4$ gelijk aan 0.80 (Lord & Novick, 1968, p. 91). Onze conclusie is dat de vier-vaardigheden-toets exclusief spreekopdracht en schrijfofdracht van aanvaardbare kwaliteit is.

In Tabel 2 zijn de resultaten van de 85 personen op de drie m.c. cloze-toetsen weergegeven. Ook van deze toetsen is de betrouwbaarheid, bepaald met behulp van Kuder-Richardson-formule 20, wederom betrekkelijk hoog. Bovenstaande gegevens maken aannemelijk dat de m.c. cloze-toetsen in dit opzicht eveneens van aanvaardbare kwaliteit zijn.

Tabel 2. Gegevens over de m.c. cloze-toetsen Radionieuwsdienst A, Radionieuwsdienst B en Energie.

	Multiple-choice cloze-toetsen		
	RND A	RND B	Energie
k	100	100	46
$\bar{X}$	69.68	65.44	33.33
s_x	11.36	13.15	5.34
$\bar{p}$	0.70	0.65	0.73
KR20	0.89	0.91	0.76

Teneinde een indruk te krijgen van de samenhang tussen de onderdelen van de vier-vaardigheden-toets en de drie m.c. cloze-toetsen zijn in Tabel 3 de productmoment-correlatiecoëfficiënten gegeven. De correlatiecoëfficiënten met een waarde grote dan 0.36 zijn significant ($\alpha=0.001$). Gezien het correlatiepatroon meten de afzonderlijke onderdelen van de vier-vaardigheden-toets voor een belangrijk deel hetzelfde. Dit gegeven op zich geeft allerminst aanleiding om bij het meten van taalvaardigheid deze vier deelvaardigheden te onderscheiden. Tevens valt uit Tabel 3 op te maken, dat de m.c. cloze-toetsen in aanzienlijke mate hetzelfde meten als wat elk van de onderdelen meet, namelijk luisteren, spreken, lezen en schrijven.

Gegeven de samenhang tussen de vier deelvaardigheden is nagegaan in hoeverre er een systematisch verband bestaat tussen de score op elk van de m.c. cloze-toetsen enerzijds en

Tabel 3. Correlaties tussen de onderdelen van de vier-vaardigheden-toets en de drie m.c. cloze-toetsen.

	Luisteren	Spreken	Lezen	Schrijven	RND A	RND B
Spreken	0.61					
Lezen	0.60	0.23				
Schrijven	0.66	0.73	0.46			
RND A	0.72	0.52	0.62	0.74		
RND B	0.87	0.62	0.69	0.73	0.87	
Energie	0.68	0.49	0.56	0.47	0.75	0.72

anderzijds de score op de totale vier-vaardigheden-toets. De laatste vatten we op als een goede schatter, E, van 'taalvaardigheid' en noemen we verder de E-score. Omdat het scorebereik van de onderdelen uit de vier-vaardigheden-toets niet gelijk is en er vooralsnog geen reden is een verschillend gewicht aan de vier toe te kennen, is als E-score het gemiddelde genomen van de percentages goede antwoorden van een persoon op de vier onderdelen (cf. Jochems & Montens, 1986, p. 34-36). De productmoment-correlatiecoëfficiënt tussen de scores op de vier-vaardigheden-toets, bepaald als E-score, en de scores op de eerste m.c. cloze-toets blijkt voor deze 85 personen 0.79 te zijn. Wordt de betrouwbaarheid van beide metingen erin betrokken, dan is de correlatiecoëfficiënt na correctie voor attenuatie 0.94 (Nunnally, 1978, p. 220). De correlatie met de vier-vaardigheden-toets is voor de tweede m.c. cloze-toets 0.78 en voor de derde 0.71, hetgeen na correctie voor attenuatie in beide gevallen 0.91 oplevert.

De conclusie is evident: de verschillende m.c. cloze-toetsen en de vier-vaardigheden-toets meten in hoge mate hetzelfde. Worden de correlatiecoëfficiënten, gecorrigeerd voor attenuatie, gelezen als validiteitscoëfficiënten, dan is de predictieve validiteit van deze m.c. cloze-toetsen zeer hoog, waarbij men dient te bedenken dat het hier om de betrouwbare scores gaat en niet om de observeerbare scores. Vanuit het oogpunt van efficiëntie is een voor de hand liggende consequentie om bij ingangsmetingen en eindmetingen te volstaan met het afnemen van een m.c. cloze-toets en af te zien van de vier-vaardigheden-toets. Gebleken is, dat voor 85 personen afname en correctie van een m.c. cloze-toets van 100 items slechts circa 5% van de tijd vergde, die nodig was voor afname en verwerking van de volledige vier-vaardigheden-toets.

ONDERZOEK II: TEKSTBEGRIP

Methode.

Aan 76 buitenlandse studenten³ zijn, onmiddellijk nadat zij de cursus 'Nederlands voor buitenlanders' aan de T.U. Delft hadden doorlopen, twee toetsen voorgelegd. Als eerste is de toets 'Nederlands tekstbegrip' uit het centraal schriftelijk eindexamen LBO 1984 afgenomen. Deze toets die geacht wordt tekstbegrip te meten, bestaat uit twee teksten waarover in totaal 50 vragen worden gesteld. Bij elke vraag zijn vier antwoordmogelijkheden gegeven waarvan er steeds slechts één juist is. Er is voor een m.c. toets gekozen, omdat dergelijke toetsen in psychometrisch opzicht van betere kwaliteit bleken dan die met open vragen (cf. Luijten et al., 1984, p. 34, p. 35 en p. 40). Daardoor is de keuze uiteraard zeer beperkt. Als tweede is een m.c. cloze-toets afgenomen, namelijk Radionieuwsdienst A uit onderzoek I welke 100 items omvat.

Resultaten

Uit Tabel 4 is af te lezen, dat de betrouwbaarheid van beide toetsen een aanvaardbare waarde bereikt. De productmoment-correlatiecoëfficiënt tussen de scores van de 76 personen op de m.c. cloze-toets en hun scores op de toets Tekstbegrip bedraagt 0.64, welke waarde significant is ($\alpha=0.001$). Dit wijst erop dat er een systematisch verband bestaat tussen datgene wat door de m.c. cloze-toets wordt gemeten en datgene wat de toets Tekstbegrip meet. De correlatie na correctie voor attenuatie bedraagt 0.78.

Wordt de correlatie tussen scores op de m.c. cloze-toets en op de toets Tekstbegrip ($r_{xy}=0.64$) vergeleken met de correlaties tussen scores op dezelfde m.c. cloze-toets en elk van de vier onderdelen uit de vier-vaardigheden-toets (zie de vierde rij in Tabel 3), dan blijkt de correlatie met Tekstbegrip er niet in positieve zin bovenuit te springen. Onze bevindingen leveren dus geen ondersteuning voor de door Porter geuite veronderstelling, dat een m.c. cloze-toets vooral tekstbegrip zou meten. De samenhang tussen de m.c. cloze-toets en de toets die tekstbegrip meet, is immers globaal genomen van dezelfde orde van grootte als de samenhang tussen deze m.c. cloze-toets en elk van de onderdelen uit de vier-vaardigheden-toets die luisteren, spreken, lezen, respectievelijk schrijven meten.

Tabel 4. De resultaten van de 76 personen op de m.c. toets Tekstbegrip en op de m.c. *cloze*-toets Radionieuwsdienst A; r_{XY} geeft de correlatie tussen de prestaties op beide testen en $\hat{r}_{XY}$ de correlatie na correctie voor attenuatie.

	Tekstbegrip	RND A
k	50	100
$\bar{X}$	24.92	75.28
s_x	7.24	8.87
$\bar{p}$	0.50	0.75
KR20	0.80	0.85
r_{XY}		0.64
$\hat{r}_{XY}$		0.78

ONDERZOEK III: METEN VAN VORDERINGEN

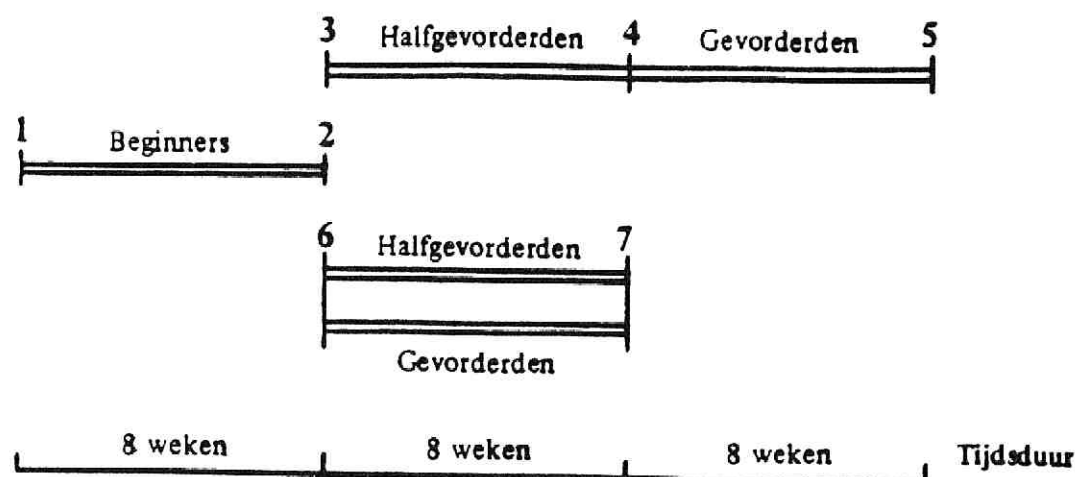
Methode

Bij 125 buitenlandse studenten⁴ die de cursus 'Nederlands voor buitenlanders' aan de T.U. Delft hebben gevolgd, zijn de vorderingen tijdens de cursus onderzocht. Omdat zowel bij de ingangsmeting als bij de eindmeting de m.c. *cloze*-toets Energie is afgenomen, kunnen vorderingen hiermee zichtbaar gemaakt worden. Deze personen hebben tussen beide metingen op verschillende manieren aan de cursus deelgenomen. We geven daarom summier de cursusopbouw weer voor zover nodig voor de interpretatie van resultaten. Zie voor een beschrijving van de cursus Sciarone en Montens (1984 of 1985) en voor het materiaal bijvoorbeeld Montens en Sciarone (1984).

Aan de hand van een score die een persoon bij de ingangsmeting bereikt, wordt bepaald op welk punt hij of zij in de cursus stapt, namelijk punt 1, 3, 4 of 6 in Figuur 1. Zodra de student punt 5 of 7 heeft bereikt, wordt de eindmeting verricht. De cursus kent twee zogenoemde stromen. In de langzame stroom doorloopt een cursist in 24 weken het traject 1-5. Bij de snelle stroom wordt in 16 weken het traject 1-7 afgelegd. Het verschil tussen beide is, dat in de snelle stroom *gelijktijdig* het cursusmateriaal voor halfgevoorderden én voor gevorderden wordt doorgewerkt en het tempo tweemaal zo hoog is vergeleken met de langzame stroom. Of een student in de snelle dan wel de langzame stroom wordt geplaatst hangt onder meer af van de score op de ingangsmeting, de vorderingen tijdens het beginnersdeel van de cursus, alsmede van uiteenlopende praktische kwesties. Het beginnersdeel is voor beide stromen identiek. Er zijn dus vijf manieren waarop een student aan de cursus deel kan nemen. In termen van Figuur 1 zijn dat de trajecten 1-5, 1-7, 3-5, 4-5 of 6-7. Merk op dat tussen ingangs- en eindmeting tenminste één periode van 8 weken valt, zodat het effect van testing gering is te achten (Campbell & Stanley, 1971, p. 179). Met dit effect wordt bedoeld dat ten gevolge van de eerste afname van de toets de scores bij de tweede afname iets hoger zullen zijn, bijvoorbeeld door herinnering.

Resultaten

Uit Tabel 5 is af te lezen, dat de prestaties op de m.c. *cloze*-toets Energie als onderdeel van de eindmeting ingrijpend verbeterd zijn ten opzichte van de prestaties op die toets als onderdeel van de ingangsmeting. De toename van de gemiddelde score van 20.26 naar 35.08 is significant (t-toets voor gecorrleerde gemiddelden; $t = 12.45$; $\alpha = 0.001$). Ook de afname van de standaarddeviatie is significant in dezelfde orde van grootte. Gelet op de toename van de gemiddelde score van de 125 studenten is het geconstateerde effect van de cursus zeer groot. In dit verband echter zijn wij niet zozeer in de *grootte* van het leereffect geïnteresseerd, temeer daar we nog niet over voldoende vergelijkingsmateriaal beschikken. Het gaat ons hier om de *opbouw* van het leereffect. Dit laatste zullen we nader analyseren.



Figuur 1. Schematisch overzicht van de opbouw van de cursus en de tijdsduur per onderdeel.

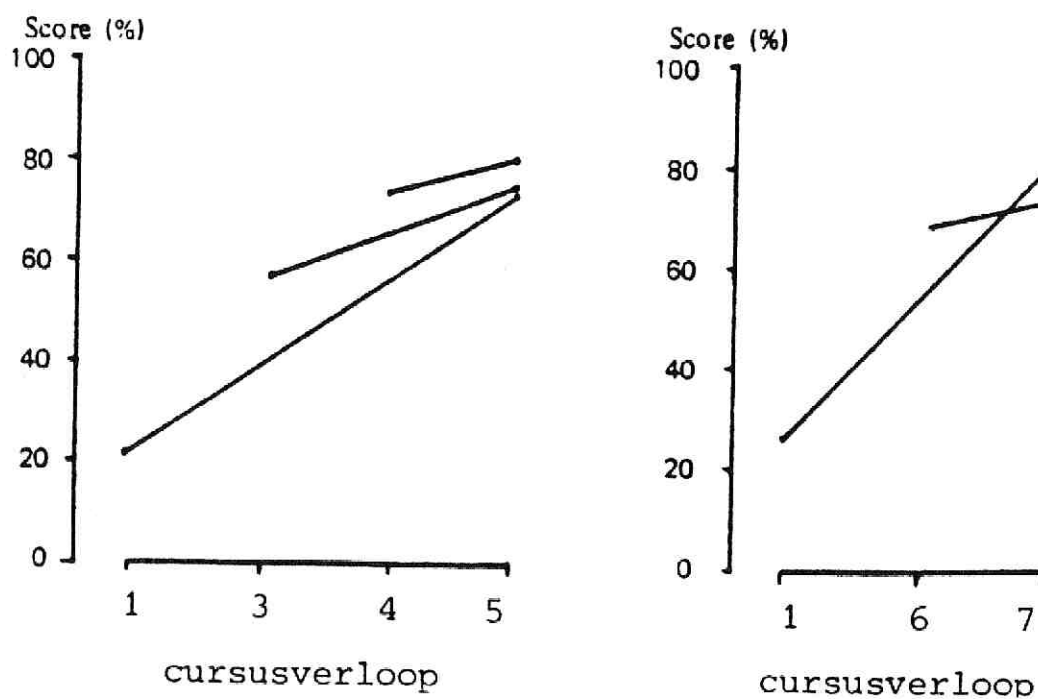
Van de 125 personen is in Tabel 6 het cursusverloop gespecificeerd. Daarbij doen zich vijf mogelijkheden voor zoals in de tweede kolom is te zien. Voor elk mogelijk cursusverloop is bepaald hoeveel personen die weg gevolgd hebben (kolom 3), alsmede wat hun gemiddelde score op de m.c. cloze-toets Energie als onderdeel van de ingangsmeting was (kolom 4) en hun gemiddelde op die toets als onderdeel van de eindmeting (kolom 5). Het verschil tussen beide gemiddelden geeft een indicatie van de leerwinst die in het betreffende traject van de cursus is geboekt. In Figuur 2 is een en ander in beeld gebracht. Hierin geeft een lijnstuk aan, welke toename in score (verticale as) gemiddeld gerealiseerd is gedurende een zeker cursusverloop (horizontale as). De interpretatie ligt voor de hand: naarmate personen verder in de cursus vorderen lijkt de bijdrage van de cursusdelen aan het leereffect geringer te worden. Deze interpretatie verdient nadere beschouwing.

Opgemerkt dient te worden dat uit de grafische weergave in Figuur 2 niet mag worden opgemaakt, dat de vorderingen in het leerproces lineair verlopen. De lijnstukken geven slechts weer welke gemeten waarden bij elkaar horen, omdat zij over dezelfde personen verkregen zijn. Maar hoe verlopen de vorderingen in het leerproces dan wel? Onze veronderstelling is deze. Stel, dat de beheersing van een taal op zeker moment gekarakteriseerd kan worden door het aantal gekende woorden van die taal door een persoon. De gekende woorden garanderen een zekere dekking van een willekeurige tekst. Anders gezegd: de gekende woorden maken een zeker percentage uit van de woorden die in de tekst voorkomen: het dekkingspercentage. Naarmate de beheersing van de taal toeneemt stijgt het dekkingspercentage. Wordt het dekkingspercentage uitgezet tegen de woorden in strikte volgorde van afnemende frequentie, dan levert dat de ideale dekkingscurve op (Sciarone, 1979; Sciarone & Montens, 1985).

Uit Figuur 3 is af te lezen, dat bekendheid met de 500 meest frequent voorkomende woorden voor een willekeurige tekst een dekkingspercentage van circa 65% kan geven; bekendheid met

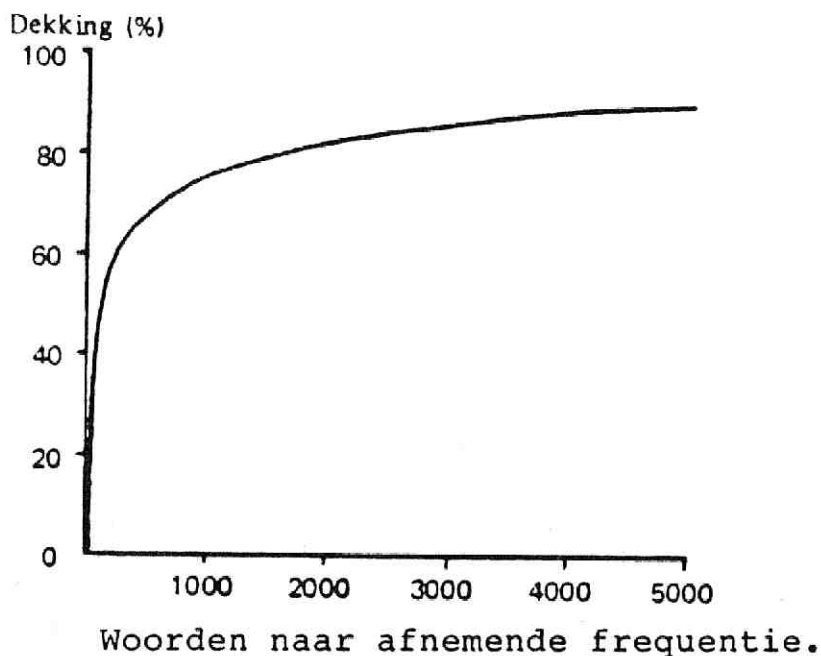
Tabel 5. Gegevens met betrekking tot de resultaten op de m.c. cloze-toets Energie als onderdeel van de ingangsmeting en als onderdeel van de eindmeting voor dezelfde 125 personen ($k=46$).

	Ingangsmeting	Eindmeting
$\bar{X}$	20.26	35.08
s_x	12.57	4.82
$\bar{p}$	0.44	0.76
KR20	0.96	0.75



Figuur 2. Grafische weergave van de vorderingen gemeten met m.c. cloze-toets Energie (verticale as) ten opzichte van het cursusverloop (horizontale as; de cijfers geven de trajecten aan overeenkomstig Figuur 1) voor de langzame stroom (links) en de snelle stroom (rechts).

de 1000 meest frequente woorden een dekkingpercentage van circa 70%. Eveneens is af te lezen, dat voor een verdere stijging van het dekkingpercentage een onevenredig grote stijging van het aantal gekende woorden nodig is en wel steeds minder frequent voorkomende woorden. Er is sprake van een *ideale* dekkingcurve omdat de geschetste relatie slechts deze vorm zal aannemen onder de ideale omstandigheid, dat de woorden in strikte volgorde van afnemende



Figuur 3. Ideale dekkingcurve bij taalverwerving in termen van dekking (verticale as) en aantal geleerde woorden naar afnemende frequentie (horizontale as).

Tabel 6. Overzicht van het cursusverloop, uitgedrukt in termen van de trajecten uit figuur 1; voor een verdere toelichting wordt verwezen naar de tekst.

Stroom	Cursus-verloop	Aantal personen	$\bar{X}$ vooraf	$\bar{X}$ achteraf	Aantal contacturen onderwijs
Langzaam	1 - 5	34	10.24	33.53	152
	3 - 5	34	26.47	34.15	96
	4 - 5	20	34.15	37.00	48
Snel	1 - 7	29	12.00	36.97	104
	6 - 7	8	31.75	34.00	48

frequentie worden geleerd (en niet meer worden vergeten). In de werkelijkheid zal bij het leren van een taal die strikte volgorde van afnemende frequentie niet voorkomen, zodat het feitelijk verloop van de curve altijd vlakker zal zijn dan in Figuur 3 is aangegeven.

Hoewel de exacte relatie tussen bedoeld dekkingspercentage en het percentage correct beantwoorde items op een m.c. cloze-toets onbekend is, kunnen de lijnstukken in Figuur 2 geïnterpreteerd worden als koorden van een min of meer afgevlakte werkelijke versie van de ideale dekkingscurve uit Figuur 3. In de eerste drie weken van het cursusdeel voor beginners worden de 1200 meest frequente woorden van het Nederlands ingevoerd. De dekking kan dan zeer snel stijgen tot circa 70%. In de navolgende cursUSDelen zal men 'veel meer' moeten leren om een verhoudingsgewijs steeds geringere toename van het dekkingspercentage te realiseren. De eindterm van de cursus ligt bij 80% dekking, waarvan wordt aangenomen dat die in staat stelt tot een redelijk goed begrip van een willekeurige tekst.

Als alternatieve verklaring voor de afnemende leerwinst met het vorderen van de cursus zou men de beperkte lengte van de toets ($k=46$) kunnen aanvoeren, waardoor een plafond in de prestaties is gelegd. Een betrekkelijk hoge score kan immers nooit meer veel verbeteren ook al wordt er veel geleerd. Deze verklaring achten wij echter weinig plausibel. Ook lange tot zeer lange m.c. cloze-toetsen met 100 tot 450 items die de afgelopen jaren aan cursisten zijn afgenomen, laten dit verschijnsel zien: een score van circa 70% of meer stijgt maar weinig gedurende de cursus ongeacht de lengte van de toets. Derhalve zijn wij van mening dat de vorderingen gemeten door de m.c. cloze-toets geïnterpreteerd kunnen worden in termen van de geschetste dekkingscurve bij taalverwerving. Anders gezegd: kennelijk bestaat er een structurele overeenkomst tussen enerzijds het verloop van scores op een m.c. cloze-toets in de cursus en anderzijds het verloop van het dekkingspercentage in het proces van taalverwerving. Wij achten dit een belangrijke indicatie voor de validiteit van de m.c. cloze-toets als taalvaardigheidstoets.

DISCUSSIE

Een m.c. cloze-toets blijkt in vergelijking met de vier-vaardigheden-toets een uitermate efficiënt meetinstrument te zijn. De efficiëntie neemt overigens nog toe als de verwerking gedeeltelijk of geheel geautomatiseerd wordt, hetgeen gezien de volstrekte objectiviteit van zo'n toets zeer wel mogelijk is. De gevonden betrouwbaarheidscoëfficiënten wijzen erop dat de betrouwbaarheid van zowel de samengestelde vier-vaardigheden-toets als die van de afzonderlijke onderdelen daarin, alsook van elk van de m.c. cloze-toetsen alleszins acceptabel is te achten. De betrouwbaarheid van de scores is bij de twee lange m.c. cloze-toetsen ($k=100$) hoger dan bij de overige toetsen. Ten slotte wijzen de bevindingen erop dat een m.c. cloze-toets een valide instrument kan zijn voor het meten van taalvaardigheid. In het navolgende zullen we de gegevens die de validiteit van de gehanteerde m.c. cloze-toetsen aannemelijk maken, nader beschouwen.

Ten eerste is uit het onderzoek gebleken dat tussen de scores op de m.c. cloze-toetsen enerzijds en anderzijds de scores voor luisteren, spreken, lezen, schrijven en tekstbegrip aanzienlijke, significante correlaties bestaan. Dit wijst erop dat een m.c. cloze-toets voor een belangrijk deel hetzelfde meet als elk van de vijf genoemde onderdelen. Nu zou men kunnen opmerken dat aan deze gegevens niet veel betekenis kan worden gehecht, omdat de in het onderzoek betrokken personen een selecte groep vormen. Naar verwachting beschikken deze cursisten over een meer dan gemiddelde intelligentie, hetgeen vermoedelijk hun prestaties positief beïnvloed heeft. Buitenlanders worden namelijk pas tot de cursus toegelaten, nadat door de T.U. Delft is vastgesteld dat de genoten vooropleiding gelijkwaardig is aan het Nederlandse V.W.O. Naar onze mening is dit geen werkelijk bezwaar. Integendeel zelfs! Indien de groep relatief homogeen is ten aanzien van intelligentie, moet het waarschijnlijk worden geacht dat niet of nauwelijks verschillen in intelligentie een rol hebben gespeeld, maar wel en in meerdere mate uitsluitend verschillen in taalvaardigheid. Met andere woorden: het feit dat de correlaties verkregen zijn over een qua intelligentie betrekkelijk homogene groep personen, vergroot juist de betekenis van de gevonden coëfficiënten (cf. Drenth, 1984, p. 260). Dit neemt overigens niet weg dat onderzoek bij andere groepen van buitenlanders gewenst is. Daarnaast kan men zich afvragen of metingen door middel van een m.c. cloze-toets geen overwaardering van het passief taalgebruik ten opzichte van het actieve impliceren. Een m.c. cloze-toets vraagt immers geen taalproductie, doch uitsluitend het aankruisen van dat woord uit drie gegeven antwoordmogelijkheden, dat 'in de tekst past' (cf. Porter, 1976, p. 153). De resultaten wijzen echter geenszins in deze richting. De correlaties van de scores op de m.c. cloze-toetsen met die voor spreken en schrijven wijken niet systematisch af van de correlaties van de scores op de m.c. cloze-toetsen met die voor luisteren, lezen en tekstbegrip (zie Tabel 3 en Tabel 4).

Ten tweede is gebleken dat er hoge correlaties bestaan tussen de scores op elk van de m.c. cloze-toetsen enerzijds en anderzijds de totaalscore op de vier-vaardigheden-toets, uitgedrukt in de zogenoemde E-score. Na correctie voor attenuatie bereiken de correlatiecoëfficiënten waarden van 0.91 en 0.94. Deze gegevens wijzen erop dat met name de lange m.c. cloze-toetsen elk in zeer hoge mate dezelfde eigenschap meten als de vier-vaardigheden-toets, opgevat als een samengestelde meting. Met andere woorden: als de vier-vaardigheden-toets taalvaardigheid meet, mag worden aangenomen dat elk van de m.c. cloze-toetsen ook in zeer hoge mate taalvaardigheid meet.

Ten derde zijn er aanwijzingen dat er een structurele overeenkomst bestaat tussen het verloop van scores op een m.c. cloze-toets in de cursus en het verloop van het dekkingspercentage in het proces van taalverwerving. Op grond hiervan is onze hypothese dat een m.c. cloze-toets dekking meet. Dat wil zeggen: de score op een m.c. cloze-toets geeft een indicatie van het dekkingspercentage dat de betreffende persoon heeft bereikt voor het type tekst waaruit de toets is afgeleid. Hiermee is een aanzet gegeven tot begripsvalidering van m.c. cloze-toetsen. Onduidelijk is echter nog hoe de gebleken samenhang tussen scores op een m.c. cloze-toets, een vier-vaardigheden-toets en het dekkingspercentage verklaard moet worden, zodat nader onderzoek gewenst is.

Op grond van bovenstaande bevindingen is onze conclusie dat de gehanteerde m.c. cloze-toetsen kwalitatief goede meetinstrumenten zijn voor het meten van algemene taalvaardigheid. Gelet op het feit dat de efficiëntie van een m.c. cloze-toets circa 20 maal groter is dan van de vier-vaardigheden-toets, verdient het derhalve aanbeveling bij ingangs- en eindmetingen de vier-vaardigheden-toets te vervangen door een m.c. cloze-toets van voldoende lengte ($k=100$) en gebleken kwaliteit. Ter voorkoming van mogelijke misverstanden: hiermee is dus niet gezegd dat alle luister-, spreek-, lees- en schrijf oefeningen in het onderwijs door m.c. cloze-oefeningen vervangen dienen te worden. Wij hebben het hier slechts over de m.c. cloze-toets als meetinstrument. Evenmin is hiermee gezegd dat in specifieke situaties het gebruik van andere, aanvullende meetinstrumenten niet zinvol zou kunnen zijn.

NOTEN

1. Hoewel Porter (1976) suggereert de 'uitvinder' van de multiple-choice cloze-procedure te zijn, is ons minstens één publicatie van oudere datum bekend waarin deze procedure wordt gehanteerd (Cranney, 1973).
2. Deze personen waren hoofdzakelijk afkomstig uit Iran, Indonesië, Marokko en Turkije en uitsluitend bij dit onderzoek betrokken.
3. Deze personen die overigens alleen aan onderzoek II hebben deelgenomen, waren hoofdzakelijk afkomstig uit Indonesië, Noorwegen en Turkije.
4. Deze personen die uitsluitend aan onderzoek III hebben deelgenomen, waren hoofdzakelijk afkomstig uit Syrië, Indonesië, Iran, Marokko en Turkije.

LITERATUUR

- Beheydt, L. (1985). Toetsconstructie bij het 'Certificaat Nederlands'. In *Toegepaste Taalkunde in artikelen* 22 (p. 129-147). Amsterdam: VU boekhandel/uitgeverij.
- Campbell, D.T. & Stanley, J.C. (1971). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.) *Handbook of research on teaching* (p. 171-246). Chicago: Rand McNally.
- Cranney, A. Garr. (1973). The construction of two types of cloze reading tests for college students. *Journal of Reading Behavior*, 5, 60-64.
- Drenth, P.J.D. (1984). *Inleiding in de testtheorie*. Deventer: Van Loghum Slaterus.
- Hinofotis, F.B. & Snow, B.G. (1980). An alternative cloze testing procedure: multiple-choice format. In J.W. Oller & K. Perkins (Eds.), *Research in language testing* (p. 129-133). Rowley, MA: Newbury.
- Jochems, W. & Montens, F. (1986). *Hoe toets je taalvaardigheid?* Meppel: Boom.
- Jonz, J. (1976). Improving on the basic egg: the M.C. cloze. *Language Learning*, 26, 255-265.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Luijten, A.J.M. et al. (1984). *Eindexamens LBO/MAVO 1984*. Arnhem: Centraal Instituut voor Toetsontwikkeling.
- Montens, F. & Sciarone, A.G. (1984). *Nederlands voor buitenlanders - de Delftse methode*. Meppel: Boom.
- Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill (sec. ed.).
- Oller, J.W. (1979). *Language tests at school, a pragmatic approach*. London: Longman.
- Oller, J.W. (1982). 'g', what is it? Albuquerque: paper University of New Mexico.
- Porter, D. (1976). Modified cloze procedure: a more valid reading comprehension test. *English Language Teaching Journal*, 30, 151-155.
- Sciarone, A.G. (1979). *Woordjes leren in het vreemde-talenonderwijs*. Muiderberg: Coutinho.
- Sciarone, A.G. & Montens, F. (1984). *Hoe leer je een taal? - de Delftse methode*. Meppel: Boom.
- Sciarone, A.G. & Montens, F. (1985). Nederlands voor buitenlanders: de Delftse methode. In *Toegepaste Taalkunde in artikelen* 22 (p. 156-165). Amsterdam: VU boekhandel/uitgeverij.
- Wijnstra, J.M. (1977). Het gebruik van de cloze procedure als maat voor schriftelijke taalbeheersing. *Tijdschrift voor Onderwijsresearch*, 6, 262-269.

Manuscript ontvangen 8-10-1986

Definitieve versie ontvangen 3-3-1987

De Betekenis van Leren: een exploratief onderzoek naar Leerconcepties met de Semantische Differentiaal

M.G.E. Pierar

Tilburgs Instituut voor Academische Studies, Stationsstraat 19, 5038 EA Tilburg.

ABSTRACT

The experiment investigated the amount in which quantitative and qualitative methods for exploring learning-conceptions obtain similar or different results. Subjects first answered an essay-question about learning, the qualitative method. Subjects then filled in a Semantic Differential questionnaire, the quantitative method. In this questionnaire objects functioned as the operationalisations of learning-conceptions. Data was analyzed with the so-called three-mode principal component analysis. The two methods obtain similar results for the objects, and different results for the subjects. The results did not show Osgood's Evaluation-Potency-Activity structure of Meaning. The most important meaning of learning for subjects seems to be an evaluative kind of meaning.

INLEIDING

De laatste tien jaar is er een toenemende belangstelling voor de wijze waarop de leerling of student aankijkt tegen het verschijnsel Leren. Onderzoek naar die wijze-van-kijken suggereert, ten eerste, dat verschillende leerlingen of studenten verschillende visies hebben over Leren, ook wel Leerconcepties genoemd, en, ten tweede, dat dit verschil in visie van invloed is op de gebruikte leerstrategie en op de leerresultaten. Zie bijv. Van Rossum en Schenk, 1983, Marton en Svenson, 1979, en Saljö, 1979. De empirische ondersteuning van deze suggesties staat ter discussie, waarschijnlijk vanwege het kwalitatieve karakter van de onderzoeken. Het onderzoek waarover in dit artikel wordt gerapporteerd was opgezet om na te gaan in hoeverre de eerste suggestie, dat er verschillende visies zouden bestaan, gerepliceerd kon worden met behulp van een kwantitatieve methode.

De essentie van onderzoek naar verschillende visies op Leren kan worden getypeerd met de stelling dat er, met betrekking tot het verschijnsel Leren, een spraakverwarring zou heersen in (en buiten) het klaslokaal. Veronderstel dat een docent de tentamenstof opgeeft aan zijn leerlingen door te zeggen: "Jullie moeten hoofdstuk 1 en 4 leren voor het tentamen". Deze uitspraak is, op zich, voor alle leerlingen eender. Volgens het onderzoek naar de verschillende visies op Leren kan de uitspraak van de docent voor elke leerling anders overkomen. Zo kan bijvoorbeeld leerling A horen dat hij/zij hoofdstuk 1 en 4 moet kunnen onthouden terwijl leerling B hoort dat hoofdstuk 1 en 4 begrepen moeten worden. Voor leerling A heeft Leren de betekenis van 'Leren = de stof onthouden' terwijl voor leerling B Leren de betekenis heeft van 'Leren = de stof begrijpen'. De oorzaak voor de verschillende interpretaties van deze twee leerlingen van deze ene uitspraak kan men zoeken in motivationele factoren, in cognitieve processen, e.d., om zo proberen te verklaren waarom de twee leerlingen een verschillende interpretatie geven. Wat die oorzaak ook mag zijn, het woord 'leren' heeft in ieder geval een ambig karakter.

Onderzoek naar de verschillende wijzen waarop mensen tegen Leren aankijken, stelt dus eigenlijk dat mensen verschillende 'talen' spreken (en horen) ook al gebruiken ze dezelfde woorden. Dit soort onderzoek heeft daardoor (minimaal) twee verplichtingen. Ten eerste, het in kaart brengen van de talen die mensen spreken als ze het hebben over het verschijnsel Leren,

niet alleen wat betreft het typische karakter van elke taal afzonderlijk maar ook wat betreft de typische verschillen tussen de talen. Dat wil zeggen, wát zijn dan de verschillende betekenissen die mensen geven aan Leren en hóe verschillen ze? Ten tweede, aan kunnen geven dat er maar een beperkt aantal van die talen bestaat en hoeveel (vanwege Occam's scheermes). Dat wil zeggen, wát is de uitputtende set van betekenissen die al die verschillende talen omvat?

Leerconcepties

Onderzoek naar Leerconcepties probeert te voldoen aan de eerste doelstelling door, voor het merendeel, kwalitatieve methoden te gebruiken. Meestal worden open-vragen of interviews gebruikt (in de trant van 'Wat versta jij onder leren...?') om aan de hand van de antwoorden te achterhalen wat de Leerconceptie is van de desbetreffende persoon. Een belangrijke assumptie van het onderzoek naar Leerconcepties is dat de antwoorden van iemand met een bepaalde Leerconceptie van een geheel andere aard zijn dan die van iemand met een andere Leerconceptie. De onderzoeksgegevens worden ook zo gepubliceerd. Men citeert uit antwoorden, of gedeelten daaruit, van personen die een bepaalde Leerconceptie zouden hebben en citeert daarna uit antwoorden, of gedeelten daaruit, van personen die een andere Leerconceptie zouden hebben (zie bijvoorbeeld Van Rossum en Schenk, 1983). Het wordt dus eigenlijk ook aan de lezer van zo'n publicatie overgelaten om te concluderen of die antwoorden wel zo verschillen, en of het typische antwoorden zijn.

Zoals gezegd, hield een Leerconceptie in de visie op Leren van een bepaalde leerling of student. Een Leerconceptie is te beschouwen als de persoonlijke betekenis van Leren van de desbetreffende persoon. Strikt genomen is een Leerconceptie een conceptualisatie of rationalisatie over het verschijnsel Leren, met andere woorden een Leerconceptie is dan een persoonlijke theorie over Leren. Echter, in een ongepubliceerd paper (Pierar, 1986) hebben we beargumenteerd dat aan de basis van elke theorie, wetenschappelijk of persoonlijk, een set van assumpties ligt die uiteindelijk niet 'bewijsbaar' zijn (logisch en/of empirisch) maar worden gekozen op grond van wel/niet betekenisvol ('meaningful') zijn. In die zin opgevat is een theorie de (logische en/of empirische) uitdrukking of uitwerking van een bepaalde betekenis die 'aan de dingen' wordt gegeven (zie ook Blocker, 1974, en Martin, 1982). Met andere woorden, hoewel een Leerconceptie strikt genomen wel een conceptualisatie of rationalisatie is, is het tegelijkertijd de uitdrukking of uitwerking van een bepaalde betekenis die gegeven wordt aan Leren, en die aan de basis ligt van die Leerconceptie. De Leerconceptie als rationalisatie is doortrokken van de betekenis die er aan ten grondslag ligt, en de vraag is dan welke betekenis (sen).

In de onderzoeksliteratuur met betrekking tot Leerconcepties wordt een onderscheid gemaakt tussen hoofdzakelijk twee Leerconcepties die worden aangeduid met de termen Surface- en Deep-leerconceptie. Eigenlijk wordt in de literatuur niet over Surface- en Deep-leerconcepties gesproken maar over Surface- en Deep-intenties of 'approaches'. Intenties of 'approaches' verwijzen naar de rol en invloed van betekenis, 'meaning', (zie bijvoorbeeld Marton en Svensson, 1979). Aangezien we zojuist aangaven dat wij een Leerconceptie zagen als de persoonlijke betekenis van Leren, veroorloven we ons de vrijheid om (toch) over Surface- en Deep-leerconcepties te spreken.

Iemand met een Surface-leerconceptie geeft aan Leren de betekenis van 'Leren = reproduceren van kennis', met als voornaamste leeractiviteit memoriseren. Zo'n iemand is bijvoorbeeld leerling A hierboven. Iemand met een Deep-leerconceptie geeft aan Leren de betekenis van 'Leren = construeren van kennis', met als voornaamste leeractiviteit (proberen te) begrijpen. Leerling B hierboven is hiervan een voorbeeld. We kunnen daarom zeggen dat er volgens de onderzoeksliteratuur twee verschillende talen bestaan om over Leren te praten, een Surface-taal en een Deep-taal. Elke Leerconceptie zou een eigen taal hebben, een eigen jargon. Typische voorbeelden van elke Leerconceptie zijn te vinden in Van Rossum, Deijkers en Hamer, 1985.

Onderzoek naar Leerconcepties suggereert dat mensen met betrekking tot Leren verschillende talen spreken, ze hebben verschillende Leerconcepties. We kunnen dat vertalen naar de

stelling dat woorden zoals Leren, of woorden die verwijzen naar een bepaald aspect van Leren zoals Onthouden of Begrijpen, verschillende betekenissen hebben voor verschillende mensen. We mogen daarom zeggen dat er individuele verschillen (zouden) bestaan wat betreft de betekenis van woorden zoals Leren, Onthouden en Begrijpen.

Zoals gezegd gebruikt onderzoek naar Leerconcepties kwalitatieve methoden om geloofwaardigheid te geven aan de gepostuleerde verschillen tussen die Leerconcepties, aan de verschillen in betekenis van het woord 'Leren' en daaraan gerelateerde woorden. We zullen een poging doen om die verschillende betekenissen eens op een andere manier te onderzoeken, en wel op een kwantitatieve manier met behulp van de bekende Semantische Differentiaal van Osgood. Het onderzoek waarover hier wordt gerapporteerd, is het eerste in een serie onderzoeken binnen het kader van het project 'De Betekenis van Leren'. Eén van de doelstellingen van dit project is het ontwerpen van een onderzoeksinstrument voor Leerprocessen in casu Leerconcepties, in de vorm van een Semantische Differentiaal. Immers, zoals het adjectief 'Semantische' al aangeeft is deze techniek juist gericht op het onderzoeken van betekenis. Door gebruik te maken van deze techniek ontstond de mogelijkheid om een eerste, theoretische ankerpunt te vinden, met wat betreft de zogenaamde E-P-A structuur van betekenis. De voornaamste aanleiding voor gebruikmaking van de Semantische Differentiaal binnen genoemd project was ook die E-P-A structuur van betekenis. De veronderstelling was dat indien Leerconcepties, met name de woorden die binnen Leerconcepties worden gebruikt, theoretisch gezien zo verschillend zijn, zij ook van elkaar (moeten) afwijken wat betreft die E-P-A structuur. Kort gezegd, indien Leerconcepties van elkaar verschillen, dan moet ook de betekenisstructuur van die Leerconcepties verschillen.

De Semantische Differentiaal

Met de Semantische Differentiaal kan de betekenis worden onderzocht van de meest uiteenlopende objecten. De betekenis van de objecten wordt onderzocht met behulp van een serie van bipolaire begrippen die een bepaalde betekenis tegenstelling uitdrukken, zoals bijvoorbeeld hard-zacht of actief-passief. Deze tegenstelling wordt gegoten in de vorm van een 7-puntsschaal, en we zullen daarom verder spreken over een betekenischaal en niet over bipolaire betekenisbegrippen. De score van een persoon op de 7-puntsschaal geeft de betekenis weer die een bepaald object heeft voor de desbetreffende persoon.

De datamatrix heeft de vorm van een blok, elk element van de matrix is de score van één persoon op één schaal voor één object. Gewoonlijk wordt deze datamatrix omgerekend naar een correlatie-matrix van de schalen waarop men een principale componenten analyse uitvoert. Het resultaat van deze analyse is een reductie van p schalen tot z componenten (met z kleiner dan p). Deze componenten zijn te beschouwen als algemene schalen die onderling onafhankelijk zijn. Elke afzonderlijke schaal is te herleiden tot één van die algemene schalen of tot een combinatie daarvan.

Eén van de belangrijkste kenmerken van de theorievorming omtrent de Semantische Differentiaal is de zogenaamde E-P-A structuur van betekenis. Betekenis zou drie-dimensionaal van aard zijn, de betekenisruimte wordt opgespannen door de drie dimensies Evaluatie (E), Potentie (P) en Activiteit (A) (Osgood, Suci en Tannenbaum, 1957). Betekenis tegenstellingen zoals bijvoorbeeld actief-passief en hard-zacht worden beschouwd als vectoren in die betekenisruimte. Deze vectoren zijn een bepaalde combinatie van drie onafhankelijke vectoren, de drie dimensies Evaluatie, Potentie en Activiteit. Objecten ontleen hun betekenis aan hun locatie in die betekenisruimte.

Een kenmerk dat met het vorige samenhangt is het universele karakter van de E-P-A structuur, deze structuur is cross-cultureel (Osgood, May en Miron, 1975). Eigenlijk wordt dus gezegd dat er, wat betreft de E-P-A structuur, zowel binnen één cultuur als tussen culturen géén individuele verschillen bestaan tussen mensen. Alle mensen verlenen betekenis aan 'de dingen' middels die E-P-A structuur. Een implicatie daarvan is dat het resultaat van een principale componenten analyse over n personen eigenlijk een reductie moet opleveren tot één component, tot één algemene persoon. Andersom geredeneerd, als er wel individuele verschillen bestaan

tussen mensen, dan moet de data-analyse meerdere componenten, meerdere algemene personen, opleveren. Snyder en Wiggins (1970) halen een aantal studies aan waaruit zou blijken dat er wel individuele verschillen bestaan tussen mensen wat betreft die E-P-A structuur. Vervolgens geven zij drie interpretaties van individuele verschillen met betrekking tot de E-P-A structuur. Ten eerste, de drie dimensies zijn niet algemeen van aard. Ten tweede, de drie dimensies zijn wel algemeen van aard maar de betekenisschalen hebben voor verschillende mensen een verschillende betekenis. Ten derde, de E-P-A structuur wordt door verschillende mensen verschillend gebruikt. Het is duidelijk dat de eerste interpretatie het meest in strijd is met de claim dat de E-P-A structuur universeel is, en de andere twee minder.

Het bestaan van individuele verschillen zou betekenen dat er een interactie bestaat tussen personen en schalen (en objecten). Eén van de grootste maar ook onderkende problemen van het onderzoek met de Semantische Differentiaal is die interactie tussen schalen, objecten en personen (zie bijvoorbeeld Osgood, Suci en Tannenbaum, 1957, Osgood, May en Miron, 1975, Heise, 1969, en Presley, 1969). Dit kan een methodologische reden hebben. Zoals gezegd wordt in het onderzoek meestal een correlatie-matrix van de schalen berekend, over de objecten en de personen, welke vervolgens wordt geanalyseerd. Aangezien een correlatie berekend wordt, wordt heel wat oorspronkelijke variantie geëlimineerd nog voordat de data-analyse wordt uitgevoerd. Daarbij kan men zich ook afvragen waarom geen principale componenten analyse wordt uitgevoerd op de personen en de objecten. Immers, principale componenten analyse is een 'blinde' analyse, het maakt voor de berekeningswijze van de componenten niet uit of we deze uitvoeren op de schalen of op de personen of objecten zolang we maar aan de beginvoorwaarden (een vierkante en symmetrische inputmatrix) voldoen. Anders gezegd, methodologisch bekeken mogen we het data-blok van een Semantische Differentiaal onderzoek analyseren vanuit alle drie de zijden van dat blok, vanuit de zijde (de modus) van de personen, van de objecten en van de schalen. Maar omdat we dan hetzelfde data-blok drie maal analyseren rijst de vraag wat de relatie is tussen die drie analyses. Het model van Tucker, het zogenaamde 'Three-mode principal component analysis model' (zie Tucker, 1966, Levin 1965, Kroonenberg, 1983), is een elegante manier om zowel het interactieprobleem als het welke-zijde-analyseren probleem op te lossen. Het geeft ook een manier aan om de relatie tussen de drie analyses te bepalen. De berekeningswijze van de componenten volgens dit model komt neer op de volgende.

Eerst berekent men de kruisproduktmatrix voor elke modus, deze matrix is vierkant en symmetrisch. Vervolgens voert men een principale componenten analyse uit op elke afzonderlijke kruisproduktmatrix, met als resultaat drie verschillende oplossingen. Deze drie analyses leveren een reductie op van n personen naar x persoonscomponenten voor de modus van de personen, van m objecten naar y objectcomponenten voor de objectmodus en van p schalen naar z schaalcomponenten voor de schaalmodus. Vervolgens berekent men met de eigenvectoren, die bij de respectievelijke componenten behoren, en de originele data-matrix de zogenaamde kernmatrix die ook de vorm van een blok heeft. Echter, deze kernmatrix heeft niet $n \cdot m \cdot p$ elementen zoals de oorspronkelijke data-matrix maar $x \cdot y \cdot z$ elementen. Elk element van de kernmatrix geeft de mate aan waarin één persoonscomponent, één objectcomponent en één schaalcomponent samenhangen, of niet. De elementen van die kernmatrix zijn te beschouwen als de mate van interactie tussen de modes, het interactieprobleem is dus in zoverre opgelost dat die interacties 'uitgerekend' zijn. De vraag of een bepaalde schaal interacteert met een bepaald onderwerp en een bepaalde persoon en in welke mate, is dan indirect te beantwoorden door na te gaan op welke componenten bijvoorbeeld een schaal en een persoon een lading hebben en vervolgens te kijken naar de ladingen van de respectievelijke componenten in de kernmatrix. Wanneer bijvoorbeeld een persoon en een schaal ieder op een component hoge ladingen hebben en het kernmatricelement dat deze twee componenten aan elkaar relateert heeft ook een hoge lading, dan mag men veronderstellen dat er een sterke interactie bestaat tussen de persoon en de schaal in kwestie.

De operationalisering van Leerconcepties

Volgens de onderzoeksliteratuur naar Leerconcepties zouden er (minimaal) twee verschillende talen zijn om over het fenomeen Leren te praten, een Surface-taal en een Deep-taal. Deze twee talen gebruiken wel dezelfde woorden maar ze bedoelen er verschillende dingen mee. Zo zou binnen de Surface-taal het woord Leren de betekenis hebben van 'Leren = reproductie van kennis' of van 'Leren = onthouden'. Binnen de Deep-taal zou dat zijn 'Leren = constructie van kennis' of 'Leren = begrijpen'. Binnen deze twee Leerconcepties wordt door sommigen (bijv. Saljö, 1979, en Van Rossum, Deijkers en Hamer, 1985) een verder onderscheid gemaakt tussen drie verschillende Surface-leerconcepties en twee verschillende Deep-leerconcepties. Deze vijf hebben ieder een eigen definitie van het verschijnsel leren, te weten:

Surface-leerconcepties

- 1) Leren = kennistoename
- 2) Leren = onthouden
- 3) Leren = toepassen

Deep-leerconcepties

- 4) Leren = inzicht
- 5) Leren = begrijpen v/d werkelijkheid

Er zouden dus in totaal 5 verschillende talen zijn om over het verschijnsel Leren te praten, elke taal heeft één of meerdere kernbegrippen waarmee Leren wordt getypeerd. Bij Leerconceptie 2 is Onthouden het kernbegrip, bij Leerconceptie 4 is dat Inzicht. Nu is het niet ondenkbaar dat de kernbegrippen ook bij andere Leerconcepties gebruikt worden maar ze zijn dan niet meer een kernbegrip. Iemand met Leerconceptie 3 of 4 zal wel gebruik maken van het kernbegrip Onthouden van Leerconceptie 2 maar dit woord neemt een ondergeschikte positie in ten opzichte van de kernbegrippen Toepassen en Inzicht. Met behulp van kernbegrippen kunnen we Leerconcepties operationaliseren voor een Semantische Differentiaal, deze kernbegrippen fungeren als de objecten waarvan de betekenis wordt onderzocht.

Vraagstellingen

Aangezien het een exploratief onderzoek betrof, werden niet hypothesen geformuleerd maar richtlijnen om een zinvolle interpretatie van de resultaten mogelijk te maken. Deze richtlijnen werden geformuleerd in de vorm van een vraag. Doelstelling van het onderzoek was na te gaan welke kernbegrippen en welke schalen bruikbaar zijn om met behulp van de Semantische Differentiaal Leerconcepties te onderzoeken. Naast de Semantische Differentiaal werd ook gebruik gemaakt van een essay-vraag over Leren, dit om inzicht te krijgen in de relatie tussen deze twee onderzoeksbenaderingen van Leerconcepties.

De volgende vragen werden geformuleerd:

1. Komen de kernbegrippen uit de data-analyse naar voren in minimaal 2 en maximaal 5 verschillende groepen, waarbij elke groep van kernbegrippen typisch is voor een bepaalde Leerconceptie?
2. Vertonen de schaalcomponenten na de data-analyse de bekende E-P-A structuur?
3. Is er bij de onderzochten een zinvolle relatie te ontdekken tussen het resultaat van hun data-analyse en hun antwoord op de essay-vraag?
4. Welke verschillen bestaan er volgens de kernmatrix tussen (de componenten van) de onderzochten wat betreft (de componenten van) de kernbegrippen en schalen, d.w.z. hebben voor verschillende subjecten dezelfde kernbegrippen een andere betekenis, of niet?

METHODE

Subjecten

De vragenlijst werd ingevuld door 55 personen, 11 H.B.O. en 44 W.O. studenten. Van de 11 H.B.O. studenten waren er 7 mannelijk en 4 vrouwelijk, van de 44 W.O. studenten waren er 23 mannelijk en 21 vrouwelijk. Er deden dus 30 mannelijke en 25 vrouwelijke studenten mee.

Studenten waren verdeeld over alle studiejaar. Deelname aan het onderzoek werd beloond met f 7.50 per uur.

Vragenlijst

In totaal werden 21 objecten of kernbegrippen gebruikt. De kernbegrippen werden gehaald uit ruwe data van onderzoek van Van Rossum. Deze ruwe data waren de antwoorden van studenten op de essay-vraag 'Wat versta jij onder Leren?', bij deze antwoorden was aangegeven (door Van Rossum) wat de Leerconceptie van de desbetreffende student was. Bij elke Leerconceptie werd gezocht naar de meest kenmerkende woorden die door de desbetreffende personen werden gebruikt. Deze woorden werden samengevoegd in een lijst die verder werd aangevuld met een paar andere woorden waarvan de auteur het vermoeden had dat ze wel een kernbegrip waren maar in de eerder genoemde ruwe data niet goed naar voren kwamen. In Tabel 1 is aangegeven met getallen achter het kernbegrip tot welke specifieke Leerconceptie (1 t/m 5) het kernbegrip behoort indien duidelijk was tot welke Leerconceptie een kernbegrip behoorde, in totaal waren dit er 12. Met S (= Surface) en D (= Deep) wordt aangegeven tot welke meer globale Leerconceptie dit kernbegrip zou kunnen behoren, in totaal waren dit er 7. Verder werden aan de lijst de woorden Leren en Studeren toegevoegd, dit om te zien bij welke kernbegrippen deze twee terecht zouden komen.

Er werden 35 schalen gebruikt, 15 uit de onderzoeksliteratuur over de Semantische Differentiaal (met name Kouwer, 1958, en Jansen en Smolenaars, 1966), en 20 zelf-geconstrueerde schalen. Deze laatste schalen werden gemaakt op grond van het voorkomen van één van de schaaladjectieven, of beide schaaladjectieven, in de ruwe data van het onderzoek van Van Rossum. De schalen liepen van -3 tot +3. In Tabel 2 is alleen achter de schalen uit de literatuur aangegeven met E (= evaluatie), p (= potentie) en A (= activiteit) tot welke dimensie een schaal behoort.

De 35 schalen werden bij elk kernbegrip aangeboden zodat de vragenlijst $21 \times 35 = 735$ score-momenten bevat. Bij elk kernbegrip moest eerst op alle 35 schalen een score worden gegeven voordat het volgende kernbegrip werd aangeboden. De volgorde waarin de kernbegrippen werden aangeboden was a-select wat betreft hun theoretische volgorde. De volgorde waarin de schalen werden aangeboden was gerandomiseerd voor elk afzonderlijk kernbegrip. Verder werd bij elk kernbegrip ook de presentatie van alle 35 schalen gerandomiseerd, d.w.z. Linkerwoord-Rechterwoord of Rechterwoord-Linkerwoord. De uiteindelijke volgorde waarin de kernbegrippen en schalen werden aangeboden was voor elke onderzochte dezelfde.

Apparatuur

De vragenlijst had niet de normale 'papier en potlood' vorm maar werd afgenomen via het scherm en toetsenbord van een microcomputer, de Philips P2500. Ook de data-analyse werd hierop uitgevoerd. De besturingsprogramma's voor de P2500 voor randomisatie en afname van de vragenlijst en voor de opslag en verwerking van de data werden in BASIC geschreven door de auteur. De dataverwerkingsprogramma's werden geschreven en 'debugged' aan de hand van Harman (1976), voor principale componenten analyse en varimaxrotatie, en aan de hand van Tucker (1966) en Kroonenberg (1983), voor het berekenen van kruisproductmatrices en kernmatrix.

Procedure

De onderzochte kreeg eerst een aantal beeldpagina's met informatie te lezen. Pagina 1 gaf algemene informatie over het kader van het onderzoek. De computer bleef op het einde van deze pagina ± 10 minuten 'hangen' voordat de onderzochte naar pagina 2 kon. Op het einde van pagina 1 werd de onderzochte namelijk verzocht een essay-vraag over Leren te beantwoorden op een vel papier. De antwoorden op de essay-vragen werden beoordeeld op Leerconceptie door Van Rossum, onafhankelijk van de resultaten van de vragenlijst welke door de auteur werden geanalyseerd en geïnterpreteerd. Op deze, ruwe wijze kon met de ene onderzoeksmethode worden nagegaan wat de betrouwbaarheid was van de andere methode.

Pagina 2 verwees naar de volgende pagina's. Pagina 3 en 4 bevatten de lijsten met kernbegrippen respectievelijk schalen. Pagina 5 bevatte instructies over de wijze van invullen van de vragenlijst, pagina 6 t/m 8 bevatte instructies over de bediening van de computer. Pagina 9 verwees naar de meteen daarna volgende oefening. Deze laatste was bedoeld om de onderzochte te oefenen in het invullen en in het bedienen van de computer.

Aangezien een computerscherm maar 25 regels heeft en er 35 schalen werden gebruikt, werden de schalen bij de vragenlijstafname in drie gedeelten op het scherm afgebeeld, $12+12+11=35$ schalen. De volgorde $12+12+11$ was voor elk kernbegrip dezelfde, de combinaties $12+11+12$ en $11+12+12$ werden niet gebruikt. Figuur 1 geeft de lay-out van de vragenlijst op het scherm.

TOEPASSEN					
BETEKENT VOOR MIJ					
-3	-2	-1	0	+1	+2 +3
		OPPERVLAKKIG	X	DIEP	
		GESLOTEN		OPEN	
		MOEILIJK		MAKKELIJK	
		VLUG		LANGZAAM	
		ENERGIEK		RUSTIG	
		PRETTIG		NAAR	
		GEBONDEN		VRIJ	
		FLEXIBEL		RIGIDE	
		TIJDELIJK		LANGDURIG	
		IDEAAL		REEEL	
		HOOG		LAAG	
		VERANDERLIJK		CONSTANT	
Hulp nodig? Druk op de H-toets					
Omhoog? Druk op de ESC-toets					
Omlaag? Druk op de RETURN-toets					

Figuur 1: De lay-out van de vragenlijst op het scherm. De x tussen oppervlakkig en diep geeft de plaats aan waar de scores op de schalen door de computer worden afgebeeld.

De onderzochte kon correcties aanbrengen in zijn/haar scores maar enkel bij de 12 (of 11) afgebeelde schalen. Schalen die van het scherm waren verdwenen konden niet meer gecorrigeerd worden. Op elk moment van de vragenlijstafname kon de onderzochte de computer om hulp vragen met de H-toets. In dat geval werden de pagina's 5 t/m 8 opnieuw op het scherm afgebeeld. Na het beëindigen van de Hulproutine bracht de computer de onderzochte terug naar waar hij/zij was gebleven voordat om hulp werd gevraagd, en kon verder worden gegaan met invullen. Na de laatste schaal bij het 11de kernbegrip kreeg de onderzochte de keuze om een korte pauze te nemen of om door te gaan met invullen.

Data-analyse

Over elke modus werd een kruisproductmatrix berekend. Elke matrix werd zodanig geschaald dat de som van de elementen op de diagonaal gelijk was aan het aantal variabelen, aan 55 (onderzochten), 21 (kernbegrippen) of 35 (schalen). Vervolgens werd op elke matrix een principale componenten analyse uitgevoerd, alle componenten met een eigenwaarde groter dan 1 werden behouden. Statistische toetsen voor het aantal componenten zijn voor kruisproductmatrices niet voorhanden (Kroonenberg, persoonlijke mededeling).

Met de eigenvectoren van de drie modes en de data-matrix werd eerst een kernmatrix berekend. Vervolgens werd voor elke modus een varimax-rotatie uitgevoerd op de desbetreffende componenten. Deze transformatie voor elke modus dient te worden verwerkt in de kernmatrix (zie Tucker, 1966, en Snyder en Wiggins, 1970) zodat het uiteindelijke resultaat een geroteerde kernmatrix is. Deze kernmatrix werd zodanig geschaald dat de som van alle gekwadeerde kernmatrixelementen in het maximale (en ideale) geval gelijk zou zijn aan $(55 \text{ onderzochten} * 21 \text{ kernbegrippen} * 35 \text{ schalen}) = 40425$.

RESULTATEN

Kernbegrippen

Tabel 1 geeft het resultaat van de analyse voor de kernbegrippen, een twee componentenoplossing, na varimax-rotatie. Alle kernbegrippen die behoren tot de Deep-leerconcepties (D, 4 en 5) zijn op de eerste component terecht gekomen. Ook Leren en Studeren zijn op deze component terecht gekomen hoewel Studeren ook een betrekkelijk hoge lading heeft op de tweede component. Deze tweede component bevat enkel kernbegrippen die behoren tot de Surface-leerconcepties (S, 1, 2 en 3) en het kernbegrip Studeren. Echter, Toepassen, Nut en Onthouden laden ook op de eerste component. Aangezien de kernbegrippen van de Deep-leerconcepties

Table 1: Factor-matrix van 21 kernbegrippen na Varimax-rotatie

Kernbegrippen	(LC) ^a	Componenten ^b	
		I	II
Inzicht	(4)	.96	
Creativiteit	(5)	.92	
Zelf Uitzoeken	(D)	.91	
Begrijpen	(4)	.9	
Oplossen van Problemen	(5)	.89	
Persoonlijke Betrokkenheid	(5)	.87	
Verbanden Leggen	(5)	.85	
Bezig zijn	(D)	.81	
Leren		.79	
Nieuwsgierigheid	(D)	.72	
Studeren		.68	.40
Tijd om na te denken	(D)	.64	
Toepassen	(3)	.63	.28
De Grote Lijn	(4)	.53	
Nut	(3)	.47	.42
Reproduceren	(2)		.7
Feitenkennis	(1)		.65
Studieverplichtingen	(S)		.59
Stof aangereikt krijgen	(S)		.51
Details	(S)		.5
Onthouden	(2)	.47	.47
Eigenwaarden na Varimax-rotatie		9.93	2.72
Eigenwaarden in %		45.87	12.95
Cumulatieve Eigenwaarden in %		45.87	58.83

a) Getallen en letters tussen haakjes geven de Leerconceptie aan waartoe een kernbegrip behoort. Getallen verwijzen naar specifieke Leerconcepties (1 t/m 5), letters naar globale leerconcepties (Surface of Deep). Leerconceptie 1 t/m 3 zijn Surface-leerconcepties, Leerconceptie 4 en 5 zijn Deep-leerconcepties.

b) Componentescores tussen $-.25$ en $.25$ zijn weggelaten.

alle op de eerste component terecht zijn gekomen, en die van de Surface-leerconcepties bijna allemaal op de tweede component, zullen we de eerste de Deep-component noemen en de tweede de Surface-component.

Table 2: Factor-matrix van 35 schalen na Varimax-rotatie

Schalen	(dim) ^a	Componenten ^b				
		I	II	III	IV	V
waardevol-waardeloos	(E)	1.11				
zinvol-zinloos		1.11				
positief-negatief	(E)	1.07				
goed-slecht	(E)	1.03				
prettig-naar	(E)	.97				
boeiend-saai	(A)	.93				
actief-passief	(A)	.89				
slim-dom	(E)	.88				
succesvol-falend	(E)	.88				
echt-schijn		.85				
geslaagd-mislukt	(E)	.85				
sterk-zwak	(P)	.83				
diep-oppervlakkig	(P)	.76		.25		
bewust-onbewust		.76				
flexibel-rigide		.63	.28	.4	— .29	
hoog-laag		.63				
open-gesloten		.61		.36	— .33	
groot-klein	(P)	.57				
bekend-onbekend		.54		— .39		
tijdelijk-langdurig		— .49	.46			
energiek-rustig		.46				
veranderlijk-constant	(A)	.26	.75			
wisselend-gelijkmatig	(A)	.31	.68			
voorspelbaar-willekeurig	(A)		— .46	— .33		
ideaal-reëel				.62		
vrij-gebonden		.44		.62		
los-vast			.28	.52		
objectief-subjectief		— .34	— .35		.6	
technisch-persoonlijk		— .57		— .25	.6	
absoluut-relatief			— .44		.46	
expliciet-impliciet					.45	
makkelijk-moeilijk						.72
zwaar-licht						— .64
hard-zacht						.55
Eigenwaarden na Varimax-rot.		15.47	2.27	2.09	1.79	1.47
Eigenwaarden in %		44.21	6.49	5.96	5.09	4.19
Cumulatieve Eigenwaarden in %		44.21	50.7	56.68	61.77	65.97

a) De letters tussen haakjes geven de dimensie aan waartoe een schaal behoort volgens de onderzoeksliteratuur. E = evaluatie, P = potentie en A = activiteit. Schalen waar geen letters achter staan zijn zelf-geconstrueerde schalen.

b) Componentscores tussen —.25 en .25 zijn weggelaten.

Table 3: Factor-matrix van de 55 onderzochten na Varimaxrotatie, 2 componenten

nr ^a	LC ^b	Componenten ^c	
		I	II
8	2	1.01	
2	2	.84	
3	4	.77	.55
35	3	.76	.44
49	2	.73	.46
20	1	.71	
43	4	.71	.32
39	4	.7	.4
10	3	.69	
44	3	.68	.45
42	3	.66	.25
24	1	.66	.28
32	1	.65	.31
45	3	.65	.51
34	3	.63	.28
29	3	.61	.46
13	3	.6	.45
48	3	.6	.33
9	3	.59	.42
38	3	.58	.3
30	5	.58	.35
54	4	.57	.42
33	2	.57	.3
31	3	.53	
37	3	.52	.37
52	4	.52	.27
4	3	.49	.48
27	3	.45	.43
46	3	.45	
40	3	.45	.39
11	2	.44	
14	2	.42	
7	3	.39	.29
18	3	.36	.34
51	4	.35	
6	2	.35	
22	2	.35	
53	2	.28	
16	5		1.08
55			.99
28	5	.51	.83
23	5	.32	.77
26	5	.38	.62
1	3	.56	.62
36	3		.61
25	4	.56	.6
17	3	.4	.56
41	3	.48	.54
5	5		.52
50	3	.41	.48
15	4	.37	.48
19	1	.44	.44
47	3	.35	.43
21	3	.31	.33
12	5	.29	.31
Eigenw na Varimax		16.02	11.16
Eigenw %		29.13	20.3
Cum. in %		29.13	49.42

a) Getallen verwijzen naar het nummer van de onderzochte.

b) Getallen geven Lerconcepties aan van onderzochten. 55 heeft geen essay-vraag beantwoord.

c) Componentescores tussen —.25 en .25 zijn weggelaten.

Schalen

Tabel 2 geeft het resultaat voor de schalen, een vijf componentenoplossing, na varimax-rotatie. De E-P-A structuur komt uit dit resultaat niet naar voren. Alle E- en P-schalen hebben hoge ladingen op de eerste component, de ladingen van de A-schalen liggen verspreid over de eerste drie componenten. De overige schalen liggen verspreid over alle vijf de componenten. Aangezien de vijf componenten zich niet eenvoudig laten benoemen, zullen we ernaar verwijzen met schaalcomponent I t/m V.

Subjecten

In eerste instantie was het resultaat voor de 55 onderzochten een 7 componentenoplossing. Daarom werd alleen op de eerste twee varimax uitgevoerd, met name om te zien of het resultaat een zinvolle interpretatie mogelijk maakte wat betreft de resultaten van het kwalitatieve onderzoeksgedeelte. Dat wil zeggen, of de ene component de groep van personen met een Deep-leerconceptie zou bevatten en de andere de groep met Surface-leerconceptie. Tabel 3 bevat het resultaat van de twee componentenoplossing na varimax-rotatie. Het merendeel van de onderzochten is verspreid over de twee componenten. Een persoon is dan, als variabele, te typeren als een combinatie van twee persoonscomponenten, van twee algemene personen.

Het probleem bij een componentmatrix van onderzochten is dat er geen directe manier bestaat om de persoonscomponenten te typeren aan de hand van de personen, zoals bij de objecten en de schalen. De kernmatrix is een indirecte manier om de persoonscomponenten te typeren, door middel van hun samenhang met de schaal- en de objectcomponenten.

Kernmatrix

Tabel 4 bevat de geroteerde kernmatrix voor 2 persoonscomponenten, 2 kernbegripcomponenten en 5 schaalcomponenten. De kruisproduktensom van deze matrix is 18288.5, ongeveer 45.24% van het totaal (40425). Dat wil zeggen, met deze kernmatrix is ongeveer 45% van de oorspronkelijk data te 'verklaren'.

Tabel 4: Geroteerde kernmatrix voor 2 persoonscomponenten

Persoons-component	Kernbegrip-component	Schaalcomponenten				
		I	II	III	IV	V
I	Deep	88.46	15.63	6.77	-19.75	10.12
	Surface	45.7	2.05	-9.55	-.75	9.21
II	Deep	76.37	14.84	14.72	-13.11	-8.11
	Surface	-2.94	-4.48	-26.49	10.19	-8.34

Uit de tabel valt af te leiden dat de eerste schaalcomponent de sterkste interactie heeft met de persoons- en de kernbegripcomponenten. De specifieke relatie tussen de drie moduscomponenten zouden we dan als volgt kunnen typeren. Persoonscomponent I brengt met schaalcomponent een kwantitatief (meer/minder) onderscheid aan tussen de twee kernbegripcomponenten, persoonscomponent II een kwalitatief (hoog/neutraal) onderscheid. Laten we op grond van Tabel 2 de eerste schaalcomponent, onder voorbehoud, een evaluatieve component noemen. Dit is in zoverre terecht dat alle E-schalen zeer hoge ladingen hebben op deze component. Persoonscomponent I is dan te typeren als een evaluatieve 'quantificeerder' van de twee kernbegrip groepen, en II een evaluatieve 'differentieerder'.

DISCUSSIE

Kernbegrippen

Eén van de vragen van dit exploratieve onderzoek had betrekking op de structuur van de kernbegrippen, op de mate waarin deze 'typisch' waren voor een bepaalde Leerconceptie. De vraag was of er minimaal 2 en maximaal 5 groepen van kernbegrippen zijn, waarbij elke groep kenmerkend is voor een bepaalde Leerconceptie. Uit de resultaten (zie tabel 1) kwamen twee groepen van kernbegrippen naar voren die we aangeduid hebben met een Deep en met een Surface groep.

Onthouden, een kernbegrip van Leerconceptie 2, laadt evenveel op de Deep- als op de Surface-component. Mogelijkerwijs raakt eerder Memoriseren in plaats van Onthouden de kenmerkende (en differentiërende) betekenis van Leerconceptie 2.

Ook Toepassen en Nut kwamen anders terecht dan verondersteld was. Een interpretatie kan zijn het mogelijk ambigue karakter van Toepassen. Volgens Van Rossum (persoonlijke mededeling) is Toepassen binnen Leerconcepties 3 'Toepassen later, in de praktijk', terwijl het binnen Leerconceptie 4 en 5 eerder is 'Toepassen nu, in de huidige leersituatie'. Het kernbegrip Toepassen differentieert dan niet tussen onderzochten en schalen, beter zou zijn geweest het gebruik van de kernbegrippen Toepassen Nu en Toepassen Later, of iets dergelijks.

Verder waren ook de resultaten van de kernbegrippen Leren en Studeren interessant. Leren was meer dan Studeren een Deep kernbegrip, Studeren had ook nog een Surface aspekt. De twee kernbegrippen zijn dus zeker niet synoniem. Als we de eerste, evaluatieve, schaalcomponent uit de kernmatrix hierbij betrekken dan lijken de onderzochten Leren veel positiever te waarderen dan Studeren. Interessant hieraan is dat personen met Leerconceptie 4 en 5 zich in hun antwoorden op essay-vragen met betrekking tot Leren vaker negatief uitlaten over Studeren, het is ondermeer schoolser, terwijl dit in de antwoorden van personen met Leerconceptie 1 t/m 3 vrijwel afwezig is (Van Rossum, persoonlijke mededeling).

We willen ook nog wijzen op het opmerkelijke resultaat dat Studieverplichtingen en Stof aangereikt krijgen, belangrijke kenmerken van de onderwijssituatie van studenten, tot de Surface groep van kernbegrippen behoren. Surface kernbegrippen werden door de eerste persoonscomponent minder sterk gewaardeerd dan de Deep kernbegrippen, door de tweede zelfs neutraal. Men zou daaruit kunnen concluderen dat personen hun onderwijssituatie niet zo positief beoordelen. Andere zaken zoals bijvoorbeeld Zelf uitzoeken en Tijd om na te denken vinden ze waardevoller.

Schalen

De richtlijn voor de schalen was dat ze de bekende E-P-A structuur zouden hebben. De resultaten waren op dat punt vrij duidelijk, de schalen hadden deze structuur niet (zie tabel 2). De eerste component bevatte wel alle E-schalen maar ook P- en A-schalen, wat volgens de theorie omtrent de Semantische Differentiaal niet zou mogen. We hebben in de resultatensectie deze component Evaluatie genoemd, onder voorbehoud. In ieder geval is opmerkelijk dat alle E-schalen erg hoog scoren op deze eerste component. Deze heeft, naast eventuele andere betekenissen, met name een evaluatieve betekenis. Verder zou men, met enige goede wil, de tweede en vijfde component nog kunnen typeren als twee verschillende soorten van Activiteit maar bij de derde en vierde component ligt dat interpreteren binnen de E-P-A structuur veel moeilijker. Tenminste, een simpele typering is niet zo direkt zichtbaar.

Voor dit resultaat, de afwezigheid van de E-P-A structuur, zijn verschillende verklaringen te geven. Eén oorzaak kan zijn de interactie tussen onderzochten, kernbegrippen en schalen. Studenten gebruiken betekenis-schalen in de context van het verschijnsel Leren misschien anders dan zij, of anderen, zouden doen in de context van bijvoorbeeld het verschijnsel Mode of Crisis. Dit kan betekenen dat de E-P-A structuur niet algemeen of universeel is maar context- (en subject-) afhankelijk.

Een tweede oorzaak kan methodologisch van aard zijn. Gewoonlijk worden de data omgerekend naar een correlatie-matrix van de schalen, over personen en objecten, waardoor heel wat

oorspronkelijke variantie verdwijnt. Hier is dat niet gedaan, de analyse werd uitgevoerd op een kruisproduktmatrix, bij elke modus, waardoor het theoretisch gemiddelde (de score 0 op een schaal) en de afwijking van dat gemiddelde nog steeds aanwezig zijn in de te analyseren matrices. Het gevolg daarvan is dat, bij elke modus, de eerste component van een principale componenten analyse ook een richtingsvector bevat van het oude middelpunt, de score 0, naar het nieuwe middelpunt, de gemiddelde score (de gemiddelde score van de 55 subjecten was bijna -1). Bij correlatie-matrices wordt de richtingsvector bij voorbaat geëlimineerd. De keuze voor de analyse van kruisprodukt- of van correlatie-matrices is dan afhankelijk van een theoretische interesse (wel of niet zoveel mogelijk oorspronkelijke variantie behouden voor de data-analyse), met methodologische consequenties. Kroonenberg (1983) bijvoorbeeld heeft een heranalyse uitgevoerd op de data van Osgood en Luria (1972) maar maakte daarbij geen gebruik van correlatie-matrices, zoals Osgood en Luria, maar van kruisprodukt-matrices. De E-P-A structuur die werd gerapporteerd door Osgood en Luria, kwam in zijn analyse niet naar voren. Ook weken zijn resultaten wat betreft personen en objecten af van die van Osgood en Luria.

Subjecten

Aan Tabel 3 is te zien dat de indeling van onderzochten op Leerconceptie, op grond van hun antwoord op de essay-vraag, geen één op één relatie vertoont met de persoonscomponenten, zelfs niet in die richting gaat. Hiervoor zijn verschillende verklaringen te geven.

Eén verklaring kan het beoordelen zijn in onderzoek naar Leerconcepties op één specifieke Leerconceptie. In dat onderzoek wordt wel een onderscheid gemaakt tussen lagere en hogere Leerconcepties, Surface resp. Deep en Leerconceptie 1 t/m 3 resp. 4 en 5, waarbij de hogere de lagere bevatten maar niet omgekeerd. Dat wil zeggen, personen met hogere Leerconcepties hebben ook de lagere tot hun beschikking maar personen met een lagere niet de hogere Leerconcepties. Gewoonlijk, en zo is het hier ook gedaan door Van Rossum, wordt de Leerconceptie dan bepaald aan de hand van de hoogste te ontdekken en/of meest expliciete Leerconceptie. Volgens tabel 3 kunnen de meeste onderzochten opgevat worden als specifieke combinaties van twee persoonscomponenten. Veronderstel nu dat we elke persoonscomponent gelijkstellen met één bepaalde Leerconceptie, dat elke persoonscomponent het prototype is van iemand met een specifieke Leerconceptie. Het merendeel van de onderzochten is dan volgens tabel 3 een combinatie van twee persoonscomponenten, en hebben (dus) niet één maar twee Leerconcepties. Daaruit zou men kunnen afleiden dat de toewijzing van maar één specifieke Leerconceptie aan een persoon op grond van zijn/haar antwoord op een essay-vraag, een te ruwe benadering is. Een persoon behoort volgens tabel 3 tot meer dan één categorie. Een implicatie daarvan voor onderzoek naar Leerconcepties kan zijn dat het zich niet zozeer zou moeten richten op de meest fundamentele en/of meest expliciete Leerconceptie maar op een inventarisatie van alle te ontdekken (brokstukken van) Leerconcepties in antwoorden van een persoon. Met andere woorden, dat de beoordeling op Leerconcepties meer gedifferentieerd zou moeten zijn; zodat wordt aangegeven welke Leerconcepties allemaal zijn te halen uit het antwoord van de desbetreffende persoon.

Een tweede verklaring zou kunnen liggen in het verschillend interpreteren van kernbegrippen door enerzijds onderzochten en anderzijds onderzoeker(s). Hiervoor zijn op z'n minst drie tegenargumenten te geven. Ten eerste, op grond van tabel 1 kan men afvragen of een indeling in 5 verschillende Leerconcepties, in 5 verschillende groepen van kernbegrippen, wel nodig is. Volgens de onderzochten lijken twee groepen wel voldoende. Hierbij dient wel te worden opgemerkt dat de tweedeling van de onderzochten in een Surface en een Deep groep van kernbegrippen afwijkt van de theoretische tweedeling in een Surface en een Deep groep. Personen nemen in hun Surface groep de kernbegrippen samen van Leerconceptie 1 en 2 (op Onthouden na), en in hun Deep groep die van Leerconceptie 3 t/m 5. De theorie neemt de kernbegrippen van Leerconceptie 1 t/m 3 samen in de Surface groep, en die van Leerconceptie 4 en 5 in de Deep groep. Dat wil zeggen, de tweedeling van de onderzochten komt vrijwel overeen met de theoretische tweedeling, op Toepassen en Nut (kernbegrippen van Leerconceptie 3) en Onthouden na. Dit brengt ons bij het tweede tegenargument.

Ten tweede, volgens tabel 1, en (dus?) volgens de onderzochten, behoren Toepassen en Nut eerder tot de Deep groep van kernbegrippen, volgens de theorie meer tot de Surface groep. Ook blijkt dat Onthouden eerder behoort tot beide groepen en niet enkel tot de Surface groep. We hebben zojuist bij de discussie van de kernbegrippen, hierboven, al een manier gegeven om binnen het kader van Leerconceptietheorieën het theoretisch afwijkend gedrag te verklaren van de kernbegrippen Toepassen (een ambigue kernbegrip) en Onthouden (geen kenmerkend kernbegrip van Surface-leerconceptie 2). Met andere woorden, ook hier lijkt er dus geen direct verschil van interpretatie aanwezig tussen enerzijds onderzochten en anderzijds onderzoeker(s).

Ten derde, persoon 55 was de auteur. In tabel 3 is te zien dat zijn scores op de 2 persoonscomponenten in beide gevallen vrijwel samenvallen met die van persoon 16 en in mindere mate met die van persoon 23. De interpretatie van de onderzoeker, in dit geval de auteur, staat dus niet haaks op die van onderzochten. Zijn interpretatie is eerder gelijk aan die van een bepaalde sub-groep van onderzochten, van de sub-groep persoonscomponent II.

Een derde verklaring dat er geen één op één relatie bestaat tussen enerzijds persoonscomponenten en anderzijds de Leerconceptie van de desbetreffende persoon, kan methodologisch van aard zijn. Het onderzoek had twee fasen, een kwalitatieve respectievelijk kwantitatieve fase. Tijdens de kwalitatieve fase moesten kernbegrippen door de onderzochten zelf gegenereerd worden en ook de verschillen tussen deze kernbegrippen. Tijdens de kwantitatieve fase werden kernbegrippen kant en klaar aangeboden en op een indirecte manier, via de schalen, met elkaar vergeleken. Men mag aannemen dat het vergelijken in de kwalitatieve fase op een directe wijze gebeurde, via de kernbegrippen onderling. Kort gezegd, de kwantitatieve methode geeft kernbegrippen en vraagt naar indirecte verschillen tussen kernbegrippen; de kwalitatieve methode verkrijgt alleen zelf gegenereerde kernbegrippen en vraagt naar directe verschillen tussen kernbegrippen. Het gevolg hiervan kan zijn dat onderzochten bij de kwantitatieve methode ook kernbegrippen krijgen aangeboden waar ze, bij het beantwoorden van de essay-vraag, niet direct aan zouden denken. Deze methode geeft dan de mogelijkheid om hoger te scoren in Leerconceptie dan met een kwalitatieve methode.

Hier staat tegenover dat de kwalitatieve methode de mogelijkheid schept op lager te scoren. Immers, een persoon moet zelf een antwoord geven, en kan een lang maar ook een kort antwoord geven. In het geval van een kort antwoord ontstaat een interpretatieprobleem voor de onderzoeker die een kwalitatieve methode hanteert, vanwege een gebrek aan (voldoende) informatie. In zo'n geval is de kans op een beoordelingsfout, in het toewijzen van een Leerconceptie aan een bepaald persoon, ook het grootst. Aangezien nogal wat onderzochten een kort antwoord hadden gegeven, lijkt het wenselijk om bepaalde maatregelen te nemen om subjecten voldoende te laten opschrijven. Men kan daarbij denken aan kwantitatieve maatregelen, bijvoorbeeld minimaal een A-4 vel volschrijven, maar ook aan kwalitatieve maatregelen, bijvoorbeeld het verder laten uitwerken van bepaalde (zelf gegenereerde) kernbegrippen.

Tot slot kan nog vermeld worden dat de resultaten geen aanleiding geven om, zoals Osgood, het bestaan te veronderstellen van enkel één algemeen persoon (de universele persoon) maar eerder het bestaan suggereren van individuele verschillen (van twee universele personen).

Kernmatrix

In de Inleiding was over de kernmatrix gezegd dat deze matrix de mate van interactie aangeeft tussen personen, objecten en schalen. In tabel 4 is te zien dat deze interactie bij de twee algemene personen (en dus eigenlijk bij alle onderzochten) het sterkst is voor de Deep kernbegrippen en de eerste schaalcomponent, die we onder voorbehoud Evaluatie hadden genoemd. Dit is te interpreteren als een vorm van Sociale Wenselijkheid, zaken zoals bijvoorbeeld Creativiteit of Inzicht zullen altijd wel positief gewaardeerd worden. Bij Surface kernbegrippen ligt die interactie genuanceerder; één groep van onderzochten waardeert ze nog in sterke mate maar de andere groep waardeert ze neutraal. Hier is bij de eerste groep van personen misschien nog sprake van Sociale Wenselijkheid, bij de andere groep niet. De interactie van de kernbegrippen en de persoonscomponenten met de overige vier schaalcomponenten zijn alle veel kleiner dan de interactie met de eerste schaalcomponent.

Over de kernmatrix is ook te zeggen dat hij een betrekkelijk complexe structuur aangeeft tussen de drie modes. Van de 20 kernelementen zijn er welgeteld 3 bijna gelijk aan 0. De relaties zouden simpeler zijn als er meer van die nul-elementen zouden zijn. Er bestaat een manier om een kernmatrix verder te transformeren naar een meer simpele structuur, met het zogenaamde ALS (Alternating Least Squares) algoritme (Kroonenberg, 1983). In het ideale geval is de kernmatrix na de ALS-transformatie een diagonale matrix (Kroonenberg, 1983). Opgemerkt dient te worden dat in zo'n ALS-algoritme het aantal componenten van een modus nog verder gereduceerd kan worden, dat men na het algoritme een kleinere kernmatrix heeft dan ervoor. Ook de (overgebleven) componenten van de drie modes veranderen door deze transformatie. In eerste instantie was in de data-analyse ook het ALS-algoritme opgenomen maar dit bleek, om technische redenen, niet haalbaar op de P2500 micro-computer. De structuur van de kernmatrix, zoals weergegeven in tabel 4, had door het uitvoeren van het ALS algoritme anders kunnen zijn geweest. Echter, de sterke interactie voor alle onderzochten tussen Diep kernbegrippen en de eerste schaalcomponent (waar we onze aandacht hoofdzakelijk op concentreerden) zou waarschijnlijk niet verdwenen zijn maar hoogstens worden genuanceerd.

CONCLUSIES

De resultaten met betrekking tot de kernbegrippen ondersteunen de stelling uit de onderzoeksliteratuur dat er (minimaal) 2 groepen van kernbegrippen zijn, dat er (minimaal) 2 verschillende 'talen' zijn om mee te praten over het verschijnsel Leren. Verder kwam de gerapporteerde tweedeling, qua kernbegrippen, redelijk overeen met de theoretische tweedeling. De resultaten van de schalen wijken af van de bekende E-P-A structuur uit de literatuur omtrent de Semantische Differentiaal. De schalen hadden niet de E-P-A structuur. De resultaten van de onderzochten suggereren dat het merendeel van hen niet één maar verschillende Leerconcepties zouden hebben, dat ze dus verschillende 'talen' kunnen spreken. De kernmatrix, ten slotte, laat zien dat de belangrijkste en meest onderscheidende betekenis van het verschijnsel Leren ligt in de evaluatieve sfeer.

In de discussiesectie over de onderzochten hebben we wat betreft kwalitatief onderzoek gewezen op het interpretatieprobleem bij korte antwoorden van personen. We hebben een aantal maatregelen gegeven om dat probleem te vermijden. Omdat het hier een exploratief onderzoek betreft, zullen in het vervolgonderzoek deze maatregelen genomen worden. Echter, de maatregelen zijn natuurlijk met name relevant voor kwalitatief onderzoek sec. Naast deze maatregelen dient men, volgens ons, bij kwalitatief onderzoek naar Leerconcepties ook meer differentiërend te werk te gaan bij de beoordeling van antwoorden van personen op essayvragen. In plaats van één Leerconceptie toe te wijzen aan een persoon, is het wellicht zinvol om nader te analyseren welke (brokstukken van) Leerconcepties een persoon vermeldt in zijn antwoord.

LITERATUUR

- Blocker, G. (1974). *The Meaning of Meaninglessness*. The Hague: Martinus Nijhoff.
- Harman, H.H. (1976). *Modern Factor Analysis* (3rd ed.). Chicago: University of Chicago Press.
- Heise, D.R. (1969). Some methodological issues in semantic differential research. *Psychological Bulletin*, 72, 406-422.
- Jansen, M.J. en A.J. Smolenaars (1966). Kort verslag inzake een interlandelijk gestandaardiseerde Semantische Differentiaal. *Nederlands Tijdschrift voor de Psychologie*, 21, 211-216.
- Kouwer, B.J. (1958). Een bewerking van de polariteitsprofielenmethode van Osgood. *Nederlands Tijdschrift voor de Psychologie*, 13, 1-14.
- Kroonenberg, P.M. (1983). *Three-Mode Principal Component Analysis Theory and Applications*. Leiden: DSWO Press.
- Levin, J. (1965). Three-Mode Factor Analysis. *Psychological Bulletin*, 64 (6), 442-452.

- Martin, J.E. (1982). Presentationalism: Toward a Self-Reflexive Psychological Theory. In: W.B. Weimer en D. S. Palermo (eds) *Cognition and the Symbolic Processes*, Vol. 2. Hillsdale: Lawrence Erlbaum Associates.
- Marton, F. en L. Svensson (1979). Conceptions of Research in Student Learning. *Higher Education*, 8, 471-486.
- Osgood, Ch.E. en Z. Luria (1972). A Blind Analysis of a Case of Multiple Personality using the Semantic Differential. In: J.G. Snider en Ch.E. Osgood (Eds.) *The Semantic Differential Technique*, a source book. Chicago: Aldine and Atherton.
- Osgood, Ch.E., W.H. May en M.S. Miron (1975). *Cross-Cultural Universals of Affective Meaning*. Urbana: University of Illinois Press.
- Osgood, Ch.E., G.J. Suci en P.H. Tannenbaum (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Pierar, M.G.E. (1986). De nieuwe kleren van de Cognitieve Keizer. Ongepubliceerd paper.
- Presley, A.S. (1969). Concept-scale Interaction in the Semantic Differential and its implications for Factor-Scores. *British Journal of Psychology*, 60, 1, 109-113.
- Van Rossum, E.J., R. Deijkers en R. Hamer (1985). Students' Learning conceptions and their interpretation of significant Educational concepts. *Higher Education*, 14, 617-641.
- Van Rossum, E.J. en S. Schenk (1983). De Relatie tussen Leerconceptie, Studiestrategie en Leerresultaat. *Pedagogische Studiën*, 60, 252-262.
- Saljö, R. (1979). Learning in the Learner's perspective II: Differences in awareness. Reports from the Institute of Education, University of Göteborg.
- Snyder, F.W. en N. Wiggins (1970). Affective Meaning Systems: A Multivariate Approach. *Multivariate Behavioral Research*, 5 (4), 453-468.
- Tucker, L.R. (1966). Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika*, 31(3), 279-311.

Manuscript ontvangen 25-8-1986

Definitieve versie ontvangen 6-4-1987

Notities en Commentaren

COMMENTAAR OP PIERAR "DE BETEKENIS VAN LEREN"

Pieter M. Kroonenberg

Vakgroep W.E.P., Rijksuniversiteit Leiden

Op de bijdrage "De betekenis van het leren: Een exploratief onderzoek naar leerconcepties met de semantische differentiaal" (Pierar, 1987) is ernstige kritiek te leveren. In het hierna volgende zal ik uitsluitend ingaan op mijn hoofdbezwaren tegen de analyses zoals die uitgevoerd en gemotiveerd zijn.

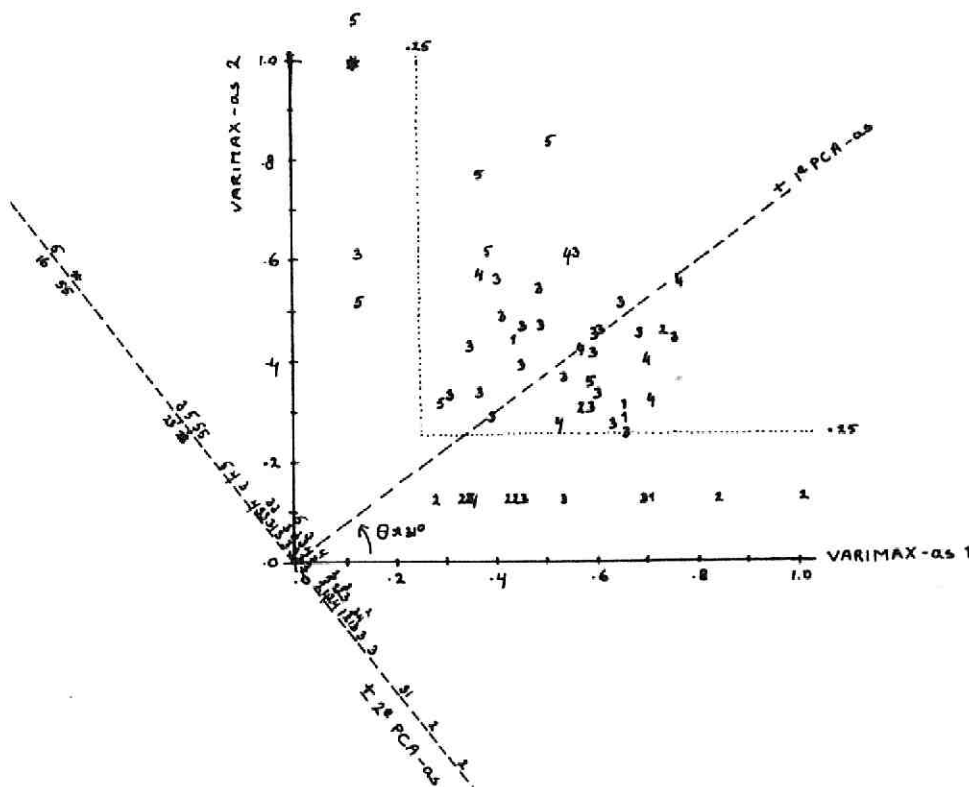
In het artikel wordt een driewegprincipale componenten analyse uitgevoerd met behulp van een door de auteur geschreven programma dat gebruik maakt van de Methode I zoals beschreven in Tucker (1966). Zoals al door de auteur wordt opgemerkt zijn er later methoden ontwikkeld die betere eigenschappen hebben, zoals de alternerende kleinste kwadratenversie van de techniek (zie Kroonenberg & De Leeuw, 1980) en de op regressietechnieken geënte techniek GEPCAM ontworpen door Weesie en Van Houwelingen (1983). Het zou me evenwel verbazen als deze andere methoden fundamenteel een andere oplossing geven.

Een eerste deel van mijn kritiek concentreert zich op problemen rond de componenten van de semantische differentiaal. In Tabel 2 presenteert de auteur een varimax-geroteerde oplossing met vijf componenten. Bij de sectie Schalen van de Resultaten merkt de auteur over deze tabel op dat "De E-P-A structuur komt uit dit resultaat niet naar voren. Alle E- en P-schalen hebben hoge ladingen op de eerste component, de ladingen van de A-schalen liggen verspreid over de eerste drie componenten." In de sectie Schalen van de Discussie gaat hij verder met "De resultaten van de schalen wijken af van de bekende E-P-A structuur uit de literatuur omtrent de Semantische Differentiaal. De schalen hadden niet de E-P-A structuur".

Hoe waar het ook mag zijn dat Tabel 2 niet de EPA structuur laat zien, dit is nog geenszins een bewijs dat die structuur ook niet bestaat. In de eerste plaats, is het de lezer onmogelijk om na te gaan in hoeverre de beslissing van vijf componenten hier debet aan zou kunnen zijn. Geen enkele informatie over de ongeroteerde oplossing is aanwezig (er worden bijvoorbeeld geen percentages verklaarde variantie van de PCA oplossing verschaft), en door het ontbreken van de lage ladingen in de tabel kan die ook niet gereconstrueerd worden. Op zijn minst had de auteur kunnen proberen om een analyse te doen met alleen de 'erkende' EPA schalen om het verschijnsel nader te onderzoeken. Het ontbreken van een EPA-structuur is per slot van rekening een belangrijke conclusie. Verder kan men zich afvragen of een varimaxrotatie de meest geëigende rotatie is als men na wil gaan of er inderdaad sprake is van een verwachte structuur, misschien zou een scheve of doelrotatie wel beter geweest zijn. Daarbij speelt overigens net zo goed de keuze van het aantal componenten een grote rol. Ik wil op voorhand niet zeggen dat de auteur ongelijk heeft, maar ik vind dat een gedegen bewijsvoering ontbreekt voor de geciteerde beweringen, terwijl in principe het materiaal wel voorhanden is om een gefundeerde uitspraak te doen. Ten Berge (1986a, b) geeft overigens een fraaie discussie over de mogelijkheden en onmogelijkheden om met rotaties structuren in PCA-oplossingen te vinden, of het bestaan ervan te logen.

Een tweede bezwaar tegen (een onderdeel van) de analyse is de (automatische?) varimax-transformatie van de personenruimte. Zowel bij de kernbegrippen als bij de semantische differentiaal ligt het zoeken naar 'simpele structuren' wel voor de hand. Een gedegen redenering ontbreekt, echter, voor het zoeken naar 'simpele structuren' bij personen (zie bijvoorbeeld Tucker, 1972, p. 11, 12, voor mogelijkheden en argumenten voor en tegen rotaties van personenruimten). Een nadere beschouwing van de personenruimte laat zelfs zien dat in vergelijking met een PCA-oplossing, de varimax-getransformeerde ruimte in dit geval ongunstig werkt voor een van de door de auteur opgeworpen hypothesen. Helaas was de auteur

niet in staat om op korte termijn de ongeroteerde PCA-oplossing of de ontbrekende elementen uit Tabel 3 te leveren, zodat ik mij moest behelpen met een reconstructie van de PCA-oplossing via een visuele weergave van de personenruimte, afgeleid uit Pierar's Tabel 3 (zie Figuur 1). In deze figuur heb ik de punten zo goed en zo kwaad als het kon aangegeven, en daarna met het oog de hoofdassen bepaald. Het is duidelijk dat vrijwel alle personen positieve, en ruwweg gelijke, waarden hebben op de eerste PCA-as, hetgeen niet verwonderlijk is aangezien de personenruimte niet gecentreerd was – redenen om deze te centreren liggen overigens ook niet voor de hand. Dit betekent dat in eerste orde alle personen een oordeel hebben over de relaties tussen de kernbegrippen en de schalen dat lijkt op dat van de gemiddelde persoon. Dit is natuurlijk niet het hele verhaal want er is ook nog een tweede persoonscomponent, en als we de (geschatte) projecties van de personen op deze as beschouwen zien we dat hier een – zij het vaag – contrast aanwezig is tussen de personen met een leerconceptie '5' en die met een leerconceptie '2' en '1'. Dit zou kunnen duiden op een contrast tussen gebruikers van een Dieptestructuur ('Dieptetaal') en gebruikers van een Oppervlaktestructuur ('Surfacetiaal'), iets wat de auteur graag wou



Figuur 1: Varimax componenten en geschatte principale componenten van de personen.
(Personen zijn aangegeven met hun Leerconceptie, zie Pierar's Tabel 3,* is dhr. Pierar. Ladingen $<.25$ ontbraken, en zijn gelijk gemaakt aan .125. Projecties op de tweede geschatte principale as zijn weergegeven, alsmede persoonsnummers van vier personen met Leerconceptie '5'.)

ontdekken bij zijn personen. Immers in de Discussie onder het hoofdje Subjecten meldt de auteur 'dat de indeling van onderzochten op Leerconceptie, op grond van hun antwoord op de essay-vraag, geen één op één relatie vertoont met de persoonscomponenten, zelfs niet in die richting gaat'. Misschien zou een discriminatieanalyse op de componenten enig licht kunnen werpen op mijn indruk, gevormd op basis van Figuur 1, dat er hier misschien wel sprake is van tenminste enige relatie.

Aan het eind van de sectie Subjecten van de Discussie wordt verder nog vermeld, dat 'de resultaten geen aanleiding geven om, zoals Osgood, het bestaan te veronderstellen van één enkel algemeen persoon (de universele persoon) maar eerder het bestaan suggereren van individuele verschillen (van twee universele personen)'. Evenwel op grond van een geroteerde varimaxoplossing kan zoiets niet geconstateerd worden, ongeacht of de opmerking waar is of niet. Ook hier duikt weer het verschijnsel op dat men op basis van de waarden van (geroteerde) componenten niet zonder meer kan besluiten dat bepaalde structuren niet aanwezig zijn. Een gedegen onderzoek is hierbij geboden. Ten overvloede zij hier nog vermeld dat Osgood geheel andere analyses uitvoerde, zodat ook nog aangetoond moet worden dat hier geen sprake is van een methodeneffect. Te meer daar sommige auteurs bij een drieweganalyse wel een EPA-structuur gevonden hebben (Snyder, & Wiggins, 1970; Levin, 1965), en anderen niet (Kroonenberg, 1983, 1985).

Naar mijn mening is de uitvoering (rapportage?) van de bepaling van een persoonskenmerk gebrekkig. In de sectie Procedure van de Methodebeschrijving wordt vermeld dat 'De antwoorden op de essay-vragen werden beoordeeld op Leerconceptie door Van Rossum, onafhankelijk van de resultaten op de vragenlijst welke door de auteur (Pierar, pmk) werden geanalyseerd en geïnterpreteerd'. Gezien men in ieder geval met zijn tweeën was, had op zijn minst de intersubjectieve betrouwbaarheid bepaald kunnen worden. Nu is controle over de kwaliteit van de Leerconceptie-oordelen nauwelijks mogelijk. Het belang hiervan wordt geïllustreerd door een uitwijding in zes paragrafen over het ontbreken van een relatie tussen de persoonscomponenten en de bovengenoemde bepaling van de leerconceptie van een persoon.

Opvallend in Figuur 1 is de extreme positie die de auteur zelf inneemt. In de tekst zegt hij hierover 'dat zijn scores op de 2 persoonscomponenten in beide gevallen vrijwel samenvallen met die van persoon 16 en in mindere mate met die van persoon 23. De interpretatie van de onderzoeker, in dit geval de auteur, staat dus niet haaks op die van de onderzochten. Zijn interpretatie is eerder gelijk aan die van een bepaalde sub-groep van onderzochten, van de sub-groep persoonscomponent II' (sectie Subjecten van de Discussie). Mijn indruk is eerder dat de auteur samen met persoon 16, en zeker niet met 23, zich duidelijk onderscheidt van de rest van de personen, tenzij men het principe huldigt: 'Two is a crowd'.

Tabel 1. Ongeroteerde Kernmatrix

<i>'Ruwe' Kernmatrix</i>		Schaalcomponenten				
Persoons						
Comp.	Structuur	I	II	III	IV	V
I	Diepte	115.2	21.4	13.4	-23.7	4.5
	Opperv.	37.7	-0.5	-21.8	4.6	3.6
II	Diepte	19.9	4.7	9.1	- 1.1	-12.2
	Opperv.	-26.1	-4.9	-17.8	9.1	-11.9
<i>Percentage Verklaarde Kwadratensom</i>		Schaalcomponenten				
Persoons						
Comp.	Structuur	I	II	III	IV	V
I	Diepte	.328	.011	.004	.014	.001
	Opperv.	.035	>.000	.012	.001	>.000
II	Diepte	.010	.001	.002	>.000	.004
	Opperv.	.017	.001	.008	.002	.003

De kernmatrix laat nog eens zien hoe aanvechtbaar een aantal beslissingen zijn geweest. Door de kernmatrix tegen te roteren met de geschatte transformatiematrix op basis van Figuur 1 ($\theta \approx 31$ graden) krijgen we een indruk hoe de kernmatrix er ongeveer uitgezien moet hebben vóór de varimaxrotatie van de personenruimte (zie Tabel 1). Door alle elementen van de kernmatrix te kwadrateren en ze elk te delen door de totale kwadratensom van de herschaalde data (= 40425), krijgt men een overzicht van de bijdrage van elk element van de kernmatrix aan de gezamenlijke verklaarde 'variantie' (eigenlijk kwadratensom) van de data door het model (zie eveneens Tabel 1). Met behulp van de kernmatrix kan worden afgeleid in hoeverre de kwadratensommen verklaard door de afzonderlijke analyses op elke weg, overeenkomen met de door alle wegen tezamen verklaarde kwadratensom in de drieweganalyse. Deze vergelijking is numeriek weergegeven in Tabel 2. (De percentages voor de ongeroteerde persoonscomponenten ontbreken bij gebrek aan gegevens). Uit Tabel 2 blijkt dat een beslissing ten gunste van vijf componenten betekent dat men componenten meeneemt die ongeveer één procent of minder van de totale kwadratensom verklaren, terwijl uit Tabel 1 blijkt dat men spreekt over componentcombinaties, die minder dan 0.4% verklaren.

Tabel 2. Percentage Verklaarde Variantie voor de Componenten van de Drie Wegen

Weg/Mode	Afzonderlijke Analyses*						In Gezamenlijke Analyse **					
	1	2	3	4	5	Tot.	1	2	3	4	5	Tot.
Personen	—	—				49.4	37.3	7.9				45.2
Schalen	44.2	6.5	6.0	5.1	4.2	66.0	38.9	1.2	2.6	1.7	0.8	45.2
Kernbegrip.	45.9	13.0				58.8	40.5	4.7				45.2

* Gegevens voor persoonscomponenten ontbreken.

** Berekend met behulp van de kernmatrix uit Tabel 1.

Zoals dergelijke analyses vaak laten zien hebben de personen meer gemeen (eerste persoonscomponent verklaart 37% van de totale kwadratensom), dan dat ze verschillen (tweede component verklaart 8%, zie Tabel 2), en personen met een '5' leerconceptie onderscheiden zich voornamelijk van personen met een '1' of een '2' leerconceptie in de negatieve evaluatie van de oppervlaktestructuur (zie Tabel 1). Merk overigens op dat Figuur 1 laat zien dat vrijwel alle personen ongeacht hun leerconceptie een positieve lading hebben op de eerste component. Het is mogelijk om een verdere typering van de persoonsassen te geven aan de hand van de kernmatrix, maar dat lijkt me in dit kader niet precies de bedoeling. Naar aanleiding van de opmerking in de sectie Kernmatrix van de Discussie dat de kernmatrix 'een betrekkelijk complexe structuur aangeeft tussen de drie modes. Van de 20 elementen zijn er welgeteld 3 bijna gelijk aan 0', kan er nog gezegd worden dat vóór rotatie van de persoonsruimte er slechts 7 van de 20 elementen waren die een bijdrage aan de kwadratensom leverden die groter was dan 1%, en wanneer men geen enkele varimaxrotatie had toegepast zou dit aantal waarschijnlijk nog kleiner zijn geweest. Dit houdt in dat men zich er terdege op moet bezinnen wat men wil, voordat bepaalde rotaties worden uitgevoerd in een drieweganalyse.

Concluderend zou ik willen zeggen, dat het artikel op vele punten onvolkomenheden bevat, of toch op zijn minst onuitgewerkte analyses. Verder worden er beweringen gedaan die door het materiaal misschien wel, maar dan toch niet door de gepresenteerde analyses ondersteund worden, ook al is het onderzoek, zoals herhaaldelijk wordt opgemerkt, exploratief van aard.

REFERENTIES

- Kroonenberg, P.M. (1983). *Three-mode principal component analysis: Theory and applications*. Leiden: DSWO Press.
- Kroonenberg, P.M. (1985). Three-mode analysis of semantic differential data: The case of a triple personality. *Applied Psychological Measurement*, 9, 83-94.
- Kroonenberg, P.M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least-squares algorithms. *Psychometrika*, 45, 69-97.
- Levin, J. (1965). Three-mode factor analysis. *Psychological Bulletin*, 64, 442-452.
- Pierar, M. (1987). De betekenis van het leren: Een exploratief onderzoek naar leerconcepties met de semantische differentiaal. *Tijdschrift voor Onderwijsresearch*, 12.
- Snyder, F.W., & Wiggins, N. (1970). Affective meaning systems: A multivariate approach. *Multivariate Behavioral Research*, 5, 453-468.
- Ten Berge, J.M.F. (1986a). Some relationships between descriptive comparisons of components from different studies. *Multivariate Behavioral Research*, 21, 29-40.
- Ten Berge, J.M.F. (1986b). Rotation to perfect congruence and the cross-validation of component weights across populations, *Multivariate Behavioral Research*, 21, 41-64; 262-266.
- Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279-311.
- Tucker, L.R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 37, 3-27.
- Weesie, J., & Van Houwelingen, H. (1983). GEPCAM user's manual: Generalized principal components analysis with missing values. Intern rapport, Instituut voor Mathematische Statistiek, Rijksuniversiteit Utrecht.

Ontvangen 6-5-1987.

Notities en commentaren

SEMANTISCHE STRUCTUURKENMERKEN, COGNITIEVE SCHEMATA EN AANSCHOUWELIJKE MODELLEN BIJ AANVANKELIJKE REDACTIE-OPGAVEN

L. Verschaffel (1)

Onderzoekscentrum voor Onderwijsleerprocessen, Katholieke Universiteit Leuven

In december 1986 verschenen in het Tijdschrift voor Onderwijsresearch twee bijdragen over de invloed van semantische structuurkenmerken op het oplossen van aanvankelijke rekenvraagstukjes: een kritisch commentaar van Tavecchio, Beishuizen, van den Berge en Bleek (1986) op een vroeger artikel van Wolters (1984) in ditzelfde tijdschrift, gevolgd door een wederwoord van Wolters (1986). In de betreffende teksten wordt het hele domein van aanvankelijke redactie-opgaven m.i. nogal eng benaderd, wanneer men rekening houdt met het thans beschikbare bestand aan onderzoeksgegevens. Bovendien vergen sommige verwijzingen van laatstgenoemde auteur naar onze artikelen (De Corte & Verschaffel, 1985a; De Corte, Verschaffel & De Win, 1984; Verschaffel, 1984, 1985) enige precisering en/of nuancering. Derhalve wil ik op mijn beurt, met het oog op het stimuleren en vooruithelpen van de discussie en het onderzoek in dit boeiende deel van de onderwijskunde, de bijdragen van Tavecchio e.a. (1986) en Wolters (1986) van enige kanttekeningen voorzien. Hiervoor baseer ik me zowel op de onderzoeken die ik de laatste jaren samen met E. de Corte op dit terrein verricht heb, als op recente literatuurgegevens terzake.

1. Semantische structuurkenmerken van aanvankelijke redactie-opgaven

In de jaren zeventig heeft er zich, zoals Wolters (1986) terecht opmerkt, in het onderzoek over eenvoudige optel- en aftrekvraagstukjes een belangrijke verschuiving voltrokken. Een van de meest opvallende kenmerken van de nieuwe studies is de belangstelling voor de semantische structuurkenmerken die aan redactie-opgaven ten grondslag liggen, in plaats van de meer oppervlakkige linguïstische en rekentechnische aspecten die in het ouder onderzoek veruit de meeste aandacht kregen (De Corte & Verschaffel, 1982; Verschaffel, 1984).

Deze verschuiving kwam voor het eerst duidelijk tot uiting in het werk van de Pittsburghse onderzoekers Greeno, Heller en Riley (Heller & Greeno, 1978; Riley & Greeno, 1978). Uitgaande van de hypothese dat in een competent oplossingsproces het semantisch verwerken van de opgavetekst een cruciale rol speelt, gingen Greeno c.s. op zoek naar de verschillende semantische structuurtypes die aan eenvoudige optel- en aftrekvraagstukjes ten grondslag liggen. Op basis van een analyse van een paar honderd dergelijke opgaven die ze aantroffen in rekenboeken en schoolvorderingentests, hebben zij de drie basiscategorieën onderscheiden die ook in de bijdragen van Tavecchio e.a. (1986) en van Wolters (1986) centraal staan, nl. vraagstukjes met een oorzaak-verandering-, een combinatie- en een vergelijkingsstructuur. Bij deze indeling dienen evenwel een paar kanttekeningen geplaatst.

Ten eerste is deze indeling niet helemaal nieuw. Uit eigen literatuuronderzoek (Verschaffel, 1984) is namelijk gebleken dat deze driedeling ook reeds voorkomt in oudere, praktijkgeoriënteerde publikaties. Ter illustratie verwijs ik naar de bijdrage van Reckzeh (1956) in *The Arithmetic Teacher*, getiteld 'Addition and subtraction situations'.

Ten tweede schijnen de resultaten van meer recente theoretische analyses en empirische onderzoeken er op te wijzen dat de indeling van Greeno c.s. voor aanvulling vatbaar is. Zo

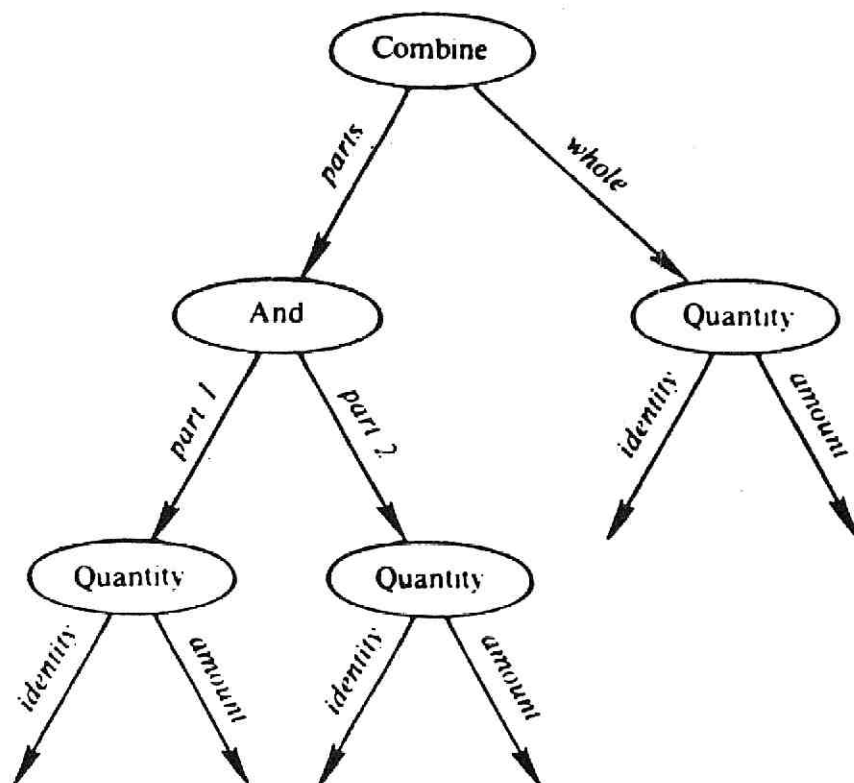
wordt door Carpenter en Moser (1982) naast de drie voormelde basiscategorieën van semantische relaties – oorzaak-verandering, combinatie en vergelijking – nog een vierde type onderscheiden dat als de samensmelting van oorzaak-verandering en vergelijking te beschouwen is, m.n. 'equalizing' ('gelijkmaken'). Een voorbeeld van zo'n opgave is: 'Piet heeft 3 appels. An heeft 8 appels. Hoeveel appels moet Piet verkrijgen opdat hij evenveel appels zou hebben als An?' Evenzo kunnen enkelvoudige vraagstukjes opgesteld worden waarin elementen uit het combinatie- en het vergelijkingsschema ineengevlochten worden: 'Piet heeft 3 appels. An heeft 6 appels. Piet en An hebben samen evenveel appels als Karel. Hoeveel appels heeft Karel?' (Verschaffel, 1984). Ten slotte ontwikkelde Vergnaud (1982) een classificatieschema bestaande uit zes basiscategorieën. Drie daarvan vallen samen met de categorieën uit de indeling van Greeno c.s.; de overige drie lijken daar niet toe herleid te kunnen worden. Bij wijze van voorbeeld vermeld ik een opgave uit een van deze categorieën: 'Piet heeft 3 appels meer dan Jan. Jan heeft 6 appels meer dan Karel. Hoeveel appels heeft Piet meer dan Karel?' Plaatsgebrek verhindert me dieper in te gaan op de vraag of deze drie bijkomende categorieën uit het classificatieschema van Vergnaud (1982) stuk voor stuk nieuwe semantische structuurtypes vertegenwoordigen in de zin zoals bedoeld door Greeno c.s. Belangrijk is hier evenwel dat wij zowel in de literatuur als in ons eigen onderzoeksmateriaal eenvoudige rekenvraagstukjes aangetroffen hebben die qua semantische structuurkenmerken duidelijk te onderscheiden zijn van die uit het classificatieschema van Greeno c.s. (Verschaffel, 1984).

Ten derde dient nogmaals beklemtoond dat de semantische structuur zeker niet het enige opgavekenmerk is dat van invloed is op het oplossen van eenvoudige rekenvraagstukjes. In Wolters' (1986, p. 320) bijdrage is reeds verwezen naar het onderzoek van De Corte e.a. (1984), waaruit gebleken is dat ook de mate waarin de semantische relaties tussen bekende en onbekende kwantiteiten in de tekst op een duidelijke, ondubbelzinnige wijze geëxpliciteerd staan, van belang is. In een meer recente studie (De Corte & Verschaffel, 1987a) vonden we ook duidelijke steun voor de invloed van de volgorde waarin de bekende hoeveelheden in de opgavetekst vermeld staan (bijvoorbeeld: 'Piet en An hebben samen 9 appels. Piet heeft 3 appels. Hoeveel appels heeft An?' versus 'Piet heeft 3 appels. Piet en An hebben samen 9 appels. Hoeveel appels heeft An?'). Een ernstige tekortkoming van veel onderzoek van de laatste jaren is dat deze factoren onvoldoende onder controle gehouden zijn. Daardoor is het enerzijds onduidelijk in hoeverre de gevonden verschillen in moeilijkheidsgraad, oplossingsstrategieën en fouten tussen semantische structuurtypes ook medebepaald zijn door deze bijkomende taakkenmerken. Anderzijds vormt het niet controleren van deze kenmerken tevens een plausibele verklaring voor de onderling tegenstrijdige resultaten van sommige studies wat betreft de invloed van de semantische structuurkenmerken op de genoemde parameters (zie De Corte & Verschaffel, 1985b).

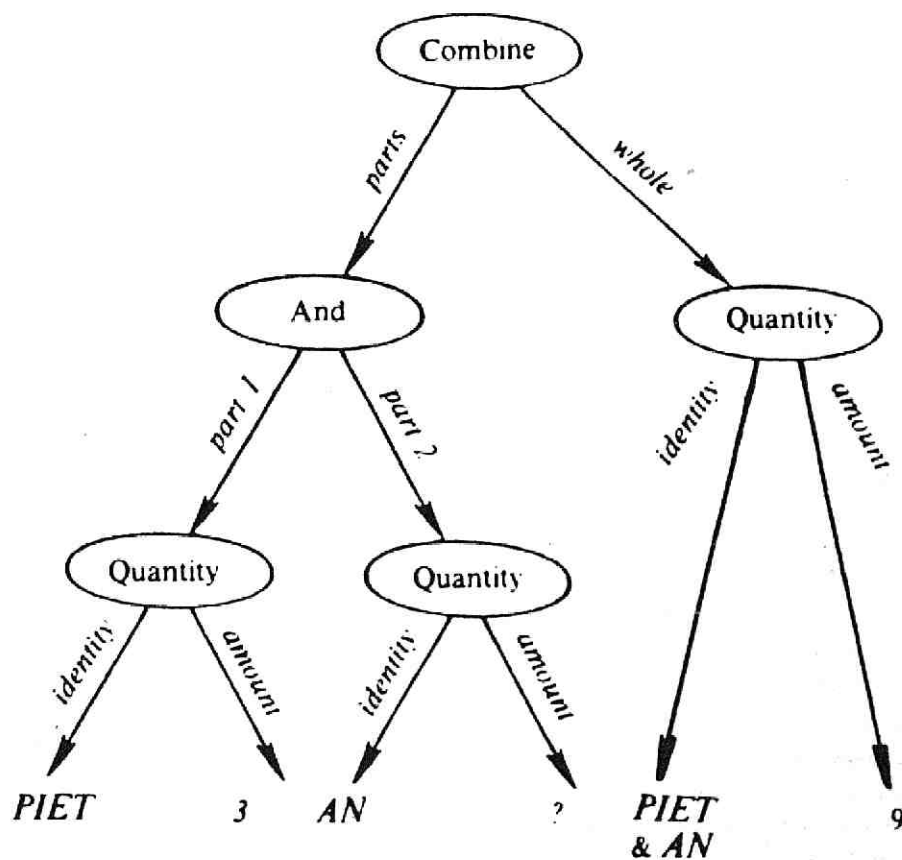
2. Cognitieve schemata

Greeno c.s. beschouwen de drie vermelde categorieën van redactie-opgaven – oorzaak-verandering, combinatie en vergelijking – niet louter als abstracte semantische structuren. In hun (computer)model van het vaardig oplossen van aanvankelijke redactie-opgaven gaan zij er bovendien van uit dat in het cognitief apparaat van de competente probleemoplosser 'psychische' correlaten van deze semantische structuurtypes bestaan in de vorm van cognitieve schemata (De Corte & Verschaffel, 1983). Het begrijpen van een vraagstukje door de competente oplosser vatten zij op als het activeren van een van de drie vermelde schemata, dat dan als leidraad fungeert bij de constructie van een globale, interne representatie van de essentiële concepten en relaties uit de tekst. Figuur 1 geeft een schematische voorstelling van het combinatie-schema en van de representatie die een competente oplosser met behulp van dit schema opbouwt van een concrete opgave zoals 'Piet heeft 3 appels. An heeft ook wat appels. Piet en An hebben samen 9 appels. Hoeveel appels heeft An?' De uiteindelijke probleemrepresentatie van de competente oplosser is dus in feite niets anders dan het (laatst) geactiveerde schema, waarvan de 'lege hokjes' opgevuld zijn met elementen uit de opgavetekst. Ook hierbij plaatsen we enkele kanttekeningen.

1a



1b



Figuur 1a en b. Schematische voorstelling van het abstracte combinatieschema en van de uiteindelijke probleemrepresentatie van een concrete combinatie-opgave door een competente probleemoplosser (naar Riley e.a., 1983).

Ten eerste dient toegegeven dat dit onderdeel van theorie van Greeno c.s. op een aantal punten nogal vaag is. Zo geeft zij geen toereikend antwoord op vragen als: hoe moeten we de kenniselementen die deel uitmaken van zo'n cognitief schema exact opvatten? Hoe komen deze schemata precies tussen in de opbouw van de probleemrepresentatie uitgaande van de opgavetekst? Hoe komen zij tot ontwikkeling bij leerlingen en wat is het aandeel daarin van schoolse, voorschoolse en buitenschoolse ervaringen? Hoewel in meer recente publikaties (Riley, Greeno & Heller, 1983; Kintsch & Greeno, 1984) dieper ingegaan wordt op deze vragen, blijven er toch nog heel wat onduidelijkheden omtrent allerlei aspecten en implicaties van hun model (Verschaffel, 1984).

Ten tweede, hoewel er ook enige empirische gegevens aan ten grondslag liggen, is de theorie van Greeno c.s. toch hoofdzakelijk gebaseerd op een rationele taakanalyse. Wel hebben in de loop van de voorbije jaren tal van onderzoekers nuttig gebruik gemaakt van dit model bij de interpretatie van hun eigen onderzoeksresultaten. Zo hebben wij deze schematheorie zelf aangegrepen bij de verklaring van een ganse reeks empirische bevindingen uit onze studies, m.n.:

- opgaven met een verschillende semantische structuur hebben ook een verschillende moeilijkheidsgraad (De Corte & Verschaffel, 1982);
- de semantische structuur van een opgave oefent ook duidelijk invloed op de aard van de oplossingsstrategieën bij kinderen (De Corte & Verschaffel, 1985b);
- knappe oplosers van aanvankelijke rekenvraagstukjes zijn wel beter in staat de globale opgavesituatie adequaat na te vertellen dan hun zwakkere klasgenootjes, maar niet de details (De Corte & Verschaffel, 1982);
- kinderen vertellen moeilijke en onvertrouwde vraagstukjes vaak verkeerd na in termen van een ander, meer vertrouwd schema (Verschaffel, 1984);
- knappe leerlingen zijn soms in staat een vraagstukje correct op te lossen vooraleer ze de opgavetekst helemaal gelezen of gehoord hebben (De Corte & Verschaffel, 1987b);
- vooral beginnende basisschoolleerlingen lossen vraagstukjes foutief op als gevolg van het onvoldoende expliciteren van de semantische relaties tussen de bekende en de onbekende kwantiteiten in de opgavetekst (De Corte e.a., 1984).

Dat een veelheid van empirische gegevens ermee verklaard kunnen worden verleent weliswaar steun aan het hypothetisch model van Greeno c.s., maar geldt nog niet als een voldoende bewijsvoering voor het bestaan van cognitieve schemata zoals beschreven door deze auteurs en voor het belang van dergelijke schemata in een competent probleemoplossingsproces. Daarom hebben wij in de loop van de voorbije jaren ook verscheidene speciale opdrachten en technieken toegepast om meer rechtstreekse informatie te verzamelen over de kennis van de diverse semantische structuurtypes bij kinderen en over de rol daarvan in het oplossen van vraagstukjes. Ik vermeld er enkele:

- het aanbieden van 'slechte' rekenvraagstukjes (bijvoorbeeld: 'Piet heeft 3 appels. An heeft 6 appels meer dan Piet. Hoeveel appels heeft An meer dan Piet?');
- kinderen zelf vraagstukjes laten opstellen;
- het werken met z.g. vlekopgaven (dit zijn vraagstukjes waarvan een zin of een zinsdeel onder een vlek verborgen is; aan de leerlingen wordt gevraagd of zij kunnen raden wat er onder deze vlek geschreven staat);
- het werken met puzzles (drie fiches met vraagstukjes worden in stukken geknipt; de leerling moet met deze stukken de drie vraagstukjes opnieuw samenstellen);
- het registreren van oogbewegingen tijdens het lezen en oplossen van eenvoudige redactie-opgaven (De Corte & Verschaffel, 1987b).

In dit verband kan ook verwezen worden naar de studie van Morales, Shute en Pellegrino (1985), waarin groepen basisschoolleerlingen – een vijftigtal derde- en zesdeklassers – niet enkel gevraagd werd een serie eenvoudige optel- en aftrek-vraagstukjes op te lossen, maar tevens om deze opgaven in groepen in te delen. De voornaamste bedoeling van deze classificeringstaak was namelijk 'to test the hypothesis that older children who are more accurate in problem solution

have refined conceptual schemata corresponding to the major problem classes described by Riley et al. (1983) and Briars and Larkin (1984)' (Morales et al., 1985, p. 56). Overeenkomstig de hypothese bleken de zesdeklassers de opgaven niet enkel beter op te lossen dan de derdeklassers, maar ook anders in te delen. Meer bepaald vertoonden de indelingen van deze oudere, meer competente probleemoplossers globaal genomen een sterke overeenkomst met de classificaties uit de recente schematheorieën over het leren oplossen van aanvankelijke rekenvraagstukjes; de indelingen van de derdeklassers daarentegen waren veeleer gebaseerd op de oppervlaktestructuur van de opgaven. Tevens werd binnen de groep derdeklassers een verband tussen oplossings- en sorteergedrag vastgesteld, in die zin dat degenen die de vraagstukjes het best oplosten, ook de hoogste sorteescores behaalden. Derhalve is de conclusie van Wolters (1986, p. 125) dat uit het onderzoek van Morales e.a. (1985) '*onontkoombaar* is komen vast te staan dat bij basisschoolleerlingen van *alle* leeftijden de semantische structuurkenmerken een beslissende invloed uitoefenen op het oplossen van redactie-opgaven' (Wolters, 1986, p. 323) niet helemaal terecht. Uit deze studie blijkt wel dat er een verband bestaat tussen de vaardigheid in het oplossen van eenvoudige optel- en aftrekvraagstukjes en het beschikken over kennis betreffende de semantische structuurtypes die aan deze opgaven ten grondslag liggen.

3. Onderwijzen van aanvankelijke rekenvraagstukjes

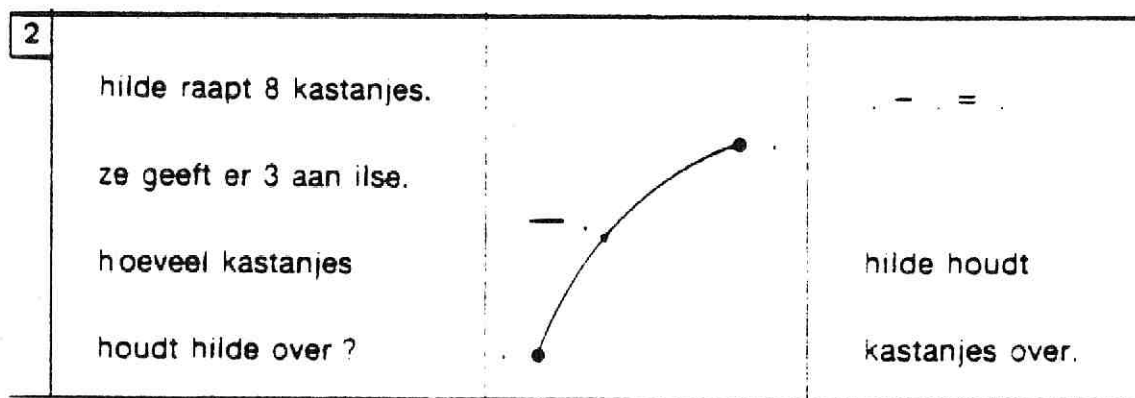
Het recent speurwerk over aanvankelijke redactie-opgaven heeft ook aanleiding gegeven tot het stellen van kritische vragen in verband met de huidige aanvankelijke rekendidactiek en het formuleren van suggesties voor de verbetering daarvan.

Zo hebben ook wij enige jaren geleden een kritische analyse doorgevoerd van de wijze waarop in zes Vlaamse rekenmethoden voor het eerste leerjaar het onderdeel vraagstukken aangepakt wordt (De Corte & Verschaffel, 1984). Deze analyse bevestigt Wolters' (1986, p. 322) bewering dat er in vele rekenmethoden een sterk onevenwicht bestaat wat betreft het soort vraagstukjes waarmee gewerkt wordt. Verscheidene door ons onderzochte methoden bevatten uitsluitend oorzaak-veranderingsopgaven waarvan de eindhoeveelheid onbekend is en combinatievraagstukjes waarvan de totaalhoeveelheid gezocht moet worden. Andersoortige oorzaak-veranderings- en combinatie-opgaven en vraagstukjes met een vergelijkingsstructuur kwamen weinig of helemaal niet voor. In tegenstelling tot wat Tavecchio e.a. (1986, p. 315) beweren, weten mensen die de recent ontwikkelde semantische typologie niet kennen, blijkbaar weinig variatie in aanvankelijke rekenvraagstukjes aan te brengen (tenzij we zouden aannemen dat zij bepaalde opgaventypes bewust niet in hun rekenmethode opnemen, hetgeen erg onwaarschijnlijk is).

Kenschetsend voor de door ons bestudeerde rekenmethoden is verder dat daarin een oplossingsmethode aangeleerd wordt waarin alle opgaven in termen van een en hetzelfde diagram (bijvoorbeeld een pijl of een venn-diagram) weergegeven moeten worden alvorens een bijpassende rekenkundige bewerking of formule-opgave te noteren (zie Figuur 2).

Op basis van de gedane bevindingen hebben we toen enerzijds gesteld dat het alleszins noodzakelijk lijkt om in de praktijk het opgavenaanbod te verbreden en tevens de vraag gesteld of het niet beter zou zijn leerlingen aan te zetten om bij de onderscheiden opgaventypes ook verschillende diagrammen te tekenen die nauwer aansluiten bij de semantische structuur van de diverse types, bijvoorbeeld: een pijl voor oorzaak-veranderingsopgaven, vergelijkingsstroken voor vergelijkingsopgaven en het deel-geheel-schema voor combinatie-opgaven. Deze aanpak had tot doel bij de leerlingen de diverse cognitieve schemata tot ontwikkeling te brengen, en hen langs deze weg te helpen bij het opbouwen van een adequate probleemrepresentatie. Een aanpak waarin leerlingen van meetaf aan slechts een uniform diagram voor alle aanvankelijke rekenvraagstukjes bijgebracht wordt, leek ons voor dit doel minder opportuun (De Corte & Verschaffel, 1984; zie ook Wolters, 1986, p. 321).

Deze ideeën werden reeds uitgetoetst in een aantal onderwijsexperimenten (Lindvall, Tamburino & Robinson, 1982; De Corte & Verschaffel, 1985a; Willis & Fuson, 1985). Hoewel de resultaten hoopgevend zijn, bieden ze toch geen voldoende basis om er definitieve conclusies



Figuur 2. Voorbeeld van de aanbiedingswijze van eenvoudige rekenvraagstukjes in een Vlaamse rekenmethode.

uit te trekken. De redenen hiervoor zijn het exploratief karakter van deze studies en een aantal tekortkomingen op methodologisch gebied. Tot slot vermeld ik nog een paar vragen en problemen die gerezen zijn naar aanleiding van deze trainingsstudies en waag ik het om enkele aanbevelingen te doen voor verder onderzoek.

Ten eerste werden in deze onderwijsexperimenten uitsluitend vraagstukjes aangeboden zoals die waarmee ook in de constaterende studies van de laatste jaren gewerkt is. De vraag rijst evenwel hoe de kinderen gereageerd zouden hebben op opgaven die niet volgens een vast stramien opgesteld zijn en (daardoor) minder eenduidig in een van de drie semantische structuurtypes onder te brengen zijn. Zoals Tavecchio (1986, p. 315) en Wolters (1986, p. 323) opmerken, is het zelfs voor ervaren onderzoekers vaak allesbehalve eenvoudig om een bestaand vraagstukje een plaats te geven in het classificatieschema van Greeno c.s., zeker wanneer het opgesteld is door iemand die daar niet zo goed mee vertrouwd is. Hoe meer kans is er dan niet dat jonge basisschoolleerlingen daar grote moeite mee zullen hebben...

Ten tweede, een mogelijk gevaar van zo'n leergang waarin men kinderen verschillende diagrammen leert tekenen voor de onderscheiden semantische structuurtypes, is dat zij een soort oppervlakkige 'sleutelwoordstrategie' gaan ontwikkelen, die hierop neerkomt dat zij zich bij de keuze van het geschikte diagram niet baseren op de globale betekenisstructuur van het vraagstukje, maar op de aanwezigheid van geïsoleerde steeds terugkerende sleutelwoorden uit de opgavetekst, zoals 'meer', 'bij', 'samen'.

Om beide redenen lijkt het me derhalve wenselijk om in vervolgonderzoek veel zorg te besteden aan de constructie en de formulering van de opgaven waarmee zowel tijdens de training als bij de effectmeting gewerkt wordt, maar ook aan de wijze waarop het invoeren van en leren toepassen van de diagrammen concreet geschiedt.

- (1) L. Verschaffel is bevoegdverklaard navorser bij het Belgisch Nationaal Fonds voor Wetenschappelijk Onderzoek.

De auteur dankt Prof. Dr. E. de Corte voor de vele waardevolle commentaren en suggesties bij een eerdere versie van dit artikel.

LITERATUUR

- Briars, D.J., & Larkin, J.H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction*, 1, 245-296.
- Carpenter, T.P., & Moser, J.M. (1982). The development of addition and subtraction problem solving skills. In T.P. Carpenter, J.M. Moser, & T. Romberg (Eds.). *Addition and subtraction: A cognitive perspective* (pp. 2-24). Hillsdale, NJ: Erlbaum.
- De Corte, E., & Verschaffel, L. (1982). Oplossingsprocessen van eersteklassers bij eenvoudige redactie-opgaven. In E. de Corte (Ed.), *Onderzoek van onderwijsleerprocessen: Stro-*

- mingen en actuele onderzoeksthema's. *Bijdragen tot de Onderwijsresearchdagen 1981* (pp. 177-193). 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
- De Corte, E., & Verschaffel, L. (1983). Representatieproblemen van jonge kinderen bij aanvankelijke redactie-opgaven. In E. de Corte & P. Span (Eds.), *Studies over onderwijsleerprocessen. Bijdragen aan een Symposium ter gelegenheid van tien jaar Belgisch-Nederlandse samenwerking* (pp. 33-58). Leuven: Helicon.
- De Corte, E., & Verschaffel, L. (1984). Redactie-opgaven in Vlaamse rekenmethoden voor de eerste klas. In E. de Moor (Ed.), *PANAMA Cursusboek 2. Beschouwingen over reken-/wiskundemethoden* (pp. 15-21). Utrecht: Stichting Opleiding Leraren/Vakgroep Onderzoek Wiskunde-onderwijs en Onderwijs Computercentrum, Rijksuniversiteit Utrecht.
- De Corte, E., & Verschaffel, L. (1985a). Working with simple word problems in early mathematics instruction. In L. Streefland (Ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education. Vol. 1. Individual contributions* (pp. 304-309). Utrecht, The Netherlands: Research Group on Mathematics Education and Educational Computer Center, Subfaculty of Mathematics, University of Utrecht.
- De Corte, E., & Verschaffel, L. (1985b). Oplossingsstrategieën van eersteklassers bij aanvankelijke redactie-opgaven over optellen en aftrekken. *Pedagogische Studiën*, 62, 125-138.
- De Corte, E., & Verschaffel, L. (1987a, April). *The influence of some non-semantic factors on solving addition and subtraction word problems*. Paper presented at the annual meeting of the American Educational Research Association, Washington.
- De Corte, E., & Verschaffel, L. (1987b). Oogbewegingen van eersteklassers tijdens het oplossen van redactie-opgaven. *Pedagogische Studiën*, 64, ter perse.
- De Corte, E., Verschaffel, L. & De Win, L. (1984). De invloed van tekstkenmerken op de representatie- en oplossingsprocessen van jonge kinderen bij eenvoudige optel- en aftrek vraagstukjes. *Pedagogisch Tijdschrift*, 9, 527-538.
- Heller, J.I., & Greeno, J.G. (1978, May). *Semantic processing of arithmetic word problem solving*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago.
- Kintsch, W., & Greeno, J.G. (1985). Understanding and solving arithmetic word problems. *Psychological Review*, 92, 109-129.
- Lindvall, C.M., Tamburino, J.L., & Robinson, L. (1982, March). *An exploratory investigation of the effect of teaching primary grade children to use specific problem-solving strategies in solving simple arithmetic story problems*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- 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.
- Reckzeh, J. (1956). Addition and subtraction situations. *The Arithmetic Teacher*, 3, 94-97.
- Riley, M.S., & Greeno, J.G. (1978, May). *Importance of semantic structure in the difficulty of arithmetic word problems*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago.
- Tavecchio, L.W.C., Beishuizen, M., Van den Berge, J.N., & Bleek, M.W. (1986). Wolters' ongerechtvaardigde conclusies. Een kritisch commentaar op 'De functie van deelgeheel-schema's in het rekenonderwijs: een terugblik' (Wolters, 1984). *Tijdschrift voor Onderwijsresearch*, 11, 312-317.
- Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.), *Addition and subtraction. A cognitive perspective*. Hillsdale, NJ: Erlbaum.
- Verschaffel, L. (1984). *Representatie- en oplossingsprocessen van eersteklassers bij aanvankelijke redactie-opgaven over optellen en aftrekken. Een theoretische en methodologische bijdrage op basis van een longitudinale, kwalitatief-psychologische studie*. Niet-gepubliceerd doctoraatsproefschrift, Seminarie voor Pedagogische Psychologie, Faculteit der Psychologie en Pedagogische Wetenschappen, K.U. Leuven.

- Verschaffel, L. (1985). *Recente ontwikkelingen in het Amerikaans onderzoek over het (leren) oplossen van aanvankelijke redactie-opgaven. Verslag van een studiereis naar de Verenigde Staten, oktober 1984.* (Rapport nr. 31). Leuven: Afdeling Didactiek, Faculteit der Psychologie en Pedagogische Wetenschappen, K.U. Leuven.
- Willis, G.B., & Fuson, K.C. (1985). Teaching representational schemes for the more difficult addition and subtraction verbal problems. In S. Damarin & M. Shelton (Eds.) *Proceedings of the Seventh Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, Columbus, Ohio.
- Wolters, M.A.D. (1984). De functie van deel-geheel-schema's in het rekenonderwijs: een terugblik. *Tijdschrift voor Onderwijsresearch*, 9, 71-83.
- Wolters, M.A.D. (1986). De invloed van semantische structuurkenmerken op het oplossen van redactie-opgaven. *Tijdschrift voor Onderwijsresearch*, 11, 318-324.

Ontvangen 9-2-1987.

Ontvangen publikaties Tijdschrift voor Onderwijsresearch

- Boomsma, Y.: Item selection by mathematical programming. Arnhem, CITO, 1986 (Bulletinreeks nr. 47).
- Boonman, J.H. & W.A.M. Kok: Kennis verwerven uit teksten. Onderzoek naar factoren die van invloed zijn op het verwerven van kennis uit teksten door leerlingen van de achtste groep van de basisschool. Utrecht, VOU, 1986 (Dissertatie).
- Boshuizen, H.P.A. & H.G. Schmidt: The developing structure of medical knowledge. Onderzoek van Onderwijs, nr. 32. Maastricht, Rijksuniversiteit Limburg, 1986.
- Brink, W.P. van den & P. Koele: Statistiek. Deel 3 Toepassingen. Meppel, Boom, 1986.
- Bronneman-Helmerts, R.: Samenhang rond de kinderopvang. Rijswijk, Sociaal Cultureel Planbureau, 1986 (Stukwerk nr. 37).
- Bruinsma, C.: Systematic Errors in Survey-interviews. A methodological study on the effect of significant others and other sources of Bias on evaluations of job qualities and union actions. Amsterdam, VU Uitgeverij, 1987 (Dissertatie).
- Cate, Th.J. ten: Leren in groepen zonder docent. Amsterdam, Universiteit van Amsterdam, 1986 (Dissertatie).
- Dijk, W. van & P. Span: De opvang van begaafde leerlingen. Een gevalstudie. Utrecht, Faculteit Sociale Wetenschappen, 1986.
- Enkevort, G. van, K. Harry, P. Morin & H.G. Schütze (Eds.): Distance Higher Education and the adult learner. Innovations in Distance Education. Occasional Paper of the Dutch Open University, nr. 1. Assen, Van Gorcum & Comp b.v., 1986.
- Gijtenbeek, J. & F.S.J. Riemersma: Opleidingen in de gezondheidszorg en allochtone bevolkingsgroepen. Den Haag, SVO, 1985 (Selecta-reeks).
- Graaff, E. de, M.L. Schreurs & Y. Winants: Kennismaking met probleemgestuurd onderwijs in de Gezondheidswetenschappen. Onderzoek van Onderwijs, nr. 31. Maastricht, Rijksuniversiteit Limburg, 1986.
- Grob, P., R. Husken & P. van der Kley (Red.): Eigen-wijs in Onderwijs: beleving, verzuim en verlaten van school. Nijmegen, ITS, 1986 (OOMO-reeks).
- Hamaekers, A., M. Lkoundi & J.M.F. Teunissen: Materiaalontwikkeling voor het onderwijs in de arabishe taal en cultuur. Utrecht, Vakgroep Onderwijskunde, 1986.
- Harbers, H.: Sociale wetenschappen en hun speelruimte. Groningen, Wolters-Noordhoff, 1986 (Dissertatie).
- Hout Wolters, B. van: Markeren van kerngedeelten in studieteksten. Een proces-produkt benadering. Lisse, Swets & Zeitlinger b.v., 1986 (Dissertatie).
- Jager, A. & R. van 't Klooster: Allochtonen leren lesgeven. De situatie van allochtone studenten op de lerarenopleiding. Utrecht, VOU, 1986.
- Janssen, P.J., G. van der Steene & H. Vertommen: Intelligentie. Tussen meten en weten in theorie en praktijk. Festschrift voor Prof. Dr. J. Stinissen. Leuven, Universitaire Pers, 1986.
- Jansen Schoonhoven, A., J. van der Linden, M. Posthuma de Boer, P. Span & M. van Susante: Hoogbegaafde leerlingen. Een onderzoek naar het onderkennen en onderwijzen van hoogbegaafde leerlingen in het voortgezet onderwijs. Deel 2 en Deel 2A (bijlagen). SVO-project 51904. Utrecht, Vakgroep Onderwijskunde, 1987.
- Jochems, W. & F. Montens: Hoe toets je taalvaardigheid? Meppel, Boom, 1986.
- Keizer, H.: Cognitieve leerdoel voor itembanken: twee opvattingen toegepast bij het vak biologie. Arnhem, CITO, 1986 (Bulletinreeks nr. 52).
- Kok, W.A.M. (Red.): Computer ondersteund onderwijs binnen het aanvankelijk rekenen. Verslag van een onderzoek naar de mogelijkheden van het gebruik van een micro-computer als onderwijsleermiddel in de onderbouw van de basisschool. Utrecht, VOU, 1986.
- Ledoux, G. & R. Kabdan: OECT in het speciaal onderwijs. Samenvatting van de resultaten van een onderzoek naar ervaringen met en opvattingen over onderwijs in eigen taal en cultuur in het speciaal onderwijs. Amsterdam, SCO, 1986 (SCO Cahier nr. 37).
- Meesters, D. & H. van Os: Flexibiliteit van zaakvakmethoden. Ontwikkeling en toepassing van een analyse-instrument m.b.t. open-/geslotenheid en differentiatiecapaciteit van zaakvakmethoden. Utrecht, VOU, 1986.

- Meij, H. van der: Questioning. A study on the questioning behavior of elementary school children. Den Haag, SVO, 1986 (Selectareeks).
- Pennings, A.H. e.a.: Letterwerk. Computer Ondersteund Onderwijs aan kinderen met leerstoornissen (SVO-project 310.4192). Utrecht, Vakgroep Orthopedagogiek en Klinische Pedagogiek/Vakgroep Onderwijskunde, 1987.
- Pennings, A.H. e.a.: Letterwerk. Handleiding bij het programma letterwerk (SVO-project 310.4192). Utrecht, Vakgroep Orthopedagogiek en Klinische Pedagogiek/Vakgroep Onderwijskunde, 1987.
- Scheerens, J.: Enhancing educational opportunities for disadvantaged learners. A review of Dutch research on compensatory education and educational development policy. Amsterdam, North-Holland Publishing Company, 1987.
- Span, P., W.A.M. Kok & F.M. Stevens: Computers en begaafde leerlingen in het voortgezet onderwijs. Utrecht, Vakgroep Onderwijskunde, 1986.
- Wesdorp, H., E. Teyse, Fr. Daems & R. Rymenans: De positie van het onderwijs in het Nederlands en de rol van de overheid in Nederland en Vlaanderen. Vlaams-Nederlands overleg onderwijsresearch en -documentatie (OORD), Nederlandse Taalunie, Den Haag, 1986.
- Wolbert, R.G.M., I. Schaap-Hummel & P. Span: Individualisering en differentiatie in de basisschool. Een verkennend onderzoek naar de stand van zaken m.b.t. individualisering en differentiatie in groep 1 t/m 4 van de basisschool + bijlagen (SVO-project 716.453). Utrecht, Vakgroep Onderwijskunde, 1986.

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie

Kees Beintema (RUL), Peter van den Dool (SCO), Dirk de Jong (LICOR), Ewoud Roede (SCO) en Peter van der Sijde (UT, TO, secretaris).

Redactie-adres

Peter van der Sijde, Faculteit der Toegepaste Onderwijskunde, Vakgroep Instructietechnologie, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893560.

Jrg. 11, nr. 3

Juni 1987

Inhoudsopgave

Van de redactie	176
Van de bestuurstafel	176
Symposium "Het programmeren van onderwijsonderzoek"	177
Mededelingen	178
Proefschriften	180
Commentaar	182
Nieuwe publikaties	183

Van de redactie

Dit nummer van het VOR-Bulletin heeft een zeer beperkte omvang. De reden hiervoor is niet dat er een gebrek aan copy is, maar aan bladzijden. Het VOR-Bulletin heeft per jaar 72 pagina's tot zijn beschikking, dat wil zeggen 12 pagina's per aflevering. De eerste twee nummers van deze jaargang overschreden dit aantal.

Ten overvloede maken wij u erop attent, dat u kunt bijdragen aan het Bulletin door te reageren op reeds gepubliceerde bijdragen en copy aan te leveren voor de vaste rubrieken.

Van de bestuurstafel

1. Het VOR-bestuur heeft – naar de SVO, i.a.a. Kamercommissie O&W – bezorgdheid uitgesproken over het gebrek aan middelen om nieuw OTG/ANGO-onderzoek voor 1988 te kunnen financieren.
2. Naar aanleiding van gerezen problemen zal nader aandacht besteed worden aan het vraagstuk van privacy-bescherming van subjecten en onderzoekbelangen: wellicht kunnen gedragsregels voor onderwijsonderzoekers nuttig zijn.
3. Met de voorzitter van de AERA, Shavelson, zal overlegd worden over de relatie van de VOR met de AERA, ook i.v.m. de participatie van VOR-leden aan de Annual Meetings.
4. Over de verschijningsdata van de ORD-bundels wordt overleg geopend met de uitgever ten einde de verschijning te vervroegen.
5. Het bestuur voert een ledenwerfactie; enkele honderden potentiële leden werd een wervende folder toegezonden. Deze folder is aan te vragen bij mevr. dr. B. van Hout Wolters, Universiteit Twente, Toegepaste Onderwijskunde (IST), Postbus 217, 7500 AE Enschede.

Symposium "Het programmeren van onderwijs-onderzoek"

Tijdens de Onderwijsresearch Dagen te Groningen werd door de redactie een symposium georganiseerd ter afsluiting van de reeks bijdragen in dit Bulletin over het programmeren van onderzoek. Voor het symposium waren sprekers uitgenodigd die het beleid, het veld en de tussenpersonen vertegenwoordigden. Het symposium werd door ± 45 personen bijgewoond.

Naar aanleiding van het symposium wordt een publicatie samengesteld. Deze kan worden besteld bij de redactiesecretaris (telefonisch en schriftelijk). De prijs zal ongeveer f 17,50 bedragen en de verschijningsdatum is gesteld op augustus 1987. Hoe meer mensen intekenen, hoe goedkoper zal het boekje worden!

Programmering van onderzoek

Onder redactie van de VORBulletin-redactie bestaande uit
K. Beintema, P. C. van den Dool, E. Roede & P. C. van der Sijde

Inhoud:

Redactie Inleiding

Roede, Beintema, Van den Dool, De programmering van onderzoek.

Van den Dool, Programmering en de competentieontwikkeling van onderzoekers.

Roede, De OTG's en de programmering van onderzoek.

Van den Dool, Problemen verkennen of problemen maken? De rol van de probleemverkenner in het programmeringsproces van SVO.

De Jong, Ook bij programmering van onderzoek zijn korte lijnen effectief.

Koornstra, Programmering van onderzoek in universitair perspectief.

Van Hoesel, Een programmeringsstructuur voor de onderwijsresearch.

Van Oyen, Naar een opbrengstgericht onderzoeksbeleid. Minder maar betere programmering van onderzoek.

Schepel, Kotte, Over programmering gedacht, gesproken, en ... gezwezen.

Lodewijks, Landelijke programmering van beleids- en praktijkgerichte onderwijsresearch.

Van der Linden, Moet eerste-geldstroom onderzoek geprogrammeerd worden?

Zijlstra, Pleidooi voor een programmeringsbeleid.

Hetebrij, Van Amersfoort, Monsma, Programmering van researchprogramma's aan de hand van twee voorbeelden.

Blok, titel bijdrage nog onbekend

Mededelingen

Samenwerkingsproject PPON van Cito en SVO:

Landelijke peilingen onderwijsniveau starten met rekenen

Een paar honderd scholen hebben dezer dagen een uitnodiging gekregen om mee te doen aan de eerste van een reeks peilingen van het niveau van het basisonderwijs in ons land. Deze eerste peiling betreft het rekenonderwijs. De uitkomsten moeten duidelijk maken wat scholen hun leerlingen op rekengebied proberen te leren en wat leerlingen feitelijk van het rekenonderwijs opsteken. De peiling dit voorjaar is gericht op leerlingen in het laatste leerjaar. In de herfst volgt een peiling, gericht op leerlingen halverwege het basisonderwijs.

De uitnodiging aan de scholen om mee te doen, gaat uit van het project 'Periodieke peiling van het onderwijsniveau', afgekort PPON. Dit is een samenwerkingsproject van Cito (Instituut voor Toetsontwikkeling) en SVO (Instituut voor Onderzoek van het Onderwijs). De opdracht tot het project werd gegeven door het Ministerie van Onderwijs en Wetenschappen. Het project is ondergebracht bij het Cito in Arnhem. SVO draagt er zorg voor dat universitaire en andere onderzoekers ondersteunende of aanvullende 'dieptestudies' uitvoeren. De uitkomsten van het project zullen er vooral toe moeten bijdragen dat de feitelijke situatie in het onderwijs in de discussies over de kwaliteit van het onderwijs niet uit het oog verloren wordt.

Volgende peilingen

Na de dit jaar te realiseren twee peilingen voor rekenen volgen in de komende jaren peilingen voor achtereenvolgens de 'leer- en vormingsgebieden': Nederlandse taal (1988), verkeerseducatie (1988), sociale wereldoriëntatie (1989), bewegingsonderwijs (1989), kunstzinnige vorming (1989/1990), natuuronderwijs en gezondheidseducatie (1990) en Engelse taal (1990). Daarbij is per gebied telkens sprake van een peiling aan het einde en een peiling halverwege het basisonderwijs. Tegen het einde van 1990 moet zo een eerste volledige ronde van peilingen zijn afgerond. Aan de hand van de ervaringen in de komende jaren zal een besluit worden genomen over de opzet van een volgende ronde. Opeenvolgende peilingsrondes zijn vooral van belang om vast te kunnen stellen of er door de jaren heen wijzigingen optreden in het leer- en vormingsaanbod en de onderwijsleerresultaten.

Werkwijze

Ten behoeve van de peiling van het niveau van het rekenonderwijs is een gedetailleerd overzicht gemaakt van de 'leer- en vormingsdoelstellingen' op het gebied van het rekenen. Het overzicht is vervolgens aan de hand van de commentaren van diverse deskundigen vervolmaakt. Daarna zijn er aan onderwijsgevendende voor te leggen vragenlijsten gemaakt, waarmee nagegaan kan worden welke leer- en vormingsdoelen feitelijk in de onderwijspraktijk worden nagestreefd. En er werden vragen en opdrachten ontworpen waarmee nagegaan kan worden wat kinderen feitelijk van het basisonderwijs opsteken. Sommige van die vragen en opdrachten zijn schriftelijk en kunnen groepsgewijs afgenomen worden. Andere zijn mondeling en moeten individueel door leerlingen beantwoord worden. Opgaven van het meerkeuzetype komen in het ontworpen materiaal nauwelijks voor.

Steekproef

Om betrouwbare gegevens over het niveau van het rekenonderwijs te krijgen, is het voldoende dat een steekproef van 350 scholen medewerking verleent aan het invullen van de vragenlijsten door onderwijsgevendende en het maken van de vragen en opdrachten door leerlingen. De afname van vragen en opdrachten hoeven de onderwijsgevendende niet zelf te verzorgen. De scholen die aan een peiling meedoen, worden bezocht door geoefende toetsassistenten met onderwijserva-

ring. Niet iedere leerling hoeft alle vragen en opdrachten te maken. Met het invullen van de vragenlijsten en het afnemen van de vragen en opdrachten is ongeveer één ochtend gemoeid. De verwerking van de vragenlijsten en de toetsresultaten gebeurt anoniem.

De samenstelling van de steekproef van scholen is overigens zodanig dat in de rapportage diverse groepen van scholen en leerlingen onderscheiden kunnen worden, bijvoorbeeld de categorie scholen met een meer dan gemiddeld percentage arbeiderskinderen maar nog betrekkelijk weinig buitenlandse kinderen en de categorie scholen met veel buitenlandse kinderen.

Verslaggeving

Per peiling publiceert PPON een verantwoording voor wetenschappelijk geïnteresseerden. Daarnaast verschijnt er per peiling een samenvattend rapport van de uitkomsten. Door de gekozen onderzoeksopzet zijn daarin geen gegevens te vinden over afzonderlijke scholen of individuele leerlingen. Alle scholen die aan een peiling deelnemen, ontvangen een exemplaar van het betreffende rapport. En tenslotte zal PPON per leer- en vormingsgebied een voorlichtingsbrochure verzorgen met daarin de belangrijkste uitkomsten van de peilingen. Alle scholen voor primair onderwijs krijgen een exemplaar van elke brochure toegestuurd. Ook voor andere geïnteresseerden zijn exemplaren van de brochures te verkrijgen.

Informatie

Wie meer informatie wil over het project PPON of over de peiling met betrekking tot het rekenonderwijs, kan terecht bij de afdeling Voorlichting en documentatie van het Cito, postbus 1034, 6801 MG ARNHEM, telefoon (085) 521208. Bij deze afdeling kan ook een folder met beknopte algemene informatie over PPON worden aangevraagd.

Proefschriften

Rijlaarsdam, G.C.W.: Effecten van leerlingrespons op aspecten van stelselvaardigheid.

Korte samenvatting van het onderzoek

In een veldexperiment (HAVO, VWO, derde klas) zijn twee stelonderwijsprogramma's met elkaar geconfronteerd die op een hoofdvariabele van elkaar verschilden. In het experimentele programma becommentarieerden leerlingen elkaars opstellen, in de controle conditie gaf de docent feedback. In beide condities werd met dezelfde reactie- en commentaarformulieren gewerkt. Beide programma's waren ook voor het overige identiek. De programma's bestonden uit vier cursusblokken, die elk 10 à 12 lessen in beslag namen. De cursusblokken waren identiek gefaseerd. Ze bestonden uit voorbereidende lessen, reflectieopdrachten, commentaarlessen, commentaarverwerklessen en herschrijflessen.

Verwacht werd dat het experimentele programma ('peer evaluation') superieur zou zijn op drie aspecten van stelvaardigheid. Onderscheiden werden schrijfprestatievariabelen (Doelgerichtheid, Publiekgerichtheid, Opbouw en Stijl), attitudevariabelen, samenhangend met schrijfangst (Angst-voor-het-niet-Kunnen-Schrijven, Houding-jegens-het-Beoordeeld-worden; Houding-jegens-Schrijven) en diverse schrijfprocesvariabelen. Voor alle variabelen werden instrumenten geconstrueerd.

Het experiment werd uitgevoerd op acht verschillende scholen waar elf docenten ieder twee derde klassen onderwees, die aselekt aan een conditie waren toegewezen. Voor, tijdens en na de totale cursus schreven de leerlingen een betogend opstel, uitsluitend voor het onderzoek. De taak voor de voor- en tussentoets was identiek, die voor de natoets was soortgelijk. Tevens werd het attitude instrument afgenomen. Tijdens de voortoets- en tussentoetsperiode werd aan twee leerlingen per klas een hardopschrijfprotocol ontlokt. De implementatie van de onderwijsprogramma's werd gecontroleerd door docent- en leerlinglogboeken, observaties en vragenlijsten.

De resultaten waren betrekkelijk eenduidig. Het verschil in de onderwijsprogramma's resulteerde niet in verschillen op Schrijfprestatie- en Psychologische Variabelen. Op de analytisch gescoorde items voor Doelgerichtheid en Publiekgerichtheid bleek geen verschil tussen de condities. Wel nam het aantal opstellen toe die een heldere lijn bevatten, en waarin een vorm van interactie met de lezer werd gebruikt. In beide condities waren de scores op de tussentoets en natoets voor alle variabelen duidelijk hoger dan op de voortoets. Een uitzondering vormden de scores op de variabele Publiekgerichtheid op de natoets. Wellicht ten gevolge van een opdracht-effect waren deze scores zelfs lager dan de score op de voortoets. Op de Schrijfproces Variabelen deden zich ook wel verschillen voor. In de experimentele conditie nam het aantal zelfinstructies toe, in de controle conditie af; in de controle conditie nam het aantal keren toe dat leerlingen zich keuzen voorlegden op het punt van formuleringen; in de experimentele conditie nam het aantal af. In beide condities werd er op de natoets minder gegenereerd, meer gestructureerd en vaker een woord of zin veranderd.

Er zijn enkele exploraties verricht waaruit voor vrijwel alle Schrijfprestatievariabelen op de tussentoets en de natoets een effect van de factor Docent bleek. Ook bleek een effect van de factor Sexe, maar er kon geen consistent patroon aangetoond worden van statistisch significante hoofdeffecten van sexe, noch van interactie-effecten van Sexe en Programma. Uit enkele correlatieberekeningen bleek dat de Psychologische Variabelen niet of nauwelijks samenhangen met Stelprestatie Variabelen op de drie toetsmomenten; maximaal 9% van de variantie van de scores op de categorie Stijl wordt verklaard door scores op Psychologische Variabelen. Schrijfprestatiescores van leerlingen bleken instabiel: correlatie tussen voortoets- en tussentoetsscores waren laag.

In de discussie worden vraagtekens geplaatst bij de resultaten van het onderzoek. Wellicht leren leerlingen wel van het geven van feedback, maar krijgen zij minder goede feedback van

medeleerlingen dan van docenten. Uit controle via vragenlijsten bleek echter dat in beide condities weinig leerlingen ontevreden waren over de duidelijkheid, zorgvuldigheid en behulpzaamheid van de feedback die zij kregen. Misschien doet feedback er niet zo toe: er bleek geen significante samenhang tussen stelprestatiescores en ontevredenheid over de feedback. Een derde mogelijke verklaring voor het uitblijven van verschillen is dat leerlingen in de experimentele conditie weliswaar criteria voor goede teksten verwerven, maar dat kennis niet direct leidt tot het toepassen in schrijftaken. Ervan uitgaande dat de groei die geconstateerd werd in stelvaardigheid niet alleen toe te schrijven is aan louter cognitieve groei, werd geconcludeerd dat peer feedback en docent feedback gelijkwaardige concurrenten zijn.

- Rijlaarsdam, G.C.W.; *Effecten van leerlingenrespons op aspecten van stelvaardigheid* (vert. symmary H.S. Lake, Bussum/Amsterdam). Stichting Centrum voor Onderwijs onderzoek van de Universiteit van Amsterdam (SCO-rapport; nr. 88). Tevens diss. U.v.A. Amsterdam, 1986.
- Rijlaarsdam, G.C.W.; *Effecten van leerlingenrespons op aspecten van stelvaardigheid*. Bijlagen-rapport 1: Onderwijsprogramma's. Het experimentele en het alternatieve stelonderwijs-programma; leerlingenboek en docentenhandleiding. S.C.O., Amsterdam, 1986.
- Rijlaarsdam, G.C.W.; *Effecten van leerlingenrespons op aspecten van stelvaardigheid*. Bijlagen-rapport 2: Instrumenten. S.C.O., Amsterdam, 1986.

Kort commentaar: G. Beukhof

Het proefschrift beschrijft een empirisch onderzoek van een vaak – ten onrechte – vergeten deel van het moedertaalonderwijs, nl. het stelvaardigheidsonderwijs. De nadruk ligt hierbij niet zo zeer op het traditionele tekstbestuderen maar meer op tekst*productie*.

Het onderwijs is uitgevoerd bij de Onderzoeksgroep Taalonderwijs van S.C.O. (UvA), onder leiding van dr. H. Wesdorp¹⁾. De auteur stelt terecht dat de theorie van de moedertaaldidactiek nog in de kinderschoenen staat. Ook het theoretisch gehalte van de didactiek van het stellen is laag. Ondanks deze beperkingen levert deze dissertatie een schat aan informatie op. De auteur heeft in zijn onderzoek (en in de beschrijving ervan) vele facetten van het stelvaardigheidsonderzoek en -onderwijs meegenomen. De veelheid aan informatie stelt hoge eisen aan de lezer van de dissertatie.

Waarderend kan gesproken worden over de uitvoerige en goed beargumenteerde beschrijvingen van het theoretisch kader, de onderzoeksliteratuur, de samenstelling van de experimentele onderwijsprocedure, de instrumentatie, de methoden van procedures en de resultaten.

In de discussie wordt een poging gedaan de resultaten te verklaren en consequenties te trekken voor de onderwijspraktijk.

Naast waardering passen toch enkele kanttekeningen:

1. Vergelijkend onderzoek tussen didactische werkvormen is een hachelijke zaak (docent feedback vs leerlingresponse). Beter is het *de taak* als uitgangspunt te nemen en op grond hiervan bepaalde leerlingkenmerken (en ook leerprocessen) te veronderstellen. In onderwijspsychologisch onderzoek wordt hiervoor de zgn. correspondentie-analyse voorgesteld.
2. Op grond hiervan is de keuze van leerprocesvariabelen en leerlingkenmerken (alleen sexe, intro-extraversie en angst) nogal discutabel.
3. In de analyse van de literatuur wordt geen melding gemaakt van de leeftijd van de proefpersonen in de diverse onderzoeken, evenmin over het onderwerp waarover de stelvaardigheidsoopdrachten gingen.
4. Aansluitend op het gestelde in 1.: de auteur stelt dat de tweekamp tussen leerlingresponse en docent feedback een onbeslist karakter heeft door a) ongelukkige keuze van de afhankelijke variabelen en b) gebrekkige statistische onderwerpen.
De vraag is of het – op grond van *taakvariabelen* – aannemelijk is dat er verschillen zijn!
5. De literatuur m.b.t. metacognitie wordt slechts zijdelings aangebracht.
6. Het onderwijsexperiment vergde één schooljaar. Hoewel getracht is alle (storende) variabelen zoveel mogelijk onder controle te houden resp. te elimineren, blijft het moeilijk (zo niet onmogelijk) de gevonden resultaten eenduidig te interpreteren.
7. De auteur stelt dat 'de resultaten verrassend eenduidig zijn, maar onverwacht, gezien de hooggespannen verwachtingen van de practici en de onderwijskundige verklaringen'.
(Inter)nationale onderzoeksliteratuur laat zien dat tutoring en/of peer teaching niet zonder meer leidt tot hogere leerprestaties. Soms wel tot andere leerprocessen en attitudes.
8. Terecht stelt de auteur dat meer onderzoek nodig is naar de kwaliteit en de effectiviteit van feedback (docent feedback en leerlingresponse). Bovendien zal onderzoek geïnitieerd moeten worden naar kennis van de criteria en de schrijfprestaties.
9. Voor de praktijk betekent dit voorlopig – volgens de auteur – dat de keuze tussen docent feedback en leerlingresponse afhankelijk is van de vaardigheden van de docent, de eigenschappen van de klas en de rol van de leerlingen in het onderwijsleerproces. Maar: wisten we dit ook niet? Dit onderzoek levert nog geen concrete voorschriften op voor de onderwijspraktijk. Het onderzoeksgebied van het stelonderwijs is daarvoor nog te jong. Dit proefschrift kan een belangrijke stap voorwaarts zijn voor systematisch onderzoek op verschillende deelterreinen van het stelvaardigheidsonderzoek.

¹⁾ Zie Onderzoeksgroep Taalonderwijs, *Activiteiten en perspectieven*. Amsterdam: S.C.O., U.v.A., 1986.

Nieuwe publicaties

- Lieshout, C.F.M. van, et al. (1986). Zelfstandigheidsontwikkeling in het basisonderwijs. Nijmegen: ITS.
- Berg, J. van de, en Tillaart, H. van de (1986). Nieuwe werknemers. Nijmegen: ITS.
- Grob, R., Husken, R., en Kley, P. van der (1986). Eigen-wijs in onderwijs: beleving, verzuim en verlaten van school. Nijmegen: ITS.
- Vries, B. de (1986). Bemiddelen op de stagemarkt. Nijmegen: ITS.
- Pelkmans, A., en Kuijk, J. van (1986). Samenwerking tussen scholen voor voortgezet onderwijs. Band 1: Tekst. Nijmegen: ITS.
- Pelkmans, A., en Kuijk, J. van (1986). Samenwerking tussen scholen voor voortgezet onderwijs. Band 2: Bijlagen. Nijmegen: ITS.
- Römkens, L., en Vries, B. de (1986). Arbeidsorganisaties en hun stagebeleid. Nijmegen: ITS.
- Beijaard, D. (1986). Didactisch agrarisch praktijkschoolonderwijs in Nederland. Deelverslag 2: explicitering en systematisering van kenmerken van het didactisch handelen. Wageningen: Vakgroep Pedagogiek en Didactiek, L.H. (f 27,50).
- OOMO (1987). Jaarverslag 1986. Nijmegen: ITS.
- Boland, T. en Mommers, M.J.C. (1987). Verschillen tussen doorstromers en achterblijvers aan het eind van het basisonderwijs. Deelrapport I. Nijmegen: Katholieke Universiteit Nijmegen.
- Boland, T. en Mommers, M.J.C. (1987). Schoolvorderingentoetsen aan het eind van het basisonderwijs: verschillen en overeenkomsten. Deelrapport III. Nijmegen: Katholieke Universiteit Nijmegen.
- Visser, J.J.C.M. en Voeten, M.J.M. (1987). Het advies voor voortgezet onderwijs: de invloed van leesvaardigheid in leerjaar drie en zes, sociaal milieu en geslacht. Deelrapport V. Nijmegen: Katholieke Universiteit Nijmegen.
- Franken, M. en Schilten, K. (1986). Onderwijs aan vrouwen: een vak apart (DOZ Boek). Utrecht: De Som (f 12,50).
- Merriënboer, J.J.G. van en Krammer, H.P.M. (1986). Leren programmeren in het voortgezet onderwijs. Enschede: Universiteit Twente.
- SLO catalogus 86/87 Basis- en speciaal onderwijs, opleidingen. Enschede: SLO.
- SLO catalogus 86/87 Voortgezet onderwijs, opleidingen, volwassenen-educatie. Enschede: SLO.
- Koetsier, C.P., Driel, C. van, en Janzen, R.W. (1986). Simulatielessen en het oefenen van nabespreken in het toekomstige 12-maandsprogramma van de afdeling didactiek van de biologie. Utrecht: Afdeling Didactiek van de Biologie, RUU.
- Koetsier, C.P., Driel, C. van, en Janzen, R.W. (1986). Theoretisch kader voor de ontwikkeling van een instrument voor het zichtbaar maken van verworven leraarsbekwaamheden in het toekomstige 12-maandsprogramma van de afdeling didactiek van de biologie. Utrecht: Afdeling Didactiek van de Biologie, RUU.
- Graaff, E. de, Schreurs, M.L., en Winants, Y. (1986). Kennismaken met probleemgestuurd onderwijs in de gezondheidswetenschappen. Maastricht: Rijksuniversiteit Limburg (OvO 31).
- Boshuizen, H.P.A. en Schmidt, H.G. (1986). The developing structure of medical knowledge. Maastricht: Rijksuniversiteit Limburg (OvO 32).
- Cate, Th.J. ten (1986). Leren in groepen zonder docent. Amsterdam: Faculteit Geneeskunde (dissertatie, f 20,-).
- Hout Wolters, B. van (1986). Markeren van kerngedeelten in studieteksten; een proces-product benadering (dissertatie, KUB). Lisse: Swets & Zeitlinger.
- Knol, D.L. (1986). Inventarisatie van automatische itemselectie procedures voor Raschschalen. Enschede: Universiteit Twente (f 8,-).
- Buning, G.R., Eggen, T.J.H.M., Kelderman, H., Linden, W.J. van der (red.) (1986). Het gebruik van het Raschmodel voor een decentraal toetsservicesysteem. Enschede, Universiteit Twente (f 25,-).
- Kogut, J. (1986). Review of IRT-based indices for detecting and diagnosing aberrant response patterns. Enschede, Universiteit Twente (f 6,50).
- Knol, D.L. (1986). Een overzicht van meerdimensionale itemresponsmodellen. Enschede, Universiteit Twente (f 6,50).
- Linden, W.J. van der (1986). The use of test scores for classification decisions with threshold utility. Enschede, Universiteit Twente (f 5,-).
- Kelderman, H. (1986). Item bias detection using the loglinear Rasch model observed and unobserved subgroups. Enschede, Universiteit Twente (f 8,-).

- Boekkooi-Timminga, E. (1986). Simultaneous test construction by zero-one programming. Enschede, Universiteit Twente (f 5,-).
- Linden, W.J. van der en Boekkooi-Timminga, E. (1986). A zero-one programming approach to Gulliksen's matched random subtest method. Enschede, Universiteit Twente (f 5,-).
- Burg, E. van der, Leeuw, J. de en Verdegaal, R. (1986). Homogeneity analysis with k sets of variables: an alternating least squares method with optimal scaling features. Enschede, Universiteit Twente (f 10,-).
- Linden, W.J. van der en Eggen, T.J.H.M. (1986). An empirical Bayes approach to item banking. Enschede, Universiteit Twente (f 6,50).
- Boekkooi-Timminga, E. (1986). Algorithms for the construction of parallel tests by zero-one programming. Enschede, Universiteit Twente (f 5,-).
- Eggen, T.J.H.M. en Linden W.J. van der (1986). The use of models for paired comparisons with ties. Enschede, Universiteit Twente (f 5,-).
- Rikers, J.H.A.N. en Pelgrum, W.J. (1986). Informatiewinning uit onderwijsgegevensbestanden met de microcomputer. Enschede, Universiteit Twente (f 15,-).
- Verhagen, P.W. en Martens, W.Th. (1986). Formatieve evaluatie van korte films: de beproeving van een instrument. Enschede, Universiteit Twente (f 10,-).
- Delft, T.J.M. van en Krogt F.J. van der (1987). Externe contacten van scholen: een literatuuronderzoek en empirische verkenning naar het contactprofiel en de grensorganisatie van scholen in het M.T.O. Enschede, Universiteit Twente (f 18,-).
- OTG Nieuwsbrief 14. (OTG Onderwijsleerprocessen).
- Kuijk, J. van en Kessel, N. van (1986). Ophef over opheffing of fusie. Nijmegen: ITS.
- Kuijk, J. van en Kessel, N. van (1986). Ophef over opheffing of fusie. Bijlage: casestudies. Nijmegen: ITS.

Vocaalsubstitutiefouten bij bekende en onbekende eenlettergrepige woorden

Margo G.H. Jansen en Anja W.M. Knuver
Vakgroep IDOK, RU Groningen

ABSTRACT

According to several writers on the subject of spelling, children at the early stages of learning to spell adopt a phonological strategy. Later on there is a shift from phonemic strategies to 'higher level' strategies that use visual as well as lexical information. In this paper we are trying to show that this shift begins earlier than is generally assumed, by studying two types of spelling errors, of which the relative frequency of occurrence can be viewed as evidence for the use of visual information. The discrepancy between the frequency of the two types of spelling errors proves to vary systematically with the familiarity of the words and the spelling ability and experience with the written language of the subjects. The substantial interpretation of these results, however, is fraught with difficulties.

1. INLEIDING

Het leren spellen wordt in de literatuur beschreven als een ontwikkelingsproces. Beginnende spellers worden geacht andere spellingsstrategieën te gebruiken, en daardoor ook andere spellfouten te maken, dan bijvoorbeeld oudere leerlingen. Voor het kunnen bijsturen en eventueel corrigeren van de leerprocessen die zich voordoen als gevolg van het spellingsonderwijs, is kennis van de ontwikkeling van de spellingsvaardigheid van groot belang. Uit de literatuur kan worden afgeleid dat wordt aangenomen dat de ontwikkeling van de spellingsvaardigheid globaal verloopt van een op het individuele foneem gerichte regelstrategie bij de beginnende speller tot een meer visueel-orthografische strategie bij de oudere en volwassen speller (Marsh e.a., 1980; Frith, 1980; Verhoeven, 1983; Verhoeven, 1985; Kruidenier, 1985).

Het blijkt dat sommige kinderen al beginnen met 'schrijven' voordat ze formeel spellingsonderwijs krijgen. Ook deze zogenaamde spontane spellingen zijn aan zekere wetmatigheden onderhevig, waaruit blijkt dat kinderen inzicht hebben in de fonologische kenmerken van woorden. Onderzoek op dit terrein werd gedaan door Read (1975). Soortgelijk onderzoek is ook bij Nederlandse leerlingen gedaan (Van Rijnsoever, 1985). In het aanvankelijk spellingsonderwijs leren de kinderen dat er een systematische relatie is tussen klanken en schrijftekens. De leerinhouden die in dit stadium aan de orde komen vertonen twee belangrijke grondkenmerken: het gaat om woorden die frequent voorkomen en waarvoor geldt dat de relatie tussen de klank en het schrijfteken eenduidig is. Om een spellingsprobleem te kunnen oplossen moet een leerling over de volgende deelvaardigheden beschikken: De leerling moet een auditief aangeboden spelwoord in de afzonderlijke fonemen kunnen segmenteren en de leerling moet de juiste grafemen aan de fonemen kunnen koppelen en deze in de goede volgorde van links naar rechts kunnen neerschrijven. Deze procedure noemen we de fonologische strategie.

Het zal duidelijk zijn dat de fonologische strategie faalt bij het spellen van woorden waarbij er geen éénduidige klank-tekenrelatie is. Wanneer we 'paart' voor /paart/ schrijven, voldoet deze spelling aan de klank-teken correspondentieregels, maar niet aan andere afspraken die er voor hebben gezorgd dat de correcte schrijfwijze 'paard' is.

Aangezien leerlingen na verloop van tijd beter gaan spellen dan op grond van een zuiver fonologische strategie verwacht mag worden, is het aannemelijk dat ook andere informatie dan

klank-tekencorrespondentieregels een rol gaat spelen. Bij veel auteurs vinden we de opvatting dat de fonologische strategie geheel of gedeeltelijk plaats maakt voor een visueel-motorische strategie. Zo veronderstelt Simon (1976) bijvoorbeeld, dat voor sommige korte hoogfrequente woorden een grafemische representatie in het geheugen aanwezig is die bij het spellen direct kan worden gecopieerd. Over de aard van dit 'woordbeeld' bestaat onduidelijkheid. Volgens Ehri (1980) moet het woordbeeld niet worden opgevat als een uit het hoofd geleerd geometrisch figuurtje, maar als een sequentie van lettertekens die een systematisch verband houdt met de fonologische kenmerken van het woord. Soortgelijke opvattingen vinden we ook bij Nederlandse auteurs (Mommers, 1985; Reitsma, 1985). Replicaties van Ehri's experimenten met Nederlandse leerlingen werden uitgevoerd door Langereis en Denys (1985), die op basis van hun resultaten o.a. concludeerden dat onervaren spellers de beschikking hebben over grafemische informatie over het gehele woord.

Sommige auteurs veronderstellen dat alle spellers zich geleidelijk ontwikkelen in de richting van een visueel-motorische strategie, anderen nemen aan dat er ook sprake is van individuele verschillen, op basis van ervaring, geschooldheid en spellingsvaardigheid van de speller (Baron, 1980; Assink, 1983; Reitsma, 1985; Verhoeven, 1985).

Hoe de veronderstelde verschuiving van de ene naar de andere strategie plaats vindt en wanneer een en ander optreedt is eveneens onduidelijk. Verhoeven (1985) vindt dat er in de periode van de vijfde klas van het lager onderwijs tot en met het tweede leerjaar van het voortgezet onderwijs een verschuiving optreedt van regelstrategie naar woordbeeldstrategie. Ook Marsh e.a. (1980) nemen aan dat de overschakeling omstreeks het vijfde leerjaar plaats vindt. Volgens andere auteurs echter zijn er aanwijzingen, dat al betrekkelijk vroeg in de ontwikkeling van de spellingsvaardigheid van een visuele oriëntatie op het woordbeeld sprake is (Beers, 1980; Kruidenier, 1985; Jansen & Kruidenier, 1986).

2. PROBLEEMSTELLING

In dit onderzoek, dat een vervolg is op eerder onderzoek van Jansen en Kruidenier (1986), wordt getracht aan te tonen dat de overgang van een eenzijdige regelstrategie naar een visuele oriëntatie begint in een vroeger stadium van de ontwikkeling van de spellingsvaardigheid dan in het algemeen wordt aangenomen. Dit gebeurt aan de hand van een tweetal vocaalsubstitutiefouten, te weten het schrijven van een open klinkerklank met een enkel klinkerteken (*strat* voor *straat*) en vice versa het schrijven van een gesloten klinkerklank met een dubbel klinkerteken (*paan* voor *pan*) in éénlettergrepige woorden, in het vervolg aangeduid met weglatings- en toevoegingsfouten. Kruidenier heeft geconstateerd dat er een groot en gedurende een bepaalde periode in de spellingsontwikkeling zelfs toenemend, verschil is in relatieve frequentie van beide foutentypen. In het bijzonder deze toename wordt door Kruidenier (Kruidenier, 1985; Kruidenier & Koster, 1986) geïnterpreteerd als een indicatie voor het zich ontwikkelen van een visueel-orthografische naast een puur fonologische strategie.

Spellingsstrategieën

Ter nadere toelichting geven we eerst een overzicht van mogelijke spellingsstrategieën, waarbij we ons baseren op het door Verhoeven (1985) geïntroduceerde categorieënsysteem: Het toepassen van regels, het gebruiken van analogieën, het letter-voor-letter (leren) opsommen (van spelwoorden), het gebruiken van het woordbeeld en het hanteren van geïntegreerde schrijfbewegingssequenties.

Voor onze probleemstelling is met name het onderscheid in deelstrategieën binnen de eerst genoemde regelstrategie van belang:

- a. sequentiële (of contextvrije) regels;
- b. conditionele (of context gevoelige) regels;
- c. paradigmatisch bepaalde regels;
- d. syntagmatisch bepaalde regels.

Bij de sequentiële regeltoepassing gaat het om de toepassing van éénduidige foneem-grafeemcorrespondenties, die ongeacht de context gelden (Marsh et al., 1980). Voor het Nederlands betreft dit de regels die de ongeveer 40 fonemen die we onderscheiden koppelen aan resp. grafemen, grafeemcombinaties, 'sets' van grafemen en 'sets' van grafeemcombinaties. De toepassing van de sequentiële of contextvrije regels beschouwen we als een beginstrategie. Bij conditionele regels gaat het eveneens om de toepassing van foneem-grafeemcorrespondenties, maar nu is de toepassing afhankelijk van contextuele kenmerken.

Het contextvrije karakter dat bij een bepaalde groep van foneem-grafeemcorrespondenties wordt aangeleerd is tot op zekere hoogte kunstmatig. De sequentiële regels kunnen worden gezien als de consequentie van een didactisch uitgangspunt, een verzameling duidelijke regels die in de aanvangsfase wordt aangeleerd en slechts een beperkte geldigheid heeft, namelijk met betrekking tot de zogenaamde klankzuivere woorden. Ook Verhoeven (1985, p.64) wijst er al op dat het onderscheid tussen beide substrategieën problematisch is. Hij tracht dit probleem op te lossen door onderscheid te maken tussen normale, ongemarkeerde gevallen (de 'aa' in 'straat') en gemarkeerde gevallen (de 'a' in 'straten').

Kruidenier veronderstelt dat de overgang van het toepassen van de in de beginfase het spellingsonderwijs aangeleerde verzameling van (vereenvoudigde) sequentiële regels naar conditionele regels al begint voordat het onderwijs ze expliciet aan de orde stelt, onder de invloed van visuele informatie. Zo wordt het enkele klinkerteken in een open lettergreep (halen, lenen, boten) pas aan het eind van het eerste leerjaar als spelprobleem aan de orde gesteld. Tevoren is de leerling echter, tijdens het lezen bijvoorbeeld, wel met deze orthografische patronen geconfronteerd. Op een gegeven moment zal de speller als schrijfwijze voor de open klinkerklank naast het dubbelteken ook het enkele teken als acceptabel alternatief gaan ervaren. Dat wil uiteraard niet zeggen dat zij dan ook direct een duidelijk beeld heeft van de condities waaronder het enkele dan wel het dubbele teken geldig is. Dit kan gemakkelijk leiden tot een overgeneralisering van het 'nieuwe' alternatief en daarmee tot een verschil in de frequentie waarmee toevoegings- en weglatingsfouten bij het schrijven van klinkers optreden als gevolg van een beginnende visueel-orthografische oriëntatie die leidt tot het ontdekken en nog relatief vaak fout toepassen van spellingsregels die nog niet zijn onderwezen. Ook werd de veronderstelling uitgesproken dat het met name goede spellers zouden kunnen zijn die het eerst met deze nieuwe regels geconfronteerd worden. Zij zouden dan een relatief groot aandeel moeten leveren aan weglatingsfouten vergeleken met de slechte presteerders.

De bovenstaande veronderstellingen zijn in tegenspraak met die van Assink en Verhoeven (1984) die uit hun onderzoek naar orthografische kennis concludeerden dat leerlingen niet of nauwelijks vormen en/of regels leren op grond van leeservaring.

Een aantal onderzoekers hebben zich inmiddels met het probleem van de vocaalsubstitutiefouten beziggehouden. In het reeds eerder genoemde artikel van Kruidenier (1985) wordt een onderzoek beschreven waarin 49 tweede klas leerlingen een aantal éénlettergrepige woorden met een open klinkerklank (23) en een gesloten klinkerklank (33) kregen aangeboden. Weglatingsfouten kwamen veel vaker voor dan toevoegingsfouten. Hetzelfde trad op bij heranalyse van het bij eerste klassers verzamelde materiaal van Verhoeven (1982). In een replicatie van dit onderzoek werd ook de door Kruidenier zo belangrijk geachte verandering met de tijd van de verhouding tussen weglatings- en toevoegingsfouten opnieuw gevonden (Koster & Kruidenier, 1986). Ook Kruidenier en Jansen vinden in hun onderzoek bij leerlingen in het begin van de tweede klas dat weglatingsfouten veel vaker voorkomen, n.l. ongeveer 8x vaker dan toevoegingsfouten. Een verdere analyse met behulp van het loglineaire model van de kruistabel met subjecten als rij- en foutentype als kolomvariabele leidde tot de conclusie dat het additieve model (één vaardigheidsparameter voor de subjecten en één moeilijkheidsgraadparameter voor de foutensoorten dragen additief bij tot het aantal fouten dat een subject maakt in een foutencategorie) geen adequate beschrijving van de geobserveerde tellingen opleverde. Uit deze resultaten werd afgeleid dat de veronderstelling dat de beide soorten fouten *niet* zijn terug te voeren op een één-dimensionele vaardigheid werd ondersteund. Tevens waren er aanwijzingen, dat het met name de relatief goede spellers zijn die de interferentiefouten maken (Jansen &

Kruidenier, 1986). De auteurs wijzen erop dat vergaande conclusies echter niet zijn aan te bevelen, gezien het kleine aantal subjecten en de beperkte set van spelwoorden waarvan gebruik is gemaakt. Een probleem daarbij is het onder controle houden van de storende factor woordvertrouwdheid. Het gebruik van de momenteel beschikbare frequentielijsten, biedt voor dit probleem vermoedelijk geen voldoende oplossing. In ons vervolgonderzoek zullen we de invloed van woordvertrouwdheid nagaan, door gebruik te maken van bestaande en niet-bestaande woorden. Welke invloed we van 'woordvertrouwdheid' zouden mogen verwachten is vooralsnog problematisch. Het zal duidelijk zijn, dat een leerling geen 'woordbeeld' kan hebben van een niet-bestaand woord, al is het mogelijk dat bij het spellen van een onzinwoord gebruik gemaakt wordt van bestaande woorden of woordsegmenten die, qua klank, lijken op het onzinwoord. Enerzijds kunnen we minder fouten verwachten, als we aannemen dat niet-bestaande woorden meer een beroep zullen doen op een puur fonologische strategie en daarom minder aanleiding zullen vormen tot 'foute' regeltoepassing dan bestaande. Anderzijds kunnen we juist meer fouten verwachten omdat er minder mogelijkheden zijn bij niet-bestaande woorden om woordbeeldinformatie te gebruiken.

Methodeverschillen

Uit onderzoek op het gebied van lezen is gebleken dat de kwaliteit zowel als de aard van de leesprestaties varieert met de gebruikte leesmethoden. (Bus & Jansen, 1984; Reitsma et al., 1981). Het ligt in de rede dat ook bij de spellingsprestaties methodeverschillen van belang zijn, ook al is het moeilijk van te voren aan te geven welke kenmerken dat zullen zijn. De noodzaak om rekening te houden met de mogelijke invloed van de in het onderwijs gebruikte taal-leesmethode op de ontwikkeling en het gebruik van spellingsstrategieën, bleek ook uit het moeilijk te interpreteren school X test interactieeffect dat werd gevonden in het onderzoek van Jansen & Kruidenier (1986).

3. OPZET

Onderzoeksmateriaal

Het onderzoeksmateriaal omvatte een vragenlijst voor de leerkrachten en een viertal instrumenten voor de leerlingen. Voor afname aan de leerlingen werden de volgende instrumenten geconstrueerd:

- een korte vragenlijst;
- een niveaudictee;
- vier woordherkenningslijsten;
- vier experimentele dictee's.

De experimentele dictee's kwamen als volgt tot stand. Uit de streeflijst woordenschat van Kohnstamm (Kohnstamm & De Vries, 1981) werden 40 éénlettergrepige woorden gekozen, voor de helft woorden met een open, respectievelijk gesloten klinkerklank. Dit waren de z.g. bekende woorden. Daarnaast werden 40 onzinwoorden bedacht, eveneens twintig met een open en 20 met een gesloten klinkerklank. Aan deze 80 werden nog 20 andersoortige woorden, onzin-zowel als bestaande woorden toegevoegd, die dienden om de onderzoeksopzet niet te doorzichtig te maken voor deelnemers aan het onderzoek en die bij de analyse van de resultaten verder geen rol speelden. Vervolgens werden er korte zinnetjes geformuleerd, waar de woorden deel van uitmaakten. Hieruit werden 4 z.g. kerndictee's gevormd. Voorts werden de woorden aselekt verdeeld over vier z.g. woordherkenningslijsten. Naast de vier experimentele dictee's werd een vijfde dictee samengesteld dat bedoeld was voor het vaststellen van het algemene spellingsvaardigheidsniveau van de leerlingen.

De vragenlijst voor de leerlingen bestond uit drie ja/nee vragen, 'houd je van lezen?', 'heb je veel boeken?', en 'ga je wel eens naar de bibliotheek?', die verondersteld werden enig verband te hebben met de mate waarin de leerling ervaring heeft met de geschreven taal. Ook werd aan de leerlingen gevraagd sekse en leeftijd in te vullen.

De vragenlijst voor de school bevatte o.a. vragen naar het schooltype, de gebruikte leesmethode en eventuele andere materialen gebruikt bij lees- en spellingsonderwijs.

Proefpersonen

Uit het adressenbestand van de Nederlandse basisscholen werd een aselechte steekproef getrokken van 50 scholen. Van deze 50 waren er 33 tot medewerking bereid. Aan het onderzoek werd deelgenomen door iets minder dan 800 leerlingen uit groep 4 (klas 2).

Afname en scoring

De afname van de dictee's vond plaats in het begin van het schooljaar. De dictee's werden mondeling aangeboden door de leerkrachten. We gingen er vanuit dat deze gang van zaken bevorderlijk is voor een goede verstaanbaarheid van de dicteewoorden voor de leerlingen. Dit laatste werd door ons geacht van groot belang te zijn, vooral vanwege de aanwezigheid van de onzin-woorden. Een nadeel is de eventuele aanwezigheid van storende verschillen in uitspraak tussen de leerkrachten. Naast de genoemde dictee's werd aan iedere leerling de vragenlijst, het z.g. niveau-dictee en één van de vier woordherkenningslijsten afgenomen. De leerlingen kregen op deze wijze de woorden uit de experimentele dictee's visueel aangeboden, met de opdracht de woorden die ze niet kenden door te strepen.

Bij de scoring werden de bestaande woorden waarop weglatingsfouten konden worden gemaakt en de woorden waarop toevoegingsfouten gemaakt konden worden samengevoegd tot twee verschillende tests. Hetzelfde gebeurde voor de niet-bestaande woorden. Bij de codering van de antwoorden werden de volgende categorieën onderscheiden: goed, toevoegings- of weglatingsfout (afhankelijk van het soort woord), toevoegings- of weglatingsfout gecombineerd met een andere fout en andersoortige fout. Foutenscores per leerling werden bepaald door fouten over de woorden per betreffende categorie op te tellen. Bij het niveau-dictee werd per leerling het totaal aantal goedgeschreven woorden geteld. Bij de leerlingvragenlijsten werd het aantal 'ja'-antwoorden geteld. Bij de woordherkenningslijsten werden twee categorieën gehanteerd: niet-bekend en bekend.

4. RESULTATEN

Uit de gegevens bleek dat van de 33 scholen er 25 Veilig Leren Lezen, 6 Letterstad en 2 nog een andere methode gebruikten bij het leesonderwijs. Een ruime helft daarvan gebruikte daarnaast ook ander materiaal. De steekproef bevatte respectievelijk één Montessori- en één Jenaplan-school.

Op het niveaudictee vonden we een verwaarloosbaar verschil tussen de leerlingen die lezen leerden met behulp van Veilig Leren Lezen en Letterstad. Het verschil in gemiddelde tussen jongens en meisjes, ten voordele van de meisjes, bleek significant op 5%-niveau. Zie hiervoor ook Tabel 1.

Uit de antwoorden op de leerlingvragenlijst konden we afleiden dat 84% van de leerlingen van lezen zei te houden, 90% van mening was veel boeken te bezitten en 75% wel eens naar de bibliotheek ging. 59% gaf een 'positief' antwoord op alle drie de vragen. Ook bij de vragenlijst vonden we een significant verschil tussen jongens en meisjes en eveneens een niet-significant verschil tussen VLL- en Letterstad-leerlingen. Het aantal 'ja'-antwoorden werd beschouwd als een, zij het zwakke, indicatie voor ervaring met geschreven taal. De onderzoekers zijn er zich tevens van bewust dat ze vraagtekens kunnen zetten bij de resultaten van een vragenlijst bij zulke jonge leerlingen. Voorts is het duidelijk dat ook al zou de lijst 'serieus' zijn ingevuld de interpretatie aan twijfel onderhevig is.

De z.g. woordherkenningslijsten bevatten de in de experimentele dictee's aangeboden woorden. Elke lijst, bestaande uit een kwart van het totaal aantal woorden, was aan ruwweg een kwart van de leerlingen afgenomen, met als opdracht om de woorden die ze *niet* kenden op de lijst door te strepen. Hiermee konden we nagaan of de niet-bestaande woorden, bij visuele

aanbieding, als zodanig werden herkend. Het merendeel van de niet-bestaande woordjes werd beoordeeld als niet-bekend door tenminste 90% van de leerlingen. Een merkwaardige maar achteraf begrijpelijke uitschieter, die overigens niet bij de analyses werd betrokken, was het woordje 'ries'. Eén van de klassen bleek een leerling te bevatten die zo heette. Voor de bestaande woorden gold het omgekeerde, 46 van de 50 woorden werd door meer dan 90% als bekend beoordeeld. Bij de bestaande woorden bleken de leerlingen wat zekerder van hun zaak dan bij de niet-bestaande.

Tabel 1 bevat ook de gemiddelden en standaarddeviaties van het aantal vocaalsubstitutiefouten, gesommeerd over de vier typen spelwoorden.

In Tabel 2 vinden we gemiddelden en standaarddeviaties voor de fouten per type woord. De bekende woorden bleken 'gemakkelijker' dan de niet-bestaande woorden. Ook nu werd gevonden dat vocaalsubstitutiefouten van het type 'straat' → 'strat' veel frequenter voorkwamen dan 'pan' → 'paan', zowel bij de bekende als bij de onbekende woorden. Voorts werden er ook betrekkelijk grote verschillen tussen de woorden onderling gevonden. Zo neemt bij de niet-bestaande woorden het woordje 'vor' verreweg het grootste aantal toevoegingsfouten voor zijn rekening. Verwarring met het bestaande woord 'voor' zou wel eens de reden kunnen zijn.

Uit de geconstateerde discrepantie tussen de relatieve frequenties van de beide soorten fouten kunnen we echter niet concluderen dat er dus ook sprake is van verschillende onderliggende dysfunctionerende vaardigheden.

In de hierna volgende analyse werd nagegaan of de verhouding tussen de weglatings- en de toevoegingsfouten samenhangt met spellingsvaardigheid gemeten met het niveaudictee, ervaring met de geschreven taal en de bekendheid van de spellingswoorden. Ook werd nagegaan of de gebruikte taal-leesmethode daarbij van belang was. We beperkten ons hier tot leerlingen die op een school zaten waarbij of Veilig Leren Lezen of Letterstad werd gebruikt. De oorspronkelijke scores op het niveaudictee werden gedichotomiseerd op de mediaan. Uit de vragenlijst werd een eveneens dichotome variabele gevormd, met als categorieën 'veel' (drie 'ja'-antwoorden) en 'weinig' ervaring met de geschreven taal. Vervolgens werd een vier-weg variantie-analyse uitgevoerd, met ervaring, spellingsvaardigheid, methode en woordbekendheid elk met twee niveau's als de onafhankelijke variabelen en een maat voor de verhouding tussen de beide foutentypen als de afhankelijke variabele. We hebben dus te maken met een herhaalde-metingen-opzet met methode, ervaring en spelvaardigheid als factoren tussen-de-personen en woordbekendheid als factor binnen-de-personen.

Het construeren van een geschikte afhankelijke variabele is een probleem op zich. In principe willen we twee relatieve foutenfrequenties met elkaar vergelijken onder een aantal verschillende condities. Als we de relatieve frequenties van de beide foutentypen benoemen als $p^{(1)}$ en $p^{(2)}$ en de 'verhouding' die we in beschouwing willen nemen als δ , dan kunnen we δ op een aantal verschillende manieren parametriseren (Cox, 1970). B.v.:

- (i) $\delta = p^{(2)} - p^{(1)}$;
- (ii) $\delta = \log \{ p^{(2)} / p^{(1)} \}$;
- (iii) $\delta = \log \{ [p^{(2)} / (1 - p^{(2)})] / [p^{(1)} / (1 - p^{(1)})] \}$.

In ons geval komen (ii) en (iii) praktisch op het zelfde neer, omdat de relatieve foutenfrequenties $p^{(1)}$ en $p^{(2)}$ in het merendeel van de gevallen kleine waarden aannemen. Cox (1970) geeft een aantal criteria op grond waarvan gekozen kan worden uit de alternatieve mogelijkheden. Wij besloten voor parametrisering (iii), de z.g. log-odds transformatie. We kunnen deze ook als volgt schrijven:

$$\delta_{ijklm} = \theta_{ijklm}^{(2)} - \theta_{ijklm}^{(1)}, \quad i = 1, \dots, n_{ijklm}$$

waarbij θ de logit is van p en waarbij $p_{ijklm}^{(1)}$ staat voor de foutenkans corresponderend met toevoegingsfouten ('a' → 'aa') voor een willekeurig subject i onder de combinatie van condities j ,

k, l en m. Een bijkomend probleem is wat we moeten doen met subjecten die gegeven een bepaalde combinatie van condities nul of n_{ijklm} fouten hebben gemaakt. In beide gevallen n.l. is θ niet gedefinieerd. We kunnen deze moeilijkheid omzeilen door de schatting van θ als volgt te modificeren:

$$\theta_i = \log \left(\frac{F_i + \frac{1}{2}}{n - F_i + \frac{1}{2}} \right),$$

waarbij F het aantal door subject i gemaakte fouten is en n het aantal spelwoorden. De correctiefactor 0.5 zorgt ervoor dat θ ook is gedefinieerd wanneer F gelijk is aan nul of aan n (Cox, 1970).

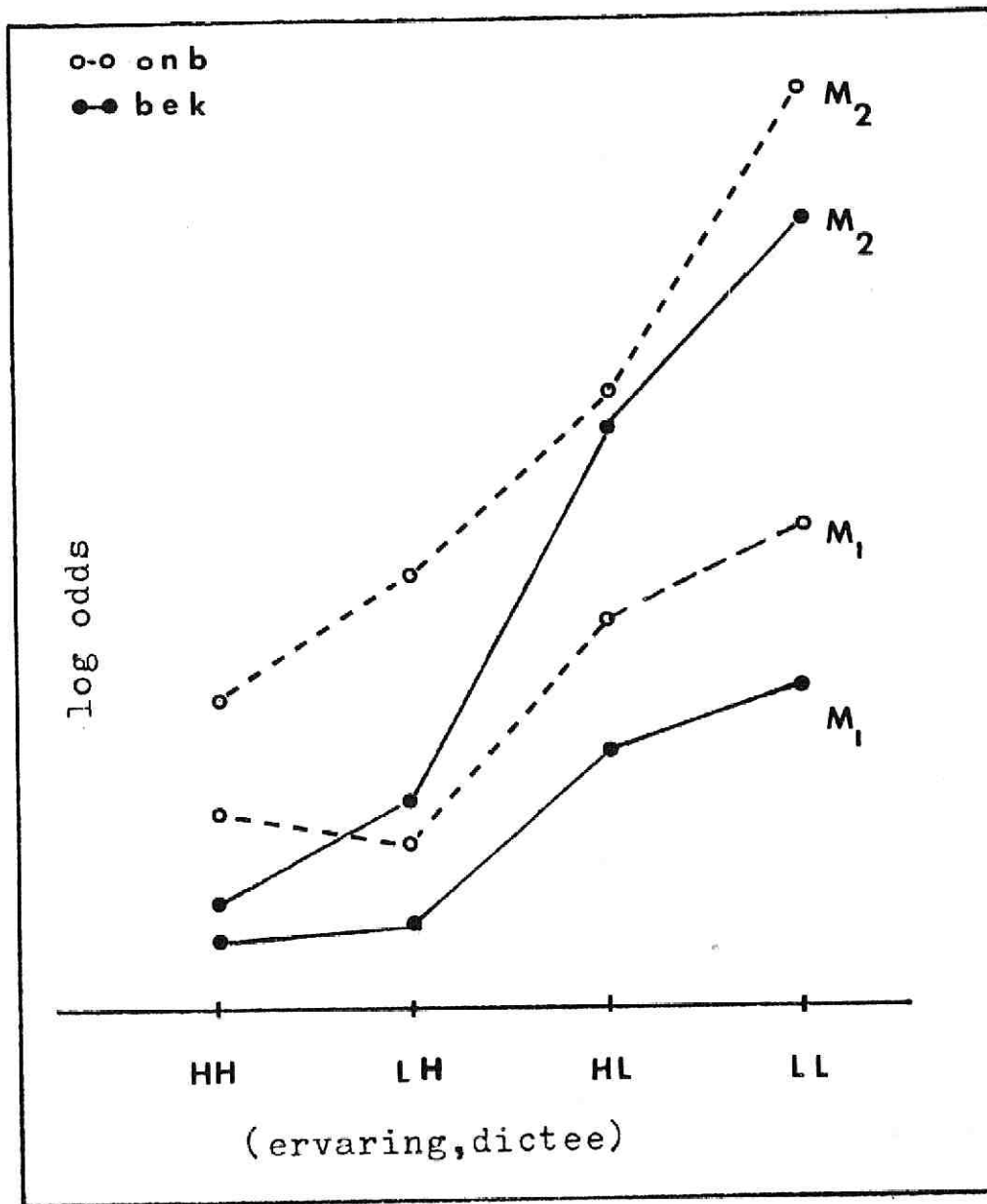
Tabel 3. Resultaten van een variantieanalyse op de logodds van de toevoegings- en weglatingsfouten als afhankelijke variabele.

Bron	SS	df	MS	F
Tussen de subjecten:				
methode	57.2	1	57.2	39.2*
niveau dicteescore	119.2	1	119.2	82.9*
vragenlijstscore	17.5	1	17.5	5.2
m x d	16.1	1	16.1	11.2*
m x v	4.9	1	4.9	3.4
d x v	3.7	1	3.7	2.5
residu	1003.5	699	1.44	
binnen de subjecten:				
woordbekendheid	24.80	1	24.80	5.9*
methode x bekendheid	.18	1	.18	.4
dicteescore x bekendheid	.04	1	.04	.1
vragenlijstscore x bekendheid	.07	1	.07	.2
residu	325.12	702	.46	

* significant bij $\alpha = .01$

Tabel 3 bevat de resultaten van de analyse. Voor het vormen van de F-ratio's werden op basis van voorafgaande analyses alle derde- en hogere orde interacties gelijk gesteld aan nul en de bijbehorende kwadratsommen samengevoegd met de binnen-de-cellen kwadratsommen. Uit de tabel kunnen we opmaken dat van de vier hoofdeffecten er drie significant zijn op 1-procentsniveau. Het effect van 'ervaring' is significant op 5-procentsniveau. Van de interactie-effecten is alleen de interactie methode x spelvaardigheid significant. Bij de interpretatie is enige voorzichtigheid echter geboden. De aantal observaties per cel in het design zijn sterk wisselend. Voorts had de gekozen parametrisering niet steeds het variantie-stabiliserende effect dat er van werd verwacht.

Een verdere illustratie van de resultaten wordt gegeven door Figuur 1. Hierin kunnen we aflezen dat bij zowel de VLL- als de Letterstad-leerlingen het de relatief goede spellers, met relatief veel ervaring met de geschreven taal zijn die weinig weglatingsfouten maken. Slechte spellers maken niet alleen absoluut, maar ook relatief de meeste weglatingsfouten. Deze bevinding is in tegenspraak met eerder gerapporteerde gegevens. Voorts is de over de leerlingen gemiddelde discrepantie tussen de frequentie van weglatingsfouten en toevoegingsfouten groter bij de niet-bestaande woorden dan bij de bestaande. De Letterstad-leerlingen maken gemiddeld meer vocaalsubstitutiefouten dan de VLL-leerlingen en het verschil in relatieve frequentie tussen de beide soorten fouten is ook wat groter.



Figuur 1: De gemiddelde log-odds van de toevoegings- en weglatingsfouten afgezet tegen weinig en veel ervaring gecombineerd met hoge en lage goedscores op het niveaudictee.

5. CONCLUSIES EN DISCUSSIE

Uit de resultaten van ons onderzoek kunnen we een aantal conclusies trekken. Ook nu vinden we een duidelijke discrepantie tussen de beide fouttypen, vocaalsubstitutiefouten van het type 'stroop' wordt 'strop' komen veel vaker voor dan 'hulp' wordt 'huulp' e.d. De inhoudelijke interpretatie blijft echter problematisch en voor meerdere uitleg vatbaar. Een mogelijke en tamelijk triviale verklaring zou kunnen zijn dat open klinkerklanken om deze of gene reden moeilijker zijn te onderscheiden dan gesloten klinkerklanken. Een tweede, eveneens wat triviale verklaring, zou kunnen zijn dat vocaalsubstitutiefouten van het type aa → a vergelijkbaar zijn met andere weglatingsfouten en dat beide een indicatie zijn voor een relatief primitief ontwikkelingsstadium, waarbij leerlingen vooral bij wat langere en/of onbekende woorden a.h.w. 'de draad kwijtraken' en daardoor grafemen gaan weglaten. Beide boven gegeven interpretaties

laten echter de door Kruidenier e.a. geconstateerde toename van weglatingsfouten in het tweede leerjaar onverklaard (Kruidenier, 1985; Koster & Kruidenier, 1986). Kruidenier brengt met name dit surplus aan weglatingsfouten in verband met een beginnende visueel-orthografische oriëntatie die leidt tot het ontdekken en nog relatief vaak fout toepassen van spellingsregels die nog niet zijn onderwezen. Met het oog hierop leek het Jansen en Kruidenier niet onredelijk om te veronderstellen dat het met name goede spellers zullen zijn die het eerste met deze nieuwe regels geconfronteerd worden. Zij zouden dan een relatief groot aandeel moeten leveren aan weglatingsfouten vergeleken met de slechte presteerders. Dit laatste wordt echter door de resultaten niet bevestigd. Leerlingen die een hogere score hadden op het niveaudictee maakten niet alleen minder vocaalsubstitutiefouten in het algemeen, maar ook relatief minder weglatingsfouten. M.a.w. het zijn de slechte, niet de goede spellers die relatief de meeste weglatingsfouten maken, in tegenspraak zowel met bovengenoemde veronderstelling als met eerder gerapporteerde resultaten.

Bij het spellen van niet-bestaande woorden ligt het voor de hand dat er sprake zal zijn van een meer éénzijdige oriëntatie op de klank van de te spellen woorden dan bij bestaande woorden. Van een niet bestaand woord kan de leerling geen orthografisch beeld in het woordgeheugen hebben, ook al is het misschien wel mogelijk om gebruik te maken van orthografische beelden van op het onzin-woord lijkende bestaande woorden of van segmenten die op zich wel zinvol zijn. Een fonologische strategie, waarbij de leerling op basis van eenduidige foneem-grafeemcorrespondenties foneem voor foneem in grafemen vertaalt ligt bij niet-bestaande woorden het meest voor de hand. Maakt de leerling fouten dan zou dit kunnen worden verklaard op grond van onvoldoende beheersing van de auditieve deelvaardigheden klanksegmentatie en foneem/grafeemkoppeling. De veronderstelling dat bestaande woorden over het algemeen 'makkelijker' moeten zijn dan niet-bestaande, omdat de leerling daarbij geen of tenminste minder steun heeft van een 'woordbeeld', wordt bevestigd. Kijken we naar de fouten, ongeacht het type fout, dan zien we dat niet-bestaande woorden gemiddeld vaker fout geschreven worden. Het aantal vocaalsubstitutiefouten ligt gemiddeld eveneens wat hoger bij de niet-bestaande woorden.

Ook is de verhouding tussen de twee onderscheiden typen afhankelijk van de factor woordbekendheid. Bij de niet-bestaande woorden vinden we niet alleen absoluut maar ook relatief meer weglatingsfouten.

Opvallend is ook de duidelijke samenhang tussen enerzijds het aantal en de verhouding tussen de twee soorten substitutiefouten en de gebruikte taal-leesmethode. Dit verschijnsel kunnen we niet (geheel) verklaren uit een verschil in spellingsvaardigheid tussen de leerlingen, omdat we geen verschil van betekenis vinden tussen Letterstad- en VLL-leerlingen op het z.g. niveaudictee. We moeten hierbij bedenken dat verschillen tussen leerkrachten hier ook een rol kunnen spelen. Het lijkt ons echter niet onaannemelijk dat de oorzaak vooral moet worden gezocht in verschillen in aanpak tussen beide methoden. Om welke aspecten het precies gaat vraagt uitgebreider onderzoek, waarbij ook andersoortige spellingsfouten zouden moeten worden betrokken.

Het is op grond van de hierboven besproken gegevens nog steeds aannemelijk dat het surplus aan weglatingsfouten inderdaad te maken heeft met door b.v. lezen verkregen visuele informatie, die kan leiden tot verwarring bij het schrijven van de lange klinker. Deze veronderstelling is compatibel met de observatie dat weglatingsfouten relatief vaak worden gemaakt door de slechtere spellers en bij de onbekende woorden. De rol die de gebruikte methode hierbij speelt vereist nog verder onderzoek.

LITERATUUR

- Assink, E.M.H., (1983). *Leerprocessen bij het spellen*. (dissertatie) Utrecht, RUU.
 Assink, E.H.M. & Verhoeven, G. (1984). Waarom kunnen sommige kinderen woorden spellen waarvoor ze de regels nog niet geleerd hebben? *Spektator*, 13, 334-345.

- Assink, E. & Verhoeven, G., (red) (1985). *Visies op spelling*, Groningen: Wolters-Noordhoff.
- Barron, R.W. (1980). Visual and phonological strategies. In U. Frith, (Ed.) *Cognitive processes in spelling*. London: Academic Press.
- Bryant, P.E. & Bradley, L. (1980). Why children sometimes write words which they do not read. In U. Frith, (Ed.) *Cognitive processes in spelling*. London: Academic Press.
- Bus, A.G. & Jansen, G.G.H. (1984). Een longitudinaal onderzoek naar leren lezen in de eerste klas. *Pedagogische Studiën*, 61, 485-495.
- Cox, D.R. (1970). *The analysis of binary data*. London: Chapman & Hall.
- McCullagh, P. & Nelder, J.A. (1983). *Generalised linear models*. London: Chapman & Hall.
- Ehri, L.C. (1980). The development of orthographic images. In U. Frith, (Ed.) *Cognitive processes in spelling*. London: Academic Press.
- Frith, U. (1980). Unexpected spelling problems. In U. Frith, (Ed.) *Cognitive processes in spelling*. London: Academic Press.
- Henderson, E.H. & Beers, J.W. (Eds.) (1980). *Developmental and cognitive aspects of learning to spell*. Newark: International Reading Association.
- Jansen, G.G.H. & Kruidenier, J. (1986). Substitutiefouten bij open en gesloten klinkerklanken in éénlettergrepige woorden. *Tijdschrift voor Onderwijsresearch*, 11, 255-264.
- Kohnstamm, G.A. & De Vries, A.K. (1981). *Nieuwe streeflijst woordenschat voor 6-jarigen*. Lisse: Swets & Zeitlinger.
- Koster, K.B. & Kruidenier, J. (1986). Cognitieve conflicten bij het leren spellen door eerste klassers. *Tijdschrift voor Taalbeheersing*, 8, 268-277.
- Kruidenier, J. (1985). Classificatie van spelfouten nader geanalyseerd. *Tijdschrift voor Taalbeheersing*, 7, 56-64.
- Langereis, M.P. & Denys, A.V.M.J. (1985). Het orthografisch woordbeeld bij de onervaren speller. In K.B. Koster & N. Kwantes (red.). *Schoolbegeleiding, leerproblemen en schooluitval*. Lisse: Swets & Zeitlinger.
- Linden, J. van der, (1986). De rol van het woordbeeld en de woordfrequentie bij het lezen en spellen. *Tijdschrift voor Taalbeheersing*, 4, 251-267.
- Marsh, G.M., Friedman, V., Welch, V. & Desberg, P. (1980). The development of strategies in spelling. In U. Frith (Ed.) *Cognitive processes in spelling*. London: Academic Press.
- Mommers, M.J.C. (1985). Spellingsonderwijs in de onderklassen van de basisschool volgens de structuurmethode. In E. Assink & G. Verhoeven (red.) *Visies op spelling*. Groningen: Wolters-Noordhoff.
- Reitsma, P. (1985). De verwerving van orthografische kennis. In E. Assink & G. Verhoeven (red.). *Visies op spelling*. Groningen: Wolters-Noordhoff.
- Reitsma, P., Komen, N. & Kappinga, T. (1981). Methoden voor aanvankelijk lezen: een vergelijking van leesresultaten na een jaar. *Pedagogische Studiën*, 58, 174-189.
- Rijnsoever, R.J. van (1980). Spontane spellingen. In E. Assink & G. Verhoeven (red.) (1985) *Visies op spelling*. Groningen: Wolters-Noordhoff.
- Simon, D.P. (1976) Spelling - A task analysis. *Instructional Science*, 5, 277-302.
- Verhoeven, G. (1985) *De strategieën van de speller*. Groningen: Wolters-Noordhoff.
- Verhoeven, L.Th. (1982) *Verantwoording toetsen: aanvankelijk lezen en spellen*. (Bulletin reeks nr 13). Arnhem: Cito.
- Verhoeven, L.Th. (1983) Foutenclassificatie bij beginnende spellers. *Tijdschrift voor Taalbeheersing*, 5, 131-138.

Manuscript ontvangen 16-4-1987

Definitieve versie ontvangen 9-6-1987

De toepassing van het weddenshapsmodel in een curriculum-evaluatie-onderzoek

G. van den Berg

RION, Instituut voor Onderwijsresearch, Rijksuniversiteit Groningen

ABSTRACT

Starting from the idea that in evaluation research the professional reputation of curriculum developers is at stake, the betting model was a source of inspiration for the design of the research project. This model is developed for situations in which only the 'one-shot experiment' can be done. The assumption underlying the model is that there are differences of opinion about the effectiveness of one treatment. In the research project there were several 'treatments'. These 'treatments' may be seen as the various answers to existing differences of opinion about the way education should be structured. Despite the limitations in time and money the experiences with the model used may be considered positive. To determine the relative effectiveness of the curricula the analysis of variance is a suitable technique and the level of statistical significance is an arguable relative standard. The benefit of predictions in comparative curriculum evaluation is limited. The predictions are only suitable for determining whether the methods come up to the expectations of their developers.

INLEIDING

Op grond van een analyse van een aantal gangbare evaluatie-modellen kwam Scheerens (1983) tot de bevinding dat het weddenshapsmodel van Hofstee (1980) één van de sterkste evaluatie-modellen is. Toch wordt over de haalbaarheid van het weddenshapsmodel in kringen van evaluatie-onderzoekers verschillend gedacht. Zo beweren tegenstanders van het weddenshapsmodel dat niemand bereid zal zijn zijn/haar reputatie op het spel te zetten of zijn/haar handelingsmarge in te laten perken, zonder de uitkomsten van het evaluatie-onderzoek te kennen (zie bijv. Stokking, 1984). Weliswaar zullen beleidsmakers, leerplanontwikkelaars, onderwijsbegeleiders, e.d. niet staan te trappelen om zich vooraf te binden aan de in evaluatie-onderzoek te hanteren evaluatiecriteria en -standaarden maar dit mag evaluatie-onderzoekers er niet van weerhouden alles in het werk te stellen hen vast te leggen op vooraf geformuleerde criteria en standaarden.

Aan de hand van een door het RION uitgevoerd evaluatie-onderzoek naar de relatieve doelmatigheid van Economie-methoden voor het MAVO zal een poging gedaan worden een aantal elementen van de weddenshapsmethodologie te waarderen, met name de elementen die betrekking hebben op de toepasbaarheid van het weddenshapsmodel in curriculum-evaluatie. Een evaluatie van de weddenshapsmethodologie als zodanig is niet aan de orde. Dit zou om een veel omvangrijker en ingewikkelder opzet vragen.

ACHTERGRONDEN VAN HET ECONOMIE-MAVO-PROJECT

Op verzoek van het Instituut voor Leerplanontwikkeling (SLO) is begin 1984 met subsidie van het Instituut voor Onderzoek van het Onderwijs (SVO) een onderzoek gestart met de vraag: hoe

doelmatig is de SLO-benadering op het vakgebied Economie voor de MAVO? (Van den Berg & Harskamp, 1985). In de eindfase van dit onderzoek is aan de SVO, via de zgn. ANGO-procedure, het verzoek gericht een vergelijkend evaluatie-onderzoek te starten naar de relatieve doelmatigheid van vier onderwijsleerpakketten. Dit onderzoek is, na goedkeuring door de SVO, begin 1985 gestart en in de loop van 1986 afgesloten. Formeel is er weliswaar sprake van twee onderzoeken, maar in feite zijn ze te beschouwen als twee fasen van één onderzoek: een vooronderzoek (SVO 1151) en een hoofdonderzoek (SVO 4213).

In deze bijdrage zal ik me beperken tot het hoofdonderzoek (zie ook Van den Berg, Brokken & Prins, 1986 en Van den Berg, 1987). Allereerst wil ik, zij het kort, ingaan op de context van het onderzoek.

Economie is in het schooljaar 1984-1985 het vak Handelskennis als keuzevak op het eind-examen MAVO gaan vervangen. De projectgroep Economie-MAVO van de SLO heeft zich van juli 1978 tot eind 1983 beziggehouden met leerplanontwikkeling voor dit nieuwe vakgebied. Allereerst heeft het projectteam een onderwijsleerplan voor Economie ontworpen dat door de minister is vastgesteld bij circulaire van 31 augustus 1982. In het Leerplan Economie-MAVO zijn ook de vakdoelstellingen vastgesteld. Deze zijn:

1. de leerling moet zich bewust worden van het economisch leven in ons land en (voor zover dat nodig en mogelijk is) in andere landen;
2. de leerling moet zich bewust worden van zijn eigen positie in dat economische gebeuren;
3. de leerling moet praktische kennis en vaardigheden verwerven die nodig zijn ten behoeve van zijn economisch handelen;
4. de leerling moet zelfstandig een (gemotiveerd) standpunt kunnen bepalen over actuele economische vraagstukken.

De vakdoelstellingen houden volgens het leerplan Economie in dat de docent bij kennisverwerving veel aandacht dient te besteden aan houdingsdoelstellingen en aan een actieve rol van leerlingen. Hierbij past volgens de SLO (1981, 19), 'een interactiemodel - vraag-leergesprek, klasgesprek en discussie - naast groepswork en zelfwerkzaamheid van de leerlingen'.

Naast het leerplan, dat algemene uitgangspunten en vakdoelstellingen voor het Economie-examen bevat, heeft het SLO-team een exemplarisch onderwijsleerpakket ontwikkeld. Dit pakket (Economisch bekeken) is met enige wijzigingen door de uitgeverij Malmberg op de markt gebracht. Daarnaast hebben ook andere uitgeverijen onderwijsleerpakketten Economie voor MAVO op de markt gebracht, te weten: Wolters-Noordhoff (Economavo), Versluys (Economie voor MAVO) en Muusses (Dag in, dag uit). Ook de auteursgroepen van deze pakketten zeggen zich te baseren op het voornoemde, door de minister vastgestelde, leerplan Economie-MAVO, maar geven er elk toch een andere interpretatie en uitwerking aan.

HET ONDERZOEKSMODEL

Uitgaande van de opvatting dat in evaluatie-onderzoek de professionele reputatie van curriculum-ontwikkelaars in het geding is, pleit Hofstee (1982) ervoor evaluatie-onderzoek op te zetten als een weddenschap; d.w.z. onderzoek waarin meningsverschillen uitgangspunt zijn. Dit model is ontwikkeld voor situaties, waarin uitsluitend het 'one-shot' experiment kan worden toegepast en het meningsverschil zich beperkt tot de effectiviteit van één treatment (bijv. een bepaalde instructiemethodiek). Hofstee (1980) heeft zijn weddenschapmodel in de eerste plaats ontwikkeld om meningsverschillen binnen de wetenschap in een empirische discussie te kunnen beslechten. Omdat meningsverschillen alleen kunnen bestaan tussen mensen (van vlees en bloed) wordt in het weddenschapmodel derhalve de persoonlijke reputatie van wetenschappers c.q. wetenschappelijke onderzoekers in het geding gebracht. Om een geschil te kunnen beslechten, zijn er regels nodig die door de partijen worden gedeeld.

De weddenschapsmethodologie impliceert derhalve dat de partijen:

1. overeenstemming moeten bereiken over het meest geschikte onderzoeksdesign;
2. overeenstemming moeten bereiken over de keuze en de operationalisatie van zowel de afhankelijke als de onafhankelijke variabelen;
3. van te voren moeten specificeren onder welke condities welke effecten zullen optreden; en
4. de mogelijke uitkomsten van het onderzoek moeten voorspellen.

Dit zijn de vier essentiële kenmerken van het weddenschapsmodel. Als aan deze kenmerken wordt voldaan, dan zijn de partijen er ook aan gebonden de (toekomstige) onderzoeksresultaten te accepteren. Het in het evaluatie-onderzoek Economie-MAVO gehanteerde model wijkt op een aantal punten af van het weddenschapsmodel. Ten eerste is niet de reputatie van onderzoekers in het geding maar die van curriculumontwikkelaars. Ten tweede is in het onderhavige evaluatie-onderzoek sprake van vier 'treatments', nl. vier onderwijsleerpakketten.

Bij de start van het project is, in overeenstemming met de betrokken partijen, het evaluatie-onderzoek opgezet als vergelijkende productevaluatie. Al in de voorbereiding van het evaluatie-onderzoek is met de partijen overeenstemming bereikt over de onderzoeksopzet. Dit betekent dat aan één essentieel kenmerk van het weddenschapsmodel al bij de start van het project was voldaan. In het hierna volgende zullen we aandacht besteden aan de andere drie essentiële kenmerken van het model.

Evaluatiecriteria en -standaarden

Het gaat er bij curriculum-evaluatie in de eerste plaats om dat het bereiken van een gesteld doel adequaat vastgesteld moet kunnen worden (De Groot, 1982). Dit betekent dat een instrument voor het vaststellen van leereffecten betrouwbaar en valide moet zijn.

Hoewel de curricula in principe de evaluatiecriteria bevatten voor de evaluatie van de curricula, levert het formuleren van evaluatiecriteria in de praktijk van curriculum-evaluatie toch vaak grote problemen op. Vaak zijn curriculumdoelen globaal geformuleerd of blijven soms zelfs geheel impliciet. Vaak ook zijn de curriculumdoelen niet expliciet gekoppeld aan de middelen.

Middel-doel-analyse moet weliswaar worden beschouwd als het grondprincipe voor de koppeling van curriculumontwikkeling, curriculumimplementatie en curriculumevaluatie, maar is meestal niet toereikend om het criteriumprobleem op te lossen. Om te komen tot adequate evaluatiecriteria (waaraan de ontwikkelaars zich niet kunnen onttrekken) zal met hen onderhandeld moeten worden. Een adequate criteriumformulering mag dan in eerste instantie de verantwoordelijkheid zijn van beleidsmakers, curriculumontwikkelaars e.d., maar in de praktijk onttrekken zij zich nogal eens aan deze plicht (zie ook Hoebe, 1985). Hetzelfde geldt voor het formuleren van evaluatiestandaarden. Curriculumontwikkelaars zouden moeten aangeven bij welke waarden van een criteriumvariabele nog wel en bij welke waarden niet meer sprake is van een beoogd gevolg. Daarbij zijn belangrijke kritische vragen: (1) wanneer heeft een leerling er voldoende profijt van gehad? en (2) hoeveel leerlingen hebben er nu voldoende profijt van gehad? Om daar zicht op te krijgen heeft men een maatstaf, een standaard nodig (vgl. De Groot, 1982).

De doelmatigheid van onderwijsleerpakketten kan afhankelijk van de gehanteerde standaard op verschillende manieren worden vastgesteld.

Allereerst kan worden vastgesteld of het gebruik van het onderwijsleerpakket leidt tot het bereiken van de doelen, zoals die door de auteur(s) zelf worden verwacht. Voorspellingen van de auteurs over de doelmatigheid van het eigen pakket kunnen hierbij als standaard worden gebruikt. Uit deze voorspellingen kunnen weliswaar conclusies worden getrokken wat betreft de door de partijen zelf ingeschatte doelmatigheid van elk curriculum afzonderlijk, maar niet over de relatieve doelmatigheid zelf. In vergelijkend evaluatie-onderzoek hebben voorspellingen van de auteurs dus een beperkte betekenis.

Vervolgens kan worden vastgesteld of het ene pakket doelmatiger is dan het andere (de zgn. relatieve doelmatigheid). Als maat voor verschillen in relatieve doelmatigheid kan statistische significantie worden gehanteerd. Afhankelijk van het gekozen significantieniveau kan dan worden bepaald met welke mate van zekerheid een gevonden verschil wel of niet is toe te

schrijven aan toevallige factoren. Het geeft geen antwoord op de vraag of een verschil ook betekenisvol c.q. de moeite waard is. Een antwoord op deze vraag zou men de partijen zelf kunnen laten geven. Men zou hen dan kunnen vragen wanneer een verschil voor hen betekenisvol is, bijvoorbeeld in termen van verklaarde variantie. Als de partijen over een dergelijke aanvullende standaard overeenstemming bereiken, dan kan ook de vraag beantwoord worden of het verschil in effectiviteit zodanig is dat aan het ene product de voorkeur moet worden gegeven boven een ander. Ontbreekt een dergelijke aanvullende standaard dan zal moeten worden volstaan met toetsing op statistische significantie met alle voorzichtigheid die men daarbij moet betrachten (vgl. Stokking, 1984).

Tenslotte kan worden vastgesteld of de onderwijsleerpakketten voldoen aan bepaalde minimumeisen, waarover van te voren tussen bijvoorbeeld de auteursgroepen en/of andere vakdeskundigen overeenstemming is bereikt.

Storende omstandigheden

Verschillen tussen curricula op prestatieniveau (de meest voorkomende vorm van criteriumverschillen) kunnen niet zonder meer als verschillen in effecten van deze curricula worden geïnterpreteerd. Het is mogelijk dat deze verschillen op prestatieniveau veroorzaakt zijn door andere relevante verschillen tussen de gebruikersgroepen dan alleen curriculumverschillen. Omdat de controle op storende variabelen in de praktijk van evaluatie-onderzoek bijna altijd problematisch is, heeft Hofstee (1980) overtuigend beargumenteerd dat het aanleggen van allerlei methodische controles op storende variabelen niet een automatisme mag zijn. Voor controle op storende variabelen moeten goede redenen aanwezig zijn. Bij curriculum-evaluatie is er echter een aantal categorieën van variabelen waarop in principe altijd gecontroleerd moet worden, te weten (1) het beginniveau van de leerlingen en (2) de wijze waarop het curriculum wordt gebruikt.

Ad. 1. Het begrip beginniveau wordt hier gebruikt om aan te geven wat de leerling 'meebrengt' als hij begint aan een cursus of cursusonderdeel en wat relevant is of kan zijn voor het volgen daarvan (vgl. Van Eyl, 1980). Het gaat hierbij niet alleen om voorkennis van de leerling, maar ook om de aanleg, het vermogen van de leerling om te leren. Sommigen (o.a. de Klerk, 1984) beschouwen ook attitudes (faalangst en prestatiemotivatie) van leerlingen als aspecten van beginniveau. Dat controle op voorkennis in curriculumevaluatie belangrijk is behoeft nauwelijks betoog. Wat een leerling reeds wist, kan nu eenmaal niet als effect van een curriculum worden gezien. Toch is het doen van een voormeting niet altijd zinvol, bijvoorbeeld als de leerlingen worden geconfronteerd met een voor hen volledig nieuw vak waarvan zij nog helemaal niets weten. In de praktijk is voormeting ook niet altijd mogelijk. Vaak worden evaluatie-onderzoekers ingeschakeld als het curriculum al in uitvoering is. Noodgedwongen zal men zich dan moeten beperken tot het vaststellen van de aanleg van de leerling, zoals gemeten door een intelligentietest. Ook dit is helaas niet altijd mogelijk. Voor het voortgezet onderwijs bestaat nauwelijks een goede intelligentietest die klassikaal kan worden afgenomen. Derhalve is men vaak gedwongen sociaal milieu als substituut variabele te nemen.

Ad. 2. In curriculum-evaluatie-onderzoek kan moeilijk de vraag worden ontweken hoe leerkrachten curricula in de onderwijspraktijk gebruiken. Omdat leerkrachten kunnen verschillen in de wijze waarop zij de curricula gebruiken, bestaat de mogelijkheid dat verschillen in leerprestaties niet het gevolg zijn van verschillen in curricula maar van het specifieke, afwijkende gebruik ervan. In curriculum-evaluatie-onderzoek zullen vormen van curriculumgebruik (met eventueel bijbehorende specifieke effecten) dan ook meestal relevante condities zijn. Curricula kunnen op diverse aspecten van elkaar verschillen. De belangrijkste zijn: (1) het leerstof-inhoudelijke aspect (bijv. de te behandelen leerstof uit het onderwijsleerpakket); en (2) het vak didactische aspect (bijv. het geven van klassikale en/of individuele instructie, het geven van groepsopdrachten, het laten samenwerken bij het maken van opdrachten, het geven van feedback). Verschillen tussen curricula op deze twee aspecten en met name de wijze waarop zij in de onderwijspraktijk worden gerealiseerd, zijn in curriculum-evaluatie om twee verschillende redenen van belang. Vakdidactische verschillen kunnen leiden tot verschillen in effectieve

leertijd van de leerlingen en derhalve tot verschillen in leerprestaties (zie bijvoorbeeld Creemers & Schaveling, 1985). Derhalve zou in curriculum-evaluatie onderzoek ook effectieve leertijd als storende omstandigheid (voor effecten van vakdidactische verschillen tussen curricula) meegenomen moeten worden. Echter, het adequaat meten van effectieve leertijd (door middel van observaties van individuele leerlingen) is zo'n kostenverhogende factor dat in veel (contract)onderzoek het meten van deze factor niet mogelijk is. Zo ook in het onderhavige evaluatieonderzoek. Leerstofinhoudelijke verschillen zijn van belang, omdat te verwachten is dat verschillen in leerprestaties verklaard worden door verschillen in het leerstofaanbod en/of de accenten op leerstofinhouden.

INSTRUMENTENONTWIKKELING

Allereerst zijn de onderwijsleerpakketten geanalyseerd op drie curriculumelementen, te weten: de te behandelen leerstof per cursusjaar, de wijze waarop instructie en verwerking van de leerstof dient plaats te vinden en de differentiatiemogelijkheden. Met name op het punt van instructie en verwerking zijn verschillen aan te wijzen.

Economisch bekeken (Malmberg) legt de nadruk op het klasgesprek en het groepsgewijs verwerken van de leerstof, van zowel reproductieve vragen als vragen die productief gebruik van kennis vereisen. Economie voor MAVO (Versluys) legt de klemtoon op kennisoverdracht via de docent, gericht op het reproduceren en memoriseren van kennis. Economavo (Wolters-Noordhoff) kan door de docent naar eigen inzicht worden gehanteerd wat betreft het gebruik van didactische werkvormen. De docentenhandleiding bevat hieromtrent ook suggesties. De opgaven bestaan zowel uit reproductievragen als uit productieve toepassingsvragen. Ook Dag in, dag uit (Muusses) laat de docent vrij in het gebruik van didactische werkvormen. De docentenhandleiding bevat hieromtrent geen suggesties. De opgaven in het Dag in, dag uit zijn overwegend reproductievragen. De verschillen in hoeveelheid leerstof tussen de vier onderwijsleerpakketten zijn weliswaar gering, maar er zijn wel duidelijke accentverschillen in de hoeveelheid leerstof naar thematiek en in de leerstofopbouw. Wat betreft de differentiatiemogelijkheden zijn er geen verschillen. Alle vier onderwijsleerpakketten zijn gebaseerd op het Basisstof - Herhaalstof - Verrijksstof-model.

De Economietoets

Bij het ontwikkelen van de leerlingtoets(en) voor klas 3 is ernaar gestreefd een voor alle methoden zo inhoudsvalide mogelijke toets samen te stellen: een toets waarvan (1) de inhoud overeenstemt met de curriculuminhoud en (2) de relevantie door de belanghebbenden achteraf niet kan worden betwist. Voor klas 4 is besloten de eindexamenresultaten te gebruiken. Deze resultaten zijn door het CITO beschikbaar gesteld. Voor het ontwikkelen van de leerlingtoets voor klas 3 is de volgende procedure gevolgd.

Op grond van de uitgevoerde inhoudsanalyse van de vier onderwijsleerpakketten bleek het leerstofaanbod in de vier pakketten geordend te kunnen worden in 10 thema's (zie voor een overzicht van deze thema's Van den Berg, e.a. 1986 of Van den Berg, 1987). Deze 10 thema's zijn vervolgens samen met de vier vakdoelstellingen gecombineerd in een matrix en voorgelegd aan de uitgevers c.q. auteursgroepen, met het verzoek voor hun pakket de gewenste verdeling van toetsitems over de thema's en doelstellingen aan te geven. Daartoe mocht men maximaal 20 toetsitems over de matrix verdelen. De uitgevers bleken geheel verschillende prioriteiten te leggen (zie Van den Berg, e.a., 1986). De door de uitgevers aangegeven voorkeuren voor toetsitems zijn verder uitgangspunt geweest bij de keuze van te ontwikkelen toetsitems per thema-doelstelling-combinatie en het verdere toetsconstructieproces.

Het ontwikkelen van toetsitems voor de verschillende thema's sec leverde betrekkelijk weinig problemen op. In combinatie met de vakdoelstellingen des te meer. Uit het vooronderzoek was reeds gebleken dat het meetbaar maken van de doelstellingen 2 en 4 op problemen stuitte. Zo bleek de auteursgroep van Versluys aan deze doelstellingen een heel

andere betekenis te willen geven dan het SLO-projectteam. Desondanks zijn toch verschillende pogingen gedaan om tot een verantwoorde operationalisering van deze doelstellingen te komen. In verschillende try-outs zijn toetsitems beproefd. Op grond van één de psychometrische kwaliteit van deze (sub)toetsen én de verschillen van mening tussen de auteursgroepen over de adequaatheid ervan (m.n. tussen Versluys en Malmberg) is uiteindelijk afgezien van meting van deze doelstellingen. Het gebruik van deze toetsen ware, vanuit het weddenschapmodel gezien, aanvaardbaar geweest als tussen de partijen overeenstemming zou hebben bestaan. Hiervan was echter geen sprake. Door de nood gedwongen zijn de uitgevers accoord gegaan met de argumenten van de onderzoekers om deze (sub)toetsen niet in het onderzoek te betrekken. Het niet 'meten' van deze doelstellingen betekent volgens de SLO echter wel dat de belangrijkste vernieuwende kenmerken van het Economie-onderwijs op het MAVO buiten de evaluatie blijven.

De voor klas 3 ontwikkelde prestatietoets bestond uit 30 multiple choice en 10 open vragen. De multiple choice-vragen zijn met name reproductievragen. Een drietal toetsvragen bleek de interne betrouwbaarheid dusdanig te beïnvloeden dat ze alsnog uit de toets verwijderd zijn. De multiple choice vragen zijn voor het merendeel reproductievragen. De open vragen zijn voornamelijk produktieve toepassingsvragen. De procedure bij het beoordelen van de open vragen was als volgt. Allereerst zijn van 100 leerlingen de antwoorden op de open vragen door drie deskundigen onafhankelijk van elkaar beoordeeld. Vervolgens is de inter-beoordelaars betrouwbaarheid berekend. Deze was zodanig hoog (.96) dat besloten is de leerlingen verder door steeds één deskundige te laten beoordelen.

Omdat bij de toetsconstructie rekening moest worden gehouden met de wensen van de partijen, is de toets nogal divers van samenstelling qua leerstofinhoud. Het is derhalve niet te verwachten dat de interne consistentie van de toetsen erg hoog is. De berekende betrouwbaarheidscoefficiënt, uitgedrukt in Cronbach alpha, is voor het multiple choice gedeelte .54. De Cronbach alpha voor het open gedeelte was .63; en voor de totale Economietoets .67. De betrouwbaarheid van het multiple choice gedeelte van de eindexametoets is voor het C-gedeelte .39 en voor het D-gedeelte .55.

Na de instrumentontwikkeling is de uitgevers gevraagd hun oordeel te geven over de door het RION ontwikkelde toetsen voor klas 3 en over het multiple choice-gedeelte van het eindexamen 1985. De uitgevers c.q. auteursgroepen is gevraagd (in het kader van dit vergelijkbare evaluatie-onderzoek van hun methode) de toetsen te beoordelen op een 5-puntsschaal lopend van geheel onbelangrijk tot zeer belangrijk. Deze oordelen worden gepresenteerd in tabel 1.

Tabel 1 Het belang dat de educatieve uitgevers hechten aan de toetsen, zoals deze zijn ontwikkeld, ter beoordeling van de doelmatigheid van hun respectievelijke onderwijsleerpakket

Toets c.q. attitude-schaal	Malmberg	Versluys	Wolters-Noordhoff	Muusses
<i>Klas 3</i>				
M.C.	enigszins belangrijk	belangrijk	belangrijk	enigszins belangrijk
Open	zeer belangrijk	enigszins belangrijk	belangrijk	enigszins belangrijk
<i>Eindexamen</i>				
niveau C en D	enigszins belangrijk	nauwelijks belangrijk	belangrijk	zeer belangrijk

Uit deze tabel blijkt duidelijk hoe verschillend de uitgevers oordelen over de in het onderzoek gebruikte criteriumvariabelen. Het meest opvallend zijn de verschillen van mening over het belang van de eindexamentoetsen en het open gedeelte van de toets voor klas 3 voor het vaststellen van de doelmatigheid van hun resp. onderwijsleerpakket.

De verschillen in waardering van de afzonderlijke toetsen illustreren weliswaar hoe moeilijk het is over alle effectcriteria afzonderlijk (volledige) overeenstemming te bereiken, maar het gehele toetsinstrumentarium overziend, kan toch worden gesteld dat een redelijke mate van overeenstemming tussen de uitgevers is bereikt wat betreft de aanvaardbaarheid van het toetsinstrumentarium.

Uit het voorgaande zal duidelijk zijn geworden dat de onderzoekers sterk sturend zijn opgetreden bij de toetsconstructie. Deze sturing was nodig om enige garantie te hebben dat aan alle partijen evenveel recht kon worden gedaan.

Overige onderzoeksinstrumenten

Ter vaststelling van het beginniveau is het begrip geoperationaliseerd als een combinatie van aanleg (de score op een IQ-test en een index voor sociaal milieu) en voorkennis (het rapportcijfer voor Economie in klas 2). Als intelligentietoets is gebruikt de zgn. TIB-toets (zie Visser, 1982). Na afname bleek echter de interne consistentie van deze toets zo laag (Cronbach $\alpha < .30$) dat besloten is deze toets verder buiten beschouwing te laten. Daarom is als maat voor beginniveau alleen gebruik gemaakt van het eindrapportcijfer voor Economie in klas 2 gecorrigeerd door de (sociaal) milieuscore van de leerling. Omdat de dataverzameling plaatsvond vlak voor de examenperiode, was het helaas niet mogelijk gegevens met betrekking tot het beginniveau te verzamelen bij de eindexamenleerlingen (klas 4).

Ter vaststelling van de mate van gebruik is een poging gedaan voor de didactische component een itemschaal 'interactief lesgeven' te ontwikkelen. Interactief lesgeven wordt door de SLO namelijk als een belangrijke voorwaarde gezien ter realisering van de vakdoelstellingen (zie SLO, 1981). Met de ontwikkelde schaal wordt de perceptie van de docenten gemeten van het eigen lesgedrag. De interne consistentie van deze schaal was redelijk tot goed (Cronbach's $\alpha = .71$). In het onderzoek moest worden volstaan met het meten van de perceptie van het eigen lesgedrag, omdat observatie van het feitelijke lesgedrag op praktische gronden niet mogelijk was. Wat betreft de leerstofinhoudelijke component is de docenten gevraagd aan te geven hoeveel leerstof (basisstof) men heeft behandeld. De hoeveelheid behandelde leerstof is uitgedrukt in het percentage basisstof dat is behandeld. Daarnaast is de docenten gevraagd het gemiddeld aantal lessen aan te geven dat per week aan Economie is besteed.

DE ONDERZOEKSGROEP

In totaal zijn 124 scholen benaderd voor deelname aan het onderzoek. In het onderzoek zijn zowel categorale MAVO-scholen als scholengemeenschappen met een MAVO-afdeling betrokken. Uit het scholenbestand is een gestratificeerde steekproef getrokken waarbij als stratificatiecriterium is gehanteerd de in gebruik zijnde Economie-methode. Bij de steekproeftrekking is gebruik gemaakt van door de uitgevers beschikbaar gestelde bestanden van scholen die volgens hen hun methoden gebruiken. Uit deze bestanden zijn a-select scholen getrokken. Daarbij is ernaar gestreefd de steekproef zodanig samen te stellen dat alle methoden in ongeveer gelijke mate in het onderzoek vertegenwoordigd zijn. Op deze manier zijn 124 scholen geselecteerd en benaderd voor deelname. Dertig scholen weigerden. Tijdgebrek werd in bijna alle gevallen als reden voor weigering opgegeven. Verder konden acht scholen niet aan het onderzoek meewerken vanwege ziekte van de docent, of omdat aan de school geen MAVO-afdeling verbonden was. Een aantal scholen ($n = 17$) moest uit de steekproef worden verwijderd, omdat de gegevens incompleet waren (leerling- of docentgegevens ontbraken) of omdat niet éénduidig kon worden vastgesteld welke methode door de docent werd gebruikt.

De uiteindelijke steekproef van scholen ($n = 69$) omvat in totaal 1680 leerlingen van klas 3, verdeeld over 69 docenten (dit is ruim 60% bruikbare respons); 529 eindexamen-leerlingen

met niveau C, en 2368 eindexamen-leerlingen met niveau D, verdeeld over 65 docenten (dit is ruim 50% bruikbare respons). Van de 69 docenten van klas 3 gebruikten 24 Economie MAVO (Versluys), 16 Economisch Bekeken (Malmberg), 15 Economavo (Wolters-Noordhoff) en 14 Dag in, dag uit (Muusses).

ONDERZOEKSRESULTATEN

Om te kunnen vaststellen of de onderwijsleerpakketten voldoen aan de verwachtingen van de uitgevers is gebruik gemaakt van de door hen voorspelde leerprestaties. Van doelmatigheidsbepaling met behulp van een minimumstandaard is afgezien, omdat de verschillen in waardering van de evaluatiecriteria zodanig waren dat redelijkerwijs niet mocht worden verwacht dat onderhandelingen over een dergelijke standaard tot overeenstemming zou leiden. De tijd daarvoor was te kort.

Om de relatieve doelmatigheid van de onderwijsleerpakketten te kunnen vaststellen, is gebruik gemaakt van variantie-analyse. Daarbij is statistische significantie ($p < .01$) gebruikt als relatieve standaard. Weliswaar is overwogen het percentage verklaarde variantie als aanvullende standaard te gebruiken bij eventueel gevonden statistisch significante verschillen, maar door tijdgebrek is afgezien om over deze aanvullende standaard te 'onderhandelen'.

De door de uitgevers verwachte leerprestaties

Ten behoeve van dit onderdeel van het onderzoek is uitgegaan van het zgn. slaagpercentage leerlingen met een voldoende op de toetsen.

Op het multiple choice gedeelte van de derde klas toets is een voldoende toegekend als 15 of meer van de 27 toetsvragen goed zijn beantwoord. Op de open vragen van de derde klas toets konden in totaal maximaal 40 punten worden behaald, 4 punten per onderdeel. Een voldoende is toegekend als 22 punten of meer zijn behaald. Op de eindexamentoets (zowel C- als D-niveau) is een voldoende toegekend als 17 of meer van de 30 vragen goed zijn beantwoord.

De 'Malmberg'-leerlingen scoren op alle onderdelen lager dan het gemiddelde van de totale steekproef¹⁾ en de 'Versluys'-leerlingen op alle onderdelen hoger. De 'Muusses'- en 'Wolters-Noordhoff'-leerlingen zitten steeds rond het gemiddelde.

Opvallend is ook dat Malmberg het open gedeelte van de Economietoets voor klas 3 weliswaar zeer belangrijk vindt als criteriumvariabele, maar op deze toets toch de slechtste prestaties realiseert ook en beduidend lager dan haar eigen verwachting. Versluys en Muusses vinden dit toetsonderdeel veel minder belangrijk, maar realiseren wel de beste resultaten die ook in overeenstemming zijn met hun eigen verwachtingen. Hoewel Versluys relatief gezien het meest afwijzend staat, bereikt zij over het gehele toetsinstrumentarium gezien de beste resultaten. Bij de interpretatie van deze resultaten moet wel uitdrukkelijk worden aangetekend dat nog geen rekening is gehouden met storende omstandigheden.

Vrijwel alle verwachtingen van de uitgevers bleken uit te komen. Vooral de goede voorspellingen van Wolters-Noordhoff vallen op. Malmberg en Muusses blijken elk één keer hun pretenties niet waar te maken. Malmberg maakt haar pretenties niet waar voor wat betreft het open gedeelte van de derde klas toets en Muusses niet voor wat betreft de eindexamentoets op C-niveau. Zowel Malmberg en Muusses tenderen met hun voorspellingen steeds tot overschatting (t -waarden positief). Versluys tendert daarentegen steeds tot onderschatting (t -waarden negatief).

De relatieve doelmatigheid

Met behulp van variantie-analyse is nagegaan of verschillen in leerprestaties zijn toe te schrijven aan de onderwijsleerpakketten. Als maat voor leerprestaties is gebruik gemaakt van de absolute score op de toetsen (dit wil zeggen het aantal goede antwoorden). De gegevens zijn geanalyseerd op leerlingniveau. Daartoe zijn de leerkracht- en klassekenmerken gedisaggregeerd, dit wil zeggen dat leerkracht- en klassekenmerken als contextvariabelen van de leerlingen worden

Tabel 2 Het voorspelde en het gevonden slaagpercentage en de *t*-waarde per toets

TOETS	Voorsp. slaag percentage			Gev. slaag percentage gemid.	<i>t</i> *
	min.	modus	max.		
MALMBERG					
<i>klas 3</i>					
Multiple-choice	40	45	70	49	+ 0.331
Open	40	48	60	32	+ 2.450
<i>Eindexamen</i>					
Niveau C	15	20	25	14	+ 0.986
Niveau D	22	30	35	18	+ 1.793
VERSLUYS					
<i>klas 3</i>					
Multiple-choice	45	53	59	56	- 0.668
Open	35	42	49	45	- 0.499
<i>Eindexamen</i>					
Niveau C	15	18	21	19	0.000
Niveau D	23	26	29	29	- 0.551
WOLTERS-NOORDHOFF					
<i>klas 3</i>					
Multiple-choice	45	52	57	51	0.000
Open	30	39	45	38	0.000
<i>Eindexamen</i>					
Niveau C	15	17	20	18	0.000
Niveau D	24	27	30	27	0.000
MUUSSES					
<i>klas 3</i>					
Multiple-choice	45	49	70	47	+ 0.993
Open	40	44	55	44	+ 0.323
<i>Eindexamen</i>					
Niveau C	28	42	56	21	+ 2.508
Niveau D	35	42	50	31	+ 1.782

**t* = Het aantal standaarddeviaties dat het voorspelde slaagpercentage van het gevonden slaagpercentage af ligt. De berekende *t*-waarden moeten als volgt worden geïnterpreteerd:

1. Als $t > +2$, dan is sprake van overschatting;
2. Als $t < -2$, dan is sprake van onderschatting;
3. Als $-2 < t < +2$, dan is sprake van een juiste schatting.

opgevat en dat leerlingen binnen eenzelfde klas gelijke scores krijgen voor de kenmerken.

Omdat vooraf door de uitgevers geen standaarden zijn geformuleerd om te kunnen vaststellen of een onderwijsleerpakket al dan niet voldoet aan een bepaalde minimeis, moest worden volstaan met een relatieve standaard. Daartoe is gekozen voor toetsing op significantie ($p < .01$). In tabel 3 worden de resultaten gepresenteerd van de uitgevoerde variantie-analyses.

Uit deze tabel blijkt dat de verschillen in leerprestaties op de toetsen van klas 3 tussen de vier groepen leerlingen (die met de verschillende methode onderwijs hebben ontvangen) niet significant zijn. Als echter wordt gecorrigeerd voor beginniveau en voor de werkwijze van de docent, de hoeveelheid behandelde leerstof en het aantal lessen per week, dan blijken de verschillen wel significant te zijn. Omdat is nagelaten te controleren op interactie-effecten moet grote voorzichtigheid worden betracht bij het interpreteren van deze resultaten. Om te kunnen beslissen aan welke methoden onder welke omstandigheden de voorkeur moet worden gegeven, had het gebruik van verklaarde variantie als aanvullende standaard mogelijk uitkomst kunnen brengen.

Tabel 3: Samenvatting van de uitgevoerde variantie-analyses

Toets	<i>F</i>	<i>df</i>	<i>n</i>	<i>p</i>
<i>klas 3</i>				
<i>m.c.</i>				
methode	1.19	3	1680	n.s.
methode + beginniveau	6.13	3	1193	<.01
meth. + werkw. + aant. lesuren	7.76	3	1204	<.01
meth. + werkw. + aant. lesuren + beh. leerstof	3.71	3	1204	<.01
<i>open</i>				
methode	3.14	3	1680	n.s.
methode + beginniveau	7.87	3	1193	<.01
meth. + werkw. + aant. lesuren	21.34	3	1204	<.01
meth. + werkw. + aant. lesuren + beh. leerstof	10.48	3	1204	<.01
<i>Eindexamen</i>				
<i>C-niveau</i>				
methode	2.77	3	529	n.s.
meth. + werkw. + beh. leerstof	1.71	3	227	n.s.
meth. + werkw. + beh. leerstof + aant. lesuren	0.74	3	208	n.s.
<i>D-niveau</i>				
methode	3.27	3	2368	n.s.
meth. + werkw. + beh. leerstof	2.07	3	904	n.s.
meth. + werkw. + beh. leerstof + aant. lesuren	9.66	3	869	<.01

De verschillen in leerprestaties op de eindexamentoetsen (op C- en D-niveau) tussen de groepen leerlingen (die onderwijs hebben gekregen met de verschillende methoden) kunnen niet worden toegeschreven aan de methoden zoals deze in het onderwijs worden gebruikt door de docenten en de leerlingen. Het grote aantal missing cases dwingt hier echter tot het betrachten van voorzichtigheid bij het trekken van vergaande conclusies.

SAMENVATTING EN CONCLUSIE

Omdat het formuleren van evaluatiecriteria en -standaarden in feite moet worden gerekend tot de verantwoordingsplicht van curriculum-ontwikkelaars, beleidsmakers, enz., legt het weddenschapmodel de verantwoordelijkheid ook daar waar ze thuishoort. Hoewel de ervaringen met het weddenschapmodel positief zijn te noemen, kan niet worden gesteld dat de in het evaluatieonderzoek Economic-MAVO toegepaste procedure heeft geleid tot de oplossing van het criterium- en standaardprobleem.

Rekening houdend met het feit dat de onderwijsleerpakketten zowel een aantal gemeenschappelijke als een aantal specifieke kenmerken hebben is het niet eenvoudig over alle onderdelen van het toetsinstrumentarium volledige overeenstemming te bereiken. Wanneer men de waardering van de auteurs over het totale toetsinstrumentarium in beschouwing neemt, mag toch worden geconcludeerd dat een redelijke mate van overeenstemming is bereikt. Rekening houdend met het feit dat slechts beperkte tijd beschikbaar was voor toetsconstructie, mag worden verwacht dat met behulp van de in het Economic-MAVO-onderzoek gehanteerde procedure ook in andere contexten een redelijke mate van overeenstemming te bereiken is over de te gebruiken evaluatiecriteria.

Het laten doen van voorspellingen (over de kwaliteit van het eigen pakket) is slechts geschikt om vast te stellen of het gebruik van de onderwijsleerpakketten leidt tot het bereiken van de doelen die de auteurs verwachten. De betekenis hiervan in vergelijkende productevaluatie is echter beperkt. Ter bepaling van de relatieve doelmatigheid van curricula, programma's, e.d. is variantie-analyse een geschikte analysetechniek. Het significantieniveau als relatieve standaard is echter niet toereikend als significante verschillen worden gevonden. Het is dan nodig te zoeken naar een geschikte aanvullende standaard; een standaard waarover de betrokkenen dan vooraf overeenstemming moeten bereiken. Het is te overwegen het percentage verklaarde variantie als aanvullende standaard te hanteren. In vergelijkend evaluatie-onderzoek is variantie-analyse als techniek uit de klassieke hypothesetoetsende traditie zeer geschikt voor de oplossing van het standaardprobleem en is als zodanig te gebruiken in combinatie met of als onderdeel van het weddenschapsmodel.

- 1) Het gemiddelde voor de totale steekproef was voor het mc.-gedeelte van de derde klas toets 51%, voor het open gedeelte van de derde klas toets 40%, voor de eindexamentoets op C-niveau 18% en voor de eindexamentoets op D-niveau 27%.

LITERATUUR

- Berg, G. van den & E. Harskamp (1985): *Twee onderwijsleerpakketten Economie-MAVO vergeleken* (Eindrapport SVO-project 1151; een vooronderzoek). Groningen, RION.
- Berg, G. van den, F.B. Brokken & J.W. Prins (1986): *Vergelijking van vier onderwijsleerpakketten Economie voor het MAVO* (eindrapport SVO-project 4213). Groningen, RION.
- Berg, G. van den (1987): *Effectief evalueren*; Een empirische studie naar de doelmatigheid van aanwijzingen voor de evaluatiepraktijk, diss., Lisse.
- Creemers, B.P.M. & J. Schaveling (1985): *Verhoging van onderwijseffectiviteit*: pre-advies uitgebracht aan de Wetenschappelijke Raad van het Regeringsbeleid in het kader van het project over basisvorming in het voortgezet onderwijs. 's-Gravenhage.
- Eijl, van P. (1980): Beginniveau. In: *Handboek onderwijspraktijk*.
- Groot, A.D. de (1982): 'Dekkingsproblemen' bij de constructie van examenprogramma's en curricula. In: *Handboek onderwijspraktijk*.
- Hoeben, W.Th.J.G. (1985): Onderzoek en ontwikkeling, een weerbarstige relatie. In: *Onderwijskundige Notities* 5, (2), 3-12.
- Hofstee, W.K.B. (1980): *De empirische discussie*. Theorie van het sociaal-wetenschappelijk onderzoek, Meppel.
- Hofstee, W.K.B. (1982): Evaluatie: Een methodologische analyse. In: *Tijdschrift voor Onderwijsresearch*, 7 (5), 193-202.
- Hofstee, W.K.B. (1984): Methodological decision rules as research policies: a betting model reconstruction of empirical research. In: *Acta Psychologica*, 56, 93-109.
- Klerk, L.F.W. de (1984): Het leren van praktische vaardigheden. In: L.F.W. de Klerk en A.M.P. Knoers: *Onderwijspsychologisch onderzoek*. Lisse, p. 44-53.
- Scheerens, J. (1983): Evaluatie-onderzoek en beleid: Methodologische en organisatorische aspecten. Den Haag, (SVO-reeks 68).
- Stichting voor Leerplanontwikkeling (1982/'83): *Economie MAVO eindexamen '85*, deel I t/m VI, 1981, deel V.
- Stokking, K.M. (1984): *Interpreteren en evalueren*: methodologie rondom de uitkomsten van onderzoek. Deventer.
- Visser, R.S.H. (1982): *Documentatie van tests en testresearch in Nederland*, Nijmegen.

Manuscript ontvangen 17-6-1987.

Definitieve versie ontvangen 13-7-1987.

De toepassing van een nieuw leerprocessen-model op het memoriseren van L.O.M.- en basisschool-leerlingen

K.P. van den Bos*, J. Kingma** en H. Scheper*
Rijksuniversiteit Groningen, University of Alberta

ABSTRACT

In this article we apply a new two-stage model of learning in order to investigate more precisely the nature of learning-disabled (LD) children's storage and retrieval problems in various experimental conditions of a memorization task. The LD children's performance is compared to that of normally learning (NL) children. The task involves the free recall, to a strong acquisition criterion, of a word list. Word lists either contain words from two semantic categories or from four categories. Test phase instructions are varied by either asking standard questions or offering retrieval cues in the form of category labels. In terms of global performance measures, the results show mixed profiles as a function of categorization. On the other hand, both LD and NL children's recall seems to benefit from cueing. In all of the conditions, however, LD children's recall is lower than NL children's recall. One of the main conclusions resulting from the two-stage model analyses is that the LD children's lower memorization performance can primarily be attributed to their greater difficulties in the acquisition of stable retrieval procedures (retrieval algorithms) and not so much to storage operations. The discussion sketches some instructional implications of the present research.

1. INLEIDING

Aangezien er een nauw verband is tussen leren en geheugen, is het niet verrassend dat vele onderzoekers hebben aangenomen dat onderzoek naar geheugenprestaties van leergestoorde kinderen op zijn minst een gedeeltelijk antwoord zou kunnen geven op de vraag wat er eigenlijk mis is met het leren van deze kinderen. Over het onderwerp 'geheugen bij leergestoorde kinderen' bestaat een uitgebreide literatuur.

* Vakgroep Orthopedagogiek van het in zijn ontwikkeling belemmerde kind, Rijksuniversiteit Groningen.

** Department of Psychology, University of Alberta.

Wij danken het Rekencentrum van de Rijksuniversiteit Groningen voor het berekenen van een nul-tarief voor de vele computerruns die de toepassing van het twee-stadia model met zich meebracht. Voorts zij vermeld dat dit onderzoek mogelijk was door subsidie van de Vakgroep Orthopedagogiek I, R.U.G. (Kpl 1634301), The Natural Sciences and Engineering Research Council, Canada (Grant A1235) en de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (Z.W.O.), subsidiernr. H56-314.

Hieruit blijkt dat leergestoorde kinderen als groep lager presteren dan normaal-lerende kinderen in allerlei geheugentaken (Bauer, 1979; Ceci, 1983; Dallago & Moely, 1980; Perfetti, 1985; Torgesen, 1978-1979; Torgesen & Kail, 1980; Van den Bos, 1983). De evidentie betreft zowel korte-termijn geheugentaken – zeker als deze gedefinieerd worden als taken waarin kortdurende opslag van relatief kleine informatiehoeveelheden die eenmalig zijn aangeboden is vereist en waarbij vrijwel direct na presentatie van de proefpersoon reproductie wordt gevraagd (Torgesen, 1982) – als het presteren op minder vaak afgenomen taken, zoals lange-termijn geheugentaken, waarin bijvoorbeeld relatief lange rijen met woorden of plaatjes moeten worden gememoriseerd tot een criterium van 100% correcte reproductie (Howe, Brainerd & Kingma, 1985; Van den Bos & Kingma, 1985; Van den Bos, Kingma & Bakker, 1986).

Een reeks van min of meer traditionele variabelen is in de literatuur voorgesteld om het contrast tussen de geheugenprestaties van de leergestoorde en normaal-lerende kinderen te verklaren. Naast de variabelen motivatie en aandacht (zie bijv. Ross, 1976) vinden we in veel studies (Torgesen, 1982) variabelen die ontleend zijn aan het model van Atkinson & Shiffrin (1968), waarin het onderscheid tussen basale processen en stuurprocessen ('control processes') gemaakt wordt. Voorbeelden van basale processen zijn de niet direct voor sturing toegankelijke processen, zoals de duur van een geheugenspoor en de snelheid waarmee toegang tot een geheugencode verkregen wordt. Voorbeelden van stuurprocessen zijn label- en repetitie-strategieën en processen die de semantische organisatie van het stimulusmateriaal verhogen. Hoewel in sommige studies naar memoriseerproblemen van leergestoorde kinderen de combinatie van basale processen en stuurprocessen centraal staat (Spring & Capps, 1974; Van den Bos, 1983) viel in de afgelopen 15 jaar in de meerderheid der studies de nadruk op de hypothese van een verminderd strategisch functioneren, een hypothese waar overigens slechts beperkte geldigheid voor blijkt te bestaan (vgl. Dallago & Moely, 1980; Pressley & Levin, 1977; Torgesen, 1977; Van den Bos et al., 1986; Worden, 1983).

Van meer fundamentele aard is echter de overweging, dat de gebruikte designs in bovengenoemde studies en het vrij oppervlakkige model van basale en strategische processen, nauwelijks recht doen aan de complexe aard van het memoriseerproces als geheel. Hierdoor kunnen geen details gegeven worden over het leerproces zoals dat in de tijd verloopt (Greeno, 1974) en blijven de bijdragen van zekere elementaire procescomponenten versluierd. Wij doelen hier op het belangrijke onderscheid tussen *opslag* van informatie in het geheugen en het *ophalen* van informatie daaruit. Hoewel het Atkinson & Shiffrin (1968) model spreekt over 'transfer of information to the long-term store' (hetgeen 'opslag' betekent) en over 'retrieval strategies' (hetgeen 'ophalen' impliceert) en hoewel deze termen vooral in de discussieparagrafen van diverse vergelijkende studies regelmatig worden gebruikt (Dallago & Moely, 1980; Ross, 1976; Worden, 1983), worden ze niet systematisch (c.q. model-gebaseerd) gebruikt en zijn er nauwelijks pogingen gedaan de bijdragen van deze processen aan het memoriseren empirisch vast te stellen. Evenmin was dit overigens het geval in de meeste 70-er jaren ontwikkelingspsychologische studies naar het memoriseren van normaal-lerende kinderen (Rohwer & Dempster, 1977). Bovendien kunnen de studies waarin wél een empirisch onderscheid tussen opslaan en ophalen werd nagestreefd gekritiseerd worden wegens een gebrek aan goed-gedefinieerde assumpties omtrent de meetschalen van relevante procesvariabelen (Brainerd, 1985).

Recentelijk is echter een oplossing voor deze problemen (te weten: het onvoldoende model-gebaseerde onderzoek naar het onderscheid tussen opslaan en ophalen en het 'schaalprobleem') geboden door Brainerd e.a. (zie bijv. Brainerd, Howe & Kingma, 1982; Brainerd, Howe, Kingma & Brainerd, 1984). Deze onderzoekers pasten op memoriseerprestaties van kinderen in diverse experimenten een nieuw *leerstadia*-model toe, waarvan de parameters geïnterpreteerd kunnen worden als metingen van diverse aspecten van het opslaan en ophalen. Bovendien heeft het model sterke, mathematisch gefundeerde schaalassumpties (zie Brainerd, 1985) waardoor haar parameterwaarden in een bepaald experiment op ratio-schaal niveau bepaald kunnen worden (zie de verderop in dit artikel gegeven toelichtingen).

Gegeven de succesvolle toepassing van het nieuwe leerstadia-model op memoriseerdata van normaal-lerende kinderen (Brainerd, 1985), ligt het voor de hand het eerdergenoemde proble-

matische memoriseren van leergestoorte kinderen evenééns met behulp van het nieuwe leerstadia-model te analyseren. Meer specifiek is het doel van deze studie om juist de in de traditionele literatuur 'diffuus' behandelde bijdragen van opslaan en ophalen in het memoriseren van leergestoorte kinderen te localiseren.

Aansluitend bij de in onderzoek naar het memoriseren meest gangbare paradigma's, wordt in de huidige studie een memoriseertaak met vrije reproductie-volgorde ('free recall') gebruikt. Tevens worden echter – aansluitend op Amerikaans onderzoek hiernaar bij leergestoorte en normaal-lerende kinderen – de effecten van bepaalde variaties van het stimulusmateriaal en de instructie onderzocht.

De variatie van het stimulusmateriaal houdt in dat er twee lijstsoorten zijn waarbinnen de gradatie in semantische georganiseerdheid van de te memoriseren woorden verschilt. Deze variatie is gekozen om na te gaan of leergestoorte en normaal-lerende kinderen bij het memoriseren op dezelfde manier profiteren van een impliciet verhoogde graad van semantische organisatie van het stimulusmateriaal. Op grond van eerdere experimentele resultaten (Parker, Freston & Drew, 1975; Torgesen, 1977) zou verwacht worden dat de leergestoorte kinderen 'spontaan' minder profiteren van de verhoogde graad van semantische organisatie dan de normaal-lerende kinderen. In deze eerdere onderzoeken was echter slechts sprake van effectmeting in termen van een globale score, zoals het aantal correct herinnerde items na een x-aantal studierondes. Het huidige onderzoek wil echter het effect van de variabele 'semantische organisatie' op details van het memoriseerproces (te weten de *opslag*- en *ophaal*componenten) nagaan.

De tweede variatie in de taak is die van de instructie. Deze variatie houdt in dat er twee instructiecondities zijn: een standaardinstructie en een conditie waarbinnen het kind expliciet gewezen wordt op de semantische georganiseerdheid van de te memoriseren woorden. De vraag is weer naar het effect van deze instructievariatie bij leergestoorte en normaal-lerende kinderen. In termen van metingen van het globale memoriseerproces suggereert de traditionele literatuur profijt van de expliciete instructie bij zowel de leergestoorte als de normaal-lerende kinderen (Ross, 1976; Torgesen, 1977), hetgeen overigens nog niet wil zeggen dat daarmee de prestaties van de groepen kinderen even hoog komen te liggen. Hoewel deze voorspellingen nagegaan kunnen worden, vallen ze niet onder de centrale doelstelling van deze studie. Primair immers zijn we geïnteresseerd in de vraag welke aspecten van het *opslaan* en *ophalen* onder invloed van de meer expliciete instructie verbeteren bij de groepen kinderen, en welke van de in het twee-stadia model vervatte processen in de instructiecondities eventueel verschillend zijn voor leergestoorte en normaal-lerende kinderen.

Eerst wordt nu het nieuwe leerstadia model besproken.

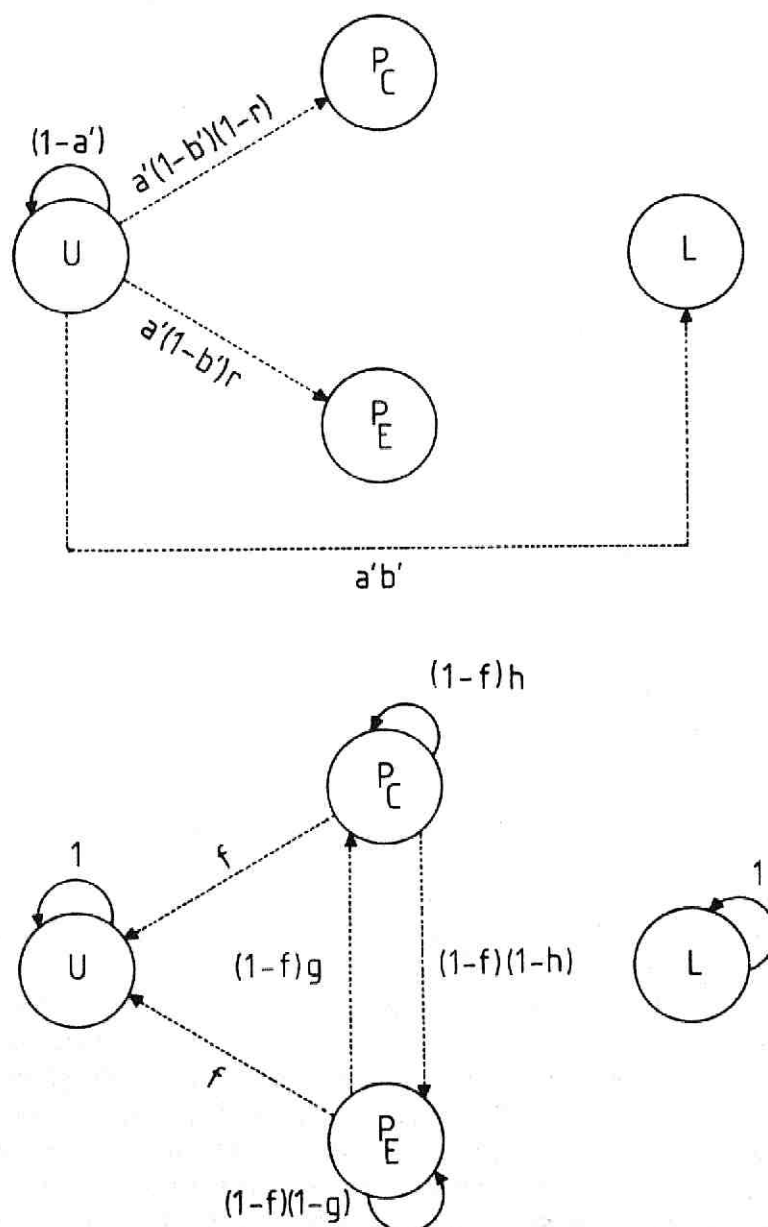
2. HET TWEE-STADIA MODEL VAN HET LEREN

Het leren van materiaal in memoriseertaken kan beschreven worden als een verandering in de mentale status van items. In het twee-stadia model worden ten aanzien van dit proces drie leertoestanden en twee leerstadia onderscheiden. De eerste leertoestand van items wordt *U* ('unlearned') genoemd. Items in deze toestand kunnen niet gereproduceerd worden. In ieder geval is van deze toestand sprake voordat het te onthouden materiaal wordt aangeboden. Maar ook na de eerste, en eventueel volgende studieronde(s) is het soms mogelijk dat een item nog steeds de *U*-toestand niet verlaten heeft. De tweede onderscheiden leertoestand wordt *P* (van 'partially learned') genoemd. Een item is in de *P*-toestand wanneer het weliswaar uit *U* is ontsnapt, maar nog niet stabiel reproduceerbaar is: soms wordt het gereproduceerd (subtoestand *P_C*; de *C* staat voor 'correct') en soms niet (subtoestand *P_E*; de *E* staat voor 'error'). De derde onderscheiden leertoestand wordt *L* (van 'learned') genoemd. In deze toestand, die uiteindelijk door de meeste items bereikt wordt, wanneer een proefpersoon aan een streng leercriterium voldoet (zie Inleiding), worden items altijd correct gereproduceerd en terugval vanuit *L* naar *P* of *U* is niet mogelijk. De twee in het model onderscheiden leerstadia zijn nu:

ontsnapping uit U en het bereiken van toestand L .

De Markov-model versie die recentelijk werd ontwikkeld door Brainerd et al. (1982) neemt aan, dat in het experiment studie- en testrondes op de volgende wijze alternieren: $S_1, T_{1A}, T_{1B}, S_2, T_2, S_3, T_3, \dots$ Hier betekent S_1 de eerste studieronde van een lijst; T_{1A} betekent de eerste testronde. Deze testronde wordt gevolgd door een tweede testronde (T_{1B}). S_2 betekent de tweede studieronde; T_2 betekent de testronde die volgt op de tweede studieronde, etc. Het is belangrijk op te merken dat in deze experimentele structuur alleen de eerste studieronde gevolgd wordt door twee testrondes. De S-T alternaties gaan door totdat voor een proefpersoon het verwervingscriterium wordt bereikt, dat in ons geval bestaat uit twee achtereenvolgende testrondes met foutloze reproducties van de items van een lijst. Voor dergelijke experimenten nu wordt het leren als volgt beschreven:

De eerste leermogelijkheid (d.w.z., studieronde S_1) heeft voor sommige items het effect dat ontsnapping uit Toestand U plaatsvindt (met een kans a'). Andere items ontsnappen echter niet uit Toestand U (met een kans $1-a'$). De items die ontsnappen uit Toestand U kunnen of Toestand



Figuur 1: Kansen van leertoestanden voor items in de eerste testronde (bovenste helft van de figuur) en in de tweede testronde (onderste helft van de figuur).

L direct bereiken (met een kans b'), of zij nemen één van de twee sub-toestanden (P_E of P_C) van de partieel-geleerde Toestand P aan (met een kans $1-b'$). De waarschijnlijkheid van het maken van een reproductiefout, of van een correcte reproductie, voor items in Toestand P in de eerste testronde is respectievelijk r of $1-r$. In het bovenste deel van Figuur 1 worden de kansen van de leertoestanden der items in testronde T_{1A} grafisch weergegeven.

Vervolgens bezien we de kansen van de mogelijke leertoestanden die de items aan kunnen nemen in de tweede testronde (T_{1B}) – zie voor een grafische weergave het onderste deel van Figuur 1. Aangezien er tussen de testrondes T_{1A} en T_{1B} geen studieronde zit, is voor items die in testronde T_{1A} niet uit U waren ontsnapt (kans: $1-a'$), de kans dat ze nog steeds (T_{1B}) in U verblijven één. Evenzo geldt dit voor naar L ontsnapte items. Hieruit is geen terugval mogelijk. Met andere woorden, de kans dat de in L verblijvende items in testronde T_{1B} nóg in L verblijven is één. Wanneer een item daarentegen in T_{1A} de P -toestand had aangenomen, kan het in T_{1B} of terugvallen naar U (met een kans van f), of in de P -toestand blijven (met een kans van $1-f$). Voorts, als een proefpersoon een P -toestand item in T_{1A} niet reproduceerde (dus als het item Toestand P_E had aangenomen) dan is de kans dat in T_{1B} wel een correcte reproductie plaatsvindt g . De kans van een $P_E \rightarrow P_C$ transitie in testronde T_{1B} is dan dus $(1-f)g$, en de kans dat het item in P_E blijft is $(1-f)(1-g)$. Als de proefpersoon een P -toestand item correct reproduceerde in T_{1A} (dus als het item Toestand P_C had aangenomen), dan is de kans dat in T_{1B} het item niet gereproduceerd wordt $1-h$. De kans van een $P_C \rightarrow P_E$ transitie in testronde T_{1B} is dan $(1-f)(1-h)$, en de kans dat het item in P_C blijft is $(1-f)h$.

Het is nu gemakkelijk in te zien wat de kansen zijn van alle mogelijke toestandscombinaties in de testrondes T_{1A} en T_{1B} in de startvector V van vergelijking (1). Let wel, dat ter vereenvoudiging de onderschriften 1A en 1B in V vervangen zijn door resp. 1 en 2.

Vergelijking (1):

$$\begin{aligned}
 V &= L_1L_2, L_1P_{E2}, L_1P_{C2}, L_1U_2, P_{E1}L_2, P_{E1}P_{E2}, P_{E1}P_{C2}, P_{E1}U_2, P_{C1}L_2, P_{C1}P_{E2}, P_{C1}P_{C2}, P_{C1}U_2, \\
 &\quad U_1P_{E2}, U_1P_{C2}, U_1U_2 \\
 &= a'b', 0, 0, 0, 0, a'(1-b')r(1-f)(1-g), a'(1-b')r(1-f)g, a'(1-b')rf, 0, a'(1-b')(1-r) \\
 &\quad (1-f)(1-h), a'(1-b')(1-r)(1-f)h, a'(1-b')(1-r)f, 0, 0, 0, 1-a';
 \end{aligned}$$

$$M = \begin{matrix} & \begin{matrix} L(n+1) & P_E(n+1) & P_C(n+1) & U(n+1) \end{matrix} \\ \begin{matrix} L(n) \\ P_E(n) \\ P_C(n) \\ U(n) \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ d & (1-d)(1-g) & (1-d)g & 0 \\ c & (1-c)(1-h) & (1-c)h & 0 \\ ab & a(1-b)e & a(1-b)(1-e) & 1-a \end{bmatrix} \end{matrix} ;$$

$$C = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Na de in vergelijking (1) gegeven vector van beginkansen, worden vervolgens in het model de kansen van de transities tussen leertoestanden gespecificeerd voor items in de *tweede studieronde* en verdere studierondes. Dit wordt samengevat in de matrix (M) in vergelijking (1). De toelichting is als volgt. Tengevolge van de tweede of een latere studieronde kan een item zich nog steeds in toestand U bevinden (met een kans $1-a$), maar ook is het mogelijk dat het daaruit nu is ontsnapt (met een kans a). Weer zijn er de mogelijkheden dat het item dan direct de toestand L bereikt (met een kans b) of dat in plaats daarvan de partieelgeleerde toestand P wordt aangenomen (met een kans $1-b$). In geval van een $U \rightarrow P$ transitie is in de *eerste* testronde in de P -toestand de kans op een falende reproductie e , en de kans op een correcte reproductie $1-e$.

Een item kan zich aan het begin van de tweede of een latere studieronde natuurlijk ook in de P -toestand bevinden en wel in P_E of in P_C . De kans dat de itemrepresentatie tengevolge van een dergelijke studieronde vanuit P_E ontsnapt naar toestand L is d . De kans dat de itemrepresentatie vanuit P_E ontsnapt naar P_C is $(1-d)g$, en de kans dat zij in P_E blijft is $(1-d)(1-g)$. Hierbij is de g de kans op een correcte itemreproductie in de aansluitende testronde. De kans dat het item in de tweede of een latere studieronde ontsnapt vanuit P_C naar L is c . De kans dat het vanuit P_C ontsnapt naar P_E is $(1-c)(1-h)$, en de kans dat het in P_C blijft is $(1-c)h$. Hierbij is h de kans op een correcte item-reproductie in de aansluitende testronde.

Vergelijking (1) wordt afgesloten met een zogenaamde vector voor correct-respons kansen (C). Hierin wordt aangegeven dat de correct-respons kans voor een item in toestand L één is, de correct-respons kans voor een item in toestand P_E nul, etc. Voorts zal bij het inspecteren van vergelijking (1) duidelijk zijn, dat de kanswaarden 'nul' in V en M staan voor 'onmogelijke' toestandscombinaties.

2.1 Interpretaties van de twee-stadia modelparameters in termen van lange-termijn geheugen-componenten

De bovengenoemde modelparameters a' , a , b' , b , c , d , e , f , g , h , en r zijn tot dusver als facetten van een abstract-theoretisch leersysteem (vergelijking (1)) gedefinieerd. In de afgelopen jaren is echter door diverse onderzoekers gepoogd de parameters in termen van kansen op psychologisch reële processen te definiëren (zie bijvoorbeeld Halff, 1977). Voorgesteld zijn aspecten van de twee belangrijkste lange-termijn geheugen processen, namelijk het *opslaan* in het geheugen van informatie en het terug kunnen keren naar de opgeslagen informatie (*ophalen*).

Allereerst wordt de gebeurtenis van ontsnapping uit U in het bovengegeven abstracte leersysteem gelijkgesteld aan de mentale gebeurtenis van het *opslaan* van een itemrepresentatie in het lange-termijn geheugen. Parameter a' wordt dan de kans dat opslag in de eerste studieronde plaatsvindt; parameter a geeft de kans van opslag tijdens studierondes *na* de eerste aan, en parameter f wordt de kans van verlies ('vergeten') van een opgeslagen itemrepresentatie vóór de tweede studieronde. De opslag van een itemrepresentatie betekent niet dat ten tijde van de testronde steeds perfecte reproducties plaatsvinden. Zowel succesvolle als niet-succesvolle reproducties komen voor. Pas het bereiken van de L -toestand betekent dat een betrouwbare of steevast succesvolle *ophaal*operatie voor opgeslagen itemrepresentaties beschikbaar is.

Aangenomen mag worden (Halff, 1977), dat het betrouwbaar kunnen ophalen berust op het hebben ontwikkeld van een gespecialiseerd *algoritme* (= het volgen van een regel of verzameling regels die tot de oplossing van het probleem – in dit geval het ophalen van een item – leidt). Het aan de L -toestand inherente betrouwbaar kunnen ophalen wordt daarom verder *algoritmisch ophalen* genoemd.

We hebben gezien dat het bereiken van de L -toestand op twee manieren mogelijk is: of rechtstreeks vanuit U , of indirect via P . Wat betreft de eerste mogelijkheid worden de parameters b' en b nu geïdentificeerd met de kans dat een ophaalalgoritme verworven wordt in dezelfde studieronde dat opslag plaatsvindt, voor b' en b respectievelijk in de eerste of latere studierondes. De tweede mogelijkheid betreft het verwerven van een ophaalalgoritme bij een $P \rightarrow L$ transitie. De aan deze transitie gebonden c en d parameters worden dan geïdentificeerd met de kans dat een ophaalalgoritme verworven wordt na respectievelijk een succesvolle en niet-succesvolle reproductie in toestand P .

In (één van) de studieronde(s) vóórdat ontsnapping uit *P* plaatsvindt, zijn de itemrepresentaties reeds opgeslagen. Van de parameters *c* en *d* wordt daarom ook wel gezegd dat ze het *post-opslag algoritmisch ophalen* meten. Hiervan onderscheiden gebruikte Brainerd (1985) voor de *b'* en *b* parameters de karakterisering 'pre-opslag algoritmisch ophalen'. Aangezien echter niet uit te maken is of binnen eenzelfde studieronde de verwerving van een ophaalalgoritme voorafgaat aan of volgt op opslag, lijkt het ons veiliger voor de *b'* en *b* parameters de karakterisering *annex-opslag algoritmisch ophalen* te kiezen.

Tot slot is aan de orde de interpretatie van de parameters *r*, *e*, *g* en *h*. Dit zijn de parameters die in het abstracte model (vergelijking (1)) de kans op correcte reproductie van items in de *P*-toestand weergeven voor diverse studierondes. We zagen dat die kans $0 < p < 1$ was. Met andere woorden, itemrepresentaties in toestand *P* zijn weliswaar opgeslagen, maar nog niet consistent en foutloos op te halen.

Wanneer een item nog niet stevast op te halen is, nemen we aan dat de proefpersoon zich van een zogenaamde *heuristiek* bedient (Halff, 1977). Een heuristiek is een oplossing-zoekende regel of strategie, die echter niet noodzakelijk tot oplossing – in dit geval 'ophalen' – leidt. Het niet-consistent ophalen wordt daarom aangeduid met *heuristisch ophalen* en het contrasteert met de verwerving van een betrouwbare ophaaloperatie (het eerder genoemde *algoritmisch ophalen*). Onderscheiden in diverse *P*-toestandsmomenten kunnen de parameters *1-r* en *1-e* enerzijds, en *g* en *h* anderzijds, respectievelijk geïdentificeerd worden met 'vroeg' en 'laat' heuristisch ophalen: de parameters *1-r* en *1-e* zijn de kansen op het succesvol kunnen ophalen

Tabel 1. *LTM-proces interpretaties van de 11 parameters van het twee-stadia model*

Parameter	Interpretatie
Opslag en Retentie	
<i>a'</i>	Waarschijnlijkheid van item- <i>opslag</i> in het lange-termijn geheugen in de eerste studieronde
<i>a</i>	Waarschijnlijkheid van <i>opslag</i> in een van de studierondes na de eerste
<i>1-f</i>	Waarschijnlijkheid van <i>retentie</i> van opslag tussen de eerste en tweede testronde
Algoritmisch ophalen	
<i>b'</i>	Waarschijnlijkheid van het verwerven van een <i>ophaal-algoritme</i> voor items in de eerste studieronde, gegeven dat de items ook opgeslagen worden in die ronde
<i>b</i>	Waarschijnlijkheid van de verwerving van een <i>ophaal-algoritme</i> voor items in een studieronde na de eerste, gegeven dat de items ook in die latere ronde worden opgeslagen (de parameters <i>b'</i> en <i>b</i> kunnen worden samengevat als metende het <i>annex-opslag algoritmisch ophalen</i>)
<i>c</i>	Waarschijnlijkheid van het verwerven van een <i>ophaal-algoritme</i> na een succes in Toestand <i>P</i>
<i>d</i>	Waarschijnlijkheid van het verwerven van een <i>ophaal-algoritme</i> na een fout in Toestand <i>P</i> (de parameters <i>c</i> en <i>d</i> kunnen worden samengevat als metende het <i>post-opslag algoritmisch ophalen</i>)
Heuristisch ophalen	
<i>1-r</i>	Waarschijnlijkheid van succesvol <i>heuristisch ophalen</i> van een item in de eerste testronde van Toestand <i>P</i> , gegeven dat het item in de eerste studieronde werd opgeslagen
<i>1-e</i>	Waarschijnlijkheid van succesvol <i>heuristisch ophalen</i> van een item in de eerste testronde van Toestand <i>P</i> , gegeven dat deze toestand in een studieronde later dan de eerste bereikt werd (de parameters <i>1-r</i> en <i>1-e</i> kunnen, omdat ze vroege <i>P</i> -toestandsmetingen betreffen, samengevat worden als metende het <i>vroeg-heuristisch ophalen</i>).
<i>g</i>	Waarschijnlijkheid dat, voor elke willekeurige opeenvolging van twee studierondes in Toestand <i>P</i> , onsuccesvol <i>heuristisch ophalen</i> (fout) gevolgd wordt door succesvol heuristisch ophalen (succes)
<i>h</i>	Waarschijnlijkheid dat, voor elke willekeurige opeenvolging van twee studierondes in Toestand <i>P</i> , succesvol <i>heuristisch ophalen</i> wordt gevolgd door succesvol heuristisch ophalen (de parameters <i>g</i> en <i>h</i> kunnen, omdat ze latere <i>P</i> -toestandsmetingen betreffen, samengevat worden als metende het <i>laat-heuristisch ophalen</i>)

van itempresentaties in de *eerste* testronde na opslag, waarbij de opslag respectievelijk plaatsvond in de eerste studieronde of latere studierondes.

De parameters *g* en *h* drukken de kansen uit op het succesvol kunnen ophalen in een testronde na de eerste P-toestand testronde, waarbij respectievelijk foute of goede itemreproductie plaatsvond in de onmiddellijk daaraan voorafgaande testronde.

In tabel 1 vindt men een samenvatting van de in deze paragraaf gegeven interpretaties van de modelparameters.

3. METHODE

Proefpersonen en design. Ten behoeve van het experiment werden uit willekeurig gekozen LOM- en basisscholen in Groningen (stad en provincie) en Enschede 344 kinderen gekozen, 172 voor ieder schooltype. Binnen ieder schooltype waren de kinderen in gelijke aantallen uit twee kalenderleeftijds 'ranges' afkomstig, te weten van 90-110 maanden en van 130-150 maanden. Deze leeftijdsgroepen waren gekozen om zekere ontwikkelingspsychologische vragen te beantwoorden. Dergelijke vragen spelen in het huidige artikel echter geen rol, zodat de leeftijdsgroepen hier 'gepooled' zijn. Binnen ieder schooltype werden de kinderen willekeurig, maar wel in gelijke aantallen en kalenderleeftijdverdelingen, aan één van de volgende vier verschillende experimentele condities toegewezen. Deze experimentele condities, die in de volgende paragraaf worden toegelicht, waren: een 2-categorieën lijst met standaardinstructie; een 2-categorieën lijst met hulp-instructie; een 4-categorieën lijst met standaardinstructie; een 4-categorieën lijst met hulp-instructie. Het experiment heeft dus een 2x2x2 factorieel design met als factoren Schooltypen (LOM en Basisschool), Lijstsoorten (2- en 4-categorieën lijsten) en Instructiesoorten (standaard- en hulp-instructieconditie).

Materialen en procedure. Als stimulusmateriaal werden twee soorten lijsten van elk 16 woorden gebruikt: een 2-categorieën lijst en een 4-categorieën lijst. De 2-categorieën lijst van zestien woorden bevatte acht woorden die in de categorie dieren zijn onder te brengen (hond, paard, koe, leeuw, geit, wolf, konijn, hert) en de resterende acht in de categorie dingen van het huis (raam, dak, vloer, kamer, trap, lamp, keuken, balkon). De 4-categorieën lijst van zestien woorden was in vier categorieën van elk vier woorden onderverdeeld, namelijk dieren (hond, paard, geit, wolf), lichaamsdelen (been, hoofd, knie, schouder), dingen van het huis (raam, dak, trap, lamp) en kledingstukken (schoen, jas, riem, pak). Bij ieder van de lijsten was elk woord op een afzonderlijk kartonnen kaartje gedrukt.

De procedure was als volgt. Alle kinderen kregen individueel één van beide lijsten aangeboden in een rustige schoolruimte. Bij binnenkomst werd het kind na algemene introductie verteld, dat er op ieder kaartje een woord stond, dat de kaartjes één voor één getoond zouden worden en dat het kind goed moest opletten, omdat het er om ging zoveel mogelijk van die woorden te onthouden. De eerste studieronde volgde. Ieder kaartje werd gedurende vijf seconden getoond. Hierbij sprak de proefleider (pl) het woord duidelijk uit, en moest het kind het woord nazeggen. Na de presentatie van de zestien kaartjes moest het kind gedurende dertig seconden een zogenaamde buffer-taak uitvoeren. Deze taak was bedoeld om korte-termijn geheugen effecten uit te schakelen. Ze bestond uit het wegstrepen van letters of groepjes letters op een lettervel, of uit simpele rekenopgaven, zoals terugtellen vanaf honderd, desnoods in groepjes van twee, vier, etc. Daarna volgde een testronde waarbij, afhankelijk van de instructieconditie waaraan het kind was toegewezen, de standaardtestvraag of de hulpinstructievraag werd gesteld. De standaard-testvraag luidde: 'Noem zoveel mogelijk woorden uit de lijst die je daarstraks hebt gezien'.

De hulpinstructievraag bestond uit het vragen naar lijstwoorden door de categorienamen te geven. Bijvoorbeeld, wanneer een kind in deze hulpinstructie-conditie de lijst leerde die bestond uit dieren en dingen van het huis, dan luidde de testvraag: 'Noem eens alle dieren op die je uit

de lijst kunt herinneren' en daarna: 'Noem eens alle dingen van het huis, of andersom. Nadat de pl de antwoorden van het kind had genoteerd, kreeg het kind weer de buffertaak gevolgd door een tweede testronde. Hierna volgde de tweede studieronde: het kind kreeg de zestien stimuluskaarten weer één voor één te zien (het pakje kaarten werd bij iedere studieronde eerst geschud, zodat de volgordes van de woorden steeds anders waren). De buffertaak volgde, en een nieuwe testronde. Deze procedure van studieronde, buffertaak en testronde herhaalde zich tot het kind het criterium (op twee achtereenvolgende testrondes alle zestien woorden correct reproduceren) bereikte.

4. RESULTATEN

4.1. Variantie-analyse op traditionele scores

Alvorens in te gaan op de resultaten van de toepassing van het twee-stadia model, geven we eerst de resultaten in termen van scores zoals die in de traditionele literatuur vaak onderwerp van analyse zijn (zie bijv. Bauer, 1979; Parker et al., 1975; Van den Bos en Kingma, 1985).

Het betreft hier de herinneringsprestaties van één studieronde (het aantal correct-herinnerde woorden in de eerste testronde) en een globale prestatie-index voor de gehele taak (het aantal studierondes om het criterium te bereiken). Gemiddelden en standaarddeviaties treft men aan in Tabel 2.

Een 2x2x2 variantie-analyse op de eerstgenoemde maat gaf voor zowel Schooltypen als Lijstsoorten en Instructiesoorten significante hoofdeffecten te zien, F-waarden respectievelijk: $F(1,336) = 9.80, 7.00$ en $7.51, p < .01$. Geen der interactie-effecten was significant, F-waarden variërend van 0.64 tot 2.93, $p > .05$.

Geconcludeerd kan worden tot een algemeen lager presteren van de LOM-kinderen. Echter, net als basisschoolkinderen, presteren zij beter op 4-categorieën lijsten dan op 2-categorieën lijsten, en presteren zij beter bij een hulpinstructie dan bij een standaardinstructie.

Een 2x2x2 variantie-analyse op het 'aantal studierondes tot criterium' leverde het volgende

Tabel 2. Gemiddelden en standaarddeviaties van twee traditionele scoringen van de herinneringsprestaties in het Schooltypen x Lijstsoorten x Instructiesoorten experiment

	Aantal correct herinnerde woorden in de eerste testronde (max = 16)		Aantal benodigde studierondes om het criterium (twee achtereenvolgende 100% correctie testrondes) te bereiken	
G.L.O.				
2 cat. lijst				
Stand. instr.	6.77	(1.89)	8.21	(1.76)
Hulpinstr.	7.00	(2.56)	5.89	(1.53)
4 cat. lijst				
Stand instr.	7.05	(1.97)	5.40	(1.69)
Hulpinstr.	7.35	(2.03)	3.73	(0.72)
L.O.M.				
2 cat. lijst				
Stand instr.	5.40	(1.85)	9.48	(2.88)
Hulpinstr.	6.31	(1.73)	7.37	(2.46)
4 cat. lijst				
Stand. instr.	6.19	(2.36)	9.83	(2.22)
Hulpinstr.	7.44	(2.46)	5.46	(1.59)

op: alle drie hoofdeffecten waren significant, F -waarden variërend van $F(1,336) = 40.29$ tot 109.73 , $p < .01$. Ook waren alle vier interactie-effecten significant, F -waarden variërend van $F(1,336) = 5.96$ tot 15.82 , $.001 < p < .05$.

Het meest interessant daarbij is de significante 3-weg interactie, $F(1,336) = 5.96$, $p < .05$, waarbij verschillen tussen gemiddelden verder werden getoetst met Tukey's HSD test. Hieruit kan geconcludeerd worden dat de LOM-kinderen in de standaardinstructie-conditie van de 4-categorieën lijst niet gemakkelijker tot criterium leerden vergeleken met hun prestaties in de standaardconditie van de 2-categorieën lijst. Dit was voor de GLO-kinderen duidelijk wel het geval. De hulpinstructieconditie bleek weliswaar voor zowel LOM als GLO-kinderen effectief (de 4-categorieën lijst was in deze conditie gemakkelijker te leren dan de 2-categorieën lijst), maar toch bleef het leren tot criterium ook onder de hulpconditie nog moeilijker voor LOM-kinderen vergeleken met GLO-kinderen.

Al met al zijn deze conclusies in redelijke overeenstemming met die van andere traditionele studies (Van den Bos, 1983).

Om onze hoofdvraag ten aanzien van de effecten van de condities (binnen en tussen de LOM- en GLO-groepen) op de *opslag-* en *ophaalprocessen* te kunnen beantwoorden richten we ons nu op de leerstadia-analyses. Zoals uiteengezet in paragraaf 2, bevat het leerstadiamodel 11 parameters waarvan nu eerst de voor het huidige experiment geldende numerieke waarden bepaald moeten worden. Meer hierover in de volgende paragraaf.

4.2. De schatting van de parameters en de bepaling van goodness-of-fit van het twee stadia model

Omdat de literatuur uitvoerige beschrijvingen bevat van de statistische methodologie die nodig is om de parameters van het twee-stadia model voor gegeven experimentele condities in *numerieke* waarden uit te drukken (Brainerd, 1985; Brainerd et al., 1982; Greeno, 1974; Greeno, James, DaPolito & Polson, 1978), gaan we er hier slechts globaal op in. De waarden van de parameters uit het theoretische model (Vergelijking 1) zijn niet rechtstreeks uit de data te berekenen. De volgende stappen worden nu genomen. Gestart wordt vanuit een model van observeerbare MARKOV toestanden die items in testrondes van experimenten met de inmiddels bekende structuur (S_1 , T_{1A} , T_{1B} , etc.) kunnen aannemen. Deze toestanden zijn als volgt gedefinieerd: Q = de toestand *per item* in alle testrondes na de laatste incorrecte reproductie in protocollen (per item) met een of meer incorrecte reproducties, en de toestand in alle testrondes voor protocollen (per item) met alleen correcte reproducties. R = de toestand in alle testrondes waarbij de itemreproductie incorrect is, maar op z'n minst één voorafgaande testronde een correcte itemreproductie opleverde. S = de toestand in alle testrondes waar de reproductie correct is, maar waarbij op z'n minst één incorrecte reproductie zich voordoet in een latere testronde. E = de toestand in alle testrondes met een incorrecte reproductie en waarbij alle eerdere testrondes eveneens incorrecte reproducties opleverden. Aangetoond is dat de kansen op paren observeerbare toestanden uit de startvector en de transitie matrix van dit model door elf parameters kunnen worden beschreven (Brainerd et al., 1982). Voorts is een *likelihood-functie* van het observeerbare proces beschikbaar, die de observeerbare parameters bevat met als exponenten de frequenties van de mogelijke paren van toestanden uit de startvector en de transitie matrix. Deze frequenties worden berekend uit de antwoord protocollen van de ppn in een experimentele conditie. Aangetoond is (Brainerd et al., 1982) dat de observeerbare parameters uit te drukken zijn als unieke functies van de theoretische parameters in Vergelijking (1). Derhalve is ook een likelihood-functie van het theoretische model (Vergelijking 1) op te stellen waarbij we de exponenten reeds kennen.

De volgende statistische-methodologische stap brengt de methode van bepaling van *maximum likelihood* (zie bijv. Levine & Burke, 1972; Restle & Greeno, 1970) met zich mee. Dit gebeurt door de likelihood functie te minimaliseren om de maximum likelihood schattingen van de elf parameters te verkrijgen. In de praktijk gebeurt dit met een computer-zoekmethode die gebruik maakt van een optimalisatie-algoritme. Wij gebruikten de simplex methode gecodeerd in OPTISEP.

De resultaten treft men aan in Tabel 3. Lage waarden (bijvoorbeeld waarden dichtbij .00) in Tabel 3 betekenen dat het proces, waar de betreffende parameter voor staat (zie Tabel 1), zich moeizaam voltrekt bij de proefpersonen in een bepaalde experimentele conditie. Hoge parameterwaarden (bijvoorbeeld waarden die betrekkelijk dicht bij 1.00 liggen) betekenen procesgemakkelijkheid.

Tabel 3. Parameter-waarden bepaald met de methode van maximale 'likelihood' voor het Schooltypen x Lijstsoorten x Instructiesoorten experiment

	Opslag en retentie van item- representaties			Verwerving van een ophaal algoritme				Heuristisch ophalen			
	<i>a'</i>	<i>a</i>	1- <i>f</i>	annex- opslag		post- opslag		vroeg		laat	
				<i>b'</i>	<i>b</i>	<i>c</i>	<i>d</i>	1- <i>r</i>	1- <i>e</i>	<i>g</i>	<i>h</i>

G.L.O.											
2 cat. lijst											
Stand. instr.	.48	.65	1.00	.28	.20	.15	.53	.83	.57	.42	.60
Hulpinstr.	.53	.62	1.00	.48	.44	.35	.69	.69	.70	.36	.69
4 cat. lijst											
Stand. instr.	.47	.64	1.00	.55	.28	.43	.64	.85	.73	.46	.69
Hulpinstr.	.59	1.00	1.00	.60	.59	.89	.87	.44	.51	.32	.69
L.O.M.											
2 cat. lijst											
Stand. instr.	.39	.50	1.00	.04	.00	.25	.27	.89	.81	.51	.68
Hulpinstr.	.47	.54	1.00	.28	.37	.30	.46	.78	.72	.49	.69
4 cat. lijst											
Stand. instr.	.44	.42	1.00	.11	.15	.22	.29	.87	1.00	.52	.73
Hulpinstr.	.73	.58	.98	.33	.00	.55	.73	.46	.93	.21	.83

Alvorens de parameterwaarden in Tabel 3 te vergelijken, diende per cel de goodness-of-fit van het toegepaste model te worden nagegaan. Conform de procedures beschreven in Brainerd et al. (1982) en Kingma & Reuvekamp (in press; a,b) werden per cel twee toetsen uitgevoerd: een noodzakelijkheids ('necessity') toets en een voldoende ('sufficiency') toets. De eerste toets onderzoekt de nulhypothese dat een één-stadium model een even goede benadering van de data geeft als het twee-stadia model. Hierbij wordt een likelihood-ratio statistiek gebruikt die asymptotisch is gedistribueerd als χ^2 (5). (Zie Brainerd et al., 1982; Equation 38). Voor ieder van de acht experimentele cellen waren de ratio's aanzienlijk groter dan de .05 kritische waarde behorend bij χ^2 (5). De nulhypothese verwerpend kan geconcludeerd worden, dat het leren tot criterium van de kinderen in het huidige experiment op z'n minst een twee-stadia proces is. Vervolgens werden de voldoende-toetsen uitgevoerd, waarin de verkregen en voorspelde distributies van verscheidene bekende leerstatistieken (aantal fouten voor het eerste succes, aantal fouten na het eerste succes, lengte van successen-ketens voor de laatste fout, laatste testronde waarin een fout voorkwam, totaal aantal fouten, leercurve) met elkaar vergeleken werden (zie Equations 47-52 in Brainerd et al., 1982). Deze toetsen, voor ieder van de acht cellen van het experiment, wezen uit dat de verkregen en voorspelde distributies goed correspondeerden. De resultaten van de noodzakelijkheids en voldoende-toetsen te zamen laten er weinig twijfel over bestaan, dat voor zowel de GLO-, als de LOM-kinderen, het 'free-recall'-leren tot criterium door een twee-stadia model in statistische zin goed beschreven wordt.

4.3 Vergelijkingen tussen de parameterwaarden

In Brainerd et al. (1982) is een serie statistische toetsen beschreven waarmee achtereenvolgens wordt uitgemaakt of (1) de verkregen parameterwaarden voor een *experiment als geheel* significant van elkaar verschillen, (2) of de parameterwaarden voor een willekeurig *conditiepaar* (uit dat experiment) *als geheel* significant van elkaar verschillen, (3a) of de waarde van een *afzonderlijke parameter* (bijvoorbeeld *a'*) in een bepaalde conditie (bijvoorbeeld GLO, 2-categorieën lijst, standaardinstructie) significant verschilt van de waarde van een andere parameter (bijvoorbeeld *a*) in dezelfde conditie, of (3b) dat de waarde van een parameter (bijvoorbeeld *a'*) in een bepaalde conditie (bijvoorbeeld GLO, 2-categorieën lijst, standaardinstructie) significant verschilt van de waarde van dezelfde parameter in een andere conditie (bijvoorbeeld LOM, 2-categorieën lijst, standaardinstructie). De basis voor zowel de experimentsgewijze toets (1), de conditiegewijze toetsen (2) en de parametergewijze toetsen (3a: 'within'; 3b: 'between') is het bepalen van 'likelihood' ratio's (zie Brainerd et al., 1982). De uitslagen zijn Chi-kwadraat (χ^2) verdeeld, en gemakkelijk kan worden nagegaan – met in achtneming van de bij een toets horende vrijheidsgraden; zie weer Brainerd et al. (1982) – of de verkregen waarde de bij $p = .05$ horende kritische waarde in de χ^2 tabel overschrijdt.

We rapporteren nu de toetsingsuitkomsten van de experimentsgewijze toets, de conditiegewijze toetsen en de parametergewijze toetsen 'between'. Wat betreft de twee laatstgenoemde toetssoorten doen we dit onder de subparagrafen: effecten van aantallen categorieën in een lijst, effecten van instructiesoorten en effecten van schooltypen.

Het toetsingsresultaat van de experimentsgewijze toets was $\chi^2(77) = 1443.71$, $p < .01$. Dit betekent dat de nulhypothese (geen verschil tussen parameterwaarden van de condities van het experiment als geheel) verworpen wordt en dat de meer specifieke stap, namelijk die van het conditiegewijze toetsen, gerechtvaardigd is.

4.3.1 Effecten van aantallen categorieën in een lijst

De resultaten van de vier *conditiegewijze* toetsen om de verschillen tussen 2- en 4-categorieën lijsten te onderzoeken waren:

$\chi^2(11) = 148.95$, $p < .01$ (GLO; standaardinstructie),

$\chi^2(11) = 204.86$, $p < .01$ (GLO; hulpinstructie),

$\chi^2(11) = 17.58$, $p > .05$ (LOM; standaardinstructie),

$\chi^2(11) = 147.16$, $p < .01$ (LOM; hulpinstructie).

Bij de significant verschillende conditieparen werden vervolgens parameter-gewijze 'between'-toetsen uitgevoerd.

GLO-kinderen. – Wat betreft de *opslag*-parameters (*a'* en *a*) werden bij de standaardinstructie geen significante verschillen tussen 2- en 4-categorieën lijsten aangetroffen. Bij de hulpinstructie waren de verschillen wel significant hetgeen betekent, dat hier (en vooral bij parameter *a*) het opslaan bij de 4-categorieën lijst gemakkelijker was dan bij de 2-categorieën lijst. De *retentie*-parameter (*l-f*) gaf voor 2- en 4-categorieën lijsten geen significante verschillen te zien. Wat betreft het *algoritmisch ophalen* (*b'*, *b*, *c* en *d*) werden bij de 4-categorieën lijst in zowel de standaard- als de hulpinstructieconditie significant hogere parameterwaarden gevonden dan bij de 2-categorieën lijst. Ten aanzien van het *heuristisch ophalen* (parameters *l-r*, *l-e*, *g* en *h*) werden in de standaardinstructieconditie bij de 4-categorieën lijst alleen voor de twee parameters *l-e* en *h* significant hogere waarden dan bij de 2-categorieën lijst verkregen, terwijl dit in de hulpinstructieconditie voor twee parameters (*l-r* en *l-e*) net andersom lag: in de 4-categorieën lijst waren de waarden van deze parameters significant lager dan in de 2-categorieën lijst.

LOM-kinderen. – Gezien het niet significante resultaat van de conditiegewijze toets bij LOM-kinderen in de standaardinstructieconditie, werden in deze conditie geen parametergewijze toetsen voor verschillen tussen 2- en 4-categorieën lijsten verricht.

Parameter-gewijze toetsen in de hulpinstructieconditie gaven voor de volgende parameters significant hogere waarden in de 4-categorieën lijst (vergeleken met de 2-categorieën lijst) te

zien: de *opslag*-parameter a' en de *algoritmisches ophalen* parameters c en d . Een 'omgekeerd' beeld vertoonde de b -parameter waarvan de waarde zich in de 4-categorieën lijst (vergeleken met de 2-categorieën lijst) verlaagde.

Geen significant verschil werd gevonden voor de *retentie*-parameterwaarden ($1-f$). Een gemengd beeld gaf het *heuristisch ophalen* te zien: de 4-categorieën lijst gaf lagere waarden dan de 2-categorieën lijst op de parameters $1-r$ en g , en hogere waarden op de parameters $1-e$ en h .

4.3.2 Effecten van instructiesoorten

De resultaten van de vier *conditie*gewijze toetsen om de verschillen tussen de standaard- en hulpinstructieconditie te onderzoeken waren:

$\chi^2(11) = 156.60, p < .01$ (GLO; 2-categorieën lijst),

$\chi^2(11) = 209.06, p < .01$ (GLO; 4-categorieën lijst)

$\chi^2(11) = 116.71, p < .01$ (LOM; 2-categorieën lijst),

$\chi^2(11) = 360.03, p < .01$ (LOM; 4-categorieën lijst).

Alle conditieparen verschilden significant en vervolgens werden parametergewijze 'between'-toetsen uitgevoerd, die de volgende resultaten gaven.

GLO-kinderen. – Wat betreft de waarden van de *opslag*-parameters (a' en a) werden in de 2-categorieën lijst geen significante verschillen als functie van instructiesoorten aangetroffen. In de 4-categorieën lijst daarentegen, waren de waarden van beide opslag-parameters (a' , en a) in de hulpinstructieconditie significant hoger dan in de standaardconditie.

Noch voor de 2-categorieën lijst, noch voor de 4-categorieën lijst werden (als voorheen) significante verschillen als functie van instructiesoorten op de *retentie*-parameter ($1-f$) aangetroffen.

Op alle vier *algoritmisches ophalen* parameters (b' , b , c en d) werden significant hogere waarden in de hulpinstructieconditie vergeleken met de standaardinstructieconditie gevonden, zowel in de 2-categorieën lijst als in de 4-categorieën lijst. De *heuristisch ophalen* parameters ten slotte, vertoonden in de hulpinstructieconditie van de 2-categorieën lijst een significant lagere waarde op $1-r$, vergeleken met de waarde van $1-r$ in de standaardinstructieconditie, en een significant hogere waarde op $1-e$.

In de 4-categorieën lijst waren de waarden op drie van de vier parameters ($1-r$, $1-e$ en g) significant lager in de hulpinstructieconditie vergeleken met de standaardinstructieconditie, terwijl er voor de h -parameter geen verschil was.

LOM-kinderen. – In de 2-categorieën lijst was wat betreft de *opslag*-parameters de waarde op één parameter (a') significant hoger onder de hulpinstructie dan onder de standaardconditie. In de 4-categorieën lijst waren de waarden van beide opslag-parameters (a' en a) in de hulpinstructieconditie hoger dan in de standaardconditie. Noch voor de 2-categorieën lijst, noch voor de 4-categorieën lijst werden (als voorheen) significante verschillen, als functie voor instructiesoorten, op de *retentie*-parameter ($1-f$) aangetroffen.

Wat betreft het *algoritmisches ophalen* (parameters b' , b , c en d) was in de 2-categorieën lijst voor de hulpinstructie op drie van de vier parameters (b' , b en d) een significante verbetering te bespeuren. Ook in de 4-categorieën lijst verbeterden als gevolg van de hulpinstructie drie van de vier algoritmisches-ophalen parameters (b' , c en d). Wat betreft het *heuristisch ophalen* zagen we in de hulpinstructieconditie significante 'verslechtingen' (vergeleken met de standaardinstructie) optreden bij de $1-r$ en $1-e$ parameters in de 2-categorieën lijst, en in de 4-categorieën lijst bij de parameters $1-r$, $1-e$ en g . In de 4-categorieën lijst werd alleen de waarde op de h -parameter significant hoger bevonden.

4.3.3 Effecten van schooltypen

De resultaten van de vier *conditiegewijze* toetsen om de verschillen tussen GLO-kinderen en LOM-kinderen te onderzoeken waren:

$\chi^2(11) = 66.10, p < .01$ (2-categorieën lijst; standaardinstructie),

$\chi^2(11) = 57.78, p < .01$ (2-categorieën lijst; hulpinstructie),

$\chi^2(11) = 233.86, p < .01$ (4-categorieën lijst; standaardinstructie),

$\chi^2(11) = 101.41, p < .01$ (4-categorieën lijst; hulpinstructie).

Voor alle conditieparen werden significante verschillen gevonden en vervolgens werden parametergewijze 'between'-toetsen uitgevoerd.

Twee-categorieën lijst. – In de standaardinstructieconditie werden wat betreft de *opslag*-parameters voor zowel a' als a hogere waarden voor GLO-kinderen aangetroffen. In de hulpinstructieconditie echter, waren wat betreft het opslaan de verschillen tussen de schooltypen niet significant. Zowel in de standaard- als in de hulpinstructieconditie was de waarde van de *retentie*-parameter ($1-f$) bij LOM- en GLO-kinderen identiek.

Wat betreft het *algoritmisch ophalen* bleken LOM-kinderen in de standaardinstructie in één geval (parameter c) een hogere waarde te behalen dan de GLO-kinderen. De waarden van de overige parameters (b' , b en d) waren significant hoger bij de GLO-kinderen. In de hulpinstructieconditie waren twee algoritmisch ophalen parameterwaarden (b' en d) significant hoger bij GLO-kinderen, en op de andere parameters (b en c) werden geen significante schooltypen-verschillen gevonden.

De waarden van de *heuristisch ophalen* parameters tenslotte, waren in de standaardinstructieconditie stevast hoger voor LOM-kinderen. In de hulpinstructieconditie gold dit voor twee van de vier parameters ($1-r$ en g), terwijl op $1-e$ en h geen significante schooltypen-verschillen gevonden werden.

Vier-categorieën lijst. – In de standaardinstructieconditie hadden de GLO-kinderen significant hogere waarden dan de LOM-kinderen op één van de *opslag*- (a) en op alle vier *algoritmisch ophalen*-parameters (b' , b , c en d). Onder de hulpinstructieconditie had de GLO-groep een significant hogere parameterwaarde op a , maar de LOM-groep op a' .

Gemiddeld gezien (over a' en a te zamen) was de opslagsuperioriteit van de GLO-groep echter niet sterker dan in de standaardconditie. De *retentie*-parameter ($1-f$) differentieerde onder geen van de instructie-condities de beide schooltypen. Evenals in de standaardinstructieconditie hadden de GLO-kinderen in de hulpinstructieconditie ten aanzien van alle vier *algoritmisch ophalen* parameters significant hogere parameter-waarden dan LOM-kinderen. Tot slot het *heuristisch ophalen*. In de standaardinstructieconditie hadden de LOM-kinderen op twee parameters ($1-e$ en h) significant hogere waarden dan de GLO-kinderen.

In de hulpinstructieconditie herhaalde deze trend zich: de waarden van de parameters $1-e$ en h waren voor LOM-kinderen significant hoger dan voor GLO-kinderen, maar nu hadden de GLO-kinderen een significant hogere waarde op parameter g .

5. DISCUSSIE

Gebruik makend van het leerstadia-model, was het doel van dit experiment nadere specificaties te leveren over aspecten van het opslaan en ophalen tijdens het memoriseren van GLO- en LOM-kinderen. Drie vragen werden gesteld: (1) Verbeteren bij deze groepen kinderen dezelfde memoriseerprocessen wanneer zij een 4-categorieën lijst versus een 2-categorieën lijst leren? (2) Welke twee-stadia model processen verbeteren bij de hulpinstructie (ten opzichte van de standaardinstructie) en verbeteren bij GLO- en LOM-kinderen dezelfde processen? (3) Welke twee-stadia model processen liggen in de vier experimentele condities ten grondslag aan GLO- en LOM-verschillen? Samengevat leverden de gerapporteerde resultaten ons de volgende antwoorden op deze vragen. Ten aanzien van (1) bleken bij de GLO-kinderen, vergeleken met

hun verbeteringen op het gebied van het opslaan, de retentie en het heuristisch ophalen, met name de processen van het *algoritmisches ophalen* te verbeteren. Bij LOM-kinderen daarentegen, speelden de verbeteringen in de verwerving van ophaal-algoritmen zich op beperkter schaal af. Zo gold bijvoorbeeld in de hulpinstructieconditie (van de 4- versus de 2-categorieën lijst) de verbetering alleen het post-opslag algoritmisch ophalen (parameters c en d).

Ten aanzien van vraag (2) gaven de resultaten voor de GLO-kinderen aan dat met name de processen van het *algoritmisches ophalen* verbeterden. Weliswaar verbeterden ook de *opslag*-processen onder de hulpinstructie, maar dit alleen in de 4-categorieën lijst.

In grote lijnen werden deze trends ook bij de LOM-kinderen gevonden. Trend-afwijkingen waren de iets sterkere opslagverbetering (gemiddeld over de parameters a' en a) als gevolg van de hulpinstructie bij LOM-kinderen in de 2-categorieën lijst, en de gemiddeld veel minder sterke verbetering (vergeleken met die van GLO-kinderen) als gevolg van de hulpinstructie op het *annex-opslag* algoritmisch ophalen (gemiddeld over de parameters b' en b) in de 4-categorieën lijst.

Dit neemt echter niet weg dat, wanneer naar de gemiddelde waarde van de vier parameters van het algoritmisch ophalen gekeken wordt, de *betrekkelijke verbetering als gevolg van de hulpinstructie* bij LOM-kinderen (gemiddelde stijging .21 en .21 in respectievelijk de 2- en 4-categorieën lijst) niet veel minder is dan die bij GLO-kinderen onder de hulpinstructie (gemiddelde stijging .20 en .26 in respectievelijk de 2- en 4-categorieën lijst). Ter contrast met de vorige analyse: de *betrekkelijke verbetering in het algoritmisch ophalen als gevolg van grotere categorisering* was bij GLO-kinderen gemiddeld .19 en .25 in respectievelijk de standaard- en hulpinstructieconditie, terwijl voor de LOM-groep in deze condities slechts gemiddelde verbeteringen van .05 en .05 gevonden werden.

Hieruit kan dus geconcludeerd worden, dat op zich een hogere graad van categorisering LOM-kinderen nauwelijks helpt in de duidelijk zwakke component van hun memoriseerproces (te weten het algoritmisch ophalen – zie de bespreking van onderzoeksvraag 3), maar dat de *hulpinstructie* waarin expliciet categorielabels aangereikt worden, veel *effectiever* is, ook al betekent dit nog niet dat LOM-kinderen ook in absolute zin het niveau van algoritmisch ophalen van GLO-kinderen bereiken (zie de nu volgende bespreking van onderzoeksvraag 3).

Tot slot onze derde onderzoeksvraag (3). Wanneer de parameterwaarden (Tabel 3) van het *opslaan* over de vier experimentele condities gemiddeld worden voor GLO- en LOM-kinderen (8 cellen per groep), levert dit de respectievelijke waarden .62 en .51 op. Doen we hetzelfde voor de *algoritmisches ophalen* parameters (16 cellen per groep), dan worden respectievelijk de waarden .50 en .27 verkregen. Hieruit kan geconcludeerd worden dat in de huidige taak het schooltypenverschil (c.q. GLO-superioriteit) met name verklaard wordt door het grotere gemak waarmee GLO-kinderen ophaal-algoritmen verwerven. Het gemiddelde schooltypenverschil bij algoritmisch ophalen (= .23) is immers ruwweg twee maal zo groot als het gemiddelde schooltypenverschil (= .11) in opslagmoeilijkheid. Tegelijkertijd kan uit genoemde waarden geconcludeerd worden, dat zowel GLO- als LOM-kinderen gemakkelijker opslaan dan (algoritmisches) ophalen in een 'free-recall' memoriseertaak, en dat het moeilijkheidsgraad-verschil tussen opslaan en algoritmisch ophalen bij LOM-kinderen groter is dan bij GLO-kinderen.

Terzijde merken we op, dat ten aanzien van de schooltypen-verschillen bij het opslaan en algoritmisch ophalen sprake is van een parallel met gegevens in het Amerikaanse onderzoek van Howe, Brainerd & Kingma (1985). Hierin moesten jongere en oudere leergestoorte en normaal lerende kinderen o.a. lijsten ongerelateerde woorden memoriseren. Wel is het zo, dat in het onderzoek van Howe et al. (1985) de 'leergestoord-normaal' verschillen in het algoritmisch ophalen met name pas in de oudere leeftijdsgroepen zichtbaar werden. Dit is echter verklaarbaar door voor de (moeilijker) taak van Howe et al. bij jongere leeftijdsgroepen bodem-effecten aan te nemen ten aanzien van het algoritmisch ophalen. In onze (gemakkelijker) taken waren – zo nemen we aan – deze bodem-effecten voor althans de jongere GLO-kinderen al afwezig, hetgeen verklaart dat de data van de samengetrokken GLO-leeftijdsgroepen superieur zijn aan die van de samengetrokken LOM-leeftijdsgroepen.

In het kader van de bespreking van schooltypen-verschillen op de verschillende procesgebieden verdient het volgende eveneens aandacht. Het betreft de trend, dat in alle vier experimentele condities LOM-kinderen gemiddeld hogere waarden hadden voor het heuristisch ophalen dan GLO-kinderen, en dat hier met name het *vroeg*-heuristisch ophalen superieur was bij LOM-kinderen. Exact dezelfde trend werd recentelijk gerapporteerd in de studie van Van den Bos & Kingma (1987), waarin het presteren van GLO- en LOM-kinderen op *woordpaar*-leertaken werd vergeleken. Het commentaar dat in genoemde studie wordt gegeven lijkt ook hier van toepassing. Het is waarschijnlijk dat het 'verschil tussen de verschillen' (de LOM-kinderen waren vooral superieur in het heuristisch ophalen van alle items die de *P*-toestand aannemen – vroeg heuristisch ophalen – en minder superieur in het heuristisch ophalen van de moeilijker items – laat heuristisch ophalen –) te maken heeft met een ander verschil, nl. dat LOM- en GLO-kinderen verschilden in de mate waarin items überhaupt de *P*-toestand aannamen. Om dit duidelijker te maken dient teruggekomen te worden op het *annex-opslag algoritmisch ophalen*.

De parameters van dit proces (b' en b) meten het verwerven van een ophaalalgoritme, waarbij de *P*-toestand wordt overgeslagen. We nemen aan dat de gemakkelijkere items hier vooral voor in aanmerking komen. We hebben gezien, dat de parameterwaarden op b' en b bij GLO-kinderen stevast hoger waren dan bij LOM-kinderen. Dit betekent ook dat GLO-kinderen minder items naar de *P*-toestand 'stuurden' en dat de items die wel de *P*-toestand aannamen de relatief moeilijk op te halen items waren. Vergeleken met GLO-kinderen, zouden LOM-kinderen daarentegen veel meer items (gemakkelijke en moeilijke) naar de *P*-toestand 'sturen'. Dit betrekkelijk groter aantal gemakkelijke items verhoogt waarschijnlijk de kans op reproductie-accuratesse in de vroege *P*-toestand, maar – nogmaals – het gaat gepaard met relatief lage waarden voor het annex-opslag algoritmisch ophaalproces. Als deze redenering correct is, zou voor onze data de volgende trend moeten gelden: waar de LOM-kinderen het meest superieur zijn aan GLO-kinderen in vroeg-heuristisch ophalen, zijn ze tevens het meest 'inferieur' aan GLO-kinderen in het annex-opslag algoritmisch ophalen. Dit blijkt voor onze data (Tabel 3) precies het geval. Het grootste verschil tussen de groepen kinderen in de gemiddelde waarden van de parameters van het vroeg-heuristisch ophalen tekent zich af bij de 4-categorieën lijst en tevens wordt bij die conditie het grootste schooltypen-verschil in annex-opslag algoritmisch ophalen gevonden.

Tot slot gaan we nog in op een belangrijk voor het onderwijs relevant aandachtspunt, dat geïmpliceerd wordt door de twee-stadia model data van het huidige experiment en van vergelijkbare studies (Howe et al., 1985a; Kingma & Van den Bos, 1987; Van den Bos & Kingma, 1987).

Het betreft de observatie dat in al deze studies de d -parameter scherp differentieert tussen normaal-lerenden enerzijds en LOM- en MLK-kinderen anderzijds. Dit betekent dat het verwerven van ophaal-algoritmes voor items, nadat deze in de voorgaande testronde niet werden gereproduceerd, veel gemakkelijker is voor de normaal-lerende kinderen. Waarschijnlijk hangt dit samen met een betere '*zelf-test*' techniek van deze kinderen. We veronderstellen dat hierbij ieder item tijdens het verloop van het experiment in de studierondes onderzocht wordt op de vraag: 'Had ik dit item tijdens de laatste testronde goed of niet?' Wanneer de proefpersoon meent dat het item goed was gereproduceerd, heeft hij/zij daarvoor minder studie-aandacht nodig dan voor een niet-gereproduceerd item. Naast een goede begeleiding van het opslagproces, en van het creëren en gebruiken van ophaal-'cues', lijkt dus expliciete instructie in dit soort zelf-feedback technieken bij leergestoorde kinderen in memoriseersituaties een belangrijk pedagogisch aandachtspunt te vormen.

LITERATUUR

- Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89-195). New York: Academic Press.
- Bauer, R.H. (1979). Memory, acquisition, and category clustering in learning-disabled children. *Journal of Experimental Child Psychology*, 27, 365-383.
- Bos, K.P. van den (1983). *Verbale en ruimtelijke informatie-verwerkingsaspecten bij leergestoorde en niet-leergestoorde kinderen*. Academisch proefschrift. Groningen: Rijksuniversiteit.
- Bos, K.P. van den, & Kingma, J. (1985). Het aanvankelijk leren en het leren tot criterium bij jongere en oudere M.L.K., L.O.M.- en G.L.O.-schoolkinderen in een memoriseertaak. *Tijdschrift voor Orthopedagogiek*, 24, 239-253.
- Bos, K.P. van den, & Kingma, J. (1987). A new look at learning-disabled and normal-achieving children's word-pair memorization ability. *European Journal of Psychology of Education*, 1, (4), 45-59.
- Bos, K.P. van den, Kingma, J., & Bakker, A.C. (1986). De invloed van leeftijd bij het memoriseren van normaal-lerende en leergestoorde kinderen. *Tijdschrift voor Orthopedagogiek*, 25, 172-181.
- Brainerd, C.J. (1985). Model-based approaches to storage and retrieval development. In C.J. Brainerd & M. Pressly (Eds.), *Basic processes in memory development: Progress in cognitive development research* (pp. 143-207). New York: Springer Verlag.
- Brainerd, C.J., Howe, M.L., & Kingma, J. (1982). An identifiable model of two-stage learning. *Journal of Mathematical Psychology*, 26, 263-293.
- Brainerd, C.J., Howe, M.L., Kingma, J., & Brainerd, S.H. (1984). On the measurement of storage and retrieval contributions to memory development. *Journal of Experimental Child Psychology*, 37, 478-499.
- Ceci, C.J. (1983). An investigation of the semantic processing characteristics of normal and language/learning-disabled children (L/LDS). *Developmental Psychology*, 19, 427-439.
- Dallago, M.L.P., & Moely, B.E. (1980). Free recall in boys of normal and poor reading levels as a function of task manipulations. *Journal of Experimental Child Psychology*, 30, 62-78.
- Greeno, J.G. (1974). Representation of learning as discrete transition in a finite state space. In D.H. Krantz, R.C. Atkinson, R.D. Luce & P. Suppes (Eds.), *Learning, memory, and thinking*. Contemporary developments in mathematical psychology, Volume 1 (pp. 1-43). San Francisco: Freeman.
- Greeno, J.G., James, C.T., DaPolito, F.J., & Polson, P.G. (1978). *Associative learning, a cognitive analysis*. Englewood Cliffs, N.J.: Prentice Hall.
- Half, H.M. (1977). The role of opportunities for recall in learning to retrieve. *American Journal of Psychology*, 90, 383-406.
- Howe, M.L., Brainerd, C.J., & Kingma, J. (1985). Storage-retrieval processes of normal and learning-disabled children: A stages-of-learning analysis of picture-word effects. *Child Development*, 56, 1120-1133.
- Kingma, J., & Bos, K.P. van den (1987). Opslag- en ophaalprocessen bij het leren van plaatjes- en woordenlijsten door twee leeftijdsgroepen basisschoolkinderen en leerlingen van scholen voor M.L.K. *Nederlands Tijdschrift voor de Psychologie*, 42, 125-136.
- Kingma, J., & Reuvekamp, J. (1987a). Markov count: A program for computing the learning statistics of two-stage Markov learning experiments. *Educational and Psychological Measurement*, in press.
- Kingma, J., & Reuvekamp, J. (1987a). Markov model: A package for computing the parameters of one-stage and two-stage learning models and hypothesis testing. *Educational and Psychological Measurement*, in press.
- Levine, G., & Burke, C.J. (1972). *Mathematical model techniques for learning theories*. New York: Academic Press.
- Parker, T.B., Freston, C.W., & Drew, C.J. (1975). Comparison of verbal performance of normal and learning disabled children as a function of input organization. *Journal of Learning Disabilities*, 8, 386-393.
- Pressly, M., & Levin, J.R. (1977). Developmental differences in subjects' associative learning strategies and performance: Assessing a hypothesis. *Journal of Experimental Child Psychology*, 24, 431-439.
- Restle, F., & Greeno, J.G. (1970). *Introduction to mathematical psychology*. Reading, Mass.: Addison Wesley.
- Rohwer, W.D., & Dempster, F.N. (1977). Memory development and educational processes. In R.V. Kail, Jr. and J.W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 407-435). N.J.: Lawrence Erlbaum Associates.

- Ross, A.O. (1976). *Psychological aspects of learning disabilities and reading disorders*. New York: McGraw-Hill.
- Spring, C., & Capps, C. (1974). Encoding speed, rehearsal, and probed recall of dyslexic boys. *Journal of Educational Psychology*, 66, 780-786.
- Torgesen, J.K. (1977). Memorization processes in reading-disabled children. *Journal of Educational Psychology*, 79, 571-578.
- Torgesen, J.K. (1978-1979). Performance of reading disabled children on serial memory tasks: a selective review of recent research. *Reading Research Quarterly*, 14, 57-87.
- Torgesen, J.K. (1982). The study of short-term memory in learning disabled children: Goals, methods and conclusions. *Advances in Learning and Behavioral Disabilities*, 1, 117-149.
- Torgesen, J., & Kail, R.V. (1980). Memory processes in exceptional children. In B.K. Keogh (Ed.), *Advances in special education*, Volume 1 (pp. 55-99). Greenwich, Conn.: JAI Press.
- Worden, P.E. (1983). Memory strategy instruction with the learning disabled. In M. Pressley & J.R. Levin (Eds.), *Cognitive strategy research*. Psychological foundations (pp. 129-153). New York: Springer Verlag.

Manuscript ontvangen 23-12-1986

Definitieve versie ontvangen 15-6-1987

On the Robustness of the 'Minimum-Chi-Square' Method for the Rasch Model

Dato N.M. de Gruijter

Educational Research Center, Leiden University

ABSTRACT

The 'minimum-chi-square' method for estimating item parameters in the Rasch model was proposed by Fischer in 1974. The method is extremely fast and accurate. It seems especially attractive for item banking purposes because it easily deals with incomplete data. An important disadvantage is, however, the fact that it is not robust under violations of the Rasch model. This is demonstrated in the present study.

INTRODUCTION

For the Rasch model a number of estimation procedures has been proposed apart from the well-known maximum likelihood procedures (De Leeuw & Verhelst, 1986), unconditional (UML), conditional (CML) and marginal maximum likelihood (MML). One of the alternative procedures for the estimation of item parameters is the 'minimum-chi-square' method, MINCHI for short, proposed by Fischer (1974). The method is based on counts n_{ij} , where n_{ij} is the number of times item i has been answered correctly and item j has been answered incorrectly. In this respect the method resembles the nonparametric procedure for ranking items on basis of dominance proposed by Cliff (1977, p. 380).

MINCHI is fast. Further, it appeared accurate in simulations, in terms of recovery of the Rasch parameters used to generate data in simulations (Fischer, 1974; Zwinderman & Van den Wollenberg, 1987). This result is somewhat surprising. Firstly, the method does not use all information contained in the data, an aspect which plays a central role in the present study. Secondly, the method is not an exact minimum-chi-square method. The fact that the counts n_{ij} are interdependent, is neglected. More recently Van der Linden and Eggen (1986) proposed to eliminate the dependencies by using only information from a subset of independent item pairs.

The method seems especially attractive in the context of item banking when a large number of item parameters is estimated on basis of incomplete data, where examinees respond to subsets of the item pool. The incomplete data result when a number of different test forms is given to a number of different groups of examinees. A particular form of incomplete data arises in individualized testing, where each examinee responds to a different set of items. The likelihood procedures always need information on test scores – total score being the sufficient statistic for the person parameter in the Rasch model – and the item selections on which the total scores are based. In contrast, item parameter estimation based on counts n_{ij} remains straightforward in the case of an incomplete design.

Van der Linden and Eggen (1986) motivated their proposal with the advantage of the estimation based on counts in connection with item banking. There is another advantage of this kind of estimation in connection with item banking over MML, which appears to be attractive otherwise (Thissen, 1982). MML can fail to produce adequate item parameter estimates with an expanding item pool or when the item selection rule changes in the course of time (see the appendix). On the other hand, MML might be a solution to the problem of estimating item parameters when the presentation of items is based on examinees' responses to items presented

previously. In that case other methods produce biased estimates (De Gruijter, 1980). In this contribution, however, it will be assumed throughout that item presentation is not contingent on responses.

In the present contribution a start is made with the study of the robustness of MINCHI when the Rasch model is not the true model. Fischer rejected a comparison of Rasch estimation methods in this situation, arguing that the question which method gives the best results, cannot be answered (Fischer, 1974, p. 271). This principle must be abandoned, however, in view of the fact that many researchers use the Rasch model as a convenient model, even when deviations from the model are apparent or suspected. A certain degree of misfit appears harmless in some applications.

It should be clear that when different Rasch methods are to be compared under deviations from the model, a criterion must be developed by which the performance of these methods can be compared. For the purpose of this study such a criterion is suggested. MINCHI and UML are compared with the help of this criterion in a simulation. For reasons of simplicity a complete design is used in the study.

MINCHI

The probability of a correct response on item i , given ability θ , can be written in the Rasch model as

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

where b_i is the difficulty parameter of item i . An alternative formulation of the model, used in the derivation of MINCHI, is

$$P_i(\xi) = \xi \epsilon_i / (1 + \xi \epsilon_i), \quad (2)$$

with

$$\begin{aligned} \xi &= \exp(\theta) \\ \epsilon_i &= \exp(-b_i). \end{aligned}$$

Under the assumption of local stochastic independence the probability of a correct response on item i , given only one correct response on items i and j , equals

$$P_i(\xi) Q_j(\xi) / [P_i(\xi) Q_j(\xi) + P_j(\xi) Q_i(\xi)] = \epsilon_i / (\epsilon_i + \epsilon_j), \quad (3)$$

where $Q_i(\xi) = 1 - P_i(\xi)$. This implies that the item parameters can be estimated independently of the person parameters. This is a central property of the Rasch model, and reversely, the only model which is continuous in terms of its parameters and for which this independence holds, is the Rasch model. This property has been derived by Fischer (1974) with an emphasis on sufficiency of the total score, while Roskam (1983) emphasized the resulting property of consistent stochastic orderings of items and persons. The quotient ϵ_i / ϵ_j can be estimated by n_{ij} / n_{ji} . In order to fix the latent scale a restriction like $\epsilon_i \epsilon_j = 1$, and $\prod_i \epsilon_i = 1$ in the general case, has to be added. The result lends itself to several estimation procedures, which can easily be applied with missing data by design. Let us look at the difference $n_{ij} \epsilon_i^{-1} - n_{ji} \epsilon_j^{-1}$. Fischer demonstrated for a fixed sum $n_{ij} + n_{ji}$ that the difference has an expectation zero and a variance $(n_{ij} + n_{ji}) \epsilon_i^{-1} \epsilon_j^{-1}$ (Fischer, 1974, p. 269). This result is used in MINCHI. In this estimation method the weighted sum of squared differences

$$F = \mathbf{e}' \mathbf{V}^{-1} \mathbf{e} = \sum_{i < j} (n_{ij} \epsilon_i^{-1} - n_{ji} \epsilon_j^{-1})^2 / [\epsilon_i^{-1} \epsilon_j^{-1} (n_{ij} + n_{ji})] \quad (4)$$

is minimized w.r.t. the parameters. The method is an approximation in its use of a diagonal variance-covariance matrix V . Differentiating and setting the results equal to zero, gives a system of equations

$$\epsilon_i^2 = \sum_j x_{ij} \epsilon_j / (\sum_j x_{ji} \epsilon_j^{-1}), \quad (i = 1, \dots, k) \quad (5)$$

where k is the total number of items and

$$x_{ij} = n_{ij}^2 / (n_{ij} + n_{ji}).$$

This system of equations can be solved very fast iteratively.

THE COMPARISON OF RASCH SOLUTIONS UNDER MODEL DEVIATIONS

When the Rasch model does hold, different Rasch solutions can be compared with respect to their estimation accuracy. A possible criterion is the mean squared error, the average of the squared deviations between the parameters and their estimates.

When the Rasch model does not hold, this criterion is not available. This does not mean that there is no criterion at all. Parameter estimates can be used for deriving predictions, and the accuracy of different predictions can be compared. One possibility is to verify how adequate a certain aspect of the data can be described on basis of the estimates which have been obtained with these data.

It is clear that the Rasch model has difficulties with substantial variation of discrimination of items: the Rasch model has no discrimination parameter. On the other hand one expects that the p -values, the item proportions correct, can be recovered adequately, when the estimation of all parameters is based on the same data as these p -values (one cannot expect to recover all p -values adequately for groups of examinees with higher or lower ability levels when the Rasch model is violated). The criterion for the comparison of different methods could be a weighted or unweighted sum of squared differences between p -values and estimates, derived on basis of the item parameter estimates.

The use of estimated p -values presents a complication in that these proportions do not only depend on the item parameter estimates, but also on the estimated distribution of person parameters or on the distribution of estimated person parameters, which can be obtained by ML. We will consider only the case of an estimated distribution of person parameters: maximum likelihood of person parameters is not possible for zero and perfect total scores, and this can prevent adequate recovery of p -values.

When the ability distribution is inadequately estimated, estimated p -values will be in error. One should try to minimize the effect of the distribution on estimated p -values. The aforementioned problem can be solved as follows. Take as criterion

$$SSE = \sum_i (p_i - \hat{p}_i(\hat{\epsilon}, \hat{\alpha}))^2, \quad (6)$$

where α is a vector of distribution parameters for a distributional family thought to be adequate in the application. One might, for example, take the normal distribution on the θ -scale in which case there are two parameters, the mean and variance of the distribution. Actually, this distribution is used for good reasons in the simulation study to be discussed. Given an estimate $\hat{\epsilon}$, one can minimize SSE w.r.t. the parameters α . This can be done for vectors $\hat{\epsilon}$ obtained with different parameter estimation methods. The values of $SSE_{\min}$, obtained in this way, can be used in order to decide which method is best in 'predicting' p -values.

THE ROBUSTNESS OF MINCHI AND UML: A SIMULATION

As mentioned above, the performance of MINCHI is investigated in this study in a complete design. A hypothetical forty-item test was constructed with probabilities correct according to the two-parameter logistic model

$$P_i(\theta) = \exp[a_i(\theta - b_i)] / \{1 + \exp[a_i(\theta - b_i)]\} \quad (7)$$

where a_i is the discrimination parameter of item i . The test was constructed according to the following specifications. For twenty-six items the value of the discrimination parameter was set equal to 1.0. The b -values for the items ranged from -2.55 to 2.55 . For seven items a was given a high value of 1.5, and for another group of seven items a was given a low value of 0.5. In both sets of items b -values ranged from -2.55 to 2.55 . A random sample of 1000 θ s was taken from a normal distribution with a mean equal to 1.7 and standard deviation equal to 1.7. Correct and incorrect responses of the 1000 hypothetical examinees to the items were generated according to the probabilities in Equation 7. This procedure was repeated so that two replications for the same group of examinees were available for further analyses. For the minimization of SSE in Equation 6 the normal distribution on the θ -scale was used. This decision is justified by the fact that a normal distribution was used for the selection of θ s. The results from the UML-analysis confirm the adequacy of this choice.

The UML parameter estimates gave a very low $SSE_{\min}$. This can be seen in Figure 1, in which the residuals $d_i = \hat{p}_i(\hat{\epsilon}, \hat{\alpha}) - p_i$ are plotted against the observed values p for the first replication. The residuals are negligible.

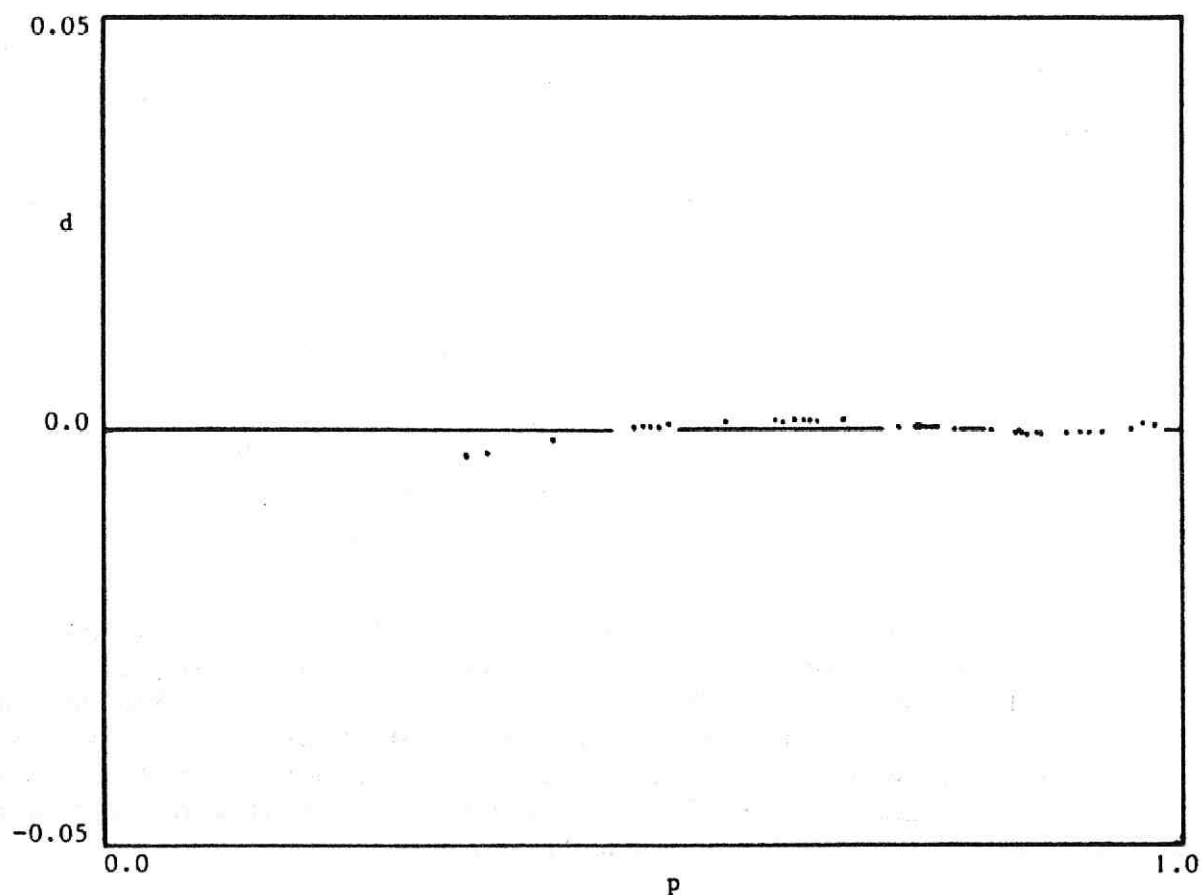


Figure 1: Plot of residuals against observed p -values for UML.

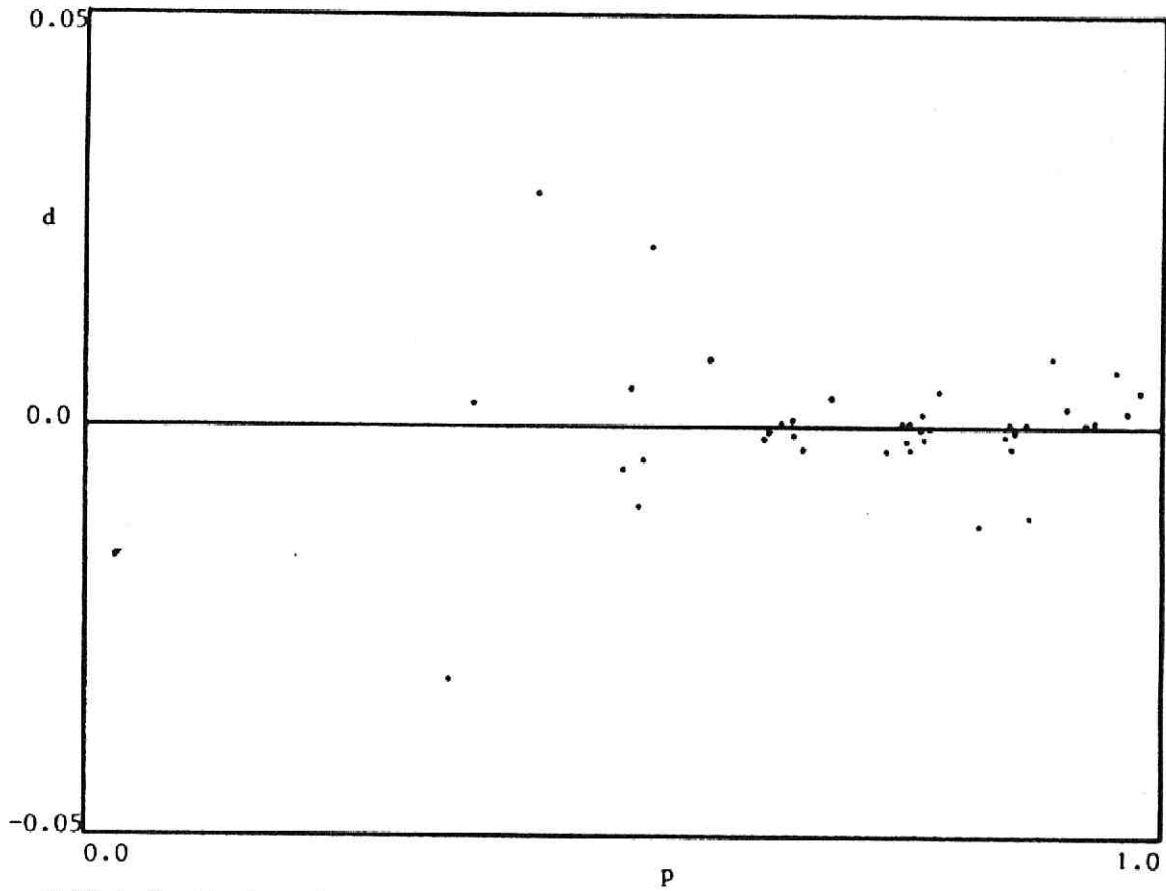


Figure 2: Plot of residuals against observed p -values for MINCHI.

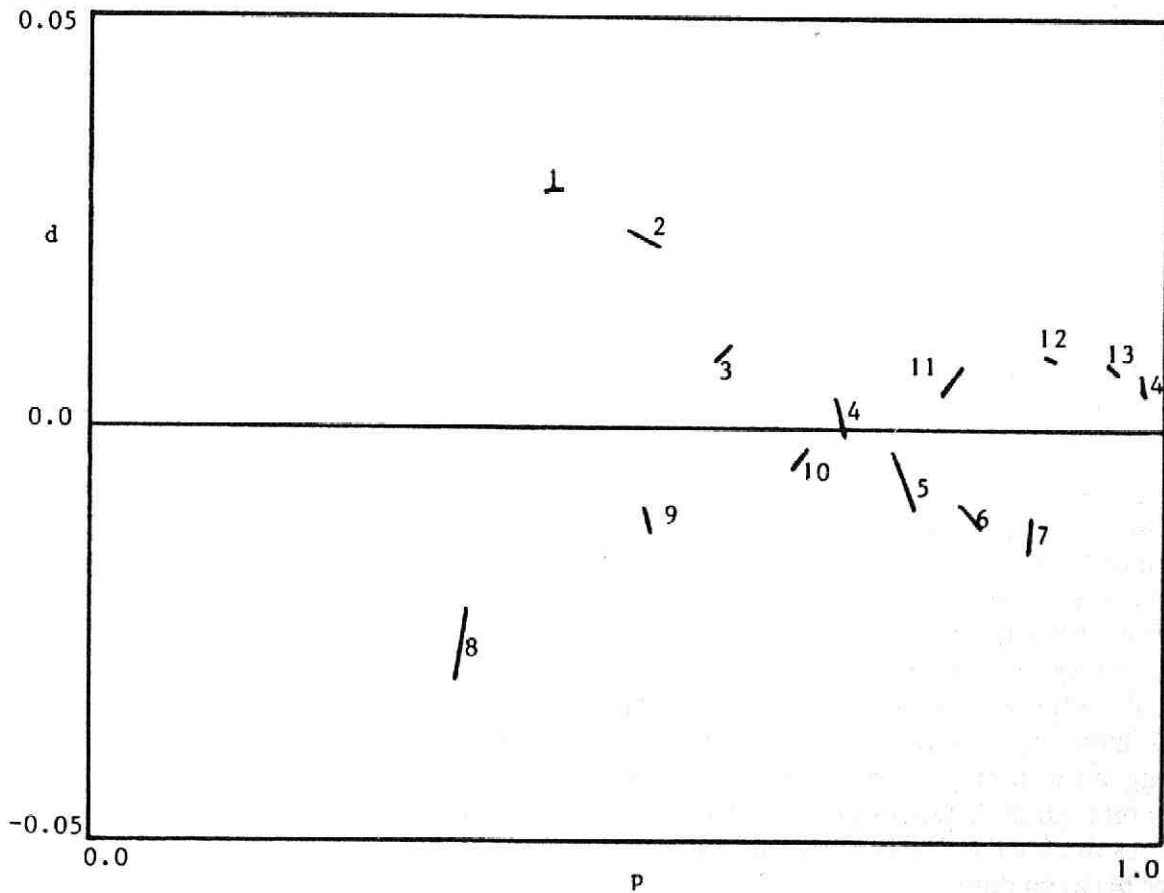


Figure 3: MINCHI-residuals for seven relatively discriminating items and seven relatively non-discriminating items (two replications).

The results for MINCHI are in strong contrast with those for UML. Figure 2 shows some relatively large residuals, notably for the lower p -values. The relatively easy and difficult items with extreme discrimination parameters are responsible for these residuals as can be seen in Figure 3. In this figure only the seven items with low a -values, items 1 through 7, and the seven items with high a -values, items 8 through 14, are shown. For each item the results of both replications are plotted and connected by a line.

From the figure two tendencies are clear. For relatively low discriminating items p is overestimated when p is relatively low, underestimated when p is relatively high. For discriminating items the reverse holds. The effects can be understood from the characteristics of MINCHI. This can be explained with the help of the next figure, Figure 4.

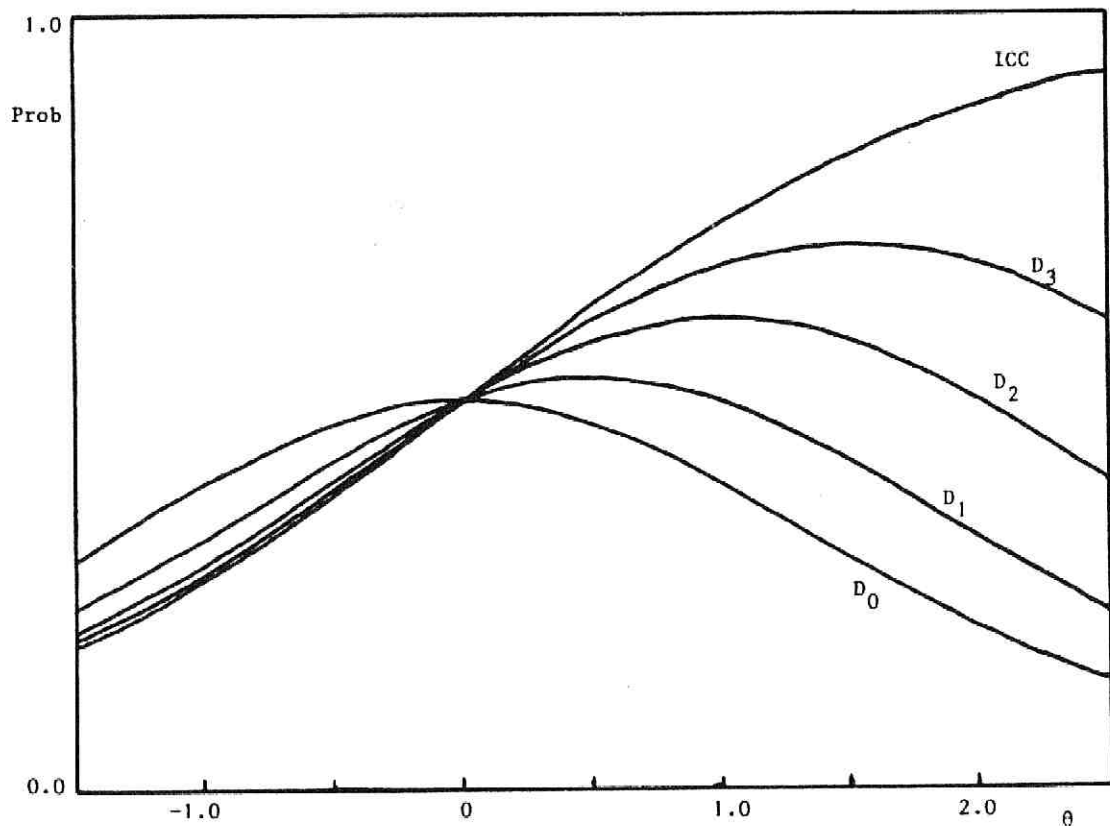


Figure 4: The ICC of a Rasch item and four curves, giving the probability of one correct out of two responses as a function of θ .

Here the item characteristic curve (ICC), the proportion correct as a function of θ , is drawn for a Rasch item. There are four other curves, labeled D_0 through D_3 . D_0 gives the probability as function of θ , that one out of two items with ICC's identical to the one displayed, is answered correctly. It gives the probability that for these two items an observation is obtained which is useful for estimation with MINCHI. This curve obtains its maximum value equal to 0.5 for θ equal to the common b -value. More to the right the probability of two correct answers increases, more to the left the probability of two incorrect answers. The other curves differ from D_0 in that the two items have different b -values. In the curve D_1 the b -values of the two items differ one logit, one of the items having a higher b than the Rasch item played in the figure. Curves D_2 and D_3 have inter-item distances of two, and three logits, respectively. The maximum of the curve always results for θ halfway between both b -values, and increases in value with an increasing distance between items.

From the figure it is clear that MINCHI uses part of the information in item responses. For

different item pairs different regions of the latent ability scale get the largest weight. When an item is relatively easy, most weight in a comparison with other items is given to abilities exceeding the item's b -value.

One can imagine what happens when a particular item deviates from the Rasch model. Take an item with a relatively low b and an a lower than average. The upper part of its ICC gets a relatively high weight. In this part of the range the curve is lower than a Rasch curve with an identical b . Consequently, the item looks more difficult than it is, when it is approximated by a Rasch curve based on MINCHI. That is precisely what happened in the simulation, as one might see in Figure 3.

DISCUSSION

MINCHI has some very attractive properties in terms of storage of information needed for an analysis and computational simplicity. However, the method is more sensitive to deviations from the Rasch model than, for example, UML. This was demonstrated in a simulation, and also in a more detailed analysis of the method itself. The result implies that the applicability of the method is reduced. As deviations from the Rasch model are more critical than with other Rasch analyses, MINCHI should be used only when it is clear that the Rasch model is adequate for a description of the data. The adequacy of the Rasch model is an empirical question, which can be answered only after an analysis of the data. So, at least in an early stage of data gathering, reduction of the information contained in item responses, using counts n_{ij} , should be discouraged. When all information is contained in counts n_{ij} , there is no way back to another analysis with the Rasch model or to another item response model.

APPENDIX

Let $\mathbf{x}$ be a response vector, where $x(i)$ equals one for a correct response to item i and zero otherwise. Further, let $\mathbf{e}$ be a vector with elements $e(i)$ equal to one when item i is selected and zero otherwise. With this notation the contribution of the response vector $\mathbf{x}$ on basis of item selection $\mathbf{e}$ to the log likelihood in MML can be written as

$$f(\mathbf{x}, \mathbf{e}) \log \int \prod_{i|e(i)=1} P_i(\theta)^{x(i)} [1 - P_i(\theta)]^{1-x(i)} h(\mathbf{e}) g(\theta) d\theta \quad (\text{A.1})$$

where $f(\mathbf{x}, \mathbf{e})$ is the observed frequency of response vector $\mathbf{x}$ with selection $\mathbf{e}$, and $h(\mathbf{e})$ is the probability of item selection $\mathbf{e}$. The factor $h(\mathbf{e})$ contains no information on item and distribution parameters and can be eliminated in further analyses.

In the general case the probability of a particular item selection is time dependent, due to the addition of items to the item pool and/or a change in $g(\theta)$. The latter change may be due to trends and short term changes in the composition of the examinee population. Only under independence, i.e. if

$$E_t h(\mathbf{e}|t) g(\theta|t) = \{E_t h(\mathbf{e}|t)\} \{E_t g(\theta|t)\}, \quad (\text{A.2})$$

A.1 remains valid. Independence is obtained when $g(\theta|t)$ does not depend on t , or when $h(\mathbf{e}|t)$ does not depend on t , i.e. when no items are added and the selection rule does not change.

When items are being added to the pool, the best strategy is to add new items only at selected occasions. Between these occasions fixed population distributions $g(\theta)$ are assumed and MML is applied to a model with several populations. Equality of populations can be tested.

REFERENCES

- Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. *Psychometrika*, 42, 375-399.
- De Gruijter, D.N.M. (1980). A two-stage testing procedure. In L.J.Th. van der Kamp, W.F. Langerak, & D.N.M. de Gruijter (Eds.), *Psychometrics for educational debates*. Chichester: Wiley.
- De Leeuw, J. & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics*, 11, 183-196.
- Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.
- Roskam, E.E. (1983). Allgemeine Datentheorie. In H. Feger & J. Bredenkamp (Eds.), *Messen und Testen* (Forschungsmethoden der Psychologie, Serie I, Band 3). Göttingen: Hogrefe, 1-135.
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175-186.
- Van der Linden, W.J., & Eggen, T.J.H.M. (1986). An empirical Bayesian approach to item banking. *Applied Psychological Measurement*, 10, 345-354.
- Zwinderman, K.H., & Van den Wollenberg, A.L. (1987). Minimum chi-square estimation in the Rasch model. Paper presented at the European Meeting of the Psychometric Society.

Manuscript ontvangen 14-5-1987

Definitieve versie ontvangen 6-7-1987

Beoordeling van de Predictieve Validiteit van een Schoolvorderingentoets door Gebruikers

M. Nijse*

Vakgroep Algemene Pedagogiek, Rijksuniversiteit Groningen

ABSTRACT

Thirty-four judges, who regularly interpret scores of elementary school children on six achievement tests, were asked to judge several aspects of the predictive validity of these tests. Aspects concerned: (1) ranking the tests according to the degree of correlation with type of secondary education chosen; (2) indicating the test with the highest correlation with year-end report marks for (a) Dutch, (b) French and English, (c) History, Geography and Biology, (d) Mathematics; (3) ranking the tests according to amount of difference between means of pupils attending various types of secondary education. Expert judgments were compared with data of 323 pupils. Experts agreed to a considerable degree on the various aspects. The average judge seemed to perform rather well. Partly, this is due to the small range of criterion correlations (.61-.71). Considering that this is not the case for all achievement tests and the imperfect agreement among judges, it is concluded that there is little reason to replace a statistical prediction by a clinical prediction.

1. INLEIDING

Ieder jaar wordt door de overgrote meerderheid van de leerlingen van klas zes van de lagere school (tegenwoordig groep acht van de basisschool) een schoolvorderingentoets afgelegd alvorens zij de school verlaten. De directeur van de basisschool brengt over deze leerlingen een advies uit betreffende de meest geschikte vorm van voortgezet onderwijs. Voor toelating tot een HAVO- en/of VWO-brugklas is de school waaraan een dergelijke brugklas is verbonden verplicht de uitslag op de schoolvorderingentoets (eventueel het resultaat van een psychologisch onderzoek of een toelatingsexamen) een rol te laten spelen bij de beslissing over het al dan niet toelaten van een leerling (Verlinden, 1968, p. 52-55).

Over de waarde die aan de uitslag op een schoolvorderingentoets wordt toegekend is weinig bekend. Van Beusekom (1979, 1984) doet verslag van enquêtes, waaruit onder meer blijkt dat toelatingscommissies weliswaar enige waarde hechten aan de schoolvorderingentoets, maar het advies van de directeur van de basisschool (het schooloordeel) aanzienlijk belangrijker vinden.

Sandbergen e.a. (1972) hebben er echter op gewezen dat de grotere waardering van het schooloordeel een gevolg kan zijn van de invloed die het toetsoordeel op het schooloordeel heeft. Leden van toelatingscommissies zijn niet de enigen van wie het oordeel over de schoolvorderingentoets interessant is. De opvattingen van directeurs van basisscholen zijn van belang omdat het uiteindelijke advies over de meest geschikte schoolkeuze kan worden beïnvloed door de behaalde resultaten. Een derde groep gebruikers wordt gevormd door de samenstellers van de schoolvorderingentoets. Hun opvattingen zijn ongetwijfeld van invloed op vorm en inhoud van de samen te stellen toets.

* De auteur dankt Prof. Dr. W.K.B. Hofstee voor zijn commentaar op een eerdere versie van dit artikel.
Adres: A-weg 30, 9718 CW Groningen.

Het onderzoek waarvan hier verslag wordt gedaan had tot doel meer inzicht te verkrijgen in de waardering die betrokkenen voor verschillende onderdelen van de schoolvorderingentoets hebben. Door de voorspellende validiteit van de schoolvorderingentoets, gebaseerd op empirische gegevens, te vergelijken met de voorspellende validiteit die uit de oordelen van gebruikers kan worden afgeleid, kan meer inzicht worden verkregen in de waarde van deze oordelen. Een dergelijk onderzoek sluit aan bij de bekende discussie over klinische versus statistische predictie (zie b.v. Wiggins, 1973, p. 156-180).

In de volgende paragraaf wordt een beschrijving gegeven van de gebruikte schoolvorderingentoets, de geënquêteerde personen, de inhoud van de enquête en de onderzochte leerlingen. In paragraaf 3 worden de resultaten van het onderzoek beschreven. Paragraaf 4 bevat de conclusies.

2. METHODE

a. Schoolvorderingentoets

De schoolvorderingentoets Noordoost-Friesland wordt ieder jaar opnieuw samengesteld door deskundigen, die bijna allen aan het basisonderwijs of voortgezet onderwijs zijn verbonden. De toets wordt tweemaal per cursusjaar afgenomen (november en februari) op een groot aantal basisscholen in Noordoost-Friesland. De toets bestaat uit zes onderdelen: *Tekstverklaring* (naar aanleiding van een leestekst worden multiple-choice vragen over de inhoud van de tekst gesteld); *Dictee* (een moeilijk woord uit een voorgelezen zin moet worden opgeschreven); *Stijl* (kennis van een idioom, spreekwoorden en gezegden); *Denksommen* (naar aanleiding van een korte tekst moet een berekening worden gemaakt); *Cijferen* (op getallen, percentages, gewone en decimale breuken moet een rekenkundige bewerking worden toegepast); *Hoofdrekenen* (idem als Cijferen, maar zonder gebruik van kladpapier). Hoewel geen psychometrische analyses op de subtoetsen worden uitgevoerd, blijkt uit een niet-gepubliceerd onderzoek de betrouwbaarheid vrij hoog: KR-20 varieert van .82 (Dictee) tot .92 (Hoofdrekenen). Tekstverklaring is een uitzondering met een KR-20-waarde van .68.

b. Geënquêteerden

In overleg met de secretaris van de Selectieraad Noordoost-Friesland werd aan 51 personen een enquêteformulier toegestuurd: twaalf directeuren van basisscholen, met wie reeds enkele jaren contact bestond in het kader van intelligentie- en schoolvorderingenonderzoek, zestien begeleid(st)ers van diverse brugklassen en 23 personen die aan de constructie van de verschillende subtoetsen meewerkten. Of deze steekproef representatief is voor alle personen die bij het gebruik van deze schoolvorderingentoets zijn betrokken, is moeilijk te zeggen. In ieder geval is geen sprake van onzuiverheid op grond van woonplaats, schooltype en betrokkenheid bij een bepaalde subtoets. De leerlingen van de betrokken basisscholen behalen gemiddelde scores die niet duidelijk afwijken van de totale gemiddelden. Er lijkt derhalve weinig reden aan de representativiteit van de geënquêteerde directeuren van basisscholen te twijfelen.

In totaal werden (eventueel na herinnering) 34 enquêteformulieren terug ontvangen. Voor directeuren van basisscholen was de respons 100%, voor brugklasbegeleid(st)ers 44% en voor toetsconstructeurs 65%. In totaal bedroeg de respons 67%. De gegeven antwoorden bleken voor de drie categorieën nauwelijks van elkaar te verschillen, zodat de drie groepen werden samengevoegd.

c. Enquête

De geënquêteerden werd (na een korte uiteenzetting van bedoeling en achtergronden van het onderzoek) verzocht de subtoets, welke naar verwachting de hoogste correlatie met het gekozen schooltype na de brugklas zou hebben, het rangcijfer één toe te kennen. De subtoets met de op één na hoogste correlatie kreeg het rangcijfer twee, enz. (Het begrip correlatie is aan geënquêteerden bekend en werd ten overvloede kort toegelicht). Wanneer men twee of meer dan twee

subtoetsen niet van elkaar kon onderscheiden, kon hetzelfde rangcijfer meerdere malen worden gegeven. De schrijver droeg zorg voor de omzetting in gemiddelde rangcijfers.

In Vraag 2 werd verzocht één van de zes subtoetsen (eventueel meer dan één) te noemen die naar verwachting de hoogste correlatie zou hebben met het eindrapportcijfer in de MAVO-brugklas voor a. *Nederlands*, b. *vreemde talen* (Frans en Engels), c. *zaakvakken* (Geschiedenis, Aardrijkskunde en Biologie) en d. *Wiskunde*. Vraag 3 en Vraag 4 waren identiek aan Vraag 2, maar hadden betrekking op HAVO-respectievelijk VWO-leerlingen. (LBO-leerlingen moesten helaas buiten beschouwing blijven, omdat de verschillende soorten lager beroepsonderwijs moeilijk met elkaar te vergelijken zijn en niet overal een cijferrapport wordt gehanteerd.)

Vraag 5 had betrekking op verschillen tussen de diverse schooltypen. Evenals bij Vraag 1 werd verzocht de zes subtoetsen de rangcijfers één tot en met zes toe te kennen. De subtoets die naar verwachting het grootste verschil tussen LBO- en MAVO-leerlingen te zien zou geven, kreeg het rangcijfer één, enz. Dezelfde vraag moest worden beantwoord voor het verschil tussen MAVO- en HAVO-leerlingen en HAVO- en VWO-leerlingen.

De voorspellende waarde van de schoolvorderingentoets die kan worden afgeleid uit de enquête-gegevens, kan worden vergeleken met de optimale voorspellende waarde die gevonden wordt voor de gegevens van 323 leerlingen die in 1980/81 de schoolvorderingentoets Noordoost-Friesland aflegden. Van de cursusjaren 1981/82, 1982/83 en 1983/84 werden gegevens over het schooltype, eindrapportcijfers en al dan niet zittenblijven verwerkt. De percentages leerlingen op de verschillende schooltypen na de brugklas zijn 44 (LBO), 30 (MAVO), 12 (HAVO) en 14 (VWO).

3. RESULTATEN

Op de antwoorden van Vraag 1 (rangordening van de subtoetsen op grond van de te verwachten Pearson-correlatie met het schooltype dat na de brugklas wordt gekozen) werd Kendall's concordantiecoëfficiënt (W) berekend. Er werd gecorrigeerd voor knopen (zie Marascuilo & McSweeney, 1978, p. 458-462). De waarde .27 blijkt significant van 0 te verschillen bij $\alpha = .01$, maar is absoluut gezien niet hoog. Er bestaat dus op het eerste gezicht een matige overeenstemming tussen de beoordelaars. Hofstee (1968) wijst er echter op dat W niet gemakkelijk interpreteerbaar is in deze situatie. In feite gaat het om de vraag naar de (geschatte) betrouwbaarheid van de gemiddelde rangordening. Hofstee geeft een algebraïsche afleiding van W die een speciaal geval van Cronbach's coëfficiënt α oplevert. De geschatte betrouwbaarheid blijkt .92 te zijn. Geconcludeerd mag worden dat de gegeven oordelen in hoge mate betrouwbaar zijn.

Denksommen wordt als belangrijkste subtoets gezien; het gemiddelde rangcijfer is 2,2. De hoge waardering is terecht: de correlatie met het gekozen schooltype is inderdaad de hoogste die wordt gevonden (.71). Tekstverklaring verschilt in waardering nauwelijks van Denksommen: het gemiddelde rangcijfer is 2,4. Deze subtoets wordt overgewaardeerd: de gevonden correlatie is met .64 de op twee na laagste. Dictee neemt op enige afstand van Denksommen en Tekstverklaring de derde plaats in met een gemiddelde van 3,4. Ook deze subtoets wordt overschat: $r = .63$. Cijferen, Hoofdrekenen en Stijl blijken ongeveer dezelfde waardering te ontvangen: de gemiddelde rangcijfers zijn 4,2, 4,3 respectievelijk 4,5. De gevonden correlaties verschillen relatief sterk van elkaar: de waarden zijn .61 (Cijferen), .67 (Hoofdrekenen) en .70 (Stijl). Kendall's tau voor de gemiddelde rangcijfers en de correlaties bedraagt slechts .07, hetgeen niet significant is op .05-niveau. Beoordelaars en werkelijkheid stemmen dus nauwelijks met elkaar overeen. Merk echter op dat de absolute verschillen tussen de correlaties niet erg groot zijn; de range loopt van .61-.71. Een toets voor afhankelijke correlaties (Steiger, 1980) toont echter aan dat de hypothese van gelijke correlaties moet worden verworpen ($\alpha = .05$).

Zoals in de Inleiding is opgemerkt, is het onderzoek verwant met onderzoek naar de vergelijking van de validiteit van klinische en statistische voorspellingen. Het kenmerkend onderscheid tussen beide soorten voorspellingen is dat klinische voorspellingen worden gedaan zonder gebruikmaking van een geëxpliciteerde procedure voor het combineren van gegevens (zie

bijvoorbeeld Sawyer, 1966). In het onderhavige onderzoek hebben beoordelaars een overall-waardering van de subtoetsen gegeven. De overwegingen die geleid hebben tot de toekenning van een bepaald rangnummer aan een bepaalde subtoets zullen variëren per beoordelaar. De uit de enquête verkregen gemiddelde rangcijfers van Vraag 1 kunnen worden opgevat als 'clinische' regressiegewichten waarmee de subtoetsscores worden gewogen. De waarde van de hieruit voortvloeiende R^2 kan worden vergeleken met de waarde van de optimale R^2 , die wordt verkregen door de subtoetsscores te wegen met 'statistische' gewichten die door de correlaties met het criterium en de intercorrelaties van de subtoetsen worden bepaald. Een vooronderstelling van deze werkwijze is dat de geënuquêteerden de informatie van de subtoetsscores op een lineaire wijze met elkaar combineren. Dat deze vooronderstelling alleszins redelijk is, wordt beargumenteerd door Wiggins (1973, p. 171-179) en Hofstee (1974, p. 85-87), die zich beiden onder meer beroepen op Goldberg (1968). De redelijkheid van de bewuste keuze voor een lineaire combinatie van de data is eveneens beargumenteerd door Wiggins (1973, p. 74-77) onder verwijzing naar een onderzoek van Ward (1954).

Omdat een laag rangcijfer een hoge waardering betekent, zijn de subtoetsscores gewogen met de reciproken van de gemiddelde rangcijfers, zodat een (gestandaardiseerd) regressiegewicht groter is naarmate een subtoets belangrijker wordt gevonden. Wanneer het schooltype na de brugklas met behulp van alle zes enquêtegewichten wordt voorspeld, blijkt R^2 .615 te bedragen. Deze waarde ligt slechts .002 lager dan R^2 die met de op de gegevens gebaseerde gewichten wordt verkregen. Het te verwaarlozen verschil zal geen verrassing zijn voor wie op de hoogte is met onderzoek naar de bruikbaarheid van vereenvoudigde regressiegewichten (zie bijvoorbeeld Pijl & Nijse, 1977). Wanneer alleen Denksommen en Tekstverklaring (de twee belangrijkste predictoren volgens de geënuquêteerden) worden gebruikt om het criterium te voorspellen, is R^2 .549 en .553 voor respectievelijk enquêtegewichten en empirische gewichten. Bij de keuze van Denksommen en Stijl, die de hoogste correlaties met het criterium hebben, blijken de overeenkomstige waarden van R^2 .569 en .575 te bedragen. Het verschil tussen Denksommen en Tekstverklaring, gewogen met enquêtegewichten, en Denksommen en Stijl, gewogen met op de gegevens gebaseerde gewichten, is slechts .026, evenmin een indrukwekkend verschil. Het komt er dus op neer dat de (subjectief gewogen) voorspelling van het schooltype na de brugklas met de twee subtoetsen die deskundigen het belangrijkste vinden nauwelijks onderdoet voor de (empirisch gewogen) voorspelling met de twee hoogstcorrelerende subtoetsen.

Vraag 2 tot en met 4 (noemen van de subtoets met de naar verwachting hoogste correlatie met eindrapportcijfers Nederlands, vreemde talen, zaakvakken en Wiskunde voor MAVO-, HAVO- en VWO-leerlingen) werd voor de verschillende schooltypen op dezelfde wijze beantwoord. De modale antwoordcategorie voor Nederlands en de zaakvakken is Tekstverklaring (door 59% respectievelijk 87% genoemd), voor de vreemde talen Dictée (49%) en voor Wiskunde Denksommen (81%). In Tabel 1 zijn de (empirisch-gevonden) correlaties die significant groter dan 0 zijn ($\alpha = .01$) vermeld.

Voor MAVO-leerlingen blijkt niet Tekstverklaring, maar Dictée de hoogste correlatie met Nederlands te hebben. Het verschil tussen deze correlaties is echter niet significant ($\alpha = .01$). Voor HAVO- en VWO-leerlingen levert geen enkele subtoets een significante correlatie op. Dictée wordt terecht gekozen als voornaamste voorspeller van het eindrapportcijfer voor de vreemde talen van MAVO-leerlingen. Voor HAVO- en VWO-leerlingen wordt deze subtoets overgewaardeerd, want de correlaties zijn niet significant. Ook andere subtoetsen correleren niet-significant met het eindrapportcijfer voor de vreemde talen ($\alpha = .01$).

Voor MAVO- en VWO-leerlingen is de keuze van Tekstverklaring als belangrijkste voorspeller van de zaakvakken terecht. Denksommen en Cijferen leveren voor VWO-leerlingen echter (vrijwel) dezelfde correlaties op; de verschillen zijn niet significant. Voor HAVO-leerlingen worden opnieuw geen significante correlaties gevonden. Voor de voorspelling van het eindrapportcijfer Wiskunde is Denksommen niet de optimale keuze voor MAVO-leerlingen: Cijferen en Hoofdrekenen leveren iets hogere, maar niet significante verschillende correlaties op. Voor HAVO-leerlingen blijkt geen enkele subtoets een significante correlatie met Wiskunde te hebben. Voor VWO-leerlingen is Denksommen inderdaad de subtoets met de hoogste correlatie.

Tabel 1. Statistisch-significante ($\alpha = .01$) correlaties tussen subtoetsen en criteria (decimaalpunten weggelaten). Tv = Tekstverklaring; Dt = Dictée; St = Stijl; Ds = Denksommen; Cij = Cijferen; Hr = Hoofdrekenen.

Criterium		Subtoets					
		Tv	Dt	St	Ds	Cij	Hr
Nederlands	MAVO	31	44	33	—	29	—
"	HAVO	—	—	—	—	—	—
"	VWO	—	—	—	—	—	—
Vreemde talen	MAVO	29	47	26	—	27	—
"	HAVO	—	—	—	—	—	—
"	VWO	—	—	—	—	—	—
Zaakvakken	MAVO	24	—	—	—	—	—
"	HAVO	—	—	—	—	—	—
"	VWO	38	—	—	38	37	—
Wiskunde	MAVO	26	25	—	26	27	27
"	HAVO	—	—	—	—	—	—
"	VWO	—	—	—	41	—	—

De antwoorden op Vraag 5 (rangordening van de subtoetsen op grond van het te verwachten verschil tussen LBO- en MAVO-leerlingen, MAVO- en HAVO-leerlingen en HAVO- en VWO-leerlingen) lenen zich, evenals de antwoorden op Vraag 1, voor de berekening van Kendall's concordantiecoëfficiënt. De waarden zijn .20 voor de vergelijking van LBO-MAVO-leerlingen, .39 (MAVO- versus HAVO-leerlingen) en .45 (HAVO- versus VWO-leerlingen). Alle waarden zijn significant ($\alpha = .01$); de geschatte betrouwbaarheden zijn .83, .92 respectievelijk .94. De overeenstemming tussen beoordelaars is met betrekking tot de vergelijking van LBO- en MAVO-leerlingen minder groot dan met betrekking tot de andere twee vergelijkingen. De t-waarden voor de vergelijkingen zijn vermeld in Tabel 2.

Dictée neemt niet de eerste maar de vijfde plaats in bij de vergelijking van LBO- en MAVO-leerlingen; Denksommen blijkt het grootste verschil tussen deze groepen op te leveren. Voor de vergelijking van MAVO- en HAVO-leerlingen en HAVO- en VWO-leerlingen worden Denksommen en Tekstverklaring als belangrijkste subtoetsen genoemd. Deze keuzen blijken niet in overeenstemming met de werkelijkheid. Kendall's tau, berekend tussen de gemiddelde rangcijfers van beoordelaars en de t-waarden van Tabel 2, bedraagt $-.33$ (LBO-MAVO), $-.33$ (MAVO-HAVO) en $.20$ (HAVO-VWO). Opvallend is dat de grootte van de t-waarden sterk afneemt wanneer de vergelijking de 'hogere' schooltypen betreft.

Tabel 2. Significante ($\alpha = .01$) t-waarden voor de vergelijking van gemiddelde scores voor subtoetsen.

Subtoets	LBO-MAVO	MAVO-HAVO	HAVO-VWO
Tekstverklaring	5,54	4,93	3,40
Dictée	7,61	3,61	—
Stijl	8,48	5,57	—
Denksommen	9,60	4,76	—
Cijferen	9,13	2,01	—
Hoofdrekenen	8,59	4,48	2,74

4. DISCUSSIE

De voornaamste conclusie uit het onderzoek is dat deskundigen (gemiddeld genomen) verschillende aspecten van de predictieve validiteit met een redelijke mate van juistheid kunnen

beoordelen. De rangorde van de beoordelaars bij Vraag 1 stemt weliswaar in geringe mate overeen met de empirische rangorde, maar consequenties voor het verlies van voorspellende waarde heeft dit verschil niet of nauwelijks. Uit analyses, die hier wegens plaatsgebrek buiten beschouwing blijven, blijkt dat bij toepassing van de aan de subtoetsen verbonden normen een subtoets voor 60 tot 66% van de leerlingen het schooltype correct voorspelt. Deze geringe verschillen leiden tot de conclusie dat, althans voor de gegevens van dit onderzoek, zelfs een willekeurig gekozen combinatie van twee subtoetsen nauwelijks slechter voorspelt dan de optimale combinatie. Andere aspecten van de predictieve validiteit, die in Vraag 2 tot en met 5 aan de orde worden gesteld, blijken eveneens met een redelijke mate van juistheid te worden beoordeeld.

De vraag of de resultaten van dit onderzoek naar andere onderzoeken kunnen worden gegeneraliseerd is moeilijk te beantwoorden, omdat geen vergelijkbaar onderzoek is verricht. Eén aspect waarmee rekening moet worden gehouden betreft de mogelijke verschillen in predictieve validiteit van verschillende subtoetsen. Ook hierover is weinig bekend. In Sandbergen e.a. (1972) en in de handleiding van de ISI-reeks (Van Boxtel, Snijders & Welten, 1982) worden grotere verschillen tussen criterium-correlaties vermeld dan voor de subtoetsen in dit onderzoek. Wanneer gebruikers zich van deze verschillen niet bewust zijn, mag een groter verschil tussen klinische en statistische voorspellingen worden verwacht.

Niet uit het oog mag worden verloren dat de uitkomsten voor dit onderzoek geldig zijn voor de gemiddelde (Vraag 1 en 5) of modale (Vraag 2-4) beoordelaar. Deze zal men in levenden lijve niet met grote frequentie tegenkomen, zoals uit Kendall's concordantiecoëfficiënten en percentages modale antwoorden kan worden afgeleid. Een conclusie hieruit is dat het onverstandig zou zijn een statistische voorspelling te vervangen door een klinische.

LITERATUUR

- Beusekom, H.A.M. van (1979). Huidige brugklasleerling heel anders dan vroeger. *Weekblad voor Leraren bij het Voortgezet onderwijs*, 12, no. 9, 291-298.
- Beusekom, H.A.M. van (1984). In basisonderwijs grote verschillen in opvattingen. *Weekblad van het Nederlands Genootschap van Leraren*, 16, no. 35, 1313-1318.
- Boxtel, H.W. van, Snijders, J.Th., & Welten, V.J. (1982). *ISI; Interesse, Schoolvorderingen, Intelligentie. Verantwoording van het prestatiegedeelte & Handleiding voor de gehele testreeks*. Groningen, Wolters-Noordhoff.
- Goldberg, L.R. (1968). Simple models or simple processes? Some research on clinical judgment. *Amer. Psychol.*, 23, 483-496.
- Hofstee, W.K.B. (1968). Estimating concordance in a set of rank orders. *Psychol. Rep.*, 23, 1279-1282.
- Hofstee, W.K.B. (1974). *Psychologische uitspraken over personen*. Deventer: Van Loghum Slaterus.
- Marascuilo, L.A., & McSweeney, M. (1978). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.
- Pijl, S.J., & Nijse, M. (1977). Een empirisch onderzoek naar de bruikbaarheid van vereenvoudigde gewichten in multiple regressieanalyses. *Methoden en Data Nieuwsbrief*, 2, 99-107.
- Sandbergen, S., Elshout, J., Akkerman, T., & Peet, A. van (1978). Enkele relaties tussen een intelligentietest en een studietoets. *Nederl. Tijdschr. Psychol.*, 27, 509-529.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychol. Bull.*, 66, 178-200.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychol. Bull.*, 87, 245-251.
- Verlinden, J.A.A. (1968). *De Mammoetwet; nieuwe fase in de ontwikkeling van ons onderwijs*. Deventer: Kluwer.
- Ward, J.H.Jr. (1954). An application of linear and curvilinear joint functional regression in psychological prediction. *Research Bulletin 54-86*, Air Force Personnel Training Research Center. Lackland Air Force Base, Texas.
- Wiggins, J.S. (1973). *Personality and prediction; principles of personality assessment*. Reading, MA: Addison-Wesley.

Manuscript ontvangen 9-6-1986

Definitieve versie ontvangen 19-5-1987

Notities en Commentaren

BEOORDELING*) VAN HET EINDRAPPORT "VERGELIJKING VAN REKENMETHODEN IN HET BASISONDERWIJS" UITGEVOERD DOOR HET RION (PROJECTLEIDER DRS. E.G. HARKAMP), TOTAAL SUBSIDIEBEDRAG f 298.768,-.

Probleemstelling en opzet

Een van de belangrijke verschillen tussen oudere en nieuwere rekenmethoden voor het basisonderwijs is de wijze van differentiatie die in de methoden is uitgewerkt. Teneinde de relatieve doelmatigheid van onderwijsleerpakketten vast te stellen en aanbevelingen te kunnen doen die van nut zijn voor scholen bij het overgaan op een nieuwe rekenmethode, heeft het voorliggende onderzoek acht methoden en hun wijze van differentiëren onderling vergeleken naar de mate van implementatie en hun effect op de leerlingprestaties.

De probleemstelling daarbij luidde:

1. Wat zijn de verschillen en overeenkomsten tussen de onderwijsleerpakketten (bijvoorbeeld verschil in differentiatiemaatregelen)?
2. Wat is de kwaliteit van die pakketten met het oog op de feitelijke ingebruikname en leerlingeffecten?
3. Onder welke condities vindt feitelijke ingebruikname plaats?
4. Waardoor worden eventuele verschillen in effectiviteit van pakketten bepaald?

In de opzet heeft – naast het hypothesetoetsend deel – het weddenschapmodel (in strikte zin) van Hofstee centraal gestaan. In dit model streven betrokken partijen (i.c. uitgevers van rekenmethoden) naar overeenstemming over evaluatiecriteria waaraan de effecten van hun methoden gemeten gaan worden en doen vervolgens voorspellingen in de vorm van een weddenschap over de mate waarin het eigen curriculum betere, slechtere of gelijke prestaties bij de leerlingen realiseert dan de concurrerende curricula. Als onderdeel van de weddenschap werd de uitgevers ook gevraagd op welke storende variabelen (die de relatie tussen methode en effect plausibel zou vertekenen) in het onderzoek gecontroleerd zou moeten worden. Gezien de vraagstelling zou ook nagegaan worden in hoeverre het curriculum-zoals-geïmplementeerd overeenkomt met het curriculum-zoals-bedoeld. Daarnaast zijn in het design opgenomen de beginsituatie van de leerlingen en de feitelijk bestede leertijd. De prestaties zouden gemeten worden door middel van al dan niet methode-specifieke CITO-toetsen, de beginsituatie door een niet-verbale intelligentietest, de implementatie-karakteristieken met een schriftelijke vragenlijst die aan de leerkrachten werd voorgelegd. De onderzoeksgroep bestond uit leerlingen van klas 6 en hun leerkrachten. In totaal zouden 500 scholen uitgenodigd worden verdeeld over zes methoden.

Samenvatting eindrapport

De probleemstelling zoals hierboven omschreven is inderdaad het uitgangspunt geweest bij de beschrijving van de resultaten in het eindrapport. Wel hebben de vragen op drie punten bijzondere accenten gekregen. De effectmeting beperkt zich niet tot al dan niet specifieke prestatie-toetsen, maar ook de rekenattitude wordt bij de resultaten betrokken. Ten tweede wordt onder de condities waaronder feitelijke ingebruikname plaatsvindt ook de factor wel/geen combinatieklas betrokken. Ten derde worden de implementatiefactoren onderscheiden in aantal lesuren, variatie in leerstof, hoeveelheid behandelde leerstof en leerstofdifferentiatie.

Bij de selectie van methoden heeft enerzijds het marktaandeel, anderzijds de implementatie-

*) De samenvatting van de beoordeling van dit door SVO gesubsidieerde onderzoek wordt ingediend onder verantwoordelijkheid van het SVO-bestuur, onder redactie van drs. A.C.C. van den Oever, medewerker onderzoeksbeoordeling van het SVO-bureau.

fase (wel of niet tot en met klas 6) een rol gespeeld. Gekozen zijn de klassieke methoden 'Niveaucursus Rekenen', 'Nieuw Rekenen', 'Op Veilig Spoor', 'Naar Zelfstandig Rekenen' en 'Naar Aanleg en Tempo'. Van de moderne methoden zijn in het onderzoek vertegenwoordigd: 'Taltaal', 'Operator Rekenen' en 'Getal in Beeld'. Ideaaltypisch kenmerken zich de traditionele methoden door nadruk op cijfervaardigheid, op rekenen binnen het metriekstelsel en geldrekenen. Centraal staat het inslijpen van basisvaardigheden, procedures en regels. De toepassingsstof is te vinden in de verrijkingsstof, alleen bestemd voor snelle leerlingen. De moderne methoden sluiten meer aan op de informele benaderingswijze van leerlingen met aanschouwelijke schema's. Problemen worden gepresenteerd in een reële context. Moderne rekenmethoden kenmerken zich verder door een grotere variatie in oefeningen en minder leerstofdifferentiatie dan de traditionele. De overeenkomsten tussen beide groepen methoden zijn aanmerkelijk. Beide gaan uit van bekende leergangen en beide hanteren overwegend groepsgewijze instructie.

Het onderzoeksmodel postuleert een effect van het type rekenmethode via implementatiefactoren op de effectvariabelen prestatie en rekenattitude. Deze relatie wordt beïnvloed door de werking van school-, leerkracht- en leerlingfactoren. Op grond van theoretische overwegingen werd een positief effect van de moderne methoden op de afhankelijke variabelen verwacht in vergelijking met klassen die één van de traditionele methoden gebruikten.

Alle (acht) uitgevers die uitgenodigd zijn hebben aan het onderzoek een bijdrage geleverd in de vorm van het bereiken van consensus over de effect-criteria, het aangeven van leerdoelen waarvoor extra toetsen ontwikkeld moesten worden en een prestatievoorspelling in termen van het percentage leerlingen dat normaal gesproken een voldoende zou hebben op elke toets, het minimale en het maximale percentage dat een voldoende zou halen. Daarnaast gaven de uitgevers het belang aan dat ze aan elk van de toetsen afzonderlijk hechtten.

De steekproef van aan het onderzoek deelnemende klassen is kleiner dan in de onderzoeksopzet voorzien, namelijk 133 zesde klassen (120 scholen). Het in het voorstel genoemde minimum van 20 klassen per methode is niet gehaald. De onderzoekers concluderen op grond van de verdeling van de CITO-resultaten wel dat de effectieve steekproef een goede afspiegeling is van de CITO-scholen. Uit de potentiële steekproef van 133 werd een reactie ontvangen van 124 leerkrachten.

Het resultaat van de inhoudelijke vergelijking tussen de verschillende rekenmethoden en de gebruikte toetsen geeft als conclusie dat er geen grote verschillen tussen de methoden zijn in relatieve aandacht voor de leerstofcategorieën en dat de CITO-toets veel rekendoelen van vrijwel alle methoden bestrijkt, terwijl de adhoc RION-toets vooral de doelstellingen van de modernere methode dekt. Het blijkt dat de RION-toets meer een beroep doet op toepassingen, de CITO-toets meer op de vaardigheid zelf. De betrouwbaarheden van de subtests van CITO en RION variëren tussen .67 en .94. Zes van de acht subtesten blijven wat betrouwbaarheid betreft onder de .80. De voorspellingen van de uitgevers over de prestaties van leerlingen die hun methoden gebruikten wijken vrij veel af van de feitelijke leerlingresultaten zowel over- als onderschattend.

De analyse op leerlingniveau laat zien dat het effect dat de verschillen in rekenmethoden op de toetsprestaties lijken te hebben gecontamineerd wordt door de invloed van het verschil in intelligentie. De intelligentiescore hangt hoog samen met de prestaties, zodat het methodeneffect door uitpartialisering van de intelligentie-factor geneutraliseerd wordt. Een uitzondering op deze conclusie is het RION-toetsonderdeel Combinatoriek dat een significant verschil tussen moderne en klassieke rekenmethoden laat zien. De methoden sorteren geen verschil in rekenattitudes.

Op klasseniveau is onderzocht of er verschillen zijn tussen methoden in de implementatiefactoren hoeveelheid behandelde stof, variatie in leerstof, aantal lessen en leerstofdifferentiatie. Alleen de hoeveelheid behandelde stof liep tussen methoden enigermate uiteen. De verschillen binnen de groep klassieke en binnen de groep moderne methoden zijn echter groter dan die tussen deze twee groepen. Het aantal combinatieklassen per school (geoperationaliseerd als het aantal leerlingen minder dan 6) is gemiddeld hoger bij traditionele scholen. Een andere onderwijsfactor, de leerling- of leerstofgerichtheid van de leerkracht varieert ook met het type

methode: de leerkrachten die moderne methoden gebruiken hebben over het algemeen een meer leerlinggerichte onderwijsopvatting.

Een verklaringsmodel is opgesteld om causale relaties tussen de diverse bovengenoemde factoren te kunnen leggen. In dit kader zijn acht hypothesen met behulp van LISREL getoetst. De intelligentie-factor heeft de grootste causale relatie met de rekenprestaties. Eveneens effect hebben het aantal rekenlessen, het aantal leerlingen dat de basisstof deed, de variatie in leerstof, de leerstofdifferentiatie en het aantal combinatieklassen. De rekenmethode (modern of klassiek) heeft geen directe invloed.

Beoordeling

De beoordeling is gebaseerd op de adviezen van drs. J.C. van Bruggen en dr. C. Hamaker.

Achtereenvolgens worden besproken de vraagstelling, de opzet, de resultaten en de bruikbaarheid.

De vraagstelling, onderverdeeld in vier deelvragen, wordt terecht in de paragraaf van het eindrapport die de conclusies bevat, nog eens expliciet aan de orde gesteld. Nagegaan wordt in hoeverre de deelvragen door de resultaten beantwoord kunnen worden. De formulering van de vragen zou duidelijker en eenduidiger gesteld kunnen worden. Dit geldt met name voor de vragen 2 en 3. Meer inhoudelijk doet zich een discrepantie voor tussen de algemene en deelvragen enerzijds die uitgaan van verschillen tussen rekenmethoden sec en de uitwerking daarvan anderzijds die in het onderzoek consequent is volgehouden in een tweedeling klassieke versus moderne methoden. Dit onderscheid nu is problematisch op inhoudelijke gronden omdat de verschillen tussen de twee groepen blijkens tabel 5.1 niet groot zijn, terwijl binnen de groepen zeker zo grote inhoudelijke verschillen te zien zijn. Voor de verschillen in didactiek tussen de twee groepen wordt in het rapport weinig argumentatie geboden. Dit probleem zou waarschijnlijk ondervangen kunnen worden door een uitvoeriger confrontatie van de vakdidactische analyse met de literatuur die over dit onderwerp, bijvoorbeeld in de dissertatie van De Jong 'Wiskobas in methoden' (1986), beschikbaar is. De vraagstelling mondt uit in een verklarend model dat in de analyse op klassenniveau door middel van LISREL getoetst wordt. Vooral uit deze modellen blijkt dat de onderzoekers de complexiteit van methodenvergelijkend onderzoek onderkennen. Op het niveau van de hoofdingeling (de clusters van variabelen) is het model volledig. Wel wordt het model niet helemaal consequent gepresenteerd in die zin dat de plaats van de 'Les'-factor in het eerste model niet gelijk is aan de plaats in het LISREL-model en dat het indirecte pad (via de CITO-resultaten) van de RION-toets in het analysemodel niet helemaal overeenkomt met de plaats die de extra toetsen in de conceptuele beschrijving hebben gekregen.

Hoewel met name twee aspecten in deze onderzoeksopzet niet aan bod komen (het longitudinale aspect en een gemeenschappelijke basis voor het vergelijkend reken- en het vergelijkend taalonderzoek door eenzelfde groep respondenten) is in grote lijnen het design geschikt voor het beantwoorden van de onderzoeksvraag. Op onderdelen is de uitwerking niet helemaal adequaat. Ten eerste treedt bij de variabelenkeuze een niet verantwoorde reductie op. Van de schoolfactoren worden alleen het aantal combinatieklassen en de onderwijsopvattingen van de leerkrachten opgenomen, van de leskenmerken alleen niveaudifferentiatie, leerstofvariantie, lessen per week en percentage leerlingen dat de basisstof doorkomt. Van de leerlingkenmerken alleen intelligentie. Deze keuzes worden niet toegelicht. Dat deze laatste variabele ook door de gebruikte rekendidactiek beïnvloed kan worden, is een methodologisch dilemma dat wel aangestipt maar niet onderzocht is.

Het tweede tevens grootste bezwaar geldt de gekozen operationalisering. Van de groep scholen van de implementatievariabelen is de operationalisering op een aantal onderdelen niet adequaat dan wel onduidelijk. Tussen methoden-inhoud en onderzoeksapparaat doet zich een Procrustes-situatie voor. Daarenboven is de operationalisering van het aantal combinatieklassen door te vragen naar het aantal leerkrachten minder dan zes, niet dekkend. Van het aantal lessen per week is de vraagvorm niet vermeld, leerstofvariantie, bedoeld als variatie tijdens de les, wordt gevraagd als variatie tijdens het cursusjaar. Bij 'behandelde leerstof' doet zich een inconsistentie voor tussen de items; met betrekking tot diezelfde variabele is het begrip 'basis-

stof variabel, afhankelijk van de methode in kwestie en is de validiteit van de differentiatielijst twijfelachtig. Deze belangrijke kritiek zorgt voor een grote scepsis over de met betrekking tot school- en implementatiekenmerken verkregen resultaten en de daarmee samenhangende verbanden.

De keuze van de onderzoeksmethoden, met name het hypothese-toetsende en het weddenshapsmodel, is een potentieel geëigende methode voor de onderhavige vraagstelling. Twee kanttekeningen dienen daar evenwel bij geplaatst te worden. Zoals ook bij de bespreking van het taalrapport is opgemerkt is een belangrijk methodisch punt het ontbreken van de klasse-observaties waardoor de validiteit van de implementatiemeting verhelderd had kunnen worden. De uitgevers kritiseren terecht deze ommissie. Daarnaast heeft het weddenshapsmodel zoals hier toegepast en geanalyseerd aanmerkelijke beperkingen: de inzet is vrijblijvend en de uitgevers beschikken over onvoldoende informatie over de uitkomsten van hun werk om gefundeerde voorspellingen te doen. Er is teveel van hen gevraagd. Deze overwegingen die nader uitgewerkt zijn in de beide adviezen leiden tot de conclusie dat aan de uitkomsten van het weddenshapsge-deelte niet al te veel belang moet worden gehecht, voor zover ze al echt nieuwe informatie opleveren.

Een laatste opmerking over de opzet betreft de steekproef. Bij de samenstelling van de steekproef wordt in het rapport niet ingegaan op de evenredigheid van de vertegenwoordiging van de methoden in vergelijking met hun marktaandeel en niet op de regionale spreiding van de scholen.

De methoden waarmee de onderzoeksdata zijn geanalyseerd zijn voor wat het hypothesetoetsende deel aangaat adequaat. Wel wordt een goed begrip van de resultaten bemoeilijkt door een onduidelijke terminologie met name bij de beschrijving van de implementatiefactoren. Het gelijkstellen van 'wijze van gebruik' met 'mate van implementatie' zorgt voor onnodige verwarring. Meer in het algemeen wordt inzicht in de analysemethode bemoeilijkt door te weinig toelichting op berekening en betekenis van de b-coëfficiënt in de pad-analyse (pagina's 64 en 76 zijn erg onduidelijk). Bij de conclusies moet worden opgemerkt dat weinig aandacht is besteed aan de beperkingen die uit dataverzameling en operationalisering voortvloeien en dat niet alle aanbevelingen (vergelijk de aanbeveling aan de leerkrachten over blok-uren) uit de voorliggende rapportage van resultaten onmiddellijk volgen.

De bruikbaarheid van de resultaten van het rekenonderzoek is vooral gelegen in het ontwikkelde instrumentarium van toetsen en in de attendering op de onmisbaarheid van uitvoeringsinstructies in de handleidingen.

Conclusie

De kritiek vervat in deze beoordeling heeft vooral betrekking op het ontbreken van een gedegen probleemanalyse, de onbeargumenteerde reductie van de leerlingkenmerken tot alleen de IQ-score, de zwakke operationalisering van de school- en implementatievariabelen en de zeer beperkte opbrengst en betekenis van het weddenshapsdeel. In vergelijking met het parallel-onderzoek 'Vergelijking taalmethoden in het basisonderwijs' bevat dit rapport meer reflectie op de problematiek en is de data-analyse meer gedetailleerd en geavanceerd.

LITERATUUR

- Jong, R.A. de: *Wiskobas in methoden*. Dissertatie. Vakgroep Onderwijs Wiskunde-onderwijs en Onderwijs-computercentrum, Utrecht, 1986.
Harskamp, E.G.: *Vergelijking van rekenmethoden in het basisonderwijs*, RION, Groningen, 1986.

Ontvangen 23-6-1987.

Notities en Commentaren

BEOORDELING*) VAN HET EINDRAPPORT "VERGELIJKING VAN TAALMETHODEN IN HET BASISONDERWIJS" UITGEVOERD DOOR HET RION (PROJECTLEIDER DRS. G. VAN DEN BERG), TOTAAL SUBSIDIEBEDRAG f 304.366,-.

Probleemstelling en opzet

Een van de kernpunten van het vernieuwingsbeleid in het basisonderwijs is de afstemming van het onderwijs op de individuele leerling.

Differentiatie is daartoe de methode. Deze doelstelling ligt vervat in de nieuw ontwikkelde onderwijsleerpakketten die opgezet zijn vanuit het principe basisstof-herhaalstof-verrijkingstof. Het voorliggende onderzoek vergelijkt oude en nieuwe taalpakketten op hun doelmatigheid, met als probleemstelling:

1. Wat zijn de verschillen tussen de onderwijsleerpakketten?
2. Wat is de kwaliteit van de pakketten gelet op de feitelijke ingebruikname en op de leerling-effecten?
3. Onder welke condities vindt feitelijke ingebruikname plaats?
4. Waardoor worden eventuele verschillen in effectiviteit bepaald?

Met betrekking tot de organisatorische vormgeving werden verschillen verwacht tussen de meer traditionele methoden, die meer cursorisch opgezet zijn, en de nieuwe methoden die meer thematisch opgebouwd zijn. De onderzoeksopzet voorzag – naast het klassieke hypothesetoetsend model in een toepassing van Hofstee's weddenschapmodel waarbij uitgevers geacht werden voorspellingen te doen over de taalprestaties van leerlingen die hun methoden gebruikten, in vergelijking met de door hen verwachte resultaten van concurrerende methoden.

Door rekening te houden met verschillen tussen leerlingen in aanvangssituatie (intelligentie en sociaal milieu), zijn de effecten tussen de verschillende methoden vergelijkbaar gemaakt. Als interveniërende variabele is in het design opgenomen de implementatievorm, waarbij het curriculum-zoals-bedoeld, wordt vergeleken met het curriculum-zoals-gerealiseerd. Bij de implementatievorm wordt de feitelijk bestede leertijd geregistreerd. De uitgevers worden, als deelnemers aan de weddenschap, ook betrokken bij de keuze van de toetsen waarmee de leerlingeffecten worden gemeten.

De respondentgroep voor de intelligentie- en effectgegevens bestaat uit leerlingen van de hoogste klas van de basisschool en daarnaast hun leerkrachten voor implementatiegegevens. Het streefgetal voor de steekproef was minstens 20 scholen per methode bij minstens zes methoden. Van de bij het onderzoek betrokken methoden zijn alle uitgevers uitgenodigd in de weddenschapsprocedure te participeren.

Samenvatting eindrapport

Onderzocht zijn acht taalmethoden, elk met een redelijk omvangrijk marktaandeel. Voor twee methoden geldt dat de leerlingen deze niet de gehele basisschoolperiode hebben gebruikt. Om de inhoudelijke verschillen te traceren is een inhoudsanalyse uitgevoerd op de dimensies cursorische of thematische opbouw, het differentiërend vermogen en de door de methode voorgeschreven instructievorm. De leerlingeffecten zijn geoperationaliseerd in toetsen die een zo breed mogelijk cognitief domein bestrijken. Gekozen is voor de CITO-eindtoets aangevuld met enige

* De samenvatting van de beoordeling van dit door SVO gesubsidieerde onderzoek wordt ingediend onder verantwoordelijkheid van het SVO-bestuur, onder redactie van drs. A.C.C. van den Oever, medewerker onderzoeksbeoordeling van het SVO-bureau.

nieuw, op instigatie van de uitgevers ontwikkelde toetsen. Voor de leerlingen bestond het instrumentarium verder uit attitudeschalen, een vragenlijst naar achtergrondgegevens en de ISI-intelligentietest. De leerkrachten is een vragenlijst voorgelegd over implementatiekarakteristieken en over achtergrondgegevens en enkele attitudeschalen.

Voor de voorspellingen van de uitgevers is in de praktijk een meer bijgestelde vorm van het weddenschapmodel gehanteerd dan oorspronkelijk in de bedoeling lag. In deze vorm deden de uitgevers alleen voorspellingen over de prestaties met betrekking tot hun eigen methode, een vergelijkende voorspelling is achterwege gebleven. Voor de taalmethoden met een relatief groot marktaandeel geldt dat het streven om 20 klassen per methode in het onderzoek op te nemen ongeveer bereikt is. Drie methoden zijn met beduidend minder klassen in het onderzoek vertegenwoordigd (6, 11 en 12).

De gevonden resultaten zijn in het onderstaande schema per variabele met een korte omschrijving samengevat.

Implementatie:	Behandelde stof:	Gemiddeld 80% van de voorgeschreven stof is aan het eind van het jaar behandeld. Geen verschil tussen methoden.
	Lestijd:	Tussen methoden nagenoeg gelijk. De traditionele vaardigheden nemen nog 71% van de totale taal-lestijd.
	Differentiatie:	Factoranalyse geeft niet de verwachte twee-dimensionale structuur aan met mastery-learning en het basisstof-herhaalstof-verrijkingstof principe als factoren. De schalen zijn onvoldoende betrouwbaar.
	Leerkracht:	De schaal is matig homogeen. Geen verschil tussen methodes.
Leerkracht:	Leerling-gerichtheid:	Schaal redelijk homogeen. Verschillen tussen methoden niet groot.
	Orde:	Bij het OSM-materiaal valt de relatief jonge leeftijd van de leerkrachten op. Zij zijn in verhouding ook het meest onervaren.
	Leeftijd:	Bij het OSM-materiaal valt de relatief jonge leeftijd van de leerkrachten op. Zij zijn in verhouding ook het meest onervaren.
Leerlingen:	Sociaal milieu:	Milieukenmerken van leerlingen die Taaltotaal gebruiken zijn relatief hoog, bij OSM relatief laag.
	Intelligentie:	Bij OSM-gebruikers gemiddeld lager.
	Leerproblemen:	Geen verschillen.

Het resultaat van het weddenschapmodel in afgezwakte vorm was dat de uitgevers de prestaties van hun eigen methode, gemeten met de CITO-toetsen, onderschatten terwijl ze de prestaties gemeten met de aanvullende RION-toetsen juist overschatten. In het algemeen zijn de voorspellingen van de uitgevers weinig nauwkeurig.

Op de klassieke hypothesetoetsende manier zijn de effecten van de verschillende methoden op de leerlingprestaties geanalyseerd. Multivariate toetsing geeft een significant effect te zien van de onafhankelijke variabele 'taalmethode' op de afhankelijke variabele 'leerprestatie'. Univariate toetsing van de effecten op de verschillende deelttoetsen wijst uit dat het effect zich beperkt tot de onderdelen 'kritisch lezen', 'synthese' en 'semantiek'. Als covariabele is de factor intelligentie in het design opgenomen, als enige indicator voor het beginniveau. Leerkrachtenmerken zijn niet als covariabelen opgenomen omdat ze tussen de diverse methoden geen verschil laten zien. De significante verschillen op de drie deelttoetsen zijn bijna uitsluitend te herleiden tot de slechte prestaties van klassen die werken met het OSM-materiaal. De conclusie wordt getrokken dat 'blijkt dat er, na correctie voor verschillen in intelligentie, geen of geringe verschillen zijn in toetsprestaties tussen methoden. De verschillen in effect zijn marginaal te noemen'.

Beoordeling

De beoordeling is gebaseerd op de adviezen van drs. J.C. van Bruggen, en van dr. C.A.J. Aarnoutse

Het beoordeelde onderzoek is een van de eerste pogingen tot methodenvergelijking gekoppeld aan productevaluatie. De kanttekeningen van kritische aard betreffen achtereenvolgens: het theoretisch kader, de opzet zoals die is uitgevoerd, de resultaten en de bruikbaarheid/relevantie voor de praktijk.

Ondanks de titel van een der hoofdstukken 'Theoretisch kader' wordt in dit rapport niet een beschrijving van de bestaande onderzoeksliteratuur geleverd en evenmin een gefundeerd model van variabelenrelaties opgesteld. Hierdoor hangen de onderzoeksvragen enigszins in de lucht en worden sleutelbegrippen niet exact omschreven. De vakdidactische analyse van de onderzochte taalmethoden is erg beknopt weergegeven, nauwelijks vergelijkend noch onderling noch met de bestaande literatuur op dit gebied.

De gekozen opzet is op zichzelf interessant door de combinatie van het klassieke hypothese toetsende model en een toepassing van Hofstee's weddenschapmodel. Bovendien is de opzet, globaal gezien, passend bij de gestelde vragen. Succesvolle toepassing van de genoemde modellen is echter sterk afhankelijk van enkele specifieke voorwaarden met betrekking tot selectie en instrumentatie. De selectie van de scholen is correct verlopen. De selectie van de taalmethoden is in zoverre discutabel dat de twee korter toegepaste taalmethoden beter uit de vergelijking verwijderd hadden kunnen worden. Zwaarder weegt de spanning die is gebleken tussen de vakdidactische inhoud van de taalmethoden en de onderzoeksinstrumenten met name met betrekking tot de leerlingprestaties.

Een aantal passages in het eindrapport getuigen hiervan. Op pagina 7 wordt geconcludeerd dat vanwege de grote complexiteit een echte analyse per deelvaardigheid niet gemaakt kan worden. Op pagina 25 moeten doelstellingen in CITO-items worden samengenomen om voldoende betrouwbaarheid te bereiken; op pagina's 27 en 28 blijkt spreekvaardigheid deel uit te kunnen maken van de evaluatie vanwege de lage correlatie van de betreffende toets met de andere toetsen en de intelligentietest.

De conclusie ten aanzien van de kwaliteit van de gebruikte instrumenten is als volgt. Ondanks het feit dat de onderzoekers serieuze pogingen hebben gedaan betrouwbare en valide instrumenten te ontwikkelen en ondanks de terechte en validiteit-bevorderende uitbreiding van het instrumentarium met specifieke toetsen, moet toch geconcludeerd worden dat de gerapporteerde betrouwbaarheid vrij laag is, dat de kans groot is dat einddoelen van methoden niet volledig door instrumenten gedekt worden en dat betwijfeld wordt of de instrumenten gevoelig genoeg zijn om de verschillen in prestaties in voldoende mate te registreren. Soortgelijke voorwaarden gelden het leerkrachtinstrument. Uit de literatuur is bekend dat een aanmerkelijke discrepantie kan bestaan tussen de reacties van leerkrachten op vragen van vragenlijsten en hun feitelijke gedrag. Het rapport besteedt aan deze kwestie, die van grote invloed kan zijn op de bevindingen ten aanzien van de implementatie-karakteristieken, geen aandacht. Een controle van de antwoorden met behulp van een aantal klasse-observaties zou in dit verband zeer verhelderend zijn geweest.

Binnen de bovengenoemde beperkingen is de hypothese-toetsende methode gedegen uitgevoerd. In het resultaten-hoofdstuk is de analyse van de relatieve doelmatigheid goed weergegeven. Ten onrechte zijn bij de conclusies weinig of geen restricties aangebracht die voortvloeien uit de beperkingen van de psychometrische kwaliteit. In het bijzonder de intelligentie-test-resultaten geldt dat niet bediscussieerd is in hoeverre zij samen met de SES- en de leerproblemen schalen een geschikte en voldoende indicator zijn voor het beginniveau van de leerlingen en dus in hoeverre de effectmetingen inderdaad onderling vergelijkbaar zijn wanneer het ISI-resultaat is uitgepartialiseerd.

De conclusies over de effecten die overwegend zijn samen te vatten onder de noemer 'methoden maken weinig verschil' missen in belangrijke mate zeggingskracht omdat geen onderscheid gemaakt kan worden tussen vakdidactische en vakinhoudelijke aspecten van de methode zelf enerzijds en verschil in implementatie door de leerkracht anderzijds. Deze beper-

king is op haar beurt het gevolg van de, op zichzelf terechte constatering, dat op basis van dit onderzoeksmateriaal geen uitspraak mogelijk is over de vraag of methoden wel worden gebruikt zoals ze bedoeld zijn.

Het weddenschapmodel van Hofstee is het tweede model waarop de opzet stoelt. Wat de voorspellingen van de uitgevers betreft kan men zich afvragen wat deze waard zijn aangezien de informatie waarover uitgevers beschikken door het ontbreken van een volledige evaluatie van de methoden niet boven het niveau van ervaren vakdidactici uitstijgt en aangezien hun voorspelling zonder gevolg in enigerlei vorm blijft waardoor de weddenschap een wat vrijblijvend karakter houdt. Deze conclusie wordt nog versterkt door de constatering dat uitgevers elkaars produkt niet zeer goed kennen waardoor van een toepassing van het weddenschapmodel in strikte zin geen sprake is.

De zeggingskracht van de uitkomsten van dit onderzoek staat of valt met de betrouwbaarheid en validiteit van de verzamelde gegevens. Uit bovenstaande opmerkingen kan worden afgeleid dat de zeggingskracht minder groot is. Onderscheiden naar groepen gebruikers, verdient het onderzoeksmateriaal zeker de aandacht van leergangontwikkelaars om daar verder mee te gaan. Voor hen zijn van belang de impliciete aanbevelingen om methoden op uitgebreide schaal uit te testen. Voor de uitgevers leveren de conclusies uit dit rapport weinig op wat ze niet al wisten. Wel is het gehanteerde instrumentarium geschikt om te gebruiken bij het op de markt brengen van nieuwe methoden. Door de summiere informatie over het gebruik van de methoden in de praktijk is het belang van de gerapporteerde bevindingen voor de schoolpraktijk niet erg groot.

Conclusie

De methodenvergelijking is niet volledig geslaagd te noemen. Deze conclusie wordt gebaseerd op het ontbreken van een theoretisch kader, de beperktheid van de opzet (vooral het achterwege blijven van klasse-observaties), de beperkte zeggingskracht van de gegevens uit het weddenschapmodel en de zwakke psychometrische status van het instrumentarium.

LITERATUUR

Batenburg, Th.A. van e.a.: *Vergelijking van taalmethoden in het basisonderwijs*, RION, Groningen, 1986.

Ontvangen 23-6-1987.

Notities en Commentaren

TAALMETHODEN ONDERZOCHT: REACTIE OP HET SVO-BEOORDELINGSRAPPORT

G. van den Berg en Th.A. van Batenburg

RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen

Behalve algemene waardering bevat de beoordeling van het eindrapport enkele punten van kritiek waarop wij hier slechts kort kunnen ingaan.

Allereerst willen we reageren op de kritische kanttekeningen bij het weddenschapmodel. In dit model staat commitment over de evaluatiecriteria en standaarden tussen de betrokkenen centraal. Wat betreft de criteria het volgende. Inhoudsanalyse biedt onvoldoende duidelijkheid over de leerdoelen die met de methoden worden nagestreefd. Toch is duidelijkheid hieromtrent nodig. Vandaar dat de ontwikkelaars is gevraagd het belang van bepaalde leerdoelen aan te geven. Deze procedure had tot doel een voor alle methoden zo inhoudsvalide mogelijk toetsinstrumentarium samen te stellen, waarvan 1. de inhoud in redelijke mate dekkend is voor de leerdoelen die elke methode nastreeft en 2. de relevantie door de belanghebbenden achteraf niet kan worden betwist. Vanuit het weddenschapmodel gezien is commitment een belangrijk, zo niet het belangrijkste, kenmerk voor de validiteit van het onderzoeksinstrumentarium. Uit het onderzoek blijkt dat de betrokkenen de leerlingtoetsen als effectcriteria aanvaarden en dat er een voldoende mate van commitment bestaat over de relevantie van de toetsen.

Wat betreft de evaluatiestandaarden het volgende. De doelmatigheid van onderwijsleerpakketten kan op verschillende manieren worden vastgesteld. In het taalonderzoek is dit op twee manieren gebeurd: (1) voldoen de onderwijsleerpakketten aan de verwachtingen van de auteurs zelf? (2) Is het ene pakket doelmatiger dan het andere (de zgn. relatieve doelmatigheid)? Ter bepaling van de doelmatigheid in de eerste betekenis zijn aan de auteurs voorspellingen gevraagd. Zij zijn immers te beschouwen als de deskundigen bij uitstek. Zij hebben de methoden ontwikkeld en weten derhalve het beste hoe hun methode moet worden geïmplementeerd en welke effecten mogen worden verwacht. De voorspellingen geven aan in hoeverre de eigen doelstellingen gerealiseerd worden. Het is echter hierbij onbekend in hoeverre persoonlijkheidsaspecten van de voorspeller een rol spelen. Voorzichtigheid is derhalve geboden bij het koppelen van voorspellingen van deskundigen aan kenmerken van de taalmethode. Ter bepaling van de doelmatigheid in de tweede betekenis is als standaard gekozen statistische significantie. Deze maat geeft aan met welke zekerheid een gevonden verschil al dan niet is toe te schrijven aan toevallige factoren.

Wat betreft de kritiek op de beknopte vakdidactische analyse, het ontbreken van een beschrijving van de bestaande onderzoeksliteratuur en het ontbreken van een duidelijk gefundeerd variabelenschema het volgende. Wij hebben bewust gekozen voor beknoptheid. Dit geldt voor de rapportage over de uitgevoerde vakdidactische analyses van de methoden. Over vakdidactische analyses is ook reeds gerapporteerd door anderen (Naber, e.a., 1983, 1986). Uitgangspunt bij de keuze van de variabelen was spaarzaamheid. Een empirische argumentatie van het variabelenschema is noodzakelijkerwijs beknopt, vanwege het ontbreken van onderzoeksliteratuur ter zake. Slechts de variabelen die vanuit de onderwijsevaluatie literatuur bekend zijn en die passen binnen het middel-doel-schema (zie bijv. Scheerens, 1983) hebben wij beargumenteerd opgenomen. Het betreft de variabelen 'het beginniveau van de leerlingen' en 'de implementatie van de methoden'. De empirische argumentatie voor reductie van beginniveau tot intelligentie is dezelfde als in het rekenonderzoek. Zie voor de argumentatie voor deze

reductie de reactie van Suhre op het beoordelingsrapport van het Rekenonderzoek in ditzelfde nummer. In het onderzoeksverslag is hierover echter niet gerapporteerd. We delen weliswaar de kritiek op het gebruik van vragenlijsten voor het meten van de variabele 'implementatie van de methoden', maar de beperkte randvoorwaarden stonden het niet toe aanvullende klasse-observaties (op zo'n 100 scholen) uit te voeren, ter controle van de antwoorden van de leerkrachten.

Tenslotte willen we ingaan op de psychometrische kwaliteit van de effectcriteria (de leerlingentoetsen). De betrouwbaarheid van de taaltoetsen van het CITO is redelijk tot goed te noemen ($\alpha = \pm .75$), voor de nieuw ontwikkelde toetsen luisteren ($\alpha = .64$) en kritisch lezen ($\alpha = .78$) eveneens. De CITO-doelstellingen zijn zo specifiek geformuleerd, dat samentrekking onder een (iets) bredere noemer dan voor de hand ligt. Zo is bijvoorbeeld het spellen van werkwoorden, het spellen van niet-werkwoorden en het gebruik van leestekens onder één noemer geplaatst (syntaxis). De items die onder deze noemer vallen zijn betrouwbaar te sommeren. Voor de toetsen stelvaardigheid en spreekvaardigheid zijn betrouwbare beoordelingsprocedures ontwikkeld. Bij alle in het onderzoek gebruikte toetsen bestonden voldoende onderlinge hoge correlaties om van validiteit te spreken. Slechts bij de toets spreekvaardigheid waren deze correlaties te laag; daarom is besloten om deze toets niet te gebruiken als effectcriterium, ondanks de relevantie van deze toets volgens de auteursgroepen van de verschillende methoden.

LITERATUUR

- Batenburg, Th.A. van, G. van den Berg en P. Edelenbos (1986): Vergelijking van taalmethoden in het Basisonderwijs (eindrapport deel I en II) Rion, Groningen.
- Naber, I., S. Nijhof en A. Schute (1983, 1986): Analyse van taalmethoden (deel I en II), KPC, 's-Hertogenbosch.
- Scheerens, J. (1983): Evaluatie-onderzoek en beleid; Methodologie en organisatorische aspecten. Den Haag, SVO-reeks 68.

Ontvangen 7-7-1987.

Notities en Commentaren

REKENMETHODEN: VERWACHT GEEN WONDEREN. REACTIE OP DE SVO-BEOORDELING VAN HET REKENONDERZOEK.

C. Suhre

RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit, Groningen.

De beoordeling van SVO bevat naast waardering een aantal punten van kritiek waarop hier wordt ingegaan.

Ten eerste wordt in de beoordeling gesteld dat een gedegen *probleemanalyse* voor vergelijking van methoden ontbreekt. De kritiek richt zich vooral op het onderscheid in traditionele en moderne rekenmethoden dat de onderzoekers hanteren. De kritiek gaat vergezeld van de suggestie dat een vergelijking tussen methoden sec niet is uitgevoerd. Deze kritiek is misplaatst. De in de subsidie-aanvraag beloofde vergelijking tussen de methoden sec (de vragen 2 en 3) is namelijk consequent uitgevoerd en beschreven in de onderzoeksverslagen. De bij de indeling in traditionele versus moderne methoden uitgevoerde analyses vormen een toegift op die methodenvergelijking. De indeling sluit aan bij vakdidactische ideeën met betrekking tot rekenmethoden (zie De Jong, 1986) en is derhalve legitiem. De geconstateerde verschillen in inhoud en didactiek geven voorts voldoende aanleiding het onderscheid traditioneel-modern te blijven gebruiken. Door ons zijn kleine inhoudsverschillen geconstateerd tussen rekenmethoden die vooral door uitgevers van moderne methoden als relevant worden ervaren. Dit blijkt uit hun vraag om extra toetsen op een aantal leerstofgebieden. Bovendien is in een afzonderlijk interimrapport (Wiersma en Willemsen, 1986) nader ingegaan op verschillen tussen beide groepen methoden in leerstofdifferentiatie en variatie in leerstof, vakdidactische kenmerken die het onderscheid traditioneel-modern typeren.

Een tweede punt van kritiek betreft de *reductie* van de leerlingkenmerken tot alleen de IQ-score. De reactie hierop kan kort zijn. Het analyse-ontwerp voorzag in toetsing van het effect van de factor methode door middel van variantie-analyses met IQ-score als controlevariabele. Voorafgaande aan dit ontwerp is nagegaan welke leerlingkenmerken onafhankelijk van de gebruikte rekenmethode van invloed zijn op de rekenprestaties. De IQ-score bleek de enige variabele die na uitpartialisatie van andere leerlingkenmerken, zoals een sociaal milieu indicatie (het beroep van de vader) en rekenattitude, sterke samenhang vertoonde met rekenprestaties. In het onderzoeksverslag is daarom volstaan met de IQ-score als controlevariabele en is verder niet ingegaan op andere (voor de methodenvergelijking minder relevante) leerlingkenmerken.

Een derde punt van kritiek betreft de zwakke *operationalisering* van school- en implementatievariabelen. In beginsel kan men het eens zijn met de kritiek dat er over de validiteit van de betreffende instrumenten beperkte empirische evidentie bestaat. Het onderzoek is gepland als survey-onderzoek omdat er slechts beperkte onderzoeksmiddelen beschikbaar waren (SVO heeft zelfs nog f 20.000,- op het oorspronkelijke budget gekort!). De meeste aandacht is besteed aan de onderzoeksprocedure, de onderhandeling met uitgevers met daaraan gekoppeld het ontwerpen van betrouwbare en valide rekentoetsen. Binnen de randvoorwaarden van het onderzoek is gestreefd naar een zo valide en betrouwbaar mogelijke leerkrachtenvragenlijst voor school- en implementatievariabelen. Deze zijn deels gemeten met behulp van feitelijke en objectief te beantwoorden vragen (aantal leerkrachten op een school en aantal rekenlessen per week). Deze vragen bezitten voldoende ('face') validiteit. Voor een ander deel zijn de variabelen gemeten door vragen waarbij ook de perceptie van respondenten (persoonsgebonden 'bias') een

rol zal spelen. Er is verder gebruik gemaakt van itemschalen voor leerstofdifferentiatie en voor de (les)attitude van leerkrachten. Beide schalen zijn overgenomen van bestaande instrumenten en bezitten in ons onderzoek dezelfde goede psychometrische kwaliteit met betrekking tot de interne consistentie. De samenhang van de twee schalen met leerprestaties in klassen is overeenkomstig eerder onderzoek. Dit lijkt enige indicatie voor validiteit te geven.

Een laatste belangrijk punt van kritiek betreft de beperkte *opbrengst van het weddenschapsmodel*. Met betrekking tot deze kritiek wordt verwezen naar de repliek van de onderzoekers Van den Berg en Van Batenburg in dit nummer op overeenkomstige kritiek aangaande het taalonderzoek.

LITERATUUR

- Jong, R.A. de (1986): *Wiskobas in methoden*, dissertatie, Vakgroep OW & OC, Utrecht.
Harskamp, E. en C. Suhre (1986): *Vergelijking van Rekenmethoden in het Basisonderwijs, eindrapport*. Groningen, RION.
Wiersma, Y. en T. Willemsen (1986): *Vergelijking van Rekenmethoden in het Basisonderwijs, interimrapport III*. Groningen, RION.

Ontvangen 7-7-1987.

ONTVANGEN PUBLIKATIES

Berg, G. van den. Effectief evalueren. Een empirische studie naar de doelmatigheid van aanwijzingen voor de evaluatiepraktijk (Dissertatie). Lisse, Swets & Zeitlinger, 1987 (Onderwijsonderzoek nr. 5).

Berg, R.M. van den & R. Vandenberghe. Strategies for large-scale change in education: dilemmas and solutions. Amersfoort, ACCO, 1986 (ISIP-Book).

Blok, H. Taal voor alledag. Feiten en meningen over het taalgebruik van LBO- en MAVO-leerlingen in alledaagse situaties. Den Haag, SVO, 1987 (Selectareeks).

Corte, E. de (Ed.). Onderwijsleerprocessen. Bevindingen van Leuvens onderwijspsychologisch onderzoek. Leuven, ACCO, 1986 (Leuvense Pedagogische Bijdragen).

Dochy, F.J.R.C. & M.L. de Volder. The use of indicators in a cyclical process for quality assurance as part of a new steering conception for higher education in the Netherlands. Maastricht, Rijksuniversiteit Limburg, 1987 (Onderzoek van Onderwijs nr. 34).

Everts, H., A. Golhof, P. Stassen & J. Teunissen. De cultureel-etnische situatie op OVB-scholen. Utrecht, Vakgroep Onderwijskunde, 1986.

Franssen, H.A.M., T.W.H.M. Stoverinck & D.A. van der Ploeg. Gericht geëvalueerd. Eindrapport SVO-project 57204. Utrecht, Vakgroep Onderwijskunde, 1987.

Glebbeek, A. & A. Jonker (Red.). Het omstreden leerplan. Sociologische beschouwingen rond de inhoud van het onderwijs. Assen, Van Gorcum, 1986.

Heyting, G.F. Autonomie en socialiteit in de opvoeding. Amersfoort, ACCO, 1987.

Juli, A. & L. van Zoelen. Een select gezelschap. Onderzoek naar de aanwijzing van project-scholen. Utrecht, ISOR, 1987.

Kallen, D.B.P. Arbeidsoriëntatie geëvalueerd. Rapportage van de externe evaluatie van de ontwikkelfase van het EG-project 'Arbeidsoriëntatie voor jongeren'. Rotterdam, Contactcentrum Onderwijs Arbeid Rijndelta, 1987.

Kemenade, J.A. van, N.A.J. Lagerweij, J.M.G. Leune & J.M.M. Ritzen (Red.). Onderwijs: bestel en beleid. Deel 2B: Onderwijs en samenleving. Groningen, Wolters-Noordhoff, 1987.

Kemenade, J.A. van, N.A.J. Lagerweij, J.M.G. Leune & J.M.M. Ritzen (Red.). Onderwijs: bestel en beleid. Deel 3: Onderwijs in ontwikkeling. Groningen, Wolters-Noordhoff, 1987.

Korsten, G. & P. Frissen. Actuele ontwikkelingen in het hoger onderwijsbeleid; trends, perspectieven en problemen. Maastricht, Rijksuniversiteit Limburg, 1987 (Onderzoek van Onderwijs nr. 33).

Leenders, F.J. & K.M. Stokking. Overdracht van vernieuwingen: een kwestie van perceptie? Verslag van een onderzoek met als voorbeeld de vernieuwing 'Zelfstandig Werken'. Utrecht, Vakgroep Onderwijskunde, 1987.

Riemersma, F.S.J. & M. Robijns. De registratie van leervorderingen en andere relevante leerlinggegevens. Deel I: Het systeem. SCO, Amsterdam, 1987.

Riemersma, F.S.J. & M. Robijns. De registratie van leervorderingen en andere relevante leerlinggegevens. Deel II: Schoolbeschrijvingen. SCO, Amstrdam, 1987.

Verstralen, H.H.F.M. Semantische modellen bij leren en probleemoplossen. Arnhem, 1987 (Dissertatie).

Vries, A.Ph. de. Onderzoekwijzer voor het landbouwonderwijs. Verslag van een probleemverkenning en aanzet voor de programmering van onderzoek en ontwikkelingswerk voor het agrarisch onderwijs. Wageningen, Landbouwuniversiteit, 1987 (Agrarische onderwijskundige reeks nr. 5).

Wetering, S. van de. Het onderwijs in eigen taal en cultuur aan marokkaanse kinderen in Nederland. Interim verslag 1: Houding en verwachtingen van enkele groepen betrokkenen. Utrecht, Vakgroep Onderwijskunde, 1986.

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie

Kees Beintema (RUL), Peter van den Dool (SCO), Dirk de Jong (LICOR), Ewoud Roede (SCO) en Peter van der Sijde (UT, TO, secretaris).

Redactie-adres

Peter van der Sijde, Toegepaste Onderwijskunde, OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 11, nr. 4

Augustus 1987

Inhoudsopgave

Congressen, Symposia, Cursussen, Trainingen	254
Mededelingen	255
Nieuwe publikaties	256

Congressen, Symposia, Cursussen, Trainingen

ACHTERGRONDEN VAN DE GROEI VAN HET SPECIAAL ONDERWIJS
LEIDEN, 30 OKTOBER 1987, STADSGEHOORZAAL

Ter gelegenheid van het 25-jarig bestaan van het Gemeentelijk Pedologisch Instituut te Leiden wordt over het genoemde onderwerp een symposium gehouden.

Uitgangspunt van het symposium is de bundel met studies die dit najaar zal verschijnen onder redactie van de hoogleraren K. Doornbos en L. Stevens. Hierin wordt dit onderwerp door deskundigen vanuit verschillende disciplines besproken (medisch, psychologisch, sociologisch, onderwijskundig, orthopedagogisch).

Op het symposium komen aan de orde:

- reacties op de probleemanalyses
- implicaties en consequenties richting wetenschap, beleid, onderwijs en maatschappij.

Inschrijven voor 1 september. Inlichtingen: Pedologisch Instituut Leiden, tel. 071-273379/273401.

ZEVENDE ONDERWIJSSOCIOLOGISCHE CONFERENTIE
LUNTEREN, 12 EN 13 NOVEMBER 1987, CONFERENTIE-OORD 'DE BLIJE WERELT'.

Het thema is: 'Onderwijs en sociale beheersing: maatschappelijke oorzaken en effecten van schoolkenmerken.

Plenaire inleidingen over de deelthema's worden o.m. verzorgd door: Prof. dr. J. Billiet (Katholieke Universiteit Leuven), Dr. P. Woods (Open University, Engeland), Dr. M. Stone (Surrey, Engeland), Prof. dr. S. Lindenberg (Sociologisch Instituut R.U. Groningen).

De conferentievoorzitter is Prof. dr. J.M.G. Leune (Erasmus Universiteit Rotterdam).

Inlichtingen: SISWO Amsterdam, 020-240075.

OTO-ONDERZOEK
TILBURG, 29 OKTOBER 1987, KATH. UNIVERSITEIT BRABANT

OTO is de afgekorte naam voor het onderzoek dat in het schooljaar 85/86 is gehouden naar de taken en de taakbelasting van leraren. Het is een breed opgezet onderzoek. Aan het onderzoek is meegewerkt door 15.000 leraren, 1500 scholen en vele beleidsmatig betrokken personen, instanties en organisaties.

Op de conferentie zullen de belangrijkste uitkomsten gepresenteerd, toegelicht en besproken worden, zowel door betrokken IVA-onderzoekers en hun adviseurs als door anderen. Deze laatsten moeten vanuit hun specifieke deskundigheden (onderzoek, beleid, schoolmanagement, etc.) de opzet, uitkomsten en conclusies van OTO evalueren en toetsen aan normen van wetenschappelijkheid en bruikbaarheid voor beleidsvorming.

Inlichtingen: IVA, Tilburg, tel. 013-662443

SYMPOSION EN INFORMATIEMARKT OVER COMMUNICATIE EN INFORMATIESYSTEMEN IN
GENT, 24 EN 25 SEPTEMBER 1987, RIJKSUNIVERSITEIT

Het *symposion* omvat drie inleidingen waarbij uit verschillende standpunten de relatie tussen onderwijs en bedrijfsleven wordt belicht. Die relatie wordt verder vooral bekeken in betrekking tot de vragen:

- a. op welke wijze speelt communicatie een rol, en
- b. welke invloed gaat er uit van informatie, kennis en kennissystemen die voor onderwijs en bedrijfsleven beide van belang zijn.

Na de inleidingen zal in werkgroepen van gedachten worden gewisseld over de ideeën,

vraagstellingen en oplossingen die de sprekers hebben voorgedragen. Er zal worden getracht de werkgroepen zodanig te vormen dat vertegenwoordigers van bepaalde sectoren van onderwijs samenkomen met vertegenwoordigers van daarmee overeenkomstige bedrijfstakken (techniek, chemie, grafische sector, handel, enz.).

De informatiemarkt zal zo goed mogelijk de verschillende aspecten van de relatie tussen onderwijs en arbeidsmarkt, zoals die op het symposion worden belicht, demonstreren en verduidelijken. De informatiemarkt is de plek waar ideeën kunnen worden besproken met vertegenwoordigers van bedrijfsleven en onderwijs. Er zal ook ruime aandacht worden geschonken aan databases waar onderwijs en bedrijfsleven beide hun voordeel mee kunnen doen.

Inlichtingen: SVO, Sweelinckplein 14, 2517 GK Den Haag.

Mededelingen

Beroepscode/Gedragsregels voor onderzoek: oproep aan leden

Aan het VOR Bestuur zijn problemen gerapporteerd in verband met de spanningsverhouding tussen de privacy van (dossier-)gegevens en hun belang voor onderzoek. In verband daarmee zal het Bestuur nadere gegevens (ook uit andere werkvelden) en gekantateerde werkwijzen verzamelen. Te zijner tijd wordt hierover gerapporteerd en zal een eventuele actie worden ondernomen. Leden die relevante ervaringen, procedures etc. kunnen melden worden verzocht hierover contact op te nemen met: mevr. B. van Hout Wolters (tel. 053-893656) of K. Beintema (tel.: 071-273399).

Nieuwe publicaties

- SWIDOC (1987). Periodiekenparade, mei 1987.
- OOMO (1987). Nieuwsbrief 1987 nr. 1.
- Schooltv gids voor het basisonderwijs (1987) nr. 1.
- DOZ-boek (1987). De taak van de docent in het nieuwe HBO. Utrecht: De SOM (030-310832, f 12,50).
- Claessen, J.F.M. (1987). Verstandig verdeeld? Deeltijd in het basisonderwijs. Heerlen: Open Universiteit (oratie).
- Scheerens, J. (1987). Het evaluerend vermogen van onderwijsorganisaties. Enschede: Universiteit Twente (oratie).
- RION (1987) Jaarverslag 1986.
- Juli, A., Van Zoelen, L. & Taes, C. (1987). Een select gezelschap, onderzoek naar de aanwijzing van projectscholen. Utrecht: ISOR, afd. Onderwijsonderzoek.
- Van de Wetering, S. (1986). Het onderwijs in eigen taal en cultuur aan Marokkaanse kinderen in Nederland. Interimverslag 1: Houding en verwachting van enkele groepen betrokkenen. Utrecht: Vakgroep Oosterse Talen en Culturen en Vakgroep Onderwijskunde.
- Leenders, F.J. & Stokking, K.M. (1987). Overdracht van vernieuwingen: een kwestie van perceptie? Verslag van een onderzoek met als voorbeeld de vernieuwing 'Zelfstandig werken'. Utrecht: Vakgroep Onderwijskunde.
- Dochy, F.J.R.C. & De Volder, M.L. (1987). The use of indicators in a cyclical process for quality assurance as part of a new steering conception for higher education in the Netherlands. Maastricht: R.L. (OvO 34).
- Korsten, G. & Frissen, P. (1987). Actuele ontwikkelingen in het hoger onderwijsbeleid; trends, perspectieven en problemen. Maastricht: R.L. (OvO 33).
- Van Kessel, N. (1987). Verlangd basisonderwijs in de gemeente Bessel. Nijmegen: ITS.
- Van Esch, W. & Roovers, M. (1987). Islamitisch godsdienstonderwijs in Nederland, België, Engeland en West-Duitsland. Nijmegen: ITS.
- COA (1987). Arbeidsoriëntatie, informatie voor scholen.
- APVO-2 (1987). Modulering in Opmars. Zeist: APVO-2 (tel.: 03404-50334).
- APVO-2 (1987). Modulieren funderen. Zeist: APVO-2.
- Van Luit, J.E.H. (1987). Rekenproblemen in het speciaal onderwijs. Nijmegen: Instituut voor Orthopedagogiek (dissertatie, f 21,50).
- Van Kessel, N. (1987). Van meester naar professional. Samenvatting. Nijmegen: ITS.
- Van Kessel, N. & Petersen, B. (1987). Verlangd basisonderwijs in de gemeente Roermond. Nijmegen: ITS.
- Hövels, B. & Verijdt, H. (1987). Naar een versterking van het leerlingwezen. Nijmegen: ITS.
- Terlouw, C. (1987). De FUNDES-procedures in onderwijsontwikkeling. Enschede: Onderwijskundig Centrum (dissertatie).

Een Verklaringsmodel voor Onderwijsproblemen van Promotie-assistenten

J.F.M.J. van Hout en J.J.G. Schmeets *

Katholieke Universiteit Nijmegen.

ABSTRACT

A survey of 166 research assistants was carried out at six Dutch universities. The purpose of the survey was to describe the situation of the research assistants and to trace the influence of the working environment on their attitudes and behaviour. This article focuses on the concerns and the stress associated with their teaching assignment. In a model, analysed with LISREL, it is demonstrated that the teaching concerns and stress are influenced by age, experience in teaching, motives for taking the job as research assistant and by the orientation towards teaching and research. In addition, concerns and stress resulting from the research assignment contribute to the teaching problems. It is also shown that the disciplines differ in concerns and stress related to the teaching assignment: research assistants in the natural sciences have less concerns and stress compared to research assistants in the social sciences and the humanities.

1. INLEIDING

In een landelijk onderzoek over promotie-assistenten (Van Hout, 1984), ook doorstromers genoemd, werd geconstateerd dat een aanzienlijk deel van hen zich zorgen maakt over de onderwijstaak. Van hen gaf 40% aan dat ze zich vaak, 56% soms en 4% (vrijwel) nooit zorgen maken over een of meer onderwijsaspecten. Tevens bleek dat er grote verschillen bestonden tussen promotie-assistenten van wis- en natuurkunde, sociale wetenschappen en letteren, zowel in aard als intensiteit van de onderwijsproblemen.

Onder een promotie-assistent verstaan we een wetenschappelijk medewerker die een tijdelijke aanstelling heeft om te promoveren en tevens belast is met een onderwijstaak. Deze personeels-categorie lijkt dus zeer veel op die van de assistent-in-opleiding, de zogenaamde AiO, maar is ook te karakteriseren als beginnend universiteitsdocent. De onderwijsproblemen van de promotie-assistenten zijn relevant met het oog op de kwaliteit van het door hen en door AiO's te verzorgen onderwijs en kunnen informatie verschaffen voor beleid ten aanzien van begeleiding en taakstelling van AiO's. Daarnaast kan kennis van de oorzaken van de onderwijsproblemen van promotie-assistenten het inzicht in de kenmerken van het beroep van universiteitsdocent vergroten en informatie verschaffen voor het opzetten van professionaliseringsprogramma's van universiteitsdocenten.

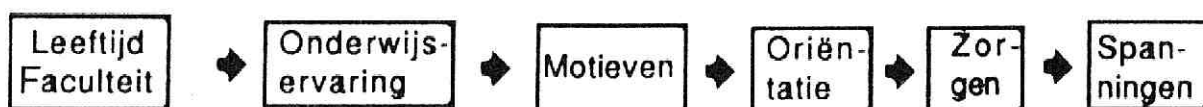
Met dit artikel wordt beoogd een eerste aanzet te geven tot de ontwikkeling van een verklaringsmodel voor onderwijsproblemen van promotie-assistenten. Voor de ontwikkeling van dit model is aansluiting gezocht bij literatuur over onderwijsproblemen van beginnende en ervaren docenten in het basis- en voortgezet onderwijs (paragraaf 2), alsmede bij literatuur over

* J.F.M.J. van Hout, Instituut voor Onderzoek van het Wetenschappelijk Onderwijs, Katholieke Universiteit Nijmegen, Montessorilaan 3, postbus 9104, 6500 HE Nijmegen.

J.J.G. Schmeets, Subfaculteit Maatschappijwetenschappen, vakgroep Methoden en Technieken, Katholieke Universiteit Nijmegen, Thomas van Aquinostraat 4, 6500 HK Nijmegen.

onderwijsproblemen van universiteitsdocenten (paragraaf 3). Tevens is het model gebaseerd op onderzoeksliteratuur over kenmerken van PhD-studenten in de VS en Engeland (paragraaf 4). Het causale model ziet er in de meest eenvoudige vorm als volgt uit (zie figuur 1). Leeftijd en de faculteit vormen de onafhankelijke variabelen. Deze variabelen beïnvloeden de onderwijservaring. De onderwijservaring heeft effect op de motieven om promotie-assistent te worden. Deze motieven beïnvloeden de oriëntaties van de promotie-assistent op de te verrichten onderwijs- en onderzoekstaken, die op hun beurt de zorgen veroorzaken. Deze zorgen zullen uiteindelijk spanningen teweeg brengen.

In paragraaf 6 wordt dit model verder uitgewerkt, nadat in paragraaf 5 de opzet van het onderzoek en de operationalisatie van de variabelen zijn uiteengezet. Na een beschrijving van de resultaten per faculteit (paragraaf 7), toetsen we het nader gespecificeerde model met behulp van LISREL (Jöreskog & Sörbom, 1985) in paragraaf 8. Paragraaf 9, tenslotte, wijden we aan discussie.



Figuur 1: Vereenvoudigde vorm van het verklaringsmodel voor onderwijsproblemen van promotie-assistenten.

2. ONDERWIJSPROBLEMEN VAN BEGINNENDE DOCENTEN IN HET BASIS- EN VOORTGEZET ONDERWIJS

Veenman (1984) analyseert 83 onderzoeken over 'perceived problems' van docenten in het basis- en voortgezet onderwijs tijdens hun eerste jaar als docent. Hij vindt een zeer constante serie problemen, waarvan orde houden, motiveren van leerlingen en het hanteren van individuele verschillen de drie top-problemen zijn. Hij onderscheidt drie benaderingen om deze problemen van docenten te verklaren: 1. 'developmental stages of concerns' van Fuller; 2. 'cognitive developmental framework' van Glassberg en Sprinthall; 3. 'teacher socialization framework' van Gehrke en Tabachnik. In al deze benaderingen doorloopt de docent fasen in denken, gedrag en waardepatronen; bepaalde patronen worden in een bepaalde ontwikkelingsfase ervaren. In deze ontwikkelingsmodellen neemt onderwijservaring een centrale plaats in. Volgens Lortie (1975) moet naast de ontwikkeling van de individuele docent tevens rekening gehouden worden met kenmerken van het beroep als docent. De kennis en vaardigheden zijn in het onderwijzersberoep weinig gesystematiseerd. Dit leidt ertoe dat deze vaardigheden pas in de praktijk goed geleerd worden. De mate van onderwijservaring blijkt dus niet alleen op grond van ontwikkelingsmodellen, maar ook op grond van kenmerken van het beroep, een belangrijke factor ter verklaring van onderwijsproblemen van beginnende docenten.

Bergen, Gerris & Peters (1987) concluderen op basis van onderzoek naar onderwijsproblemen van (beginnende) docenten 'dat zowel de ervaring als de leeftijd van docenten een belangrijke rol spelen bij het soort en het aantal problemen, die docenten tijdens hun beroepsuitoefening ervaren' (o.c., p. 143). Naarmate de ervaring en de leeftijd toeneemt, neemt het aantal ervaren probleemsituaties van het micro-taakveld af en van het meso-taakveld toe. Bij zeer ervaren en oudere leerkrachten is er daarentegen weer een toename van probleemsituaties op het micro-taakveld en een afname van probleemsituaties op het meso-taakveld.

3. ONDERWIJSPROBLEMEN VAN (BEGINNENDE) UNIVERSITEITSDOCENTEN

In tegenstelling tot het basis- en voortgezet onderwijs is relatief weinig onderzoek verricht naar onderwijsproblemen van zowel beginnende als ervaren universiteitsdocenten. De resultaten van studies van King (1973), Bruhns & Thomsen (1979), Fox (1980), Fink (1983, 1984, 1985) kunnen we als volgt samenvatten:

1. De zorgen die beginnende universiteitsdocenten zich vooraf maken hebben vooral betrekking op de organisatie van de onderwijsinhoud, de uitdrukkingsvaardigheden, het motiveren en activeren van studenten tijdens het onderwijs, en het voldoende boven de stof staan (Bruhns & Thomsen, 1979; Fox, 1980). Afgezien van het motiveren van studenten is er geen overeenkomst met bovengenoemde drie meest voorkomende problemen van beginnende docenten in het basis- en voortgezet onderwijs.
2. In navolging van Fuller (1969) onderscheiden Bruhns & Thomsen (1979) 'ego-centered', 'teaching-centered' en 'student-centered concerns' bij beginnende universiteitsdocenten. De beginnende universiteitsdocenten lijken zich vooral 'ego-centered' zorgen te maken, in tegenstelling tot universiteitsdocenten met ervaring die zich meer 'teaching-centered' zorgen maken (Bruhns & Thomsen, 1979). Fink (1985) toont aan dat beginnende universiteitsdocenten zichzelf vooral zien als vakinhoudelijke experts en van mening zijn dat het hun taak is 'to teach what they know' en niet 'to teach what they are'.
3. De zorgen die men zich maakt worden na het eerste jaar minder, maar het blijft moeilijk de eigen kwaliteit in te schatten, genoeg tijd te vinden voor voorbereiding en nieuwe ideeën over het onderwijs te ontwikkelen (Fox, 1980).
4. Beginnende universiteitsdocenten hebben vaak wel enige onderwijservaring, onder meer opgedaan als student-assistent. Deze ervaring heeft waarschijnlijk invloed op de door hen ervaren onderwijsproblemen (Bruhns & Thomsen, 1979; Fink 1983; King, 1973).
5. Kenmerken van de werkomgeving en de aard van de onderzoekstaak hebben invloed op de wijze waarop men de onderwijstaak aanpakt en de problemen die men ervaart (Fink, 1984, 1985).

Uit deze literatuur blijkt dat inhoudelijke problemen van beginnende universiteitsdocenten mogelijk niet overeenkomen met die van docenten in het basis- en voortgezet onderwijs, maar dat de bovengenoemde ontwikkelingsmodellen wel van belang kunnen zijn ter verklaring van onderwijsproblemen van universiteitsdocenten.

Over de andere verklaringsbron, het beroep van universiteitsdocent, is wel in ruimere mate literatuur voor handen, zij het niet over de Nederlandse situatie. We beperken ons tot de volgende aspecten: de relatie onderwijs-onderzoek, de spanningen bij onderwijs- en onderzoekstaken en het verschil tussen disciplines.

De relatie tussen de onderwijs- en onderzoekstaak van universiteitsdocenten is een complexe. Onderzoek naar de relatie tussen onderzoeksproductiviteit en de kwaliteit waarmee de onderwijstaak wordt vervuld heeft slechts enkele duidelijke conclusies opgeleverd, aldus Finkelstein (1984). Op de eerste plaats moet de hypothese dat er geen relatie tussen beide componenten zou bestaan verworpen worden. Ten tweede leiden de tijd en energie die de onderzoeksinspanningen vragen niet als zodanig tot een slechte vervulling van de onderwijstaak. Het verrichten van goed onderzoek vraagt weliswaar om extra tijd, maar het gaat ten koste van de beschikbare vrije tijd en niet van de tijd voor de onderwijstaak (Yuker, 1984). Goede onderzoekers blijken over het algemeen goed te scoren op de intellectuele aspecten van de onderwijstaak. Geen relatie wordt gevonden tussen de onderzoekskwaliteiten en de sociaal-emotionele aspecten van de onderwijstaak. Er is dus eerder sprake van invloed van de onderzoekstaak op de onderwijstaak, dan omgekeerd. De onderzoeksproductiviteit en de kwaliteit als docent worden, eveneens volgens Finkelstein, bepaald door de oriëntatie van de docent op onderzoek danwel op onderwijs, in combinatie met de onderzoeks- of onderwijsgerichtheid van de instelling of vakgroep waarin men werkt. Carrière binnen het universitaire systeem maken zij die reeds in het begin van hun loopbaan publiceren, aan de belangrijkste graduate-schools afstuderen en als onderzoekers worden aangetrokken bij een top-universiteit,

waar zij veel tijd aan onderzoek en weinig tijd aan onderwijs mogen besteden (Creswell, 1985; Finkelstein, 1984; Wanner, Lewis & Gregorio, 1981). Bij deze stafleden versterken de individuele onderzoeksoriëntatie en die van de omgeving elkaar en dat leidt tot succes. Zo zou het ook ten aanzien van onderwijs het geval kunnen zijn: de individuele onderwijsoriëntatie leidt in interactie met een sterke onderzoeksoriëntatie van de omgeving tot een succesvol docentschap.

De literatuur is niet eenduidig over spanningen bij de universitaire onderwijs- en onderzoekstaken. Startup (1985) meldt, dat het Engelse universitaire wetenschappelijk personeel de meeste satisfactie ontleent aan de onderzoekstaken. Volgens Gmelch, Lovrich & Wilke (1984) veroorzaakt onderwijs in het Amerikaanse universitaire systeem de meeste stress, terwijl volgens Perlberg & Keinan (1986) in Israël de onderzoekstaak de meeste stress veroorzaakt. Uit beide laatste onderzoeken blijkt dat er tussen de wetenschappelijke disciplines geen verschillen in onderwijs- en onderzoeksstress bestaan. In een onderzoek van Eble & McKeachie (1985) zegt 66% van de stafleden te veel te doen te hebben en 54% conflicten te hebben over onderwijs- en onderzoeksverantwoordelijkheden. De meeste stress wordt gerapporteerd door de jongere staf (assistant professors), hetgeen overeenkomt met de bevindingen van Perlberg & Keinan (1986).

Dat er geen verschil in spanningen tussen de stafleden van verschillende disciplines wordt aangetroffen is opmerkelijk, omdat de verschillen tussen disciplines een saillant kenmerk van het universitaire systeem en van het beroep van universiteitsdocent is (Clark, 1983). Deze verschillen hebben onder meer betrekking op de sterkte van de onderwijs- en onderzoeksoriëntatie, het publicatiegedrag en de student-docent interacties. Docenten van wis- en natuurkunde zijn meer onderzoeksgeoriënteerd, publiceren meer artikelen en zijn in hun contacten met studenten meer taakgeoriënteerd dan docenten van sociale wetenschappen en letteren (o.a. Becher, 1981; Finkelstein, 1984).

Verklaringen voor de variatie in het voorkomen van onderwijsproblemen bij (beginnende) universiteitsdocenten zouden dus gezocht kunnen worden in leeftijd en de mate van onderwijservaring op grond van eerder genoemde ontwikkelingsmodellen (o.a. Fuller, 1969) en in kenmerken van het beroep van universiteitsdocent, zoals onderwijs- en onderzoeksoriëntatie, de invloed van de onderzoekstaak op de onderwijstaak en de verschillen tussen de disciplines.

4. KENMERKEN VAN PROMOTIE-ASSISTENTEN

De hoofdactiviteit van de promotie-assistent is het werken aan een promotie-onderzoek. De afronding daarvan in een dissertatie en de kwaliteit van de dissertatie worden geacht een sterke invloed op de carrière te hebben. De onderwijstaak is in zekere zin een nevenactiviteit.

Men mag de promotie-assistenten niet zonder meer zien als een homogene groep. De motieven waarom men wil promoveren c.q. de doelstellingen die men met een functie als promotie-assistent nastreeft kunnen sterk variëren. Rudd (1975 en 1985), Walford (1980) en Winfield (1987) vinden in onderzoek bij Engelse PhD-studenten vier groepen motieven voor deze studie: 'the academic motive, the vocational motive, inertia and other reasons'. De PhD-studenten met het 'academic' motief worden door Rudd de 'dedicated scholars' genoemd; zij bestaan uit twee subgroepen, namelijk een groep die in een bepaald onderzoeksgebied geïnteresseerd is en daarin onderzoek wil doen en een groep die universiteitsdocent wil worden. Studenten met een 'academic' motief komen volgens Rudd meer voor bij sociale wetenschappen en letteren, dan bij wis- en natuurkunde. De studenten die door Rudd als 'vocational' worden aangeduid zijn op carrière (promotie en salaris) gericht, meestal buiten de universiteit. De 'inertia' groep, ook wel 'drifters' genoemd, wil het liefst de situatie als student continueren; zij heeft geen duidelijke doelstelling die verder gaat dan de PhD-opleiding zelf.

Ook in de Amerikaanse literatuur naar motieven voor de PhD-opleiding worden soortgelijke motieven genoemd. Zo stelt Baird (1976) dat Amerikaanse PhD-studenten verschillen in de mate van beroepsgerichtheid, betrokkenheid bij de 'academic way of life' en interesse in een

bepaald vak. Verwacht mag worden dat verschillen in motivatie voor de PhD-studie onder meer leiden tot verschillen in waardering van de te verrichten taken.

Ook de leeftijd van de promotie-assistenten is variabel; deze hangt samen met de voorafgaande studiecariëre en beroepsloopbaan. Buis (1983) vond een verschil in leeftijd bij promovendi; die van wis- en natuurkunde beginnen op jongere leeftijd aan het promotie-onderzoek dan die van sociale wetenschappen. Dergelijke verschillen worden ook bij PhD-studenten aangetroffen, onder andere door Winfield (1987).

Verschillen in discipline-gerelateerde aspecten die bij de universiteitsdocenten worden gevonden, zoals onderwijsoriëntatie, mogen ook bij promotie-assistenten worden verwacht op basis van mechanismen van anticiperende socialisatie en selectie (Bess, 1978; Corcoran & Clark, 1984) en/of op basis van de structurerende invloed die van het paradigma van het wetenschapsgebied zelf uitgaat (Van Hout, 1986a; Lodahl & Gordon, 1972; Whitley, 1984).

Tenslotte mogen we verwachten dat promotie-assistenten verschillen in de hoeveelheid onderwijservaring, opgedaan als student of als afgestudeerde.

5. OPZET VAN HET ONDERZOEK EN OPERATIONALISATIE VAN DE VARIABLEN

Het hier beschreven onderzoek naar onderwijsproblemen van promotie-assistenten maakt deel uit van een omvangrijker survey-onderzoek naar motieven, beleving en gedrag van promotie-assistenten, aangesteld bij de faculteit wiskunde en natuurwetenschappen (WN), sociale wetenschappen (SW) en letteren (LE) van zes Nederlandse universiteiten. Het onderzoek is in 1983 uitgevoerd. Het aantal respondenten bedroeg 166 (response-percentages 64%). In de vragenlijst zijn vragen opgenomen over entreekenmerken van promotie-assistenten (o.a. leeftijd, vooropleiding, onderwijservaring en motieven), over kenmerken van de werkomgeving (o.a. de faculteit, de onderwijs- en onderzoeksoriëntatie van de vakgroep) en over gedrag en attitudes van de promotie-assistenten (o.a. onderwijsoriëntatie, onderwijs- en onderzoeksproblemen). In het onderzoek wordt nagegaan in welke mate gedrag en attitudes worden beïnvloed door de kenmerken van de werkomgeving en de entreekenmerken. Op basis van deze gegevens zijn diverse meetmodellen geconstrueerd. Over dit onderzoek is gerapporteerd in Van Hout (1984, 1985, 1986b) en Van Hout en Schmeets (1985).

Volgens de bovenbeschreven literatuur zouden de onderwijsproblemen van promotie-assistenten verklaard kunnen worden uit enkele entreekenmerken (nl. leeftijd, onderwijservaring en motieven voor het promotie-assistentschap), uit de werkomgeving (o.a. de faculteit) en uit attitudes en gedrag (m.n. de oriëntatie op de onderwijs- en onderzoekstaak en de problemen met de onderzoekstaak). In deze paragraaf beschrijven wij de wijze waarop wij deze variabelen hebben geoperationaliseerd.

De variabele 'leeftijd' is uitgedrukt in de leeftijd in maanden. De variabele 'onderwijservaring' is een somscore van antwoorden (wel/niet) op vragen naar didactische scholing, onderwijservaring als studentassistent en onderwijservaring als afgestudeerde.

Ten aanzien van 'de motieven voor het promotie-assistentschap' zijn drie factoren verkregen met behulp van een factoranalyse over een aantal vragen met betrekking tot aspecten van motieven om te promoveren en om het promotie-assistentschap te ambiëren. De eerste factor is een op onderwijs en maatschappij gericht motief: door ons korthedshalve 'onderwijsmotief' (OW-MOT) genoemd. De tweede factor is een op carrière gericht motief, 'carrière-motief' (CAR-MOT) genoemd. De derde factor is een op de universitaire wetenschap gericht motief, 'wetenschapsmotief' (WET-MOT) genoemd (zie tabel 1). Het in paragraaf 4 vermelde 'inertiämotief' komt in ons onderzoek niet voor.

De variabele 'faculteit' bestaat uit drie categorieën, namelijk wis- en natuurkunde (WN), sociale wetenschappen (SW) en letteren (LE). In het in de volgende paragraaf beschreven verklaringsmodel wordt deze variabele ingevoerd als twee dummy-variabelen, namelijk a) wel of niet werkzaam bij de faculteit wis- en natuurkunde en b) wel of niet werkzaam bij de faculteit sociale wetenschappen.

Tabel 1: Factorladingen en communaliteiten (h^2) van de varimax geroteerde principale componentenanalyse op de vragen betreffende het onderwijsmotief (OW-MOT), het carrièremotief (CAR-MOT) en het wetenschapsmotief (WET-MOT).

	OW-MOT	CAR-MOT	WET-MOT	h^2
- ik heb functie als promotie-assistent gekozen omdat onderwijs mij aantrok	.76	-.06	.18	.61
- ik voel me onderzoeker en docent of primair docent	.74	-.27	.01	.61
- ik ben geselecteerd op grond van ervaring met onderwijstaak	.69	.14	-.09	.50
- ik wil promoveren vanwege maatschappelijke relevantie	.57	.27	-.15	.42
- na m'n promotie wil ik onderzoeker worden in het bedrijfsleven	-.52	.20	-.30	.40
- stel dat U vooraf zou weten dat U het niet verder zou brengen dan wetenschappelijk hoofdmedewerker, wat zou U er van vinden	.02	.85	-.08	.74
- stel dat U vooraf zou weten dat U het niet verder zou brengen dan wetenschappelijk medewerker, wat zou U er van vinden	-.09	.88	-.11	.80
- ik heb functie als promotie-assistent gekozen omdat ik graag wilde promoveren	.01	.49	.40	.40
- ik wil promoveren vanwege belang functie van wetenschappelijk medewerker	.15	.03	.70	.51
- na promotie wil ik wetenschappelijk medewerker worden aan de universiteit	.13	.02	.74	.57
- ik heb functie als promotie-assistent gekozen omdat ik mensen kende die er werken	-.28	.02	.41	.25
- ik wil promoveren omdat het de wetenschap verder brengt	-.04	-.12	.45	.22
% verklaarde variantie	20.6	16.0	13.7	

Met betrekking tot 'onderwijsoriëntatie' en 'onderzoeksooriëntatie' zijn een aantal vragen gesteld over de waardering van de onderwijs- en onderzoekstaak. De twee factoren die uit de factoranalyse te voorschijn komen (zie tabel 2), zijn benoemd als 'onderzoeksooriëntatie' (OZ-OR) en 'onderwijsoriëntatie' (OW-OR).

We onderscheiden twee soorten onderwijsproblemen, namelijk 'onderwijssorgen' en 'onderwijsspanningen', die corresponderen respectievelijk met de 'concerns' van Fuller (1969) en 'stress'. De variabele 'onderwijssorgen' is een één-factoroplossing (tabel 3) over een aantal vragen naar zowel 'teaching-centered', 'ego-centered' en 'student-centered' onderwijssorgen. De factor noemen we 'onderwijssorgen' (OW-ZORG).

Tabel 2: Factorladingen en communaliteiten (h^2) van de varimax geroteerde principale componentenanalyse op de vragen betreffende de onderwijsoriëntatie (OW-OR) en onderzoeksooriëntatie (OZ-OR).

	OZ-OR	OW-OR	h^2
- ik werk met plezier aan mijn onderzoek	.79	.10	.64
- ik vind van mijzelf dat ik een goed onderzoeker ben	.80	-.09	.65
- ik vind het belangrijk een goed onderzoeker te zijn	.61	.02	.38
- ik ben tevreden over mijn functioneren als onderzoeker	.71	.13	.52
- ik zou blij zijn als ik mijn onderzoekstaak zou kunnen afstoten	-.72	.04	.52
- mijn onderwijstaken verricht ik met plezier	-.16	.83	.72
- ik vind van mijzelf dat ik een goed docent ben	.11	.73	.54
- ik vind het belangrijk een goed docent te zijn	.13	.69	.49
- ik ben tevreden over mijn functioneren als docent	.17	.60	.39
- ik zou blij zijn als ik mijn onderwijstaak zou kunnen afstoten	.21	-.69	.53
% verklaarde variantie	28.1	25.6	

Ook de vragen die betrekking hadden op onderwijsstress omvatten één factor (tabel 4). Deze factor noemen we "onderwijsspanningen" (OW-SPAN).

Ook de onderzoeksproblemen zijn in twee soorten onderscheiden, namelijk 'onderzoekszorgen, en 'onderzoeksspanningen'. De beide verkregen één-factoroplossingen (zie tabel 5 en 6) noemen we 'onderzoekszorgen' (OZ-ZORG) en 'onderzoeksspanningen' (OZ-SPAN).

Tabel 3: Factorladingen en communaliteiten (h^2) van de principale componentenanalyse over uitspraken ten aanzien van onderwijszorgen (OW-ZORG)

	OW-ZORG	h^2
Ten aanzien van mijn onderwijs maak ik me zorgen of ik . . .		
- de studenten voldoende weet te motiveren	.41	.17
- een goede verstandhouding met de studenten opbouw	.41	.17
- helder en consistent uitleg	.51	.26
- op een adequate wijze door studenten gemaakte fouten bespreek	.51	.26
- de leerstof systematisch genoeg orden	.67	.45
- mijn enthousiasme en interesse overbreng	.41	.17
- studenten echt inzicht bijbreng	.63	.40
- gebruik maak van de goede onderwijsmethoden	.55	.31
- de lessen wel goed heb ingedeeld en gepland	.67	.45
- studenten voldoende aan het woord laat komen	.56	.31
- een adequaat tempo hanteer	.46	.22
- tijdens een discussie de draad verlies	.46	.22
- voldoende open sta voor kritiek van studenten	.62	.38
- voldoende zicht heb op mijn eigen onderwijskwaliteiten	.60	.36
% verklaarde variantie	29.8	

Tabel 4: Factorladingen en communaliteiten (h^2) van de principale componentenanalyse over uitspraken ten aanzien van onderwijsspanningen (OW-SPAN).

	OW-SPAN	h^2
- tijdens het onderwijs zeer nerveus zijn	.74	.54
- vlak voor het onderwijs zeer gespannen zijn	.71	.56
- uit onzekerheid zeer veel tijd aan voorbereiding van iedere bijeenkomst besteden	.69	.47
- na het onderwijs doodop zijn	.44	.20
- steeds het gevoel hebben dat ik mij 'waar moet maken' bij studenten	.77	.60
- overdreven strak vasthouden aan mijn planning van de bijeenkomst	.49	.24
- bang zijn dat ik de situatie niet onder controle kan houden	.57	.33
- bang zijn voor het oordeel van andere docenten	.71	.50
% verklaarde variantie	42.2	

Tabel 5: Factorladingen en communaliteiten (h^2) van de principale componentenanalyse over uitspraken ten aanzien van onderzoekszorgen (OZ-ZORG)

	OZ-ZORG	h^2
Ten aanzien van mijn onderzoek maak ik me zorgen . . .		
- over de maatschappelijke relevantie	.49	.24
- over de adequaatheid van de onderzoeksopzet	.62	.38
- of ik de verwachtingen die men van mij heeft wel kan waarmaken	.54	.29
- of het onderzoek wel voldoende wetenschappelijk niveau heeft	.79	.62
- over de wetenschappelijke relevantie	.73	.53
% verklaarde variantie	41.0	

Tabel 6: Factorladingen en communaliteiten (h^2) van de principale componentenanalyse over uitspraken ten aanzien van onderzoeksspanningen (OZ-SPAN).

	OZ-SPAN	h^2
- ik voel me sterk onder druk staan (b.v. vanwege strakke planning)	.70	.49
- ik ben vaak erg moe	.67	.45
- ik slaap vaak slecht omdat ik over mijn onderzoek lig te piekeren	.53	.28
- kritiek op mijn onderzoek maakt me in de war	.61	.37
- ik heb het gevoel dat ik steeds op mijn tenen moet lopen	.77	.59
- ik heb het gevoel dat ik steeds beoordeeld word	.63	.40
- ik twijfel vaak of ik wel met mijn onderzoek zal doorgaan	.58	.34
- ik twijfel vaak of ik wel geschikt ben voor de wetenschap	.54	.29
% verklaarde variantie	40.2	

6. VERKLARINGSMODEL VOOR ONDERWIJSPROBLEMEN VAN PROMOTIE-ASSISTENTEN.

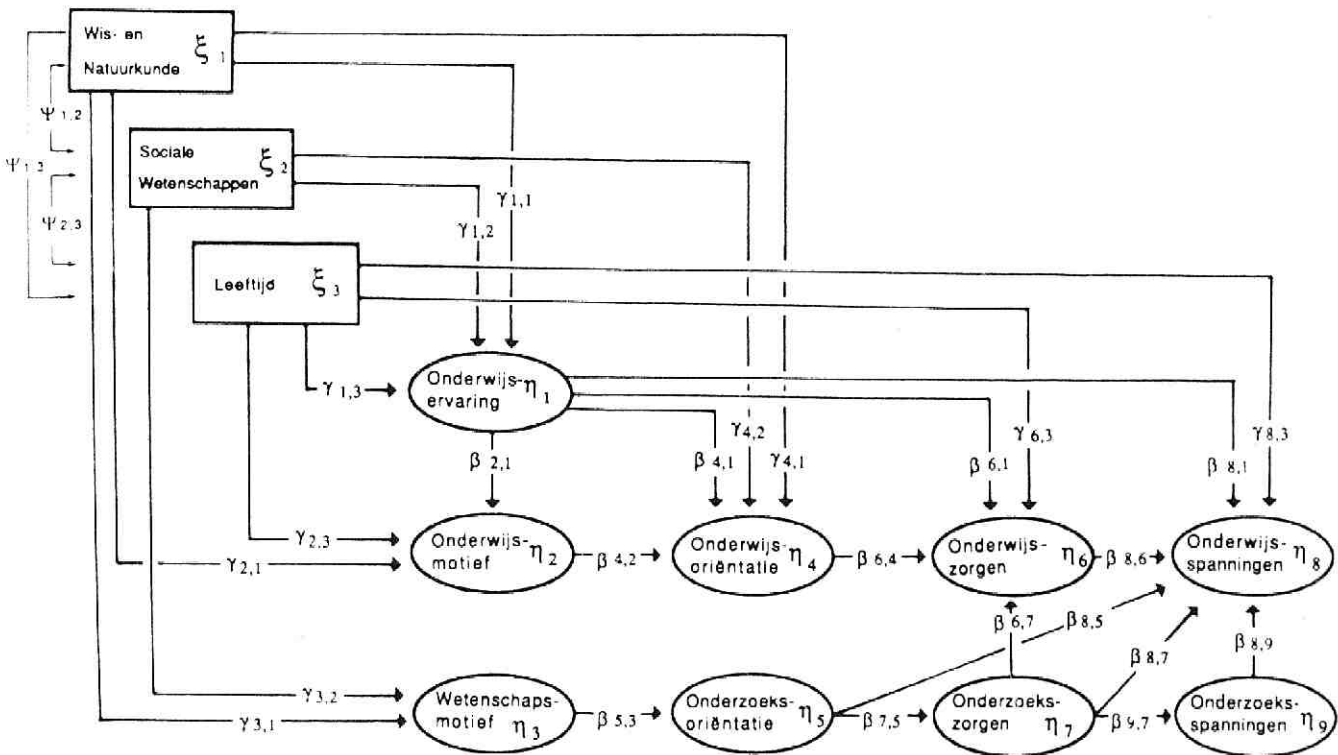
In paragraaf 1 is reeds een eenvoudige vorm van het verklaringsmodel gepresenteerd. Met behulp van de beschreven literatuur kan dit model verder uitgewerkt worden. Wat betreft de onderwijsproblemen gaan wij er vanuit dat de onderwijszorgen een voorwaarde zijn voor de onderwijsspanningen; de onderzoekszorgen zijn voorwaarde voor de onderzoeksspanningen.

Op basis van Finkelstein (1984) veronderstellen we dat problemen die promotie-assistenten ervaren met hun onderzoek zullen resulteren in problemen op het gebied van onderwijs. Onderzoekszorgen leiden tot onderwijszorgen en onderwijsspanningen; onderzoeksspanningen leiden tot onderwijsspanningen. De promotie-assistent die sterk georiënteerd is op de onderzoekstaak zal extra tijd aan onderzoek besteden (zie Yuker, 1984). Deze extra aandacht zal de onderzoekszorgen verminderen maar kan eventueel spanningen bij de onderwijstaak veroorzaken. Een sterke onderzoeksoriëntatie zal dus leiden tot meer onderwijsspanningen en minder onderzoekszorgen. Daarnaast zal een sterke onderwijsoriëntatie de onderwijszorgen verminderen.

Naarmate de promotie-assistent ouder is en meer ondervinding heeft opgedaan, nemen de prestaties als universiteitsdocent toe (zie o.a. Bruhns en Thomsen, 1979; Veenman, 1984). We verwachten dat naarmate de leeftijd en de ondervinding toenemen, de onderwijszorgen en -spanningen afnemen. Tevens zal meer ondervinding tot een sterkere onderwijsoriëntatie leiden.

De motieven c.q. doelstellingen van de promotie-assistent beïnvloeden de onderwijs- en onderzoeksoriëntatie: naarmate de promotie-assistent een sterker onderwijsmotief heeft, zal hij een sterkere onderwijsoriëntatie hebben. Naarmate hij een sterker wetenschapsmotief heeft, zal zijn onderzoeksoriëntatie sterker zijn. De doelstellingen hangen tevens samen met leeftijd, ondervinding en de faculteit. Naarmate de promotie-assistent ouder is en meer ondervinding heeft, neemt zijn onderwijs-motief in sterkte toe. In navolging van Rudd (1975) nemen we aan, dat de promotie-assistenten van wis- en natuurkunde minder 'academic' georiënteerd zijn dan de promotie-assistenten van de sociale wetenschappen en letteren, hetgeen tot uitdrukking komt in een zwakker onderwijsmotief en een zwakker wetenschapsmotief.

Naast verschillen in de beide motieven verwachten we ook verschillen tussen de faculteiten in leeftijd en ondervinding. De literatuur wijst op verschillen tussen faculteiten in onderwijs- en onderzoeksoriëntatie; wij verwachten die verschillen bij promotie-assistenten wel op de onderwijsoriëntatie maar niet op de onderzoeksoriëntatie vanwege de centrale plaats die de promotietaak voor de promotie-assistenten heeft. In tegenstelling tot onze verwachtingen, wordt in de literatuur geen verschil in 'stress' gevonden tussen de faculteiten. In het model volgen wij in eerste instantie de literatuur-bevindingen.



Figuur 2: Het verklaringsmodel voor onderwijsproblemen van promotie-assistenten.

Uit het bovenstaande volgt, dat wij de onderwijszorgen en onderwijsspanningen van de promotie-assistenten willen verklaren vanuit een complex geheel van op elkaar inwerkende variabelen. In figuur 2 zijn de door ons veronderstelde relaties tussen de variabelen in beeld gebracht. Dit model kan beschreven worden door de afzonderlijke relaties tussen de variabelen als hypothesen te beschrijven. Deze hypothesen staan hieronder nogmaals vermeld. In een model toetsen we met behulp van LISREL alle hypothesen te zamen. Alvorens daartoe over te gaan, geven we in paragraaf 7 een beschrijving van de resultaten met betrekking tot de verschillen tussen de drie faculteiten.

Hypothesen

1. De promotie-assistenten van de verschillende faculteiten verschillen naar leeftijd: die van wis- en natuurkunde zijn jonger dan die van sociale wetenschappen en letteren.
2. De promotie-assistenten van letteren hebben de meeste onderwijservaring, gevolgd door promotie-assistenten van sociale wetenschappen en wis- en natuurkunde.
3. De promotie-assistenten van verschillende faculteiten verschillen ten aanzien van het onderwijsmotief en het wetenschapsmotief. De promotie-assistenten van sociale wetenschappen en letteren scoren hoger op het onderwijsmotief (OW-MOT) in vergelijking met de promotie-assistenten van wis- en natuurkunde. De promotie-assistenten van letteren hebben een sterker wetenschapsmotief (WET-MOT) dan de promotie-assistenten van wis- en natuurkunde en sociale wetenschappen. Ten aanzien van het carrière-motief (CAR-MOT) zijn er geen verschillen tussen de faculteiten.
4. De promotie-assistenten van wis- en natuurkunde hebben een zwakkere oriëntatie op hun onderwijstaak (OW-OR) dan de promotie-assistenten van sociale wetenschappen en letteren; er zijn geen verschillen in de oriëntatie op hun onderzoekstaak (OZ-OR) tussen de faculteiten.
5. Er zijn geen verschillen tussen de faculteiten ten aanzien van zorgen en spanningen (OW-ZORG, OW-SPAN, OZ-ZORG en OZ-SPAN).

6. Naarmate de promotie-assistent ouder is, heeft hij meer onderwijservaring. Oudere promotie-assistenten scoren hoger op het onderwijsmotief (OW-MOT) en hebben minder onderwijszorgen (OW-ZORG) en minder onderwijsspanningen (OW-SPAN).
7. Naarmate de promotie-assistent meer onderwijservaring heeft, scoort hij hoger op het onderwijsmotief (OW-MOT), heeft hij een sterkere oriëntatie op zijn onderwijstaak (OW-OR), minder onderwijszorgen (OW-ZORG) en minder onderwijsspanningen (OW-SPAN).
8. Naarmate de promotie-assistent hoger scoort op het onderwijsmotief (OW-MOT), zal hij sterker georiënteerd zijn op de onderwijstaak (OW-OR). Naarmate hij hoger scoort op het wetenschapsmotief (WET-MOT), is zijn oriëntatie op de onderzoekstaak sterker (OZ-OR).
9. Hoe sterker de oriëntatie op de onderwijstaak (OW-OR), des te minder zorgen heeft men met betrekking tot de onderwijstaken (OW-ZORG).
10. Hoe sterker de oriëntatie op de onderzoekstaak (OZ-OR), des te minder zorgen heeft men met betrekking tot de onderzoekstaken (OZ-ZORG), maar des te meer spanningen met de onderwijstaak (OW-SPAN).
11. Onderzoekszorgen (OZ-ZORG) veroorzaken onderwijszorgen (OW-ZORG), onderzoeksspanningen (OZ-SPAN) en onderwijsspanningen (OW-SPAN).
12. Onderzoeksspanningen (OZ-SPAN) veroorzaken onderwijsspanningen (OW-SPAN).

7. RESULTATEN: VERSCHILLEN TUSSEN DE FACULTEITEN

Alvorens de resultaten van de modelschatting en -aanpassing te presenteren, beschrijven we de verschillen tussen de faculteiten, omdat de dummyficering de interpretatie van die resultaten bemoeilijkt. De verschillen tussen de faculteiten zijn getoetst met behulp van variantie-analyses.

Uit tabel 7 blijkt dat de faculteiten verschillen in onderwijservaring en leeftijd; de promotie-assistenten van wis- en natuurkunde zijn jonger en hebben minder onderwijservaring dan de promotie-assistenten van sociale wetenschappen en letteren.

Uit tabel 8 blijkt dat de faculteiten verschillen ten aanzien van het onderwijsmotief en het wetenschapsmotief. De promotie-assistenten van WN scoren lager op beide motieven. Op het onderwijsmotief scoren de promotie-assistenten van SW het hoogst, terwijl de promotie-assistenten van LE de hoogste score hebben op het wetenschapsmotief. Ten aanzien van het carrièremotief zijn er geen verschillen tussen de faculteiten.

In tabel 9 worden de gemiddelde scores binnen de drie faculteiten op de onderwijs- en onderzoeksoriëntatie gepresenteerd. Het blijkt dat de faculteiten verschillen ten aanzien van de onderwijsoriëntatie: WN heeft de laagste score, gevolgd door SW en LE. Ten aanzien van de onderzoeksoriëntatie verschillen de drie faculteiten niet.

Uit tabel 10 blijkt dat de faculteiten verschillen naar onderwijszorgen: de promotie-

Tabel 7: Aantal promotie-assistenten (N), gemiddelden (M) en standaarddeviaties (SD) van onderwijservaring en leeftijd binnen de faculteiten wis- en natuurkunde (WN), sociale wetenschappen (SW) en letteren (LE).

	N	OW-ervaring*		Leeftijd (in jaren)	
		M	SD	M	SD
WN	103	0.72	0.76	28.0	2.7
SW	38	1.18	0.83	31.3	2.9
LE	24	1.67	1.05	30.1	3.6
		F = 14.4 p < .001		F = 20.1 p < .001	

* Range (0-3)

Tabel 8: Aantal promotie-assistenten (N), gemiddelden (M) en standaarddeviaties (SD) van het onderwijsmotief (OW-MOT), het wetenschapsmotief (WET-MOT) en het carrière-motief (CAR-MOT) binnen de faculteiten wis- en natuurkunde (WN), sociale wetenschappen (SW) en letteren (LE) *.

	N	OW-MOT		WET-MOT		CAR-MOT	
		M	SD	M	SD	M	SD
WN	103	460	73	476	98	496	95
SW	38	580	126	505	106	511	92
LE	24	543	78	577	76	493	109
		F = 28.7 p < .001		F = 10.7 p < .001		F = 0.4 p = n.s.	

* De factorscores zijn getransformeerd, zodat het gemiddelde gelijk is aan 500 en de standaarddeviatie gelijk is aan 100.

Tabel 9: Aantal promotie-assistenten (N), gemiddelden (M) en standaarddeviaties (SD) van de onderwijsoriëntatie (OW-OR) en de onderzoeksoriëntatie (OZ-OR) binnen de faculteiten wis- en natuurkunde (WN), sociale wetenschappen (SW) en letteren (LE) *

	N	OW-OR		OZ-OR	
		M	SD	M	SD
WN	103	475	96	499	103
SW	38	516	89	494	109
LE	24	565	82	525	86
		F = 9.9 p < .001		F = 0.8 p = n.s.	

* De factorscores zijn getransformeerd, zodat het gemiddelde gelijk is aan 500 en de standaarddeviatie gelijk is aan 100.

Tabel 10: Aantal promotie-assistenten (N), gemiddelden (M) en standaarddeviaties (SD) van onderwijsspanningen (OW-SPAN), onderzoeksspanningen (OZ-SPAN), onderwijszorgen (OW-ZORG) en onderzoekszorgen (OZ-ZORG) binnen faculteiten wis- en natuurkunde (WN), sociale wetenschappen (SW) en letteren (LE). *

	N	OW-SPAN		OZ-SPAN		OW-ZORG		OZ-ZORG	
		M	SD	M	SD	M	SD	M	SD
WN	103	480	88	489	95	474	95	493	95
SW	38	566	117	526	103	546	91	515	112
LE	24	473	71	497	114	525	104	498	105
		F = 11.4 p < .001		F = 1.8 p = n.s.		F = 8.1 p < .001		F = 0.6 p = n.s.	

* De factorscores zijn getransformeerd, zodat het gemiddelde gelijk is aan 500 en de standaarddeviatie gelijk is aan 100.

assistenten van WN hebben de minste onderwijszorgen, terwijl die van SW de meeste onderwijszorgen hebben. De faculteiten verschillen ook naar onderwijsspanningen: de promotie-assistenten van SW hebben de meeste onderwijsspanningen. Er zijn echter geen verschillen tussen de faculteiten met betrekking tot de onderzoekszorgen, noch met betrekking tot de onderzoeksspanningen.

Bovenstaande gegevens betekenen het volgende voor de in paragraaf 6 geformuleerde hypothesen.

Hypothese 1 wordt bevestigd: de leeftijd van de promotie-assistenten verschilt naar faculteit; de promotie-assistenten van WN zijn het jongste.

Hypothese 2 wordt bevestigd: de onderwijservaring van de promotie-assistenten verschilt naar faculteit, de promotie-assistenten van LE hebben de meeste en die van WN de minste onderwijservaring.

Hypothese 3 wordt eveneens bevestigd: de promotie-assistenten van WN scoren het laagst op het onderwijsmotief en die van LE het hoogst op het wetenschapsmotief. Er zijn bovendien geen verschillen tussen faculteiten op het carrière-motief.

Hypothese 4 wordt gedeeltelijk bevestigd: de promotie-assistenten van WN scoren weliswaar lager op de onderwijsoriëntatie in vergelijking met de promotie-assistenten van LE, maar niet in vergelijking met die van SW. Zoals voorspeld verschillen de faculteiten niet ten aanzien van de onderzoeksoriëntatie.

Hypothese 5 wordt slechts ten dele bevestigd: de faculteiten verschillen wel ten aanzien van de onderwijszorgen en onderwijsspanningen, maar niet ten aanzien van de onderzoekszorgen en onderzoeksspanningen.

8. RESULTATEN: MODELSCHATting EN -AANPASSINGEN.

Het gepresenteerde model (zie figuur 2) hebben we geanalyseerd met LISREL VI. Het model past niet bij de geobserveerde correlaties. De chi-kwadraat is 75.9 bij 38 vrijheidsgraden.

Uit de verkregen oplossing blijkt dat twee relaties niet significant ($t < 1.96$) zijn: de relatie van leeftijd met onderwijsspanningen ($\gamma_{8,3}$) en de relatie van onderwijservaring met onderwijszorgen ($\beta_{6,1}$). Daarnaast valt het model te verbeteren door de relatie van het wetenschapsmotief naar het onderwijsmotief, de relatie van onderzoeksoriëntatie naar onderzoeksspanningen, de relatie van wis- en natuurkunde naar onderwijszorgen en de relatie van sociale wetenschappen naar onderwijsspanningen in het model op te nemen. Met andere woorden: we passen het model aan door $\beta_{6,1}$ en $\gamma_{8,1}$ te fixeren en $\beta_{2,3}$, $\beta_{9,5}$, $\gamma_{6,1}$ en $\gamma_{8,2}$ vrij te maken.

In figuur 3 staan de gestandaardiseerde coëfficiënten van dit beter passende model. De chi-kwadraat bedraagt 39.5 bij 36 vrijheidsgraden ($p = .318$). Dit model is niet het enige dat goed past bij de geobserveerde correlaties. Andere modellen zijn mogelijk, maar zijn theoretisch niet erg plausibel. We geven derhalve de voorkeur aan het gepresenteerde model in figuur 3.

Met betrekking tot de hypothesen verkrijgen we in dit model de volgende resultaten, waarbij we niet nader ingaan op de hypothesen die reeds in paragraaf 7 zijn besproken.

Hypothese 6 wordt slechts ten dele bevestigd. Oudere promotie-assistenten hebben wel meer onderwijservaring (.22) en hebben een sterker onderwijsmotief (.18). Er is echter geen verband tussen leeftijd en onderwijsspanningen, terwijl de oudere promotie-assistenten niet minder, maar juist meer onderwijszorgen blijken te hebben (.26).

Hypothese 7 wordt eveneens slechts gedeeltelijk bevestigd. Naarmate men meer onderwijservaring heeft, heeft men een sterker onderwijsmotief (.17) en een sterkere oriëntatie op de onderwijstaak (.14). Onderwijservaring blijkt echter geen samenhang te vertonen met onderwijszorgen, wel is er een verband met onderwijsspanningen (-.12).

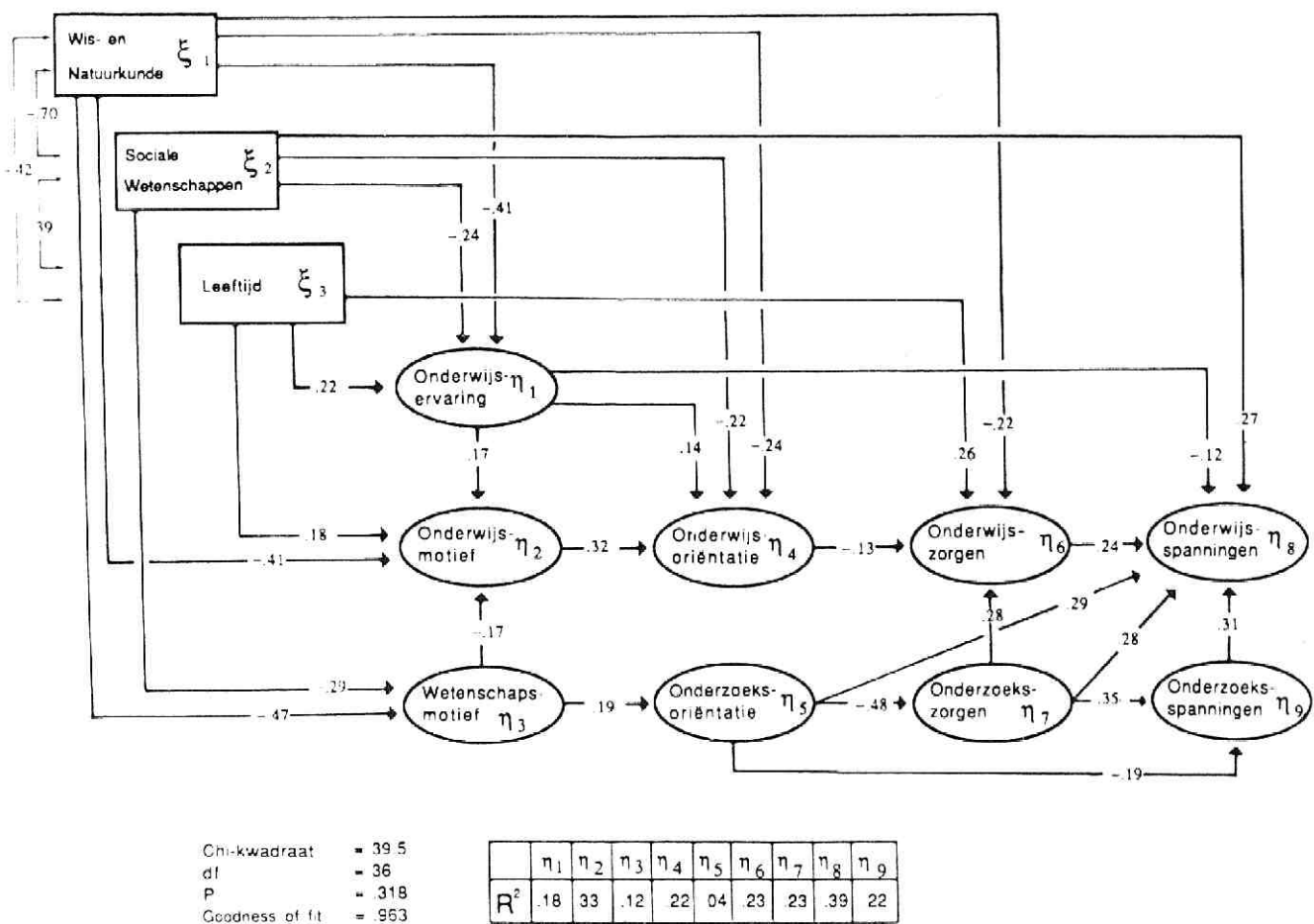
Hypothese 8 wordt bevestigd. Naarmate de promotie-assistenten een sterker onderwijsmotief hebben, zijn zij ook sterker op hun onderwijstaak georiënteerd (.32). Eveneens is het zo dat naarmate de promotie-assistenten een sterker wetenschapsmotief hebben, zij ook sterker georiënteerd zijn op hun onderzoekstaak (.19).

Hypothese 9 wordt bevestigd. Naarmate de promotie-assistenten zich sterker op de onderwijstaak oriënteren, hebben zij minder onderwijszorgen (-.13).

Hypothese 10 wordt eveneens bevestigd. Naarmate de oriëntatie op de onderzoekstaak sterker is, heeft men minder onderzoekszorgen (-.48). Ook leidt een sterkere onderzoeksoriëntatie tot een toename van onderwijsspanningen (.29).

Hypothese 11 wordt ook bevestigd: onderzoekszorgen leiden tot onderwijszorgen (.28), onderwijsspanningen (.28) en tot onderzoeksspanningen (.35).

Hypothese 12 tenslotte wordt bevestigd: onderzoeksspanningen leiden tot onderwijsspanningen (.31).



Figuur 3: Het aangepaste verklaringsmodel voor onderwijsproblemen van promotie-assistenten met de gestandaardiseerde regressiecoëfficiënten en de verklaarde variantie.

9. DISCUSSIE

Met dit artikel wordt beoogd een eerste aanzet te geven om relaties tussen kenmerken van promotie-assistenten en hun onderwijsproblemen in de vorm van een model te presenteren. Dit model is zeker niet volledig. Andere kenmerken kunnen (mede) een belangrijke rol spelen in de verklaring van de onderwijsproblemen. Met name is hierbij aan taakgerelateerde variabelen te denken, zoals taakomvang en type onderwijstaak. Binnen het wetenschappelijk onderwijs bestaat hierin echter een zeer grote variatie (Heijink e.a., 1978), die niet eenvoudig te classificeren is.

Het voorliggende model gaat uit van een verklaring van de onderwijsproblemen vanuit onderzoeksproblemen. Voor de promotie-assistenten lijkt dit bevestigd te worden. Generaliseren van dit gegeven naar de ervaren, vaste universitaire staf kan niet zonder meer plaatsvinden, gegeven de specifieke positie van de promotie-assistenten. Toch lijkt het een interessant gegeven voor het opzetten van professionaliseringsprogramma's voor universiteits-docenten.

De rol van de onderzoeksoriëntatie en die van de onderwijsoriëntatie zijn heel opvallend. Naarmate de onderzoeksoriëntatie sterker is, zijn de onderzoekszorgen minder, zoals verwacht werd. Bij onderwijs vinden we slechts een zwak verband tussen de onderwijsoriëntatie en de onderwijszorgen. Dit kan mogelijk veroorzaakt worden door een factor die ook Lortie (1975) noemt, namelijk het feit dat kennis en vaardigheden bij onderwijs weinig gesystematiseerd zijn. Een andere verklaring is dat de omgeving waarin of de condities waaronder men werkt, bepalen

of de oriëntatie tot veel of weinig zorgen leidt (zie Finkelstein, 1984). We zouden in dat geval kunnen concluderen, dat de condities ten aanzien van de onderzoekstaak voor de sterk onderzoeksgeoriënteerden wel goed zijn, terwijl de condities ten aanzien van de onderwijstaak voor de sterk onderwijsgeoriënteerden niet goed zijn. Bij de promotie-assistenten lijkt een dergelijke redenering, gezien het belang van de onderzoek/promotie-taak, plausibel, maar verder onderzoek hiernaar is noodzakelijk.

De motieven van de promotie-assistenten zijn duidelijk verschillend per faculteit en dat bepaalt mede hun onderwijs- en onderzoeksoriëntatie. De verschillen tussen faculteiten en disciplines zullen daarmee niet beperkt blijven tot de onderwijsproblemen, maar tot meerdere aspecten van het promotie-assistentschap. Beleidsmatig zou men zich kunnen afvragen of de opzet van het AiO-stelsel voldoende rekening houdt met de verschillen tussen faculteiten en disciplines.

Tot slot kan de vraag worden gesteld of het resultaat van het onderzoek aanwijzingen geeft welke verklaringsbron van onderwijsproblemen het meeste perspectief biedt, namelijk de beroepskenmerken en kenmerken van de werkomgeving, dan wel kenmerken die gerelateerd zijn aan individuele ontwikkelingsfasen. De resultaten laten zien, dat 39% van de variantie van de onderwijsspanningen en 23% van de variantie van de onderwijszorgen is verklaard.

Wat betreft de onderwijsspanningen zijn het vooral de onderzoekszorgen en onderzoeksspanningen enerzijds en de onderwijszorgen anderzijds die aan deze verklaarde variantie bijdragen. De uiteindelijke invloed van de onderzoeksoriëntatie op onderwijsspanningen is gering: er is wel een sterk rechtstreeks positief effect, maar dit effect wordt vrijwel volledig teniet gedaan door de indirecte negatieve relaties via onderzoekszorgen en onderzoeksspanningen. De invloed van faculteit op onderwijsspanningen is niet onaanzienlijk. Leeftijd en onderwijservaring daarentegen verklaren slechts een klein deel van de variantie van de onderwijsspanningen. Hieruit kunnen we concluderen, dat voor de onderwijsspanningen de kenmerken van het beroep (met name de zorgen en spanningen van de onderzoekstaak) en de kenmerken van de werkomgeving (de faculteit) een sterkere verklaringsbron zijn dan de kenmerken van de individuele ontwikkeling.

Aan de verklaarde variantie van onderwijszorgen dragen leeftijd, onderzoekszorgen en faculteit ongeveer evenveel bij. Ook hier lijken de kenmerken van het beroep en van de werkomgeving dus een aparte verklaringsbron naast ontwikkelingskenmerken van het individu. Verder onderzoek ter verklaring van onderwijsproblemen van (beginnende) universiteitsdocenten zal zich (mede) moeten richten op kenmerken van het beroep van universiteitsdocent en op kenmerken van de werkomgeving.

Wat betreft het beroep van universiteitsdocent is verder onderzoek nodig naar de relatie tussen de onderwijs- en de onderzoekstaak. De aard en de omvang van beide deeltaken zouden in de analyse betrokken moeten worden. Wat de werkomgeving betreft lijkt onderzoek naar kenmerken van de vakgroepen of subgroepen daarbinnen van belang. Hierbij kan gedacht worden aan de organisatiegraad en de onderwijs- en onderzoeksoriëntatie van de vakgroepen. Uiteraard is ook verder onderzoek nodig naar het voorkomen van onderwijsproblemen van (beginnende) universitaire docenten en de rol van onderwijservaring daarbij. Het onderzoek naar problemen van (beginnende) universiteitsdocenten staat, zeker in vergelijking met het onderzoek naar onderwijsproblemen van docenten in het basis- en voortgezet onderwijs, pas in de beginfase. Het hier getoetste verklaringsmodel voor onderwijsproblemen van promotie-assistenten laat echter zien, dat de gegevens uit onderzoek naar onderwijsproblemen van (beginnende) docenten uit het basis- en voortgezet onderwijs niet zonder meer toepasbaar zijn op die in het wetenschappelijk onderwijs, als gevolg van de specifieke kenmerken van het beroep en de werkomgeving.

LITERATUUR

- Baird, L.L. (1976). Who goes to graduate school and how they get there. In J. Katz & R.T. Hartnett, *Scholars in the making: The development of graduate and professional students* (pp. 19-48). Cambridge: Ballinger.
- Becher, T. (1981). Towards a definition of disciplinary cultures. *Studies in Higher Education*, 6(2), 109-122.
- Bergen, Th., Gerris, J. & Peters, V. (1987). De invloed van ervaring en leeftijd op de perceptie van problemsituaties door docenten tijdens hun beroepsuitoefening. In Th. Bergen, J. Giesbers & C. Morsch, *Professionalisering van onderwijsgeveenden* (pp. 143-157). Lisse: Swets & Zeitlinger.
- Bess, J.L. (1978). Anticipatory socialization of graduate students. *Research in Higher Education*, 8, 289-317.
- Bruhns, I., & Thomsen, O.B. (1979). Concerns of new university teachers. *Higher Education*, 8, 99-110.
- Buis, P. (1983). *Het schrijven van een proefschrift* (A.R. Themaserie no. 1). Den Haag: Academische Raad.
- Clark, B.R. (1983). *The higher education system: Academic organization in cross-national perspective*. Berkeley: University of California.
- Corcoran, M., & Clark, S.M. (1984). Professional socialization and contemporary career attitudes of three faculty generations. *Research in Higher Education*, 20(2), 131-153.
- Creswell, J.W. (1985). *Faculty research performance: Lessons from the sciences and the social sciences*. (ASHE-ERIC Higher Education Report No. 4). Washington D.C.: Association for the Study of Higher Education.
- Eble, K.E. & McKeachie, W.J. (1985). *Improving undergraduate education through faculty development: An analysis of effective programs and practices*. San Francisco: Jossey-Bass.
- Fink, L.D. (1983). First year on the faculty: Getting there. *Journal of Geography in Higher Education*, 7(1), 45-56.
- Fink, L.D. (1984). First year on the faculty: Being there. *Journal of Geography in Higher Education*, 8(1), 11-25.
- Fink, L.D. (1985). First year on the faculty: The quality of their teaching. *Journal of Geography in Higher Education*, 9(2), 129-145.
- Finkelstein, M.J. (1984). *The American academic profession: A synthesis of social scientific inquiry since World War II*. Columbus: Ohio State University Press.
- Fox, D. (1980). What are the principal concerns of new lecturers? In R. Oxtoby (Ed.) *Higher Education at the crossroads* (pp. 111-118). Guildford: Society for Research into Higher Education.
- Fuller, F.F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, 6, 207-226.
- Gmelch, W.H., Lovrich, N.P. & Wilke, P.K. (1984). Sources of stress in academe: A national perspective. *Research in Higher Education*, 20(4), 477-490.
- Heijink, J.A., Derks, J.A.H., Wouters, F.J.J.M., Bois, H.J.W. du, & Krijnen, G. (1978). *Taken en functies van leden van het wetenschappelijk corps*. Nijmegen: ITS.
- Hout, J.F.M.J. van (1984). *Doorstromers en assistenten-in-opleiding: Verslag van een landelijk onderzoek onder doorstromers en een aanzet tot discussie over inrichting van het AiO-stelsel*. Nijmegen: Instituut voor Onderzoek van het Wetenschappelijk Onderwijs.
- Hout, J.F.M.J. van (1985). Doorstromers en AiO's: Aandachtspunten voor de invoering van het AiO-stelsel. *Universiteit en Hogeschool*, 31(5), 225-233.
- Hout, J.F.M.J. van (1986a). De diversiteit in de cognitieve en sociale organisatie van wetenschapsgebieden en universitair onderwijs. In P. Frissen, P.M.Th. van Hoewijk & J.F.M.J. van Hout (Eds.). *De universiteit: Een adequate onderwijsorganisatie?* (p. 96-126). Utrecht: Spectrum.
- Hout, J. van (1986b). Towards a new structure for postgraduate research training in the Netherlands. *European Journal of Education*, 21(3), 275-286.
- Hout, J.F.M.J. van, & Schmeets, J.J.G. (1985). *Schaalconstructie in het doorstromersonderzoek*. Nijmegen: Instituut voor Onderzoek van het Wetenschappelijk Onderwijs.
- Jöreskog, K.G., & Sörbom, D. (1985). *Lisrel VI: Analysis of linear structural relationships by the method of maximum likelihood*. Mooresville: Scientific Software.
- King, M. (1973). The anxieties of university teachers. *University Quarterly*, 28(1), 69-83.
- Lodahl, J.B., & Gordon, G. (1972). The structure of scientific fields and the functioning of university graduate departments. *American Sociological Review*, 37, 57-72.
- Lortie, D.C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Perlberg, A., & Keinan, G. (1986). Sources of stress in academe: The Israeli case. *Higher Education*, 15, 73-88.
- Rudd, E. (1975). *The highest education: A study of graduate education in Britain*. London: Routledge & Kegan Paul.

- Rudd, E. (1985). *A new look at postgraduate failure*. Guildford: The Society for Research into Higher Education & NFER-Nelson.
- Startup, R. (1985). The changing perspective of academic researchers, 1973-1983. *Studies in Higher Education*, 8, 39-55.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143-178.
- Walford, G. (1980). Why physics students start doctorates. *Studies in Higher Education*, 5(1), 77-80.
- Wanner, R.A., Lewis, L.S., & Gregorio, D.I. (1981). Research productivity in academia: A comparative study of the sciences, social sciences and humanities. *Sociology of Education*, 54, 238-253.
- Whitley, R. (1984). *The intellectual and social organization of the sciences*. Oxford: Clarendon.
- Winfield, G. (1987). *The social science PhD: the ESRC inquiry in submission rates*. London: Economic and Social Research Council.
- Yuker, H.E. (1984). *Faculty workload: Research, theory and interpretation*. ASHE-ERIC Higher Education Research Report No. 10 Washington D.C.: Association for the Study of Higher Education.

Manuscript ontvangen 11-3-1987.

Definitieve versie ontvangen 16-7-1987.

Interpretatie en anticipatie in het fysiotherapeutisch probleemoplossen*

A.A.W. Smits
Haagse Academie voor Lichamelijke Opvoeding

P.C. van der Sijde
Universiteit Twente

ABSTRACT

Physiotherapists solve problems with a medical character. In this paper two aspects of physiotherapeutical problem solving (interpretation of information and anticipation of actions), are investigated. Three groups of subjects of different levels of expertise in physiotherapy had to solve a problem, presented in the form of a modified essay question (a shoulder problem). From the answers it appeared that there were differences between the groups with respect to (1) the interpretation of facts and (2) the anticipation of actions to undertake. Also within the groups there are differences which are discussed.

Iedere patient die bij een fysiotherapeut komt, is verwezen door een arts. De verwijzing kan een diagnose of een omschrijving van de klacht(en) inhouden. Het is de taak van een fysiotherapeut de diagnose c.q. de verwijzing van de arts te 'vertalen' in een fysiotherapeutische 'diagnose' op grond waarvan een of meer vormen van therapie kan/kunnen worden toegepast. De diagnose van de fysiotherapeut is (weefsel) specifiekler dan die van een arts. Zo kan een verwijzing van een huisarts P.H.S. (periarthritis humeroscapularis) zijn wat niet meer betekent dan een niet nader omschreven schouderaandoening. Een fysiotherapeut zal dan moeten vaststellen om welk van de vier basisweefsels (epitheel, bind-/steunweefsel, spierweefsel, zenuwweefsel) van de patiënt het gaat.

Het stellen van een diagnose wordt door sommige auteurs, niet geheel correct, opgevat als het oplossen van een medisch probleem (Boshuizen & Claessen, 1982); het is juistler dit te zien als een onderdeel van het medisch probleemoplossen, immers het probleem van een patiënt is niet opgelost door er een label met de diagnose aan te hangen. Medisch en fysiotherapeutisch probleemoplossen omvat zowel het stellen van een diagnose, als het opstellen, uitvoeren, evalueren en bijstellen van een behandelplan. Met andere woorden een medisch, respectievelijk een fysiotherapeutisch probleem omvat meer dan het stellen van een diagnose.

Medisch probleemoplossen verschilt van andere vormen van probleemoplossen door het feit dat beslissingen genomen moeten worden in een situatie waarin *niet* alle informatie vanaf het begin bekend is (Elstein, Shulman, & Sprafka, 1978; Westin, Salvesen, & Gotestam, 1986). Dit impliceert dat medische problemen in principe altijd slecht gedefinieerde problemen zijn, maar het is mogelijk binnen het 'grote' slecht gedefinieerde probleem deelproblemen te onderscheiden, waarvoor wel alle informatie beschikbaar is of kan komen [het interpreteren van röntgenfoto's (Dijkstra, Van der Stelt, & Van der Sijde, 1983) is een voorbeeld van het oplossen van zo'n deelprobleem].

Als strategie om een medisch probleem in te perken, noemen Elstein, Shulman en Sprafka (1978) het in een vroegtijdig stadium genereren van hypothesen over de aard van de klacht. Het resultaat van deze strategie is dat van het open, slecht gedefinieerde probleem een aspect wordt uitgekozen dat een gesloten, goed gedefinieerd probleem is of kan worden. Door middel van het

* Het hier gerapporteerde onderzoek werd uitgevoerd toen de tweede auteur verbonden was aan de Stichting Akademie voor Fysiotherapie Amsterdam (SAFA) te Amsterdam. Onze dank gaat uit naar P. Albers, L. de Leeuw en A. Prins.

toetsen van hypothesen op grond van gegevens die verkregen worden uit anamnese, inspectie en lichamelijk onderzoek, wordt nagegaan of er onder de initiële hypothesen een of meerdere hypothesen zijn, die door het onderzoek ondersteund kunnen worden. Wanneer dit niet het geval is, dan zullen er nieuwe hypothesen opgesteld en getoetst moeten worden. Naarmate het onderzoek vordert, moet idealiter het aantal mogelijke hypothesen afnemen. Deze vermindering van het aantal mogelijke hypothesen in het verloop van een medisch onderzoek wordt door het onderzoek van Elstein, Shulman en Sprafka (1978) ondersteund, maar Johnston, Duran, Hassebrock, Moller, Prietula, Peltovich, en Swanson (1982) vonden in hun onderzoek dat het lichamelijk onderzoek nieuwe/andere hypothesen opwerpt, naast de op grond van de anamnese verkregen hypothesen.

Het stellen van een 'diagnose' door een fysiotherapeut is evenals die van andere (para-) medische specialisten op te vatten als een hypothetico-deductieve werkwijze. De fysiotherapeut krijgt door de verwijzing van de arts informatie, die richting geeft of kan geven aan de anamnese. Het daarop volgend lichamelijk onderzoek van de fysiotherapeut is gericht op het verzamelen van gegevens, waarmee de opgestelde hypothese(n) bevestigd c.q. verworpen kan (kunnen) worden.

Zoals hierboven werd aangegeven is diagnostiseren een onderdeel van het medisch probleemoplossen en in ons onderzoek hebben wij ons hiertoe beperkt.

Uit onderzoek naar probleemoplossen blijkt, dat beginners op een bepaald vakgebied vaak veel sneller van start gaan dan de experts (o.a. Beishuizen, 1983). Dit blijkt eveneens het geval te zijn bij het oplossen van medische problemen: Gale en Marsden (1982) vonden dat beginners de verstrekte klinische informatie vaak onmiddellijk interpreteerden en evalueerden. In het onderzoek naar het oplossen van fysiotherapeutische problemen werd door Van der Sijde en Sellink (1983) voorgesteld hierin twee fasen te onderscheiden: interpretatie van gegevens en anticipatie van de te ondernemen acties op basis van de gegevens. Uit de interpretatie van klinische gegevens kunnen een of meer acties worden afgeleid die tot doel hebben in een volgende fase van het onderzoek de gestelde hypothesen nader te verifiëren. Binnen het fysiotherapeutisch onderzoek betreffen de te ondernemen acties het testen van spieren, het palperen van anatomische structuren etc.

Het doel van deze studie is de verschillen in het probleemoplosgedrag van beginnende fysiotherapie studenten, gevorderde fysiotherapie studenten en experts (praktiserende en docerende fysiotherapeuten) op een specifiek fysiotherapeutisch onderwerp – een schouderprobleem – ten aanzien van het interpreteren van gegevens en anticiperen van acties te onderzoeken. Gezien de opbouw van curricula van fysiotherapie-opleidingen (Van der Sijde, Sellink, & Wurms, 1987) is het aannemelijk dat beginnende studenten beter klinische informatie kunnen interpreteren dan fysiotherapeutische acties kunnen anticiperen (in het begin van de studie ligt de nadruk op het verwerven van medische kennis, terwijl, naarmate de studie vordert, meer aandacht wordt besteed aan het toepassen van deze kennis in de praktijk). Verder verwachten wij (in navolging van Elstein et al., 1978) dat de anamnese meer hypothesen zal opwerpen dan het lichamelijk onderzoek.

METHODE

Proefpersonen

De proefpersonen waren studenten en docenten van een fysiotherapie-opleiding. De proefgroepen waren als volgt samengesteld:

- Groep 1: Tweedejaars studenten fysiotherapie ($n=15$). Zij beschikken in principe over kennis van anatomie, kinesiologie, pathologie, gedragswetenschappen en fysiotherapie.
- Groep 2: Derdejaars studenten fysiotherapie ($n=11$). Zij beschikken in principe over een tamelijk uitgebreide theoretische kennis en praktische ervaring aangaande de fysiotherapie.

Groep 3: Praktiserende docenten fysiotherapie ($n=6$). Dit zijn docenten verbonden aan een fysiotherapie-opleiding, die tevens werkzaam zijn in het (fysiotherapie-)veld. De proefpersonen namen op vrijwillige basis deel aan het onderzoek.

Materiaal

Het materiaal, aan de proefpersonen voorgelegd, bestond uit een gestructureerde open vraag (G.O.V.) 'Schouderprobleem'. In een G.O.V. wordt een proefpersoon geconfronteerd met een casus die aan de praktijk ontleend is (Hodgkin & Knox, 1975). De informatie wordt in gedeelten aangeboden en niet direct volledig, zoals bij de traditionele casus. In een aantal opeenvolgende fasen wordt er informatie toegevoegd, waarvan sommige gegevens relevant zijn voor het fysiotherapeutisch probleem en sommige minder relevant. Een G.O.V. bestaat dus uit een serie open vragen over een zich ontwikkelende situatie, waarbij de antwoorden vrij zijn. De proefpersonen worden gevraagd een antwoord te geven naar aanleiding van de geschetste situatie, alvorens de tekst verder door te nemen.

De G.O.V. 'Schouderprobleem' is geconstrueerd op basis van medische en fysiotherapeutische literatuur. Vervolgens is deze op relevantie ten aanzien van de fysiotherapie doorgesproken met twee docenten (die verder niet deelnamen aan het onderzoek). Daarna zijn er twee proefafnamen van de G.O.V. geweest, een met een proefpersoon uit groep 1 en een met een proefpersoon uit groep 2. Tijdens deze proefafnamen werden hardop-denk protocollen vastgelegd en deze werden gebruikt om de G.O.V. 'Schouderprobleem' bij te stellen. De G.O.V. 'Schouderprobleem' is in principe door alle groepen proefpersonen op te lossen en is opgenomen in Figuur 1.

De vragen die in de G.O.V. zijn opgenomen hebben betrekking op een van de volgende twee aspecten: (1) interpretatie van de aangeboden informatie en (2) op anticipatie van acties, die op grond van de hypothesen (interpretatie) ondernomen moet worden.

Aan de G.O.V. 'Schouderprobleem' ligt een diagnostisch basisschema ten grondslag. (Figuur 2). In dit basisschema is de procedure van probleemoplossen weergegeven, die gevolgd wordt om tot een fysiotherapeutische diagnose te komen; anamnestiche en klinische informatie geven aanleiding tot hypothesen (mogelijke diagnoses). Op grond hiervan worden medische en fysiotherapeutische onderzoeken uitgevoerd, welke informatie opleveren om tot de meest waarschijnlijke fysiotherapeutische diagnose te komen of aanleiding zijn voor het verzamelen van nieuwe anamnestiche en klinische informatie. Met behulp van dit basisschema is het mogelijk een overzicht te geven van de mogelijkheden die in een G.O.V. 'Schouderprobleem' verwerkt kunnen zijn. (Figuur 3.)

Scoring

De G.O.V. werd door de eerste auteur gescoord aan de hand van een schema dat werd samengesteld op grond van informatie uit gesprekken met docenten fysiotherapie en literatuur (Godt, Malin & Wittenberg, 1981; Grond, 1981; Kingma, 1976; Oostendorp, 1976 en 1980; Stenvers & Overbeek, 1979). Een voorbeeld van een scoringslijst is in Tabel 1 afgebeeld.

In het algemeen geldt, dat iedere goede interpretatie van de informatie en elke juist gewenste actie gewaardeerd wordt met een punt. Voor de eerste vier vragen gold een maximale score van 10, voor vraag vijf een maximum van 5 en voor vraag zes een maximum van 2 punten.

Procedure

De G.O.V. 'Schouderprobleem' werd als volgt afgenomen:

Groep 1 en groep 2 beantwoordden de G.O.V. 'Schouderprobleem' tijdens een fysiotherapieles (50 minuten).

Groep 3 kreeg de G.O.V. 'Schouderprobleem' mee als een 'take home' activiteit na instructie over de specifieke werkwijze te hebben ontvangen.

De G.O.V.'s zijn in eerste instantie door de eerste auteur gescoord. Om de intrabeoordelaar

De Gestuctureerde Open Vraag Schouderprobleem

Aanwijzingen

- Beantwoord de vragen zo uitgebreid mogelijk.
- Lees, blader deze Gestuctureerde Open Vraag (GOV) niet door voor u begint.
- Ga niet terug om te veranderen, wat u hebt opgeschreven.

De heer Dalver, 35 jaar oud, komt met de verwijzing schouderandoening. Hij klaagt dat hij zijn arm nauwelijks kan heffen.

1. Waar denkt u bij deze klacht aan?

---ruimte voor antwoord---

2. Wat wilt u de heer Dalver in eerste instantie vragen?

---ruimte voor antwoord---

(pagina 2)

De heer Dalver vertelt dat hij sinds enige weken problemen, pijn ondervindt bij het zijwaards optillen van de rechterarm. Een (onge-)val kan hij zich niet herinneren. De pijnverschijnselen doen zich voor bij zijn beroep bouwvakker en bij zijn sportactiviteit, voetbalkeeper. De pijn is gelokaliseerd aan de voor-bovenzijde van het schoudergewricht. De pijn straalt uit naar de arm en soms naar de rug. 's Nachts kan de heer Dalver niet op de aangedane schouder liggen. De pijn is het hevigst in de buurt van het horizontale vlak.

3. a. Welke hypothese(n) kom(t)(en) in u op omtrent de fysiotherapeutische diagnose(n)?

b. Wat zijn andere mogelijke hypothesen daarnaast, die niet op voorhand uitgesloten kunnen worden?

---ruimte voor antwoord---

4. Wat gaat u onderzoeken en waarom?

---ruimte voor antwoord---

(pagina 3)

Inspectie levert geen bijzondere informatie op. Functie-onderzoek geeft het volgende beeld. De acromio- en sterno-claviculaire gewrichten functioneren normaal. Het gleno-humerale gewricht vertoont:

- . abductie, maar met pijnlijk traject bij ongeveer 90° afvoeren van de arm waarbij de pijn toeneemt door endorotatie;
- . verstoring van het humero-scapulaire ritme;
- . abductie tegen weerstand in is zeer pijnlijk.

Weerstand extensie elleboog-gewricht is niet pijnlijk. Op neurologisch gebied worden geen afwijkingen geconstateerd.

5. Welke extra informatie (van welke aard dan ook) zou u nog van een verwijzend arts en/of specialist kunnen gebruiken om uw diagnose te bevestigen of te verwerpen?

---ruimte voor antwoord---

6. Wat is uw diagnose en waarom?

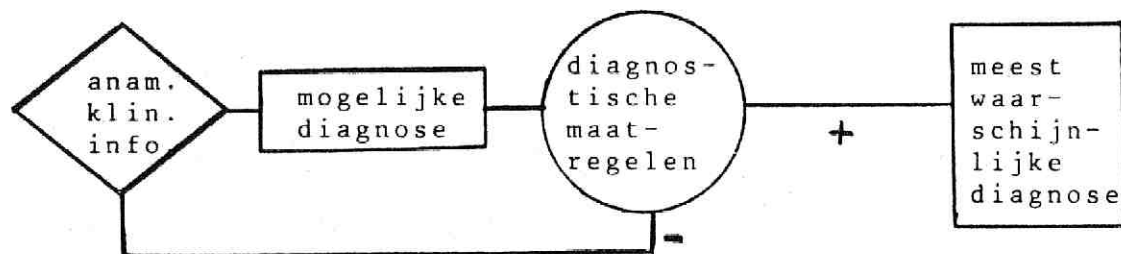
Waarop richt u zich bij de uitvoering van de fysiotherapeutische behandeling?

---ruimte voor antwoord---

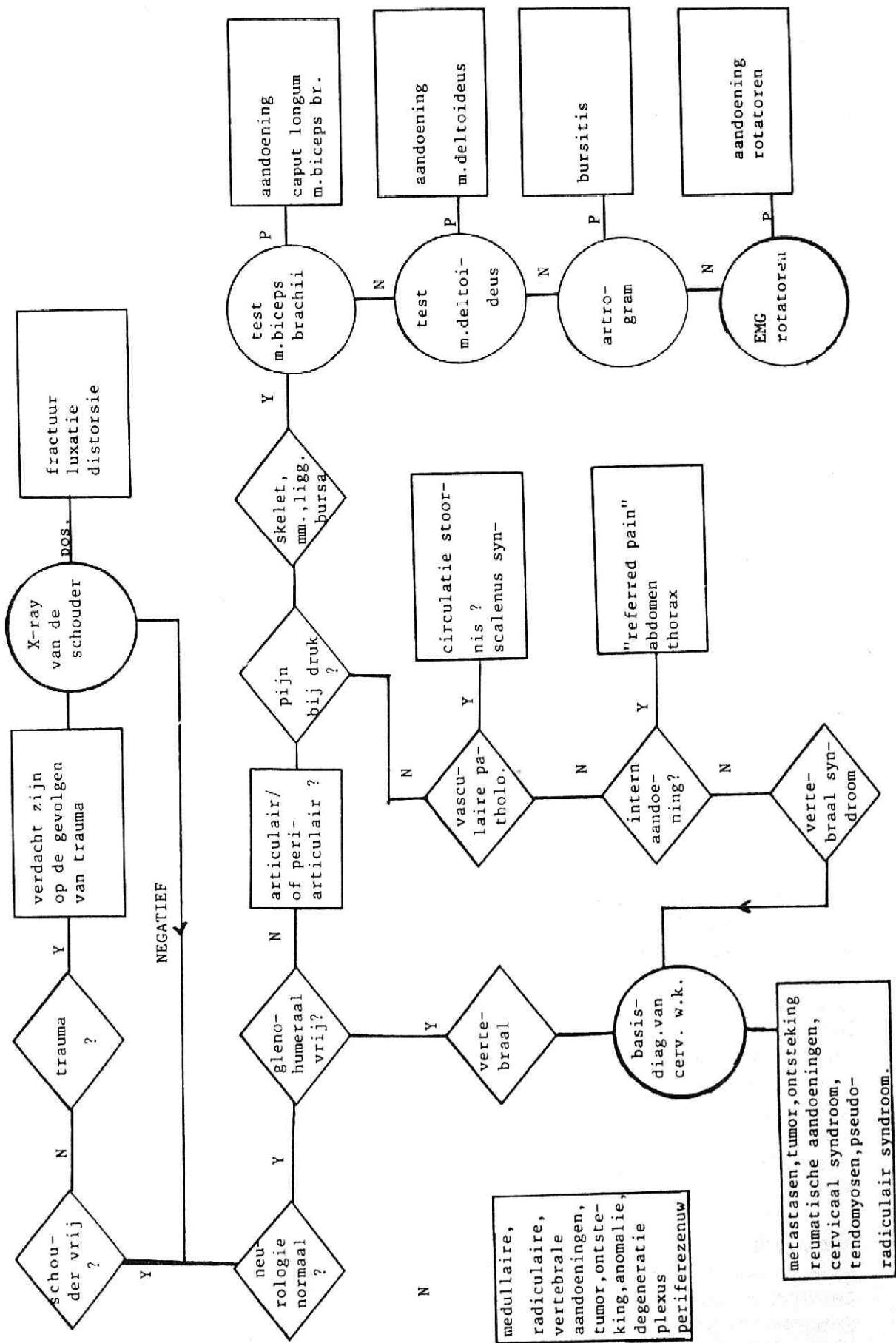
(pagina 4)

Geef a.u.b. uw kritiek, positief en/of negatief op de GOV Schouderprobleem.

Figuur 1: De gestructureerde open vraag 'Schouderprobleem'.



Figuur 2: Basis van het diagnostisch schema voor een gestructureerde open vraag.



Figuur 3: Diagnostisch schema in de gestructureerde open vraag 'Schouderprobleem'.

Tabel 1. Scoringslijst van vraag 4 van de G.O.V. 'Schouderprobleem'.

‘Wat gaat u onderzoeken en waarom?’	
Lichamelijk onderzoek, passief	1 punt
Lichamelijk onderzoek, actief	1
Inspectie	1
Palpatie	1
Compressietest (pijntoename)	1
Tractie test (pijnvermindering)	1
Translatie	1
Spierlengte onderzoek	1
T.O.C. test	1
Sneltest	1
Totaal maximaal	10 punten.

Antwoorden, die niet tot het bovenstaande behoorden, werden niet gescoord.

overeenstemming te bepalen zijn na ongeveer twee jaar de G.O.V.'s door dezelfde persoon een tweede maal gescoord. De G.O.V.'s zijn bovendien door een deskundige op dit gebied, die op geen enkele wijze bij het project betrokken is geweest, gescoord, zodat interbeoordelaarsovereenstemming eveneens is nagegaan (Cohen, 1960).

Data

Om de hypothese te toetsen, dat er in het begin van de studie beter geïnterpreteerd wordt dan acties worden geanticipeerd, wordt het quotiënt interpretatie-score : anticipatie-score onderzocht. Hoe groter het quotiënt, hoe beter is, in verhouding tot de anticipatie-score, de interpretatie-score en visa versa.

RESULTATEN

Inter- en intrabeoordelaarsovereenstemming

De kappa, dat wil zeggen de mate van overeenstemming tussen twee observatoren, bedraagt 0.43. De gemiddelde intrabeoordelaarsovereenstemming levert een kappa van 0.53 op.

Interpretatie

In de G.O.V. 'Schouderprobleem' zijn drie vragen (vragen 1, 3, 6) opgenomen, die betrekking hebben op het interpreteren van anamnestiche en/of klinische informatie. Een overzicht van de gemiddelde groepsscore (per vraag en totaal) is weergegeven in Tabel 2. Uit analyse met behulp van de Kruskal-Wallis toets blijkt dat er een significant verschil is tussen de groepen ($\chi^2_2 = 6.03, p < .05$, gecorrigeerd voor 'ties': $\chi^2_2 = 6.05, p < .05$), dat het gevolg is van een significant verschil tussen de scores van de proefpersonen in groep 1 en groep 2 (Mann-Whitney: MWU = 33, $p < .01$).

Anticipatie

Drie vragen in de G.O.V. 'Schouderprobleem' (vragen 2, 4, 5) hebben betrekking op het anticiperen van de te ondernemen acties. In Tabel 3 wordt een overzicht van de gemiddelde groepsscores (per vraag en totaal) gegeven. Uit de analyse van de gegevens met de Kruskal-Wallis toets blijkt dat er een significant verschil is tussen de groepen ($\chi^2_2 = 6.19, p < .05$, gecorrigeerd voor 'ties': $\chi^2_2 = 6.22, p < .05$), dat het gevolg is van een significant verschil tussen de groepen 1 en 2 (MWU = 34.5, $p < .01$).

Tabel 2. Overzicht van de gemiddelde groepsscore (per vraag en totaal) op de interpretatie vragen in de G.O.V. 'Schouderprobleem'.

vraag	groep 1	groep 2	groep 3
1	2.67 (1.68)*	5.45 (2.66)	3.67 (3.83)
3	3.13 (1.19)	3.82 (1.33)	5.50 (2.26)
6	1.53 (0.52)	1.64 (0.50)	1.67 (0.52)
totaal	7.33 (2.64)	10.91 (3.11)	10.83 (5.91)
% van max.	33.3**	49.6	49.2

* tussen haakjes staat de standaarddeviatie vermeld

** maximum score bedraagt 22

Interpretatie-anticipatie

Voor alle proefpersonen is de verhouding anticipatie-score : interpretatie-score berekend. Voor de proefpersonen in groep 1 was deze verhouding gemiddeld 1.53 ($sd = 0.58$), voor de proefpersonen in groep 2 was dit gemiddeld 1.47 ($sd = 0.57$) en voor de proefpersonen in groep 3 bedroeg het gemiddelde 0.84 ($sd = 0.35$). Alleen groep 1 en groep 2 verschillen niet significant van elkaar (MWU = 80, $p = n.s.$), groep 1 wel van groep 3 (MWU = $p < .05$) en groep 2 eveneens van groep 3 (MWU = 11, $p < .05$).

Voor geen van de groepen bleek er een significant verschil te zijn tussen het aantal vragen, dat door de anamnese (vraag 1) en door het lichamelijk onderzoek (vraag 3) werd opgeworpen (groep 1: MWU = 93.5, $p = n.s.$; groep 2: MWU = 38.5, $p = n.s.$; groep 3: MWU = 13.5, $p = n.s.$) en dit bleek ook het geval te zijn als alle groepen gezamenlijk werden genomen (MWU = 487.5, $p = n.s.$).

Kwalitatieve verschillen

Naast kwantitatieve verschillen tussen de groepen zijn er eveneens kwalitatieve verschillen, uiteraard tussen de groepen, maar tevens binnen de groepen. In Tabel 4 zijn de antwoorden van een laag en een hoog scorende proefpersoon uit ieder van de groepen opgenomen.

In Tabel 4 vallen een aantal zaken op, ten eerste dat de laag scorende proefpersoon uit groep 1 al vanaf het begin op het verkeerde 'been' wordt gezet. De laagscorende proefpersoon uit groep 2 noemt veel oorzaken van algemene aard, de uiteindelijke diagnose is niet specifiek. De laagscorende proefpersoon uit groep 3 geeft op vraag 1 het antwoord 'tendinitis' en stelt dit eveneens als diagnose in antwoord op vraag 6. Een tendinitis bij een P.H.S. kan gezien worden als een schot voor de boeg, gezien het feit dat er in het schoudergebied een groot aantal spieren en dus ook pezen verlopen. Verder valt bij de hoogscorende proefpersonen op dat alle drie de juiste spier noemen, terwijl de hoogscorende proefpersonen uit groep 2 en 3 tevens nog de aard van de aandoening (tendinitis) vermelden. De benadering van de hoogscorende proefpersonen is een hypotetico-deductieve wijze van werken.

Tabel 3. Overzicht van de gemiddelde groepsscore (per vraag en totaal) op de anticipatie vragen in de G.O.V. 'Schouderprobleem'.

vraag	groep 1	groep 2	groep 3
1	5.80 (2.65)*	7.64 (1.70)	5.17 (3.00)
4	4.00 (1.36)	6.36 (2.06)	4.33 (1.97)
5	0.87 (0.83)	1.00 (1.10)	1.00 (1.10)
totaal	10.67 (4.06)	15.00 (3.66)	10.50 (5.01)
% van max.	42.7**	60.0	42.7

* tussen haakjes staat de standaarddeviatie vermeld

** maximum score bedraagt 25

Tabel 4 Vergelijking van de antwoorden op de G.O.V. 'Schouderprobleem' tussen een laag en een hoog scorende proefpersoon uit elke van de drie groepen.

Vr. (*)	Gr. (**)	Laag scorende pp.	Hoog scorende pp.
1	1	Conclusie: arm kan bijna niet geheven worden	Is oorzaak musculair, neurogeen, vasculair, arthrogeen?
	2	Noemt veel, ook overbodige mogelijkheden	Denkt in eerste instantie aan acht mogelijkheden.
	3	Denkt aan een tendinitis.	Vermeldt somatische aandoening in/buiten schouder.
2	1	Zijn pijnklachten echt de oorzaak?	Aandacht voor duur, pijn (waar, soort), beroep, hobbies, overige gewrichten.
	2	Vermeldt 'anamnese' en willekeurige vragen.	Komt met een aantal vragen in een logische volgorde.
	3	Vraagt of patiënt de arm wel/niet kan heffen?	Naast algemene anamnese specifieke vraag naar duur, verloop en soort (pijn-)klachten.
3	1	Geeft 2 hypothesen.	Geeft 3 hypothesen.
	2	Noemt algemene aandoeningen, b.v. bursitis.	Geeft 3 hypothesen en 2 alternatieve hypothesen.
	3	Heeft 2 hypothesen	Heeft vier hypothesen en enige alternatieve hypothesen.
4	1	Kijkt naar gewricht en weke delen.	Functie onderzoek actief, passief, weerstand- en rektesten.
	2	Volstaan wordt met algemene termen.	Geeft verschillende oorzaken aan met redenen omkleed.
	3	Gaat localisatie van compressie pijn na.	Geeft 6 onderzoeken aan om mogelijkheden in te perken.
5	1	Is er een zenuw bekneeld geraakt?	Röntgenfoto is wenselijk vanwege beroep en hobby.
	2	Gaat op deze vraag niet in	Stelt belang in röntgenfoto en bloed-onderzoek.
	3	Laat deze vraag open.	Stelt procaine injectie in de pees m. supraspinatus voor.
6	1	Ruptuur m.biceps br..	Aandoening m.supraspinatus.
	2	Tendinitis (Spier?).	Aandoening m.supraspinatus, waarschijnlijk tendinitis.
	3	Vermeldt tendinitis	Tendinitis m.supraspinatus.

* Vr = nummer van de vraag

** Gr = expertise-groep

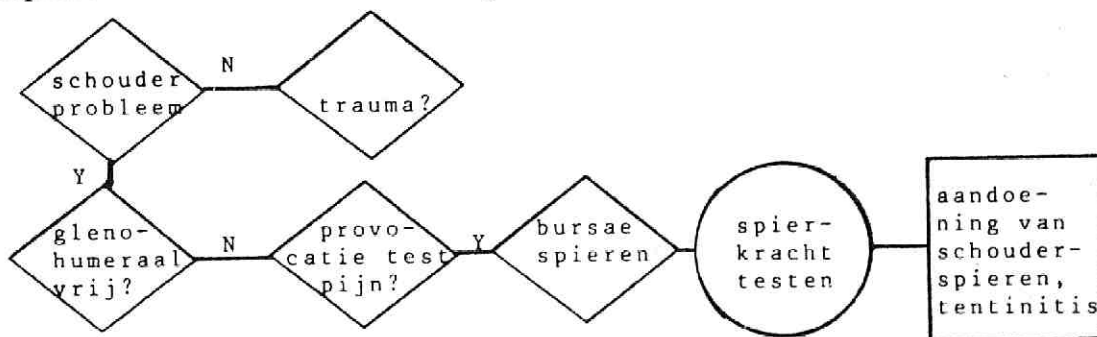
DISCUSSIE

Beginnende studenten fysiotherapie behalen geen hogere score voor interpretatie dan voor anticipatie. Enerzijds blijkt dit uit het quotiënt (anticipatie : interpretatie) en anderzijds blijkt dat uit de vergelijking van percentages van de maximum-score op de anticipatie en interpretatie-score.

Naarmate de expertise toeneemt wordt het quotiënt (anticipatie : interpretatie) kleiner en alleen voor de groep van de 'echte' experts (de praktiserende fysiotherapeuten) geldt dat er een hogere score voor interpretatie dan voor anticipatie wordt behaald. Een tweetal kanttekeningen zijn bij deze resultaten op zijn plaats. Ten eerste, de open vraag is zo geconstrueerd dat ook de groep met de laagste expertise in het onderzoek het probleem kon oplossen. Met andere woorden, studenten met relatief geringe fysiotherapeutische kennis kunnen dit probleem oplossen, echter de kwaliteit van het oplosproces is verschillend van die van de andere expertisegroepen. Bijvoorbeeld, een schouderprobleem kan naast een orthopedische ook een neurologische component bevatten. Beginnende studenten fysiotherapie beschikken over onvoldoende, onvolledige neurologische kennis en deze kennis zal dan ook veelal niet geactiveerd worden. De tweede kanttekening betreft het aantal acties dat geanticipeerd wordt; beginnende studenten fysiotherapie anticiperen acties welke op zich niet foutief zijn en/of de patiënt schade berokkenen, maar die door experts niet uitgevoerd worden, omdat deze geen of weinig informatie opleveren.

Verder blijkt uit de gegevens dat er geen verschil is in het aantal vragen c.q. hypothesen, dat door de anamnese (vraag 1) en door het lichamelijk onderzoek (vraag 3) worden opgeworpen. Met het oog op de onderzoeksresultaten van Elstein, Shulman en Sprafka (1978) en Johnson, Duran, Hassebrock, Moller, Prietula, Feltovich en Swanson (1982) kan worden opgemerkt dat, afhankelijk van het domein waaruit het probleem wordt gekozen, de anamnese meer / minder / gelijk aantal hypothesen opwerpt dan het lichamelijk onderzoek.

De aspecten van belang voor het oplossen van een 'schouderprobleem' kunnen in kaart worden gebracht. Dit diagnostisch schema, ten grondslag liggend aan de G.O.V. 'Schouderprobleem' (zie Figuur 3) is een theoretisch schema, dat min of meer het gehele fysiotherapeutisch kennisdomein afbeeldt. Geen enkele praktiserende fysiotherapeut zal dit schema in zijn geheel afwerken, en geen enkele diagnose die gesteld wordt zal op een ideaal typische wijze plaatsvinden. Op grond van ervaring en 'klinische blik' (intuïtie?) zullen bepaalde delen van het schema niet expliciet worden afgewerkt. Dat er een soort selectie heeft plaatsgevonden is ondermeer af te lezen uit de antwoorden van zes proefpersonen in Tabel 4 en anderzijds uit de omvang van de diagnostische schema's van de drie proefpersonen uit elk van de drie expertise-groepen (zie Figuur 4a, b en c). Zoals hierboven beschreven, blijkt uit Figuur 4a de beperktheid van de kennis van de beginnende student.

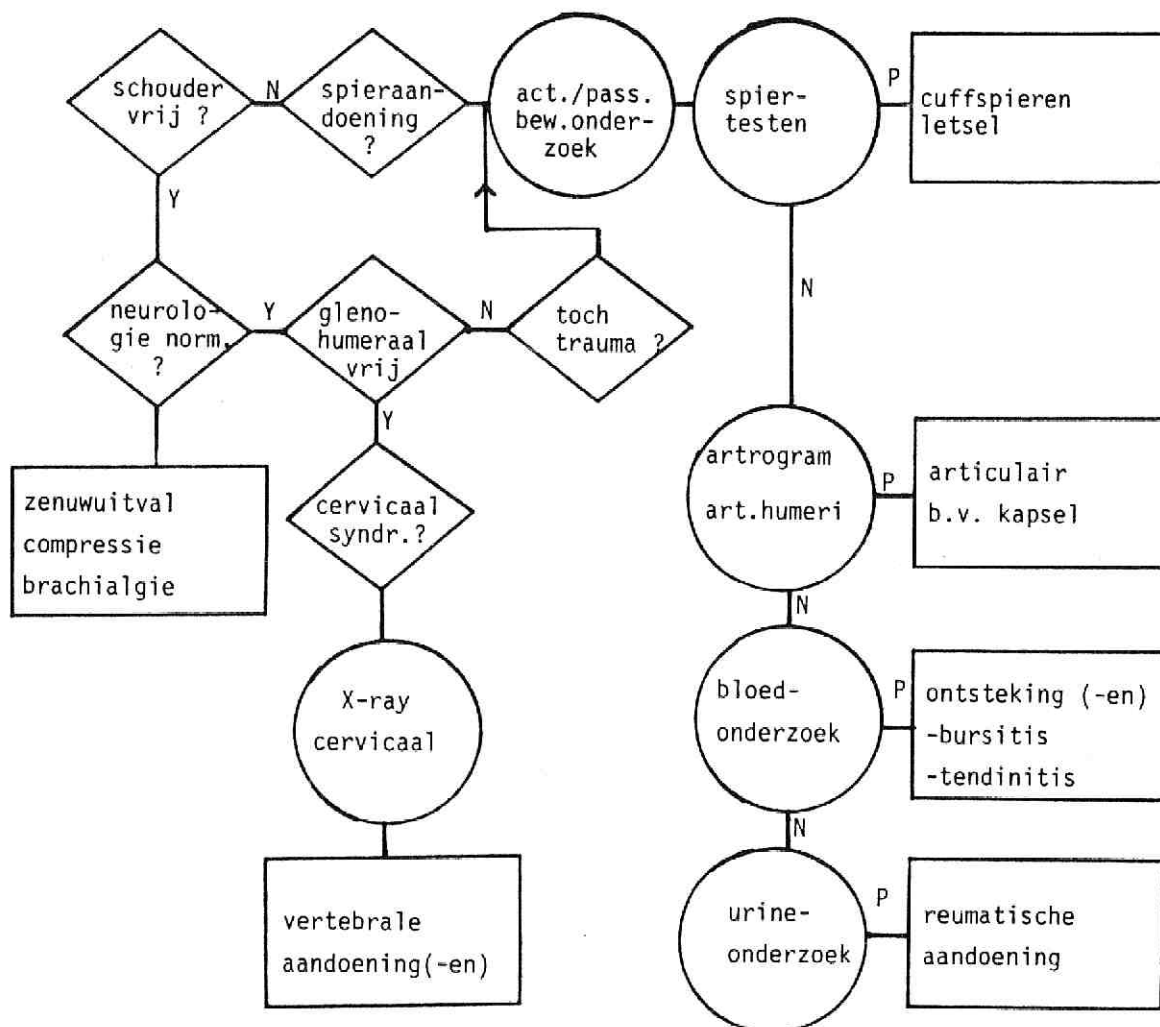


Figuur 4a: Diagnostisch schema van een tweedejaars student fysiotherapie.

Het uitgebreider worden van het kennisdomein, naarmate de expertise groter wordt, c.q. naarmate de student vordert in de studie is te zien in Figuur 4b.

In Figuur 4c is het diagnostisch schema van een fysiotherapeut weergegeven. De ervaring maakt dat er een selectie wordt gemaakt uit alle mogelijkheden.

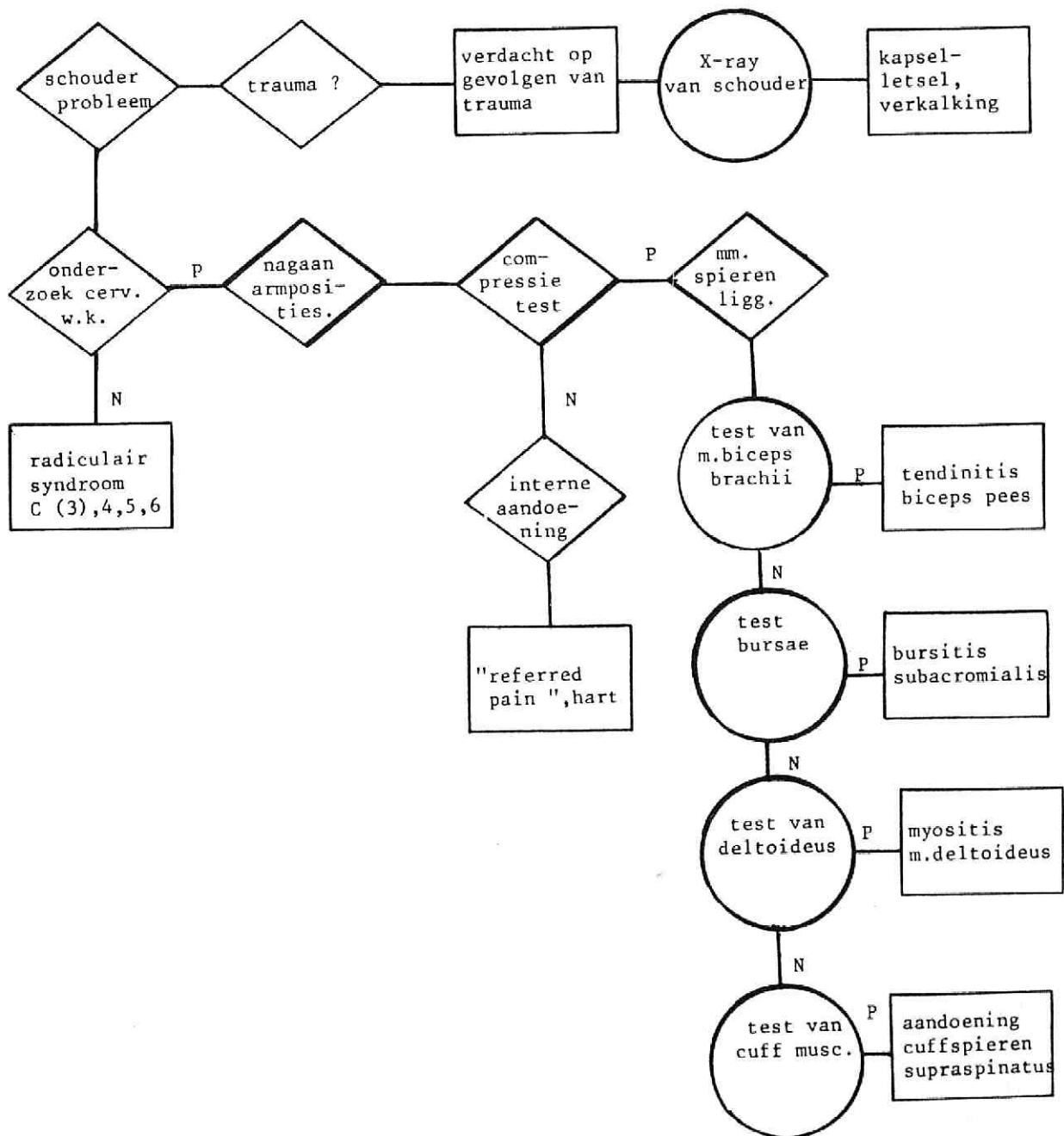
Concluderend kan worden opgemerkt, dat binnen het domein van de fysiotherapie de aspecten interpretatie en anticipatie zinvolle onderscheidingen zijn in het (medisch) probleem-oplossen. Anderzijds blijkt, wanneer het probleemoplosproces wordt beschouwd in een relatief eenvoudige vorm dat door alle expertise groepen is op te lossen, dat er grote verschillen bestaan tussen de personen, afhankelijk van de mate van expertise. In hoeverre de resultaten van dit



Figuur 4b: Diagnostisch schema van een derdejaars student fysiotherapie.

onderzoek te generaliseren zijn naar andere probleemttypen binnen de fysiotherapie (zoals problemen met een overwegend neurologisch of psycho-sociaal karakter) zal nader onderzoek moeten uitwijzen.

Uit het voorgaande blijkt de bruikbaarheid van de gestructureerde open vraag als een onderzoeksinstrument. Verder is deze manier van vragen bruikbaar als instructiehulpmiddel, dat aan te passen is aan het niveau van de studenten. De gestructureerde open vraag kan met name gebruikt worden in die situaties waarin de student (nog) geen verantwoordelijkheid kan/mag dragen. Bovendien valt het tijds- en kostenaspect positief uit in vergelijking met andere instructie-hulpmiddelen, zoals casus op video (Van der Sijde, Sellink, & Wurms, 1987) en computergestuurde casus. Door een probleem uit het dagelijks leven als uitgangspunt te nemen in een G.O.V. worden studenten gemotiveerd om tot een (goede) oplossing te komen. Vergroting van de betrokkenheid van de studenten kan verder verkregen worden door de G.O.V. in tweetallen te laten oplossen. Een G.O.V. vraagt integratie van kennis uit de verschillende basisdisciplines. Dit is voor zowel samenstellers van de G.O.V. (bezinging op de onderwijsleerdoelen) en studenten een uitdagende factor. De opgedane ervaring in het oplossen van (papieren) problemen kan het onthouden van kennis vergemakkelijken. Het veelal hieraan verbonden idee van transfer is (nog) niet aangetoond. Met behulp van 'patient management problems' zouden probleem-oplossingsstijlen te onderscheiden zijn. Probleem echter is dat er nog weinig kennis is over hoe en wanneer probleem-oplossingsstijlen door (medische) studenten worden (aan-)geleerd, aldus Westin, Salvesen, en Gotestam (1986). Een beperking van de G.O.V. is dat het geen valide simulatie is van de klinische situatie. De structuur en het



Figuur 4c: Diagnostisch schema van een fysiotherapeut.

puntensysteem kunnen het gedrag van de participanten verstoren (Newble, Hoare, & Baxter, 1982). De G.O.V. is daarom als (selectie-)test niet geschikt.

Probleemoplossen houdt altijd een complexe combinatie van allerlei soorten cognitieve vaardigheden in. Dit geldt ook voor het fysiotherapeutisch probleem in een G.O.V.. Het frequent oefenen van het oplossen van (die) problemen is onmisbaar voor de opbouw van expertise. Een G.O.V. kan hieraan een bijdrage leveren.

REFERENTIES

- Beishuizen, J.J. (1983). Al doende leert men problemen oplossen. In C. Hamaker, B. van Hout Wolters, & K.B. Koster (red.). *Onderwijsleerprocessen: Tekstverwerking, Probleemoplossen en Leermoeilijkheden*. Lisse: Swets & Zeitlinger.
- Boshuizen, H.P.A., & Claessen, H.F.A. (1982). Problems of research into medical problem solving: Some remarks on theory and method. *Medical Education*, 16, 81-87.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Dijkstra, S., Stelt, P.F. van der, & Sijde, P.C. van der (1983). Het effect van verschillende expertise niveau's op de interpretatie van tandheelkundige foto's. In Beishuizen et al. (red.), *Onderwijsleerprocessen*. Lisse: Swets & Zeitlinger.
- Elstein, A.S., Shulman, L.S., & Sprafka, S.A. (1978). *Medical Problem Solving*. Cambridge: Harvard University Press.
- Gale, J., & Marsden, P. (1982). Clinical problem solving; the beginning of the process. *Medical Education*, 16, 22-26.
- Godt, J.Th.H., Malin, J.P., & Wittenberg, A. (1981). *Das Schulter-Arm Syndrom*. Stuttgart: Thieme Verlag.
- Grond, J.Th.H. (1981). Degeneratieve tendinitis van de schouder. *Nederlands Tijdschrift voor Geneeskunde*, 125, 129-133.
- Hodgkin, K., & Knox, J.D.E. (1975). *Problem-centred Learning*. Edinburgh: Churchill Livingstone.
- Johnson, P.E., Duran, A.S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, O.J., & Swanson, D.B. (1982). *Cognitive Science*, 6, 235-283.
- Kingma, M.J. (1980). Schouderpijn. *Nederlands Tijdschrift voor Geneeskunde*, 120, 129-133.
- Newble, D.I., Hoare, J., & Baxter, A. (1982). Patient management problems. Issues of validity. *Medical Education*, 16, 137-142.
- Oostendorp, R.A.B. (1976). Fysiotherapeutische benadering van het schouderprobleem. *Nederlands Tijdschrift voor Fysiotherapie*, 12, 309-315.
- Oostendorp, R.A.B. (1980). Oorzakelijke factoren, onderzoek en behandeling van de periarthrititis humeroscapularis. *Nederlands Tijdschrift voor Fysiotherapie*, 12, 264-275.
- Stenvers, J.D., & Overbeek, W.J. (1979). Bestaat bij de frozen shoulder toch ook benige beperking? *Oefentherapie Mensendieck*, 40, 13-20.
- Van der Sijde, P.C., & Sellink, W.J.L. (1983). *Ontwikkeling van audiovisuele programmatuur voor het leren oplossen van fysiotherapeutische problemen*. Interimrapport Amsterdam: SAFA.
- Van der Sijde, P.C., Sellink, W.J.L., & Wurms, R.J. (1987). Developing audiovisuals for problem solving in physical therapy education. *Physical Therapy*, 67, 554-557.
- Westin, S., Salvesen, K., & Gotestam, K.G. (1986). Problem solving styles of medical students related to knowledge and personality in simulated clinical situations. *Medical Education*, 20, 109-116.

Manuscript ontvangen 17-2-1987

Definitieve versie ontvangen 1-7-1987

Convergentie en Determinanten van Beoordelingen van Onderzoeksaanvragen Voor en Na de Commissiebespreking

T.H.A. van der Voort, B. Pangemanann* en M.W. Vooijs
Subfaculteit Opvoedkunde, RU Leiden.

ABSTRACT

The peer review process was evaluated by analyzing the ratings that 33 research proposals received from a Dutch proposal review panel. The panel was composed of 9 university professors of various social and behavioral sciences, and the proposals covered the same field. Each single panel member rated the proposals both before (pre-ratings) and directly after the proposal's discussion (post-ratings). The proposals were rated as being good, fair or poor on two criteria: the expected quality (EQ) of the scientific investigation described in the proposal and its expected significance (ES). Before the panel meeting, the reviewers held widely divergent views on the expected quality and significance of the proposals. The average interrater correlation was as low as .08 (ES) and .17 (EQ), the result being that the panel's ratings of the proposals were quite unreliable. Due to the panel discussions, however, the disagreements among the reviewers were reduced: the average interrater correlation rose to .39 (ES) and .53 (EQ). As appears from the poor discriminant validity coefficients, the reviewers did not make a clear distinction between the proposals' expected quality and their expected significance. The panel discussion produced a sharp decline in the panel's appreciation of the proposals, a phenomenon that may be specific for the committee under consideration. The following characteristics of the proposals favored a high pre- and post-rating: (1) the scientific competence of the applicants as evidenced by their number of international publications and the membership of a prestigious academy; (2) a high number of departments acting together in the proposed investigation; (3) proposals in the field of the behavioral sciences. These characteristics predicted the pre-ratings better than the post-ratings. Apparently, the insights gained during the panel meeting gave cause to take the proposals' objective characteristics more lightly. It is questioned whether the use of panels improves the peer review process. The panel discussion does reduce the differences of opinion within panels, but that does not necessarily apply to the differences of opinion between panels.

1. INLEIDING

1.1. Overeenstemming tussen beoordelaars

Blijkens enkele Amerikaanse onderzoekingen (Cole & Cole, 1979; Cole, Cole, & Simon, 1981) en een tweetal recente Nederlandse wetenschapsstudies (Hofstee, 1983; Scheerens & Beem, 1986) bestaat er tussen beoordelaars die onafhankelijk van elkaar dezelfde subsidie-aanvragen voor onderzoek beoordelen een onthutsend gebrek aan overeenstemming. Het Amerikaanse en Nederlandse onderzoek had betrekking op verschillende soorten beoordelingsprocedures. In het Amerikaanse onderzoek ging het om postale beoordelingen van onderzoeksaanvragen

Adres: Sectie Kind & Media, Rijnsburgerweg 169, 2334 BP Leiden.

* Thans werkzaam bij het Instituut voor Medische Statistiek te Den Haag.

(*mail review*). De National Science Foundation (NSF) nodigde per aanvraag vier tot vijf beoordelaars uit een schriftelijke beoordeling uit te brengen, waaruit vervolgens door een ambtenaar ten huize van de NSF een eindadvies werd gedistilleerd. Het Nederlandse onderzoek had betrekking op commissie-beoordelingen (*panel review*). Bij de Stichting voor Onderzoek van het Onderwijs (SVO) ingediende aanvragen werden beoordeeld door een vaste commissie, die bijeenkwam om in onderling overleg een gemeenschappelijk advies uit te brengen.

Postale beoordeling

Bij hun onderzoek naar de aanvraagkenmerken die de schriftelijke beoordelingen en de uiteindelijke NSF-beslissing beïnvloedden, stuitte Cole en Cole (1979) op een onverwacht geringe overeenstemming tussen de beoordelaars. Bij elk van de tien disciplines die in het onderzoek betrokken waren, werd blijkens variantie-analyses ongeveer de helft van de beoordelingsvariantie door meningsverschillen tussen beoordelaars verklaard. Tegen de verwachting van de onderzoekers in, was de consensus bij beoordelingen van aanvragen uit de natuurwetenschappen even ver te zoeken als bij sociaal-wetenschappelijke aanvragen.

In een vervolgstudie legden Cole et al. (1981) het gebrek aan overeenstemming tussen beoordelaars bloot door 150 reeds door de NSF afgehandelde aanvragen door nieuwe clubs van postale beoordelaars te laten herbeoordelen. De aanvragen in kwestie kwamen uit drie onderscheiden disciplines, die elk met 50 aanvragen vertegenwoordigd waren. Van elke set van 50 aanvragen was de helft inmiddels reeds door de NSF toegewezen, en de andere helft was afgewezen. De correlatie tussen de over beoordelaars gemiddelde oorspronkelijke en heroordeelen varieerde bij de drie disciplines van .60 tot .66.

Door halverwege in de op basis van de herbeoordelingen opgestelde rangorde van de voorstellen een streep te zetten, kon worden vastgesteld welke aanvragen in de prijzen waren gevallen als de NSF haar beslissingen op de herbeoordelingen had gebaseerd. Bij de drie onderzochte disciplines zou dit voor 24 tot 30% van de aanvragen tot een tegengestelde beslissing hebben geleid. Hieronder bevonden zich ook gevallen die bij de eerste beoordeling in het hoogste kwintiel vielen en bij de herbeoordeling zouden zijn afgewezen.

Commissie-beoordeling

Anders dan bij postale beoordelingen, waarbij de beoordelaars doorgaans met de aanvraag wisselen, kennen commissies meestal een vaste samenstelling. Men trekt langere tijd met elkaar op, waardoor de commissie een geschiedenis krijgt. Het is daarom denkbaar dat de leden van de beoordelingscommissie gaandeweg meer op eenzelfde lijn gaan zitten. Beoordelaars die van een vaste commissie deel uitmaken, zouden daarom een sterkere onderlinge overeenstemming kunnen vertonen dan postale beoordelaars.

Voor de vaststelling van de consensus doet het uiteindelijke commissie-oordeel niet terzake, want daarin ziet men, tenzij minderheidsadviezen zijn toegestaan, van de oorspronkelijke meningsverschillen niets terug. Wel kan men nagaan in hoeverre de commissieleden het onderling eens zijn op het moment dat zij de vergaderzaal betreden (overeenstemming tussen voorbeoordelingen). Daarnaast kan men vaststellen hoe elk commissielid direct na de discussie over de voorstellen denkt (overeenstemming tussen nabeoordelingen). Om dit te onderzoeken moet elk commissielid voorafgaand aan en direct volgend op de discussie het eigen oordeel op één of meer schalen vastleggen, iets dat zelden gebeurt. Bovendien moeten de commissieleden bereid zijn die gegevens voor onderzoek vrij te geven, een nog grotere zeldzaamheid, die buiten onze landsgrenzen nog nooit is vertoond. Twee Nederlandse studies geven inzicht in de overeenstemming tussen beoordelingen die voorafgaand aan het commissie-overleg zijn gegeven. Ons eigen onderzoek werpt behalve op de convergentie van voorbeoordelingen tevens licht op de overeenstemming tussen de nabeoordelingen die de afzonderlijke commissieleden direct na de bespreking van elke aanvraag gaven.

Hofstee (1983) analyseerde de voorbeoordelingen van 54 subsidie-aanvragen voor onderwijs-onderzoek. Voorafgaand aan het overleg beoordeelde elk commissielid op vijfpuntschalen de aanvragen op een zestal beoordelingscategorieën. Daarnaast gaf men tevoren een concluderend

totaaloordeel, waarbij uit de schaalpunten 'toekennen', 'neen, tenzij' en 'afwijzen' gekozen kon worden. De correlatie tussen deze totaaloordelen bedroeg, gemiddeld over paren van beoordelaars, niet meer dan .14. De betreffende correlatie was voor de somscores over de zes categorie-oordelen een ietsje hoger: .21. Vult men de laatste correlatie-coëfficiënt in de Spearman-Brown-formule in, dan vindt men voor de beoordelaars samen een geschatte betrouwbaarheid van .58. Dit cijfer geeft een indicatie voor de hoogte van het verband dat men bij herbeoordeling door een parallelle commissie tussen beide commissie-oordelen zal aantreffen. De voorspelde correlatie ligt, zoals Hofstee (1985) in een later artikel vaststelt, inderdaad dicht bij de correlaties die Cole, Cole en Simon (1981) – zie hiervoor – feitelijk aantreffen.

Hofstee (1983) sluit de volgende weddenschap af: '(...) ongeacht bijzondere omstandigheden, (zal) in alle commissies die onderzoeksaanvragen beoordelen, de feitelijke (niet de subjectief ervaren) overeenstemming laag zijn, d.i. in 9 van de 10 gevallen in de orde van .00 tot .30, fluctuaties als gevolg van kleine steekproeven daargelaten' (p. 282). Daarnaast meent hij, dat de betreffende SVO-beoordelingscommissie in het vervolg er beter aan doet de somscores over de categorie-oordelen (in plaats van de globale totaaloordelen) als uitgangspunt voor discussie te nemen, dit omdat de somscores een hogere convergentie laten zien en niet minder met het uiteindelijke commissie-advies correleren.

Die laatste conclusie wordt niet gedeeld door Scheerens en Beem (1986), die behalve de 54 door Hofstee geanalyseerde aanvragen nog 185 andere bij de SVO ingediende aanvragen in de analyse betrokken. Volgens deze analyse was de convergentie bij de somscores niet groter dan bij de totaaloordelen, zodat beide typen oordelen met evenveel recht als discussie-uitgangspunt genomen kunnen worden. Maar zijn weddenschap won Hofstee (1983) met glans. De tussen-beoordelaar-overeenstemming viel nog lager uit dan in Hofstee's onderzoek en bewoog zich rond .10. Omdat in deze studie nieuwe beoordelingen meededen van dezelfde commissie die in Hofstee's onderzoek werd doorgelicht, en de convergentie hierbij nog lager was, komen Schererens en Beem tot de slotsom dat hun bevindingen geen steun geven aan een '(...) learning effect in the sense of increased agreement, which perhaps might be expected when committees have had more practice' (p. 72).

Verklaringen voor het gebrek aan consensus

Voorafgaand aan de discussie of bij afwezigheid van discussie (postale beoordeling) blijkt de overeenstemming tussen beoordelaars welhaast ongeloofwaardig laag uit te vallen, dit ook bij commissies met de nodige ervaring. Met Hofstee (1985) gaan we ervan uit, dat de verklaring niet in nalatigheid van de beoordelaars moet worden gezocht. We nemen aan, dat beoordelaars zeer wel beseffen wat er op het spel staat, en naar beste vermogen een zo gefundeerd mogelijk oordeel geven.

In de literatuur treft men een drietal verklaringen aan, die elkaar geenszins uitsluiten. Humphreys (1982) wijst op de psychologische drempel die onderzoeksbeoordeling kan opwerpen. Om faalervaringen te vermijden zouden met name de zwakkere vakbroeders van de indiening van een aanvraag af kunnen zien. Door deze zelf-selectie vertoont het aangeboden pakket aanvragen een geringere spreiding in kwaliteit, met het bekende gevolg dat de correlaties gedrukt worden. De paradoxale situatie doet zich voor dat bij een toenemende zelf-selectie, hetgeen op zichzelf meer reden geeft aan te nemen dat de toegekende subsidies op de juiste plaats terechtkomen, de tussen-beoordelaar-overeenstemming juist afneemt.

Een tweede verklaring is, dat bij de grove en tamelijk onbeholpen ratingschaaltjes die doorgaans gebruikt worden, antwoordtendenties kunnen optreden die ertoe leiden dat beoordelaars met een identieke opvalling toch een verschillend schaalpunt aankruisen (Cole et al., 1981).

Een derde verklaring is, dat de meningsverschillen voortvloeien uit het feit, dat beoordelaars verschillend tegen onderzoeksvoorstellen aankijken. Elke beoordelaar beschouwt de aanvragen vanuit een eigen idiosyncratische optiek. Tot op heden is het niet gelukt idiosyncratische optieken op het spoor te komen waarin een vaste lijn is te onderkennen. Elke beoordelaar lijkt

bij elke aanvraag weer andere accenten te leggen, waardoor zijn reactie moeilijk te voorspellen is (Hofstee, 1983).

Het gesignaleerde gebrek aan overeenstemming tussen beoordelaars wordt overigens niet algemeen als zorgwekkend ervaren. Swanborn (1986) heeft erop gewezen, dat niet is onderzocht wat het uiteindelijke advies van de commissie waard is. Volgens Swanborn heeft geen enkel commissielid alle wijsheid in pacht en is een ieder daarvan ook doordrongen. De voorbeoordeling waarmee men de commissie binnenstapt wordt daarom als een voorlopig oordeel gezien, en men is graag bereid dat voor een betere te geven. Het eindoordeel van de commissie zou daarom kwalitatief beter zijn dan de som van de initiële individuele opinies. Op deze optimistische visie komen we nog terug.

1.2. Beoordelingsdeterminanten

Behalve van de convergentie van de beoordelingen is tevens enige studie gemaakt van de determinanten die de beoordeling beïnvloeden. In het uitgevoerde onderzoek is vooral nagegaan welke kenmerken van de aanvragen of de aanvragers van invloed zijn op de uitslag van de beoordeling. Daarnaast is in een enkel onderzoek nagegaan of kenmerken van de beoordelaars zelf de uitslag beïnvloeden.

Kenmerken beoordelaars

Het onderzoek naar de invloed van kenmerken van de beoordelaars is beperkt gebleven tot een toetsing van wat we de *ouwe-jongens-krentenbrood-hypothese* noemen ('old boy hypothesis'). Volgens Cole, Rubin en Cole (1977) kan deze hypothese op tenminste drie manieren worden opgevat. Een eerste interpretatie is, dat beoordelaars met eenzelfde wetenschapsopvatting geneigd zijn alleen aanvragen die van dezelfde opvatting getuigen gunstig te beoordelen. In deze betekenis is de hypothese nog niet getoetst.

Ten tweede kan de hypothese op netwerken van vriendschappen slaan. Wetenschappers die elkaar goed kennen zouden elkaar de bal toespelen en elkaars aanvragen gunstig beoordelen. Scheerens en Beem (1986) onderzochten dit door na te gaan of beoordelaars aanvragen afkomstig uit de eigen instelling gunstiger beoordelen dan aanvragen uit instellingen waarmee zij geen binding hebben. Bij de meeste beoordelaars bleek dit niet het geval. Slechts één beoordelaar leek de aanvragen uit de eigen instelling 'voor te trekken'.

Ten derde kan de hypothese op de sociale positie van de beoordelaars betrekking hebben. In deze interpretatie is het vermoeden, dat beoordelaars die een gegeven positie in de wetenschapshierarchie innemen aanvragen van onderzoekers met eenzelfde positie het gunstigst zullen beoordelen, dit óók als men hen niet kent. Deze hypothese is op geen enkele manier bevestigd. Blijkens een Amerikaanse studie beoordelen gereputeerde beoordelaars de aanvragen van dito onderzoekers of instellingen niet gunstiger dan beoordelaars die een lage positie in de wetenschapshierarchie innemen (Cole et al., 1977). En Nederlandse studies laten zien, dat louter uit hoogleraren samengestelde beoordelingscommissies door professoren ingediende aanvragen niet gunstiger beoordelen dan aanvragen die door doctorandi of doctores zijn ingediend (Hofstee, 1983; Scheerens & Beem, 1986).

Kenmerken aanvragers of aanvragen

Sommigen vrezen dat het systeem van collegiale beoordeling ertoe leidt, dat de 'rijken steeds rijker worden'. Gereputeerde onderzoekers en onderzoekers uit excellente onderzoeksinstituten zouden, ongeacht de kwaliteit van hun voorstellen, een grotere kans op subsidietoewijzing hebben. Blijkens de Amerikaanse studie van Cole et al. (1977) hebben bij een gerenommeerde instelling afgestudeerde of werkzame aanvragers inderdaad een grotere kans op subsidietoewijzing. Hetzelfde geldt voor aanvragers die reeds eerder subsidies kregen, een groot aantal publicaties op hun naam hebben en vaak geciteerd worden.

Aanvragers met een goede staat van dienst blijken dus bij de NSF-procedure hun aanvragen inderdaad vaker gehonoreerd te zien. Deze bevinding geeft op zichzelf echter geen enkele reden

om de beoordelaars iets te verwijten, omdat de onderzoeksprestaties die de aanvrager in het verleden heeft geleverd een zeer geschikte voorspeller van de kwaliteit van toekomstig onderzoek kan zijn. Bovendien vereist de beoordelingsprocedure in kwestie dat er expliciet op de reeds geleverde onderzoeksprestaties wordt gelet. Het systeem valt pas iets te verwijten, wanneer het ertoe zou leiden dat onderzoekers die aan het begin van hun carrière staan niet aan de bak komen. Daarvoor bestaan echter geen aanwijzingen. De negen variabelen die Cole, et al. (1977) als indicatoren voor de wetenschappelijke status van de aanvragers hanteerden, bleken met elkaar niet meer dan 16% van de beoordelingsvariantie te verklaren.

Het onderzoek naar de invloed van kenmerken van de aanvragen zelf op de hoogte van de beoordelingen heeft weinig opgeleverd. Factoren als de lengte van de aanvraag, het aantal literatuurverwijzingen en dergelijke bleken niet of nauwelijks met de oordelen samen te hangen (Spetz, 1979; Hofstee, 1983; Scheerens & Beem, 1986).

1.3. Vraagstelling

Het huidige onderzoek heeft betrekking op de beoordelingen die in 1984 door de Commissie ad hoc Discipline-Advies Sociale Wetenschappen (DAS) over voorstellen voor zwaartepuntfinanciering zijn uitgebracht. Tijdens de bijeenkomsten waarin de commissie de zwaartepuntvoorstellen beoordeelde verleende de eerste auteur dezes hand- en spandiensten bij de verslaglegging en de rekenkundige verwerking van de schriftelijk uitgebrachte beoordelingen van de voorstellen. Hierdoor kregen we de beschikking over alle individuele beoordelingen die door de commissieleden zijn uitgebracht. De commissieleden hebben met het gebruik van de beoordelingsgegevens voor onderzoeksdoeleinden ingestemd, onder het vanzelfsprekende voorbehoud dat deze anoniem zouden worden behandeld.

De verzamelde beoordelingsgegevens zijn voor een deel gelijksoortig aan die welke door Hofstee (1983) en Scheerens en Beem (1986) zijn geanalyseerd. Net als bij de SVO-beoordelingen beoordeelde elk DAS-lid namelijk voorafgaand aan de commissiebespreking alle ingediende aanvragen. Maar het eindadvies van de commissie kwam op een geheel andere wijze tot stand. Bij de SVO-beoordelingen werd na de commissiebespreking een collectief eindadvies opgesteld, terwijl bij de DAS-procedure elk commissielid onmiddellijk volgend op de bespreking van elk voorstel voor zichzelf een herbeoordeling noteerde. Het eindadvies kwam tot stand door de individuele herbeoordelingen van de voorstellen te middelen.

Onze analyses zijn gericht op een viertal punten:

1. Ten eerste gaan we na welke objectief vaststelbare kenmerken de ingediende zwaartepunt-aanvragen vertonen.
2. Vervolgens bezien we de convergentie, betrouwbaarheid en de discriminante validiteit van de uitgebrachte voor- en nabecoordelingen. Dit biedt onder meer de gelegenheid Hofstee's voorspelling houdend dat de overeenstemming tussen de beoordelaars de .30 niet te boven zal gaan nog eens op de proef te stellen.
3. Ten derde gaan we na welke verschuivingen de commissiebespreking in de initiële oordelen van de commissieleden teweeg heeft gebracht.
4. Tenslotte stellen we vast welke objectief vaststelbare kenmerken van de aanvragen een rol bij de beoordeling hebben gespeeld.

2. PROCEDURE

De analyses zijn verricht op de herhaalde beoordelingen die door de individuele leden van de DAS-commissie over elk van de 33 behandelde zwaartepuntvoorstellen schriftelijk zijn uitgebracht.

Beoordelaars

De DAS-commissie stond onder voorzitterschap van prof. dr. R.A. de Moor (secretaris: drs.

R.D. Bennink) en telde elf leden. Aan de beoordeling van de zwaartepuntvoorstellen namen negen leden deel. Van deze negen beoordelaars waren er vijf hoogleraar in een sociaal-culturele wetenschap (2 sociologen, 2 politicologen en een cultureel antropoloog) en vier hoogleraren in een gedragswetenschap (3 psychologen en 1 pedagoog).

Aanvragen

Elke universitaire instelling die over een sociale faculteit beschikte kon maximaal vijf voorstellen voor zwaartepuntonderzoek indienen, waarbij interuniversitaire voorstellen geteld werden voor de reciproke van het aantal deelnemende instellingen. Omdat in de meeste instellingen meer dan vijf voorstellen voor indiening werden voorgedragen, vond er binnen de instellingen een voor-selectie plaats, een taak waarmee de betrokken facultaire wetenschapscommissies meestal belast werden. Begin 1984 werden alles bijeen 36 voorstellen bij de DAS-commissie ingediend, waarvan er drie op formele gronden afvielen. De commissie had als taak uit die 33 voorstellen de 12 beste te selecteren. De indieners was voorgespiegeld dat het uit te kiezen dozijn zwaartepuntvoorstellen gedurende een zestal jaren mocht rekenen op een extra-budgettaire toewijzing van ongeveer een half miljoen gulden per jaar.

Een zwaartepunt was gedefinieerd als een langdurig programma van onderzoek op een welomschreven gebied, dat beloofde een belangrijke bijdrage aan de vermeerdering van kennis op het betrokken terrein te leveren. Voorts diende een zwaartepunt een gespecialiseerde onderwijsvoorziening van hoge kwaliteit te zijn, met een centrale plaats in de post-doctorale opleiding van onderzoekers. De zwaartepuntfinanciering was, aldus de commissie, vooral bedoeld om het fundamentele onderzoek te versterken (DAS, 1984).

Beoordelingsmaatstaven

De voorstellen werden op twee maatstaven beoordeeld: de verwachte kwaliteit en het verwachte belang van het voorstel. Bij de *verwachte kwaliteit* moest een schatting gegeven worden van de te verwachten productie, gegeven de bij het voorstel betrokken personen, opzet, organisatie en de haalbaarheid van de gestelde doelen, maar ongeacht de aard van die doelen. Bij de oordeelsvorming over de verwachte kwaliteit diende te worden gelet op aspecten als:

- is het onderzoeksgebied duidelijk omschreven?
- is duidelijk welk nieuw onderzoek zal worden uitgevoerd?
- is de taak- en verantwoordelijkheidsverdeling duidelijk?
- wat is de kwaliteit van de programmaleiders?
- is er een goede balans van gevraagde middelen en gestelde doelen?

Bij de beoordeling van het *verwachte belang* ging het niet om het belang van het onderwerp maar van de te verwachten productie, waarbij de wetenschappelijke en maatschappelijke betekenis samengenomen dienden te worden. Bij de oordeelsvorming over het verwachte belang moest worden gelet op aspecten als:

- leidt het onderzoek tot een belangrijke bijdrage aan de wetenschap?
- zijn er gebruiksmogelijkheden voor beleidsdoeleinden?
- wordt interdisciplinaire samenwerking bevorderd?
- zijn er spin-offs voor het wetenschappelijk onderwijs?

Beoordelingsprocedure

De beoordeling verliep in drie ronden: (1) voorbeoordeling; (2) commissiebespreking; en (3) eindbeoordeling. Bij de voorbeoordeling die voorafgaand aan het commissie-overleg plaats vond, gaf elk commissielid over elk voorstel aan de hand van een driepuntschaal (onvoldoende/voldoende/goed) een oordeel over de verwachte kwaliteit en het verwachte belang. Bij het oordeel over het verwachte belang (B) diende men het gegeven oordeel over de verwachte kwaliteit (K) in aanmerking te nemen (zie: De Groot, 1985).

Onmiddellijk voorafgaand aan de plenaire bespreking werden de K- en B-oordelen over de groep van negen gemiddeld, hetgeen in een K- en een B-rangorde van de voorstellen resulteerde. Met deze initiële rangordes in handen vond de commissiebespreking plaats. Zodra een voorstel

besproken was vulden alle leden staande de vergadering met gebruikmaking van dezelfde driepuntschaal een nieuw K- en B-oordeel in. Commissieleden die rechtstreeks bij een voorstel betrokken waren, dat wil zeggen er zelf aan deelnamen, verlieten tijdelijk de zaal om weer terug te komen nadat de overige leden hun oordelen hadden genoteerd. Aan de beoordeling van voorstellen uit de eigen instelling waaraan niet zelf werd deelgenomen, deed elk commissielid onverkort mee.

Nadat alle voorstellen besproken waren, werd na middeling over de beoordelaars een nieuwe K- en B-rangorde vastgesteld. Na confrontatie met dit resultaat besloot de commissie haar advies te baseren op de B-rangorde (verwacht belang). De bovenste 13 voorstellen uit de B-ranglijst zijn voor financiering bij de Minister van Onderwijs en Wetenschappen voorgedragen. Zoals bekend heeft deze het advies goeddeels naast zich neergelegd.

Kenmerken van de aanvragen

De volgende objectief vaststelbare kenmerken van de aanvragen zijn in het onderzoek betrokken:

1. het aantal instellingen dat de aanvraag indiende;
2. meest betrokken discipline (de discipline die qua personele bezetting het zwaarst in de aanvraag vertegenwoordigd is);
3. aantal in de aanvraag participerende vakgroepen;
4. aantal in de aanvraag participerende hoogleraren (voor zover werkzaam bij een sociale faculteit);
5. aantal deelnemende medewerkers (inclusief hoogleraren, exclusief aangevraagd personeel);
6. aantal gepromoveerde medewerkers (inclusief hoogleraren);
7. lengte aanvraag (totaal aantal bladzijden);
8. aantal publicaties (na verwijdering rapporten en interne publicaties);
9. aantal internationale publicaties (na verwijdering interne publicaties);
10. aantal gerealiseerde dissertaties (begeleid door bij de aanvraag betrokken hoogleraren);
11. aantal subsidies verkregen uit de tweede of derde geldstroom;
12. KNAW-lidmaatschap (maakt één of meer deelnemers deel uit van de KNAW?; opgenomen als indicator voor de reputatie van de aanvragers);
13. DAS-lidmaatschap (behoren één of meer DAS-leden tot de aanvragers?; opgenomen om vast te stellen of voorstellen waarbij DAS-leden zijn betrokken gunstiger zijn beoordeeld).

3. RESULTATEN

3.1. Kenmerken van de aanvragen

Alvorens onze hoofdvragen aan te snijden, staan we eerst kort stil bij de feitelijke kenmerken die de aanvragen vertoonden. Zoals Tabel 1 duidelijk maakt, gingen de meeste voorstellen (23) van één instelling uit, terwijl in een tiental voorstellen interuniversitaire samenwerking werd gezocht. Omdat aan de indiening van voorstellen meestal een voor-selectie op facultair niveau voorafging, is het interessant te zien hoe het uiteindelijk ingediende pakket van voorstellen disciplinair verdeeld is. De voorstellen uit de hoek van de psychologie zijn er bij de prioritering die binnen de instellingen plaats vond het gunstigst afgekomen. Bij bijna de helft van de ingediende voorstellen is de psychologie de meest betrokken discipline. Gemeten naar de relatief geringe omvang die de subfaculteiten sociologie in ons land kennen, heeft ook de sociologie het facultaire selectieproces goed doorstaan. De culturele antropologie annex sociologie der niet-westerse volken en de pedagogiek zijn in het totaal van ingediende aanvragen duidelijk ondervertegenwoordigd.

Behalve getalsmatig komt de psychologie ook in een ander opzicht goed uit de bus. Bijna de helft van de in de psychologische voorstellen opgevoerde publicaties heeft een internationaal karakter (47%). In de twee cultureel-antropologische voorstellen is dit percentage nog iets

hoger (56%). Bij de sociologische en politicologische c.q. bestuurskundige voorstellen valt het percentage opgevoerde internationale publicaties duidelijk lager uit (ongeveer 30%). In de twee ingediende pedagogische voorstellen is het aantal opgevoerde internationale publicaties het geringst (14%).

Afgezien van interne publicaties en rapporten, werden in de voorstellen gemiddeld 121 publicaties van deelnemers opgevoerd. De spreiding was echter groot: het laagste aantal opgevoerde publicaties was 32 en het dikste voorstel vermeldde er 386. Gemiddeld per medewerker werden ongeveer 8 publicaties opgevoerd. Ongeveer 40% van alle opgevoerde publicaties had een internationaal karakter. Dit verschilde sterk per voorstel. In het minste geval was 9% van de opgevoerde publicaties internationaal, terwijl in het beste geval 90% van de opgevoerde publicaties in den vreemde was verschenen.

Het 'gemiddelde' voorstel ging uit van drie vakgroepen en werd gedragen door een zestiental medewerkers uit de zittende formatie (exclusief aangevraagd personeel). Van die 16 deelnemers waren er gemiddeld 4 hoogleraar, en van de resterende 12 was ongeveer de helft gepromoveerd.

Tabel 1: Verdeling Zwaartepuntaanvragen naar Discipline en het Aantal Instellingen waarvan de Aanvraag Uitgaat

Discipline	Eén instelling	Meerdere instellingen	Totaal
Psychologie	12	4	16
Sociologie	7	3	10
Politicologie/Bestuurskunde	3	—	3
Pedagogiek	1	1	2
CA/SNWW	—	2	2
Totaal	23	10	33

3.2. Convergentie, betrouwbaarheid en discriminante validiteit

Convergentie

De eerder aangehaalde voorspelling van Hofstee (1983), waarin gesteld wordt dat de overeenstemming binnen beoordelingscommissies zich in negen van de tien gevallen tussen .00 en .30 zal bewegen, slaat uitsluitend op de overeenstemming, of liever het gebrek daaraan, waarvan voorafgaand aan het commissie-overleg sprake is. In ons geval kan dus alleen de bij de *voorbeoordeling* geconstateerde overeenstemming als toets dienen. Zoals Tabel 2 uitwijst, is bij de voorbeoordeling de overeenstemming tussen de beoordelaars inderdaad zeer gering. Bij de voorbeoordeling op *kwaliteit* is de tussen-beoordelaar-overeenstemming, gemiddeld over alle paren van beoordelaars, slechts 0.17, een waarde die ruim beneden het in Hofstee's voorspelling gespecificeerde plafond van 0.30 ligt. Bij inspectie van de oorspronkelijke 9 x 9-overeenstemmingsmatrix, die om ruimte te sparen niet is afgedrukt, blijkt dat er bij veel paren van beoordelaars tussen de gegeven kwaliteitsoordelen nauwelijks verband bestaat. Vooral beoordelaar 8 ($\bar{r} = -.02$) lijkt op een geheel eigen wijze tegen de kwaliteit van de voorstellen aan te kijken: in vijf gevallen correleren zijn kwaliteitsoordelen zelfs negatief met de oordelen van de overige beoordelaars.

Was de overeenstemming tussen beoordelaars bij de voorbeoordeling op kwaliteit reeds ver te zoeken, bij de *voorbeoordeling op belang* valt de overeenstemming nog lager uit. Over alle paren van beoordelaars is de gemiddelde tussen-beoordelaar-correlatie slechts 0.08. Een blik op de, weer niet afgedrukte oorspronkelijke overeenstemmingsmatrix leert, dat bij de meeste paren de samenhang tussen de gegeven oordelen gering is, en bij iets minder dan de helft van de paren

Tabel 2: Gemiddelde Tussen-Beoordelaar-Correlaties per Beoordelaar bij de Voor- en Na-Beoordelingen op Kwaliteit (K) en Belang (B)

	Beoordelaars									
	1	2	3	4	5	6	7	8	9	1-9
K vóór	.21	.31	.21	.19	.20	.13	.17	-.02	.16	.17
K ná	.52	.61	.43	.59	.53	.48	.55	.52	.56	.53
B vóór	.08	.21	.05	.17	.06	.08	.08	-.08	.10	.08
B ná	.48	.43	.45	.30	.30	.34	.39	.36	.45	.39

is de correlatie negatief. Wanneer de commissie het 1000 pagina's omvattende pakket voorstellen ongelezen had gelaten en met een dobbelsteen over het belang van de voorstellen had beslist, had deze correlatiematrix er niet veel anders uitgezien.

Onder invloed van de commissiebespreking verandert dit beeld echter drastisch. Na de gedachtenwisseling over de aanvragen blijkt een aanzienlijk deel van de initiële meningsverschillen te zijn opgeheven. Dit geldt vooral voor de kwaliteitsbeoordelingen, waarbij de gemiddelde tussen-beoordelaar-correlatie over alle paren van beoordelaars toeneemt van 0.17 (vóór de discussie) tot 0.53 (ná de discussie). Bij de oordelen over het belang van de voorstellen neemt de overeenstemming iets minder toe: de gemiddelde tussen-beoordelaar-correlatie stijgt van 0.08 (vóór de discussie) tot 0.39 (ná de discussie).

Betrouwbaarheid

De betrouwbaarheid van de commissie-oordelen schatten we voor elk van beide beoordelingscriteria afzonderlijk. De DAS-commissie vat de kwaliteit en het belang namelijk niet als gelijkwaardige criteria op. Tussen de oordelen op beide criteria bestaat een voorwaardelijke relatie: het B-oordeel wordt *gegeven* het K-oordeel geveld, en het B-oordeel acht de commissie beslissend. Het is daarom niet gerechtvaardigd de K- en B-oordelen in een somscore samen te nemen.

Tabel 3 vermeldt de resultaten van twee typen schattingsprocedures. De betrouwbaarheid is in de eerste plaats uitgedrukt in een *coëfficiënt alpha*, die zich laat berekenen als men de aanvragen als 'proefpersonen' en de beoordelaars als 'items' opvat. Daarnaast geeft de tabel schattingen die met behulp van de *Spearman-Brown-formule* zijn berekend, een procedure die Hofstee (1983) heeft toegepast. Bij deze schattingsprocedure worden de beoordelaars eveneens als 'items' opgevat, en wordt de betrouwbaarheid geschat door met gebruikmaking van een verlengingsfactor van 9 (= het aantal 'items') de Spearman-Brown-formule op de gemiddelde tussen-beoordelaar-correlatie (= de 'tussen-item-correlatie') toe te passen (cf. Stanley, 1957).

De resultaten van beide schattingsprocedures ontlopen elkaar weinig (en formeel gezien is de Spearman-Brown-schatting, als de standaarddeviaties van de beoordelaars gelijk zijn, identiek aan alpha). Duidelijk is dat de betrouwbaarheid van het commissie-oordeel over zowel de kwaliteit ($r_{xx} = 0.65$) als het belang ($r_{xx} = 0.42$ à 0.44) van de voorstellen bij de voorbeoordeling veel te wensen overlaat. Dankzij de bespreking van de voorstellen neemt de betrouwbaarheid echter sterk toe. Bij de nabeoordeling is de betrouwbaarheid van het kwaliteitsoordeel van de commissie 0.89 à 0.91 en van de oordelen over het belang 0.83 à 0.85. Wanneer men een waarde van 0.90 als een acceptabele betrouwbaarheid ziet, voldoen de nabeoordelingen op kwaliteit, althans volgens één van beide schattingsprocedures, aan die eis.

Tabel 3: Betrouwbaarheidscoëfficiënten voor de Commissie-Oordelen over Kwaliteit en Belang

Schattingsprocedure	K-vóór	K-ná	B-vóór	B-ná
Coëfficiënt alpha	0.65	0.89	0.42	0.83
Spearman-Brown	0.65	0.91	0.44	0.85

Discriminante validiteit

Een multitrait-multimethod-analyse kan uitwijzen in hoeverre bij de beoordeling een onderscheid tussen de kwaliteit en het belang van de voorstellen wordt gemaakt. Naar het voorbeeld van Hofstee (1983) worden hierbij de beoordelaars als 'methods' en de beoordelingscriteria kwaliteit en belang als 'traits' opgevat. Omdat de oordelen over kwaliteit en belang niet geheel onafhankelijk zijn – kwaliteit is immers een noodzakelijke doch onvoldoende voorwaarde voor een gunstig oordeel over het belang – ligt een perfect onderscheid tussen de oordelen over de kwaliteit en het belang niet in de rede. Maar als het om een zinvol onderscheid gaat, moet uit de analyse blijken dat de beoordelaars dit onderscheid tenminste tot op zekere hoogte hanteren.

Om ruimte te sparen presenteren we in plaats van de complete multitrait-multimethod-matrices alleen de overall-gemiddelden van de correlaties op de diagonaal en van de driehoeken erboven en eronder. Deze zijn weergegeven in Tabel 4. Gelet op de minimale heterobeoordelaar-monocriterium-correlaties ($\bar{r} = 0.13$) is de *convergente validiteit* bij de voorbeoordeling zeer gering. Bij de nabeoordeling ($\bar{r} = 0.46$) is wel van een zekere convergente validiteit sprake.

Bij de voorbeoordeling is ook de *discriminante validiteit* uiterst gering. Blijkens de hoge monobeoordelaar-heterocriterium-correlaties ($\bar{r} = 0.76$) maken de beoordelaars weinig onderscheid tussen de kwaliteit en het belang van de voorstellen. En bij inspectie van de oorspronkelijke matrix blijkt zelfs dat twee beoordelaars tijdens de voorbeoordeling helemaal geen onderscheid maken (hun K- en B-oordelen correleren perfect). Het ontbreken van discriminante validiteit blijkt tevens uit het feit, dat niet voldaan is aan de eisen dat de heterobeoordelaar-monocriterium-correlatie hoger moet zijn dan de heterobeoordelaar-heterocriterium-correlatie en de monobeoordelaar-heterocriterium-correlatie. Bij de nabeoordeling is het beeld een ietsje gunstiger. Hierbij is aan één van de twee laatstgenoemde eisen voldaan. Tevens is de monobeoordelaar-heterocriterium-correlatie ($\bar{r} = 0.59$ in plaats van 0.76) wat minder hoog, en ook de twee beoordelaars die aanvankelijk in het geheel geen onderscheid tussen de kwaliteit en het belang maakten, hanteren nu tot op zekere hoogte dit onderscheid.

Tabel 4: Gemiddelde Multibeoordelaar-Multicriterium-Correlaties Vóór en Ná de Commissiebespreking

Type Correlatie	Gemiddelde Correlatie	
	Vóór	Ná
Monobeoordelaar-heterocriterium	.76	.59
Heterobeoordelaar-monocriterium	.13	.46
Heterobeoordelaar-heterocriterium	.13	.41

3.3. Verschuivingen in de oordelen

Individuele beoordelaars

Omdat de overeenstemming tussen de beoordelaars onder invloed van de discussie duidelijk is toegenomen, moet tenminste een deel van de beoordelaars bereid zijn geweest om gehoord de discussie het aanvankelijke oordeel te herzien. De mate waarin elke beoordelaar zich door de discussie heeft laten beïnvloeden, kan uit de hoogte van de correlaties tussen zijn voor- en nabeoordelingen worden afgeleid. Deze ziet men in Tabel 5. Men kan deze correlaties als een soort *persistentie-index* zien: hoe hoger de correlatie des te minder heeft men zich door de argumenten die over tafel zijn gegaan laten beïnvloeden. Een hoge persistentiescore behoeft echter niet per se te betekenen, dat de beoordelaar de argumenten van anderen achter zijn oor steekt. Het zou ook kunnen betekenen, dat de beoordelaar de groep naar zich toe heeft getrokken. Om die mogelijkheid na te gaan zijn in Tabel 5 tevens de correlaties per beoordelaar tussen eigen voorbeoordeling en de groeps-eindbeoordeling opgenomen. Deze correlaties kan men als een *invloed-index* zien.

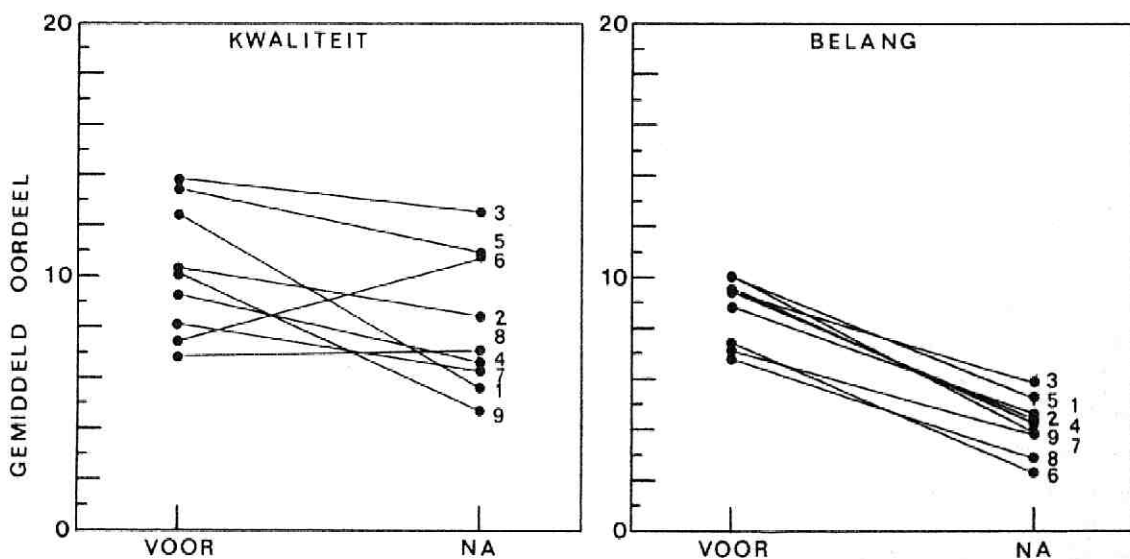
Tabel 5: Correlaties per Beoordelaar van de Voorbeoordeling met de Eigen Nabehoordeling en die van de Commissie, voor Kwaliteit en Belang

VOOR-OORDEEL VAN BEOORDELAAR	KWALITEIT		BELANG	
	Eigen na-oordeel	Eindoordeel commissie	Eigen na-oordeel	Eindoordeel commissie
1	.42	.30	.33	.08
2	.64	.59	.58	.45
3	.91	.54	.58	.41
4	.70	.44	.36	.31
5	.84	.60	.77	.57
6	.39	.41	.40	.34
7	.59	.40	.28	.20
8	.27	.01	.37	.08
9	.40	.33	.47	.24

Zoals Tabel 5 demonstreert, verschillen de beoordelaars sterk in de mate waarin zij hun initiële oordelen herzien. Sommige beoordelaars, beoordelaar 5 bijvoorbeeld, herzien hun initiële oordelen maar weinig, terwijl er ook beoordelaars zijn die hun aanvankelijke oordelen bijna integraal over boord zetten (onder andere beoordelaar 8). De persistentie-index valt bij de beoordelingen over het belang van de voorstellen ($\bar{r} = 0.46$) lager uit dan bij de kwaliteitsbeoordelingen ($\bar{r} = 0.57$). Blijkbaar heeft de commissiebespreking de grootste invloed op de meningsvorming over het belang van de voorstellen gehad.

Een hoge persistentiescore wil geenszins zeggen dat de beoordelaar blind voor de argumenten van anderen halsstarrig bij het initieel ingenomen standpunt blijft. De vermelde persistentie- en invloedsscores blijken namelijk sterk positief verbonden: de rangcorrelatie ertussen is bij de kwaliteitsbeoordeling 0.80 en bij de beoordeling op belang 0.87. De beoordelaars die het meest in hun oorspronkelijke standpunt volharden doen dit dus met een zeker 'recht', omdat zij er vaker in geslaagd zijn de groep van het eigen gelijk te overtuigen.

De vermelde correlaties zeggen nog niets over de richting waarin de oordelen van de commissieleden zich door de discussie hebben ontwikkeld. Figuur 1 geeft hierover uitsluitsel. Voor elke beoordelaar is weergegeven hoe zijn gemiddelde oordeel over de kwaliteit en het



Figuur 1: Gemiddelde Oordelen over de Kwaliteit en het Belang van de Voorstellen per Beoordelaar, Vóór en Ná de Discussie.

belang van de gezamenlijke voorstellen er voor en na de discussie uitzielt. Op twee beoordelaars na denken alle beoordelaars na de discussie minder gunstig over de kwaliteit van de voorstellen dan ervoor. Als men weet dat in de afbeelding de scores 0, 10 en 20 respectievelijk de schaalbetekenis 'onvoldoende', 'voldoende' en 'goed' hebben, is duidelijk dat sommige beoordelaars (vooral beoordelaar 1) aanzienlijk donkerder tegen de kwaliteit van het pakket voorstellen zijn gaan aankijken. Op het belang dat de commissie aan de gezamenlijke voorstellen toekent, heeft de discussie een welhaast verwoestende invloed gehad. Zonder uitzondering zijn alle beoordelaars hierover na de discussie minder gunstig gaan denken, een verschijnsel waarop we bij de Discussie ongetwijfeld terugkomen.

De commissie als geheel

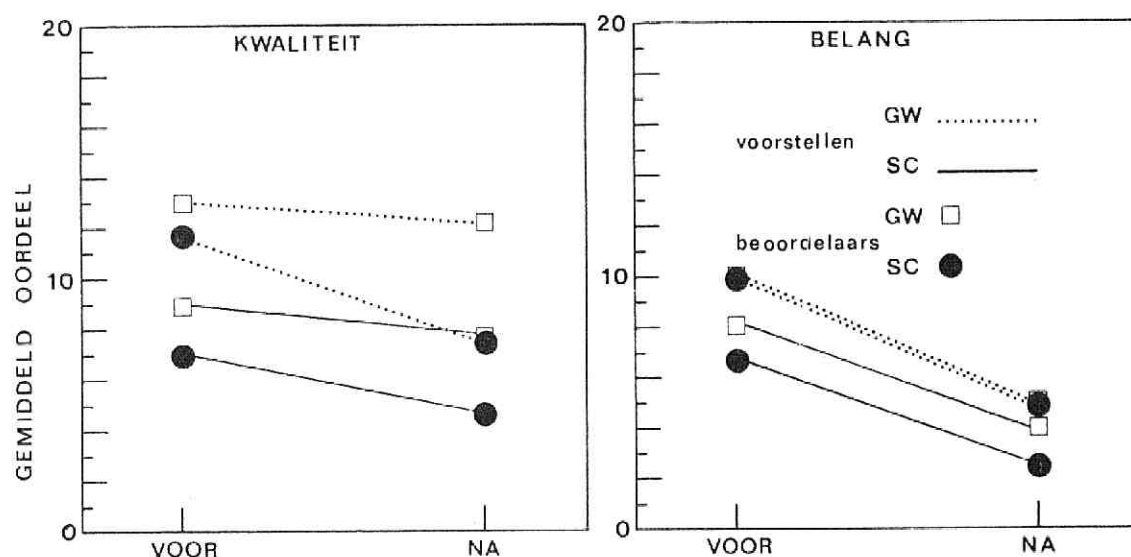
Of de gesignaleerde verschuivingen in de commissie-oordelen significant zijn, toetsen we in eerste instantie met behulp van multivariate variantie-analyse (MANOVA). In het gehanteerde ontwerp vormen de beoordelingen op de criteria kwaliteit en belang de afhankelijke variabelen, en worden drie onafhankelijke variabelen (factoren) onderscheiden: het tijdstip van de beoordeling (voor en na), het type voorstel en het type beoordelaar. De gehanteerde indeling naar type voorstel en beoordelaar is gebaseerd op hun disciplinaire achtergrond. We onderscheiden de gedragswetenschappelijke voorstellen (die primair door psychologen of pedagogen worden gedragen) van sociaal-culturele voorstellen (waaraan hoofdzakelijk sociologen, politicologen of bestuurskundigen deelnemen). Op een analoge wijze worden de vier beoordelaars uit de gedragswetenschappelijke hoek onderscheiden van de vijf beoordelaars die een sociaal-culturele discipline beoefenen.

De resultaten van deze MANOVA zijn in Tabel 6 weergegeven. Op elk van de drie geïntroduceerde factoren doet zich een significant hoofdeffect voor; en daarnaast is er sprake van één interactie-effect. Bij de MANOVA blijft in het vage of deze multivariate effecten voor beide afhankelijke variabelen opgaan. Daarom is voor elk significant multivariaat effect met behulp van univariate F-toetsen nagegaan op welke criteria het zich voordoet. Ook de resultaten van deze exercitie zijn in Tabel 6 samengevat. De meeste effecten blijken uitsluitend voor de kwaliteitsbeoordeling op te gaan. Alleen het hoofdeffect van tijdstip blijkt zowel voor de kwaliteitsbeoordeling als voor de beoordeling op belang te gelden.

Figuur 2 maakt duidelijk wat men zich bij deze effecten moet voorstellen. Het hoofdeffect van type voorstel houdt in, dat de gedragswetenschappelijke voorstellen kwalitatief hoger worden aangeslagen dan de sociaal-culturele voorstellen (ten aanzien van het belang van de voorstellen bestaat een soortgelijke tendentie, maar deze is niet significant). Het hoofdeffect van type beoordelaar houdt in, dat de gedragswetenschappelijke beoordelaars een rooskleuriger kijk op

Tabel 6: Toetsingsresultaten MANOVA van de Beoordelingen op Kwaliteit en Belang, per Type Voorstel, Type Beoordelaar en Tijdstip van Beoordeling, alsmede Resultaten Univariate F-Toetsen

Bron	MANOVA			ANOVA'S	
	F	p		F	p
Type Voorstel (A)	3.51	.05 ^x ---	{ Kwaliteit Belang	7.03	.01 ^x
				3.62	.07
Type Beoordelaar (B)	10.01	.00 ^{xx} ---	{ Kwaliteit Belang	17.97	.00 ^{xx}
				1.99	.17
Tijdstip (C)	47.01	.00 ^{xx} ---	{ Kwaliteit Belang	12.07	.00 ^{xx}
				83.10	.00 ^{xx}
AB	1.17	.32	{ Kwaliteit Belang		
AC	.50	.61			
BC	7.21	.00 ^{xx} ---		11.18	.00 ^{xx}
				.00	.99
ABC	1.37	.27			



Figuur 2: Gemiddelde Oordelen over de Kwaliteit en het Belang Vóór en Ná de Discussie, Uitgesplitst naar de Disciplinaire Achtergrond van de Voorstellen en de Beoordelaars.

de kwaliteit van de voorstellen hebben dan de sociaal-culturele beoordelaars (ook over het belang van de voorstellen oordelen de gedragswetenschappelijke beoordelaars het mildst, maar dit verschil is niet significant). Het hoofdeffect van tijdstip kennen we reeds: na de discussie wordt zowel over de kwaliteit als het belang minder gunstig geoordeeld. Bij de kwaliteitsbeoordeling hebben vooral de sociaal-culturele beoordelaars hun aanvankelijke opinie in ongunstige zin omgebogen, hetgeen tot uitdrukking komt in de significante interactie van type beoordelaar met tijdstip.

3.4. Beoordelingsdeterminanten

In Tabel 7 zijn de correlaties tussen de in het onderzoek betrokken kenmerken van de aanvragen en de commissie-oordelen over de aanvragen (somscores over beoordelaars) opgenomen. Opvallend is dat de correlaties van de aanvraagkenmerken met het initiële commissie-oordeel over de kwaliteit en het belang van de voorstellen een grote gelijkenis vertonen. Dezelfde kenmerken die de beoordelaars aanleiding geven een voorstel als kwalitatief goed aan te merken, maken zo'n voorstel in de ogen van de commissie blijkbaar ook belangrijk. Dit bevestigt onze eerdere waarneming dat bij de voorbeoordeling een gebrekkig onderscheid is gemaakt tussen het belang en de kwaliteit van de voorstellen (ontbreken van discriminante validiteit).

Drie typen aanvraagkenmerken lijken een gunstige voorbeoordeling in de hand te werken. In de eerste plaats – maar dat wisten we al – worden de gedragswetenschappelijke voorstellen gunstiger beoordeeld dan de sociaal-culturele voorstellen. Ten tweede speelt de wetenschappelijke reputatie, blijkend uit het aantal internationale publicaties per deelnemer, het aantal deelnemende gepromoveerde medewerkers en de deelname van KNAW-leden, een rol bij de beoordeling. Tenslotte is er een tendentie om 'big' als 'beautiful' aan te merken: aanvragen waarbij veel vakgroepen betrokken zijn en de dikkere aanvragen (die hun grotere omvang vooral danken aan het feit dat voor méér medewerkers ruimte voor publicaties, verkregen subsidies en dergelijke moet worden gereserveerd) worden gemiddeld genomen het gunstigst beoordeeld. De deelname van één of meer leden van de DAS-commissie aan een voorstel leidt echter niet tot een gunstiger beoordeling.

De correlaties tussen de aanvraagkenmerken en de na-oordelen van de commissie vertonen ongeveer hetzelfde beeld als bij de voorbeoordeling. Ook bij de nabehoordeling lijken de correlaties met de kwaliteitsbeoordeling sterk op die met de beoordeling op belang, hetgeen

Tabel 7: Correlaties tussen de Kenmerken en de Beoordelingen van de Aanvragen

Aanvraagkenmerken	VOORAF		ACHTERAF	
	K	B	K	B
Aantal betrokken instellingen	.23	.26	.29 ^x	.24
Meest betrokken discipline (GW = 1, SC = 2)	-.51 ^{xx}	-.40 ^{xx}	-.31 ^x	-.21
Aantal betrokken vakgroepen	.23	.39 ^{xx}	.27	.33 ^{xx}
Aantal betrokken hoogleraren	.01	.02	.06	.15
Aantal gepromoveerde medewerkers (exclusief hoogleraren)	.38 ^{xx}	.34 ^x	.24	.18
Aantal niet-gepromoveerde medewerkers	.12	.06	-.13	-.24
Lengte aanvraag (aantal bladzijden)	.32 ^x	.35 ^x	.36 ^x	.21
Aantal internationale publicaties per medewerker	.45 ^{xx}	.39 ^{xx}	.42 ^x	.31 ^x
Aantal nationale publicaties per medewerker	-.08	-.09	-.02	.14
Aantal gerealiseerde promoties	.31	.15	.01	-.04
Aantal verkregen subsidies	.00	.12	.22	.40
KNAW-lidmaatschap (nee = 1, ja = 2)	.37 ^x	.31 ^x	.44 ^{xx}	.42 ^{xx}
DAS-lidmaatschap (nee = 1, ja = 2)	-.20	-.19	-.10	.01

Noot: Bij de correlatieberekening is *pairwise deletion* toegepast om te voorzien in ontbrekende data; $x = p < .05$; $xx = p < .01$.

bevestigt dat de discriminante validiteit ook bij de nabehoordeling te wensen overlaat. Ook het soort kenmerken dat tot een gunstige beoordeling leidt is grofweg hetzelfde gebleven. Maar de correlaties zijn over het algemeen wel lager dan bij de voorbeoordeling. Kennelijk trekken de beoordelaars zich bij de nabehoordeling minder van de objectieve kenmerken van de voorstellen aan. Dit blijkt ook uit de multiple correlaties die men vindt als men op basis van de aanvraagkenmerken het commissie-oordeel tracht te voorspellen. De voorbeoordeling op belang ($R = 0.52$) en kwaliteit ($R = 0.70$) laat zich duidelijk beter voorspellen dan de nabehoordeling ($R = 0.42$ resp. 0.44). Een presentatie van de gevonden regressie-oplossingen laten we achterwege, omdat aan de in de analyse geselecteerde voorspellers slechts een beperkte betekenis kan worden toegekend. De gebruikte set van voorspellers bevat namelijk enige gecorreleerde variabelen, met als consequentie dat er voor sommige geselecteerde voorspellers gemakkelijk andere uit de analyse hadden kunnen rollen.

4. DISCUSSIE

De geringe consensus die we bij de voorbeoordeling aantreffen, geeft nieuwe steun aan de meermaals gereleveerde voorspelling van Hofstee (1983), waarin voorzien wordt dat in commissies waarvan de leden onafhankelijk van elkaar dezelfde onderzoeksvorstellen beoordelen, de gemiddelde tussen-beoordelaar-correlatie zich meestentijds tussen .00 en .30 zal bewegen. Vooralsnog moet het eerste geval waarin de overeenstemming een waarde van 0.30 overstijgt zich nog voordoen. Vanuit de wetenschap gezien is dit een bevredigende uitkomst, maar uit praktisch oogpunt zou men wensen dat de voorspelling gefalsificeerd was. De consequentie is immers, dat aan de geloofwaardigheid van commissies die onderzoeksvorstellen beoordelen ernstig afbreuk wordt gedaan. Het ondermijnt in elk geval het vertrouwen in postale beoordelingscommissies, waarbij hetgeen we hier als voorbeoordeling aanduiden tegelijkertijd eindbeoordeling is. Wanneer de DAS-commissie haar eindoordeel op haar voorbeoordeling op verwacht belang had gebaseerd, had dit tot een zeer onbetrouwbaar advies geleid. Als een nieuwe commissie die voorbeoordeling zou overdoen, mag immers verwacht

worden dat dit tot een ranglijst van de voorstellen zou leiden die slechts in de orde van grootte van 0.40 met de DAS-ranglijst zou correleren.

De in onze Inleiding besproken drie algemene verklaringen voor het gebrek aan consensus tussen onafhankelijke beoordelaars, lijken ook in het geval van de DAS-commissie onverkort op te gaan. Van zelf-selectie is zelfs in tweeërlei opzicht sprake. Behalve dat vermoedelijk vooral de zwakkere onderzoeksgroepen van de indiening van een zwaartepuntaanvraag zullen hebben afgezien, heeft er ook nog binnen de instellingen zelf een voor-selectie plaatsgehad. Ook de gebruikte grove driepuntschalen zullen de overeenstemming tussen de beoordelaars niet hebben bevorderd. Tenslotte vormden de vele gezichtspunten die bij de beoordeling van de verwachte kwaliteit en het verwachte belang in aanmerking moesten worden genomen, gevoegd bij het ontbreken van enig voorschrift omtrent het gewicht dat aan de onderscheiden gezichtspunten moest worden toegekend, een voedingsbodem waarop idiosyncratische beoordelingsoptieken welig konden kieren.

Men kan ook op zoek gaan naar bijzondere omstandigheden die het gebrek aan consensus bij de DAS-voorbeoordeling kunnen verklaren. Met name kan men zich afvragen of de DAS-leden met hun uiteenlopende disciplinaire achtergrond en expertise wel geacht mochten worden over voorstellen uit alle hoeken van de sociale wetenschappen oordeelkundig te kunnen oordelen. Zoals we later nog zullen toelichten, is dit ons inziens inderdaad aan twijfel onderhevig. Maar dat betekent nog niet dat het gebrek aan consensus aan de gemêleerde samenstelling van de commissie mag worden geweten. Amerikaans onderzoek (Cole et al., 1981) wijst immers uit, dat ook bij monodisciplinair samengestelde commissies die uitsluitend over voorstellen uit de eigen wetenschapsector oordelen, de consensus ver te zoeken is. Met Hofstee (1983) moet worden gewaarschuwd tegen pogingen het gebrek aan overeenstemming 'weg te verklaren' door op bijzondere omstandigheden te wijzen. Gelet op het alarmerend gebrek aan consensus dat tot op heden alle uitgevoerde studies laten zien, moet de conclusie veeleer luiden, dat bij welke samenstelling of precieze inrichting van het beoordelingsproces dan ook, de DAS-commissie voorafgaand aan de discussie sterk verdeeld over de voorstellen zou hebben gedacht.

Volgens Swanborn (1986) moeten we ons op dit gebrek aan overeenstemming niet blind staren. Er is immers slechts aangetoond dat de voorlopige meningen waarmee de beoordelaars de vergadering binnenkomen verdeeld zijn, en op dat moment moet de commissiebespreking haar heilzame werking nog uitoefenen. De uiteenlopende en provisorische initiële meningen worden zijns inziens in de discussie gladgestreken en omgezet in een weloverwogen en beter commissie-oordeel.

Onze analyses stellen Swanborn op tenminste twee punten in het gelijk. In de eerste plaats hebben de voor-oordelen waarmee de commissieleden de bespreking aanvangen inderdaad een sterk voorlopig karakter. De meeste DAS-leden herzien hun aanvankelijke oordelen frequent en met een opmerkelijke lenigheid van geest. En daarnaast leidt de discussie tot een aanzienlijke vermindering van de initiële meningsverschillen, met het gevolg dat ook de geschatte betrouwbaarheid van het commissie-oordeel sterk stijgt.

De toename van de betrouwbaarheid moet echter met de nodige korrels zout genomen worden, en in feite is het misleidend om de door de discussie gegroeide consensus in een betrouwbaarheidscoëfficiënt uit te drukken. De na de discussie gevonden betrouwbaarheden laten immers, omdat niet langer van onafhankelijke oordelen sprake is, niet de conclusie toe dat het na-oordeel van een parallelle commissie in een orde van grootte gelijk aan de betrouwbaarheidscoëfficiënt met het na-oordeel van de DAS-commissie zou correleren. In werkelijkheid kan die correlatie heel wat lager uitvallen, en in principe is het zelfs niet uitgesloten dat de oordelen van twee parallelle commissies vóór de discussie nauwer samenhangen dan erna. Of de discussie de overeenstemming tussen de oordelen van parallelle commissies vergroot of verkleint, hangt af van de mate waarin de discussie in beide commissies overeenkomstig verloopt. Komen in beide commissies dezelfde argumenten, met dezelfde nadruk en in dezelfde volgorde op tafel, dan is het aannemelijk dat de meningsverschillen tussen commissies worden verkleind. Worden daarentegen in beide commissies verschillende argumenten gehanteerd, met een verschillend accent en in een andere volgorde, dan mag men vrezen dat de meningsverschil-

len tussen commissies juist worden vergroot.

Hoe de verhouding van overeenkomstige en commissiespecifieke elementen bij een vergelijking van de besprekingen in parallelle commissies uitvalt, is een kwestie van speculatie. Om dat uit te zoeken zou men bij wijze van experiment meerdere commissies naast elkaar dezelfde voorstellen moeten laten beoordelen, en de besprekingen woord voor woord moeten vastleggen en analyseren. Als hiervoor medewerking te krijgen is, hetgeen juist in een land als het onze niet op voorhand hoeft te worden uitgesloten, verdient het ons inziens ook aanbeveling dat dergelijke experimenten worden uitgevoerd. Met zulke vergelijkingsstudies zou tegelijkertijd recht worden gedaan aan de recente pleidooien van De Groot (1985) en Swanborn (1986) voor het verrichten van kwalitatieve analyses van het argumentatieproces binnen, en in onze suggestie ook *tussen* commissies.

We zouden er geenszins van opkijken als dergelijke experimenten grote verschillen tussen het verloop en de inhoud van de discussie in parallelle commissies aan het licht zouden brengen. Door het gebruik van onderscheiden vergaderprocedures, groepsdynamische processen, verschillen in de interpretatie van de beoordelingscategorieën en de gewichten die er impliciet aan worden toegekend, verschillende wetenschapstheoretische pre-occupaties van de toevallige deelnemers aan de parallelle gesprekken, en zo voort, kunnen er gemakkelijk verschillen ontstaan tussen het type overwegingen dat in elke commissie gehanteerd wordt en het belang dat eraan wordt gehecht. We achten het daarom geenszins vanzelfsprekend, dat met het terugbrengen van de meningsverschillen binnen commissies ook de meningsverschillen tussen commissies worden verkleind. We kunnen de optimistische visie van Swanborn (1986) dan ook slechts delen voor zover hij stelt, dat de beoordelaars graag bereid zijn op geleide van de discussie hun mening voor een *ander* te geven. Maar of men het ook voor een *beter* geeft, hetgeen hier wil zeggen: voor een oordeel dat ook door een parallelle commissie zou worden geveld, valt ons inziens nog maar te bezien.

Gebruikmakend van de aantekeningen die de eerste auteur dezes van de DAS-bespreking heeft gemaakt, kan op diverse elementen in de DAS-commissie worden gewezen die in een andere commissie gemakkelijk anders uit zouden kunnen vallen. Met name zou het interessant zijn na te gaan hoe dwingend de argumenten zijn die ertoe hebben geleid, dat sommige voorstellen bij de nabeoordeling een groot aantal plaatsen stegen, terwijl andere enorm kelderden. Met zo'n gedetailleerde analyse wordt echter, zelfs bij een anonieme bespreking, licht inbreuk op het vertrouwelijke karakter van het commissiebesluit gemaakt. Zonder uit de school te klappen kan echter wel in algemene zin worden vastgesteld, dat de DAS-discussie alleen al bij een keuze voor een andere vergaderprocedure een ander karakter zou hebben gehad. De DAS-discussie had een vrij verloop. Er waren geen vaste aandachtspunten die bij elk voorstel consequent de revue passeerden. Elk commissielid dat daaraan behoefte gevoelde kon over elk voorstel kwijt wat hij het vermelden waard achtte, en zodra het gevoel bestond dat alles wat opmerkenwaard was gezegd was, werd naar het volgende voorstel overgegaan. Deze vergaderprocedure had meerdere consequenties. Ten eerste wisselden hierdoor de gezichtspunten (oftewel beoordelingscategorieën) waarover argumenten werden uitgewisseld sterk per voorstel. Een tweede gevolg was, dat sommige gezichtspunten veelvuldig op tafel kwamen, terwijl andere zelden of in het geheel niet aan de orde werden gesteld. Het behoeft geen betoog, dat het argumentatieproces bij een keuze voor een vergaderprocedure waarbij per voorstel systematisch een vaste reeks beoordelingscategorieën was nagelopen, tenminste voor een deel anders was verlopen.

De gehanteerde vergaderprocedure is er vermoedelijk mede de oorzaak van dat de nabeoordelingen van de commissie zoveel lager uitvielen dan de voorbeoordelingen. Een derde gevolg van de gehanteerde vergaderprocedure is namelijk, dat de overgrote meerderheid van de ter vergadering gemaakte kanttekeningen bij de voorstellen een negatieve strekking had. Een discussie die zich beperkt tot wat noemenswaard wordt geacht, lokt immers bij voorrang kritische opmerkingen uit. De dubieuze elementen in een voorstel springen nu eenmaal meer in het oog dan de punten die in orde zijn. Men vraagt niet gauw het woord om te melden, dat men de probleemstelling bevredigend beschreven acht (behalve wanneer men de uiteenzetting

opmerkelijk lucide vindt) of met de voorgestelde onderzoeksmethodiek kan instemmen. Maar als men de beschrijving ietwat vaag vindt of vragen bij de onderzoeksmethodiek heeft, voelt men zich al snel geroepen de vergadering hierop te attenderen. Evenzo staat men niet te dringen om de medebeoordelaars erop te wijzen, dat de verschillende deelprojecten of de in het voorstel samenwerkende onderzoeksgroepen een redelijk coherente indruk maken. Maar zodra daar twijfels over bestaan, wordt zelden verzuimd er melding van te maken.

In feite is er slechts één beoordelingscategorie waarover meer positieve dan negatieve opmerkingen zijn gemaakt, en dat is de bekwaamheid van de onderzoekers. Dat juist deze beoordelingscategorie vaak in gunstige zin aan de orde werd gesteld, komt wellicht doordat een positieve uittaling over de bekwaamheid van de aanvragers de weg vrijmaakt om met een geruster gemoed negatieve noten te kraken ('Op zichzelf gaat het om heel bekwame onderzoekers, maar ...'). Doordat nu de pluspunten van de voorstellen tijdens de discussie niet of niet uitputtend gememoreerd werden, terwijl de minpunten alle zonder mankeren de revue passeerden, hadden de beoordelaars ten tijde van de herbeoordeling de gebreken het meest fris voor ogen. Dit kan verklaren waarom men na afloop van de discussie doorgaans somberder tegen de voorstellen aankeek dan ervoor. De kijk op de kwaliteit van de voorstellen werd minder donker dan die op het belang, wellicht omdat bij de kwaliteitsbeoordeling de competentie van de onderzoekers werd meegewogen. Want de kwaliteit van de onderzoekers werd – zie hiervoor – zelden in twijfel getrokken.

Het negatiever worden van de oordelen kan daarnaast een uitvloeisel zijn van de omstandigheid, dat de beoordelaars niet altijd in staat waren elk voorstel op al zijn merites te beoordelen. Als men als beoordelaar onvoldoende in een onderwerp is ingevoerd, zal men eerder een gunstige voorbeoordeling geven. In zo'n geval kan men immers alleen maar marginale bezwaren tegen een voorstel koesteren, want wat de inhoud betreft is men min of meer gedwongen de aanvragers het voordeel van de twijfel te geven. Wanneer nu die twijfels ter vergadering door een meer ter zake kundige beoordelaar worden weggenomen, zal dit de na-oordelen eerder in ongunstige dan in gunstige zin veranderen. Als de beter ingevoerde collega zich positief over het voorstel uitlaat, kan men immers volstaan met het initieel reeds gunstige oordeel te handhaven, terwijl eventuele bedenkingen die worden ingebracht dwingen het initiële oordeel in ongunstige zin te herzien.

Het bedoelde fenomeen is aan de hand van het volgende fictieve voorbeeld te illustreren. Stel, er ligt een voorstel ter tafel voor een onderzoek dat beoogt nieuw licht op de etologie van dyslexie te werpen, en de aanvragers claimen dat de te verwerven inzichten tot een betere aanpak van lees- en schrijfstoornissen zullen leiden. De DAS-leden die het fijne van dit onderwerp niet weten (vermoedelijk een meerderheid), blijft weinig anders over dan marginaal te toetsen of de uiteenzetting een begrijpelijke, doordachte en consistente indruk maakt, en daarnaast zal men zich aan de hand van de publicatielijsten een indruk van de bekwaamheid (en ijver) van de onderzoekers in kwestie vormen. Als die marginale toetsing positief uitvalt, rest de niet-ingevoerde beoordelaar weinig anders dan een gunstig oordeel uit te brengen (en vermoedelijk kiest men voor de hoogste schaalpositie, want tegen een betere aanpak van dyslexie kan geen weldenkend mens bezwaar hebben). Hoort men nu ter vergadering van een beter ingevoerde beoordelaar dat het om een zinnig voorstel gaat, dan zal men zijn (gunstige) voor-oordeel handhaven. Verneemt men echter dat het überhaupt een mistige zaak is waar dyslexie nu precies vandaan komt en dat soortgelijk onderzoek reeds eerder is uitgevoerd en op niets is uitgelopen, dan is men gedwongen op gezag van de betreffende spreker(s) een ongunstig na-oordeel uit te brengen.

Als men de mate waarin een beoordelaar zich in de discussie over bepaalde voorstellen mengt, als een indicator mag zien voor de mate waarin hij zich bevoegd acht over die voorstellen te oordelen, kan worden vastgesteld dat de sociaal-culturele beoordelaars zich over een aanzienlijk smaller terrein tot oordelen bevoegd achtten dan de beoordelaars van gedragswetenschappelijke huize. Afgaand op onze aantekeningen van de DAS-bespreking deden de sociaal-culturele beoordelaars er bij de bespreking van de gedragswetenschappelijke voorstellen namelijk vaak het zwijgen toe, terwijl de gedragswetenschappelijke beoordelaars bij de discussie

over de sociaal-culturele voorstellen bijna even vaak aan het woord kwamen als hun collegae uit de sociaal-culturele sector. De gêne die de sociaal-culturele beoordelaars aan de dag legden om zich over voorstellen van buiten de eigen sector uit te spreken is niet onverklaarbaar. Want het pakket gedragswetenschappelijke voorstellen bevatte diverse aanvragen uit de sfeer van de functieleer, mathematische psychologie annex datatheorie en de 'abnormale' psychologie, terreinen die – wellicht op een enkele 'generalist' na – ook psychologen zelf niet alle tegelijkertijd geheel kunnen overzien. Omgekeerd hadden de sociaal-culturele voorstellen vaak betrekking op onderwerpen waarover men met een abonnement op een kwaliteitskrant al gauw (maar ook oordeelkundig?) kan meedenken.

Blijkens eerder gepresenteerde analyses onderscheiden de sociaal-culturele beoordelaars zich ook van hun gedragswetenschappelijke collegae doordat zij de kwaliteit van de voorstellen geringer schatten, een verschijnsel waarvoor we geen verklaring hebben. Daarnaast bleek dat de kwaliteitsoordelen van de sociaal-culturele beoordelaars onder invloed van de discussie het meest kelderden (interactie-effect). Dit kan een uitvloeisel zijn van het feit, dat de sociaal-culturele beoordelaars bij de na-beoordeling nóg minder onderscheid tussen de kwaliteit en het belang van de voorstellen maakten dan de gedragswetenschappelijke beoordelaars (bij nadere analyse blijkt namelijk, dat de monobeadelaar-heterocriteriumcorrelaties bij de sociaal-culturele beoordeelaar ($\bar{r} = .67$) duidelijk hoger zijn dan bij de gedragswetenschappelijke beoordelaars ($\bar{r} = .50$). Omdat de sociaal-culturele beoordelaars zeer weinig onderscheid tussen beide beoordelingscategorieën maakten, kan de geringe dunk die de discussie over het belang van de voorstellen deed ontstaan, hen verleid hebben tevens de kwaliteit lager te schatten.

Omdat de discriminante validiteit (ook bij de gedragswetenschappelijke beoordelaars) duidelijk te wensen overliet, kan men zich overigens afvragen of de commissie er wel verstandig aan heeft gedaan haar eindadvies op haar eindoordeel over het verwachte belang te baseren. Principieel gezien zijn er weliswaar met De Groot (1985) steekhoudende argumenten te geven om het oordeel over het verwachte belang in plaats van de verwachte kwaliteit de doorslag te laten geven. Maar nu blijkt dat beide typen oordelen min of meer uitwisselbaar zijn en de overeenstemming over de kwaliteit van de voorstellen het grootst is, was het – achteraf gezegd – empirisch gezien wellicht rechtvaardiger geweest het eindadvies op het kwaliteitsoordeel te baseren.

Onze analyse van de determinanten van de beoordeling heeft weinig nieuwe gezichtspunten opgeleverd. Behalve de wetenschappelijke reputatie van de onderzoekers, vonden we vooral determinanten die specifiek voor de DAS-beoordeling lijken te gelden. Wel is nog opmerkenwaard, dat het commissie-oordeel zich vóór de discussie duidelijk beter liet voorspellen dan erna. Kennelijk zijn er tijdens de discussie nieuwe gezichtspunten op tafel gekomen, die aanleiding gaven zich bij de oordeelsvorming minder door de objectieve kenmerken van de voorstellen te laten leiden. Men kan hierin een extra argument zien om aan Swanborns (1986) optimistische visie te twijfelen. De bewijslast voor de stelling dat voor de variantie die door de discussie aan de objectieve kenmerken is onttrokken, valide in plaats van commissiespecifieke variantie in de plaats is gekomen, berust bij de optimisten.

Tenslotte kan men zich afvragen of de wijze waarop de DAS-commissie zwaartepunten geselecteerd heeft herhaling verdient. In elk geval is duidelijk dat alvorens zo'n duizenden mensuren verslindend aanvraagcircuit in gang wordt gezet, vast moet staan dat het advies ten departemente serieus wordt genomen. Als de faculteiten opnieuw zouden worden uitgenodigd concrete zwaartepuntvoorstellen in te dienen die *inhoudelijk* worden beoordeeld, verdient het overweging die voorstellen door twee commissies te laten beoordelen, één voor de sociaal-culturele en één voor de gedragswetenschappen. Niet omdat de consensus binnen elk van die twee commissies gunstig zou afsteken bij die van de DAS-commissie, want er is geen reden dit te verwachten. Wel echter omdat aparte commissies, hoe feilbaar de uiteindelijke oordelen ook mogen zijn, geacht mogen worden met meer inhoudelijke kennis van zaken de voorstellen te kunnen beoordelen.

REFERENTIES

- Cole, J.R., & Cole, S. (1979). Which researcher will get the grant? *Nature*, 279, 575-576.
- Cole, S., Cole, J.R., & Simon, G. (1981). Chance and consensus in peer review. *Science*, 214, 881-886.
- Cole, S., Rubin, L., & Cole, J.R. (1977). Peer review and the support of science. *Scientific American*, 237(4), 34-41.
- DAS (1983). *Discipline-advies sociale wetenschappen*. 's-Gravenhage: Commissie ad hoc discipline-advies sociale wetenschappen.
- DAS (1984). *Advies over de zwaartepunten voor sociaal-wetenschappelijk onderzoek*. 's-Gravenhage: Commissie ad hoc discipline-advies sociale wetenschappen.
- De Groot, A.D. (1984). Beoordelen van onderzoeksvoorstellen: voorspellend redeneren. *VOR-Bulletin*, 8(1), 2-5.
- De Groot, A.D. (1985). Kwaliteit in de wetenschap, een meetbaar begrip? In H.A. Becker & A.F.J. van Raan (Red.), *Kwaliteit in de wetenschap, een meetbaar begrip?* Leiden: DSWO Press.
- Hofstee, W.K.B. (1983). Beoordelingen van subsidie-aanvragen voor onderwijsresearch: een psychometrische evaluatie. *Tijdschrift voor Onderwijsresearch*, 8, 273-284.
- Hofstee, W.K.B. (1985). Beoordeling van de kwaliteit van wetenschappelijk onderzoek: fundamentele kwesties. In H.A. Becker & A.F.J. van Raan, *Kwaliteit in de wetenschap, een meetbaar begrip?* Leiden: DSWO-Press.
- Humphreys, L.G. (1982). NSF peer review. *Science*, 215, 344.
- Spetz, S.H. (1979). *Quantity vs. quality in research proposals: an empirical justification for limiting the length of USOE grant applications*. Raleigh, NC: Conserva, Inc.
- Scheerens, J., & Beem, A.L. (1986). Ratings of requests for grants in educational research revisited. *Tijdschrift voor Onderwijsresearch*, 11, 67-82.
- Stanley, J.C. (1957). K-R 20 as the stepped-up mean item intercorrelation. *Yearbook of the National Council on Measurements Used in Education*, 14, 78-92.
- Swanborn, P.G. (1986). Committee judgment of research proposals: the limits of quantitative analysis. *Tijdschrift voor Onderwijsresearch*, 11, 114-117.
- Winer, B.J. (1971). *Statistical principles in experimental design*. New York, etc.: McGraw-Hill.

Manuscript ontvangen 20-5-1987.

Definitieve versie ontvangen 13-7-1987.

Open en Meerkeuze-vragen Tekstbegrip: een onderzoek naar de relatie tussen prestaties op open en meerkeuze-vragen

Huub van den Bergh
SCO, Amsterdam.

ABSTRACT

In this study the relation between achievement on open-ended and multiple-choice questions for the measurement of reading comprehension was investigated. In a randomized design 590 pupils (from vocational education (LBO), $n = 326$, and lower secondary education (MAVO), $n = 264$) read one of two tests, followed by 25 open-ended questions, or these same questions in multiple-choice format. The results indicate that pupils score somewhat higher on multiple-choice items, and that this effect is stronger for LBO-pupils than for MAVO-pupils. Furthermore, it is shown that the reliability of the multiple-choice tests was lower than that of the tests with open ended questions.

1. INLEIDING

De vraag naar de eventuele verschillen tussen de open en gesloten vraagvorm is wederom in de belangstelling gekomen door de recente examenveranderingen voor het vak Nederlands op het LBO en MAVO. De leerlingen op het LBO kregen voorheen twee tekstbegriptoetsen voorgelegd, één met louter vierkeuze-vragen en één met open vragen, terwijl de eindexamen-toets tekstbegrip op het MAVO voornamelijk uit open vragen bestond. In het kader van de harmonisatie LBO-MAVO, krijgen beide groepen leerlingen dezelfde eindexamenopdracht voor het onderdeel tekstbegrip van het vak Nederlands. Het cijfer van een kandidaat wordt dan voor 30% door de prestaties op open en voor 70% door de prestaties op gesloten vragen bepaald.

Het onderzoek naar de validiteit van gesloten vragen biedt nog weinig houvast, enerzijds door tegenstrijdige resultaten, anderzijds door statistisch-methodologische tekortkomingen. Het meeste onderzoek is verricht volgens een herhaalde-metingen-design, doch sporadisch zijn ook studies met een gerandomizeerde opzet uitgevoerd. In het onderstaande worden de resultaten die beide typen onderzoek opgeleverd hebben (kort) besproken.

In onderzoek naar de relatie tussen prestaties op open en meerkeuze-vragen met een herhaalde-metingen-design, maakt elke proefpersoon verscheidene toetsen. We onderscheiden onderzoek met twee metingen en onderzoek met meer metingen.

Bij het onderzoek met twee (herhaalde) metingen wordt een toets tweemaal afgenomen bij dezelfde proefpersonen. Het verschil tussen de toets in de eerste en in de tweede afname is de vraagvorm. In de meeste studies wordt een, voor de toets-onbetrouwbaarheid gecorrigeerde correlatie-coëfficiënt gevonden tussen .70 en 1.20 (!). Slechts in enkele studies worden lagere

coëfficiënten gevonden (Van Westrhenen, 1973). Veel auteurs trekken dan ook de conclusie dat open en gesloten vragen een beroep doen op dezelfde cognitieve vaardigheden (zie o.a. Toops, 1921; Ruch & Stoddard, 1927; Correy, 1930; Carter & Crone, 1940; Hurlburt, 1954; Cook, 1965; Heim & Watts, 1967; Choppin, 1968; Van Westrhenen, 1973). Wel is regelmatig aangetoond dat meerkeuze-vragen makkelijker zijn (een hogere p-waarde hebben) dan vergelijkbare open vragen (zie o.a. Heim & Watts, 1967; Carcelli, 1980; Benson, 1981) en dat toetsen met open vragen een hogere betrouwbaarheid hebben dan vergelijkbare toetsen met meerkeuze-vragen (Frary, 1985).

In het onderzoek met meer dan twee herhaalde metingen kan een nuancering worden aangebracht tussen studies waarin drie of meer vraagvormen onderzocht worden, en studies waarin de prestaties op toetsen met open en meerkeuze-vragen gecorreleerd worden met cognitieve-vaardigheidscores. De resultaten zijn wisselend: sommige auteurs rapporteren dat zij geen wezenlijke verschillen tussen toetsen met een verschillend vraagtype aan konden tonen (Traub & Fisher, 1977), anderen concluderen dat open en gesloten vragen duidelijk iets anders meten (Vernon, 1962; Kempa & L'Odiaga, 1984; Ward et al., 1982).

Bij het onderzoek met herhaalde metingen vallen drie kanttekeningen te maken. Ten eerste wordt 'dezelfde' toets verschillende malen afgenomen. Dat wil zeggen: de proefpersonen worden verschillende malen geconfronteerd met dezelfde vragen in dezelfde formulering; alleen de manier waarop zij het antwoord kenbaar kunnen maken verschilt van toets tot toets. Hierbij werd vrijwel altijd de toets in open vraagvorm gevolgd door de toets(en) met een alternatieve vraagvorm. Hoewel de vragen op deze wijze 'exact' gelijk zijn, valt het te betwijfelen of de vragen ook identiek zijn voor de leerling. Immers, het maakt nogal een verschil of een leerling in het tijdsbestek van één à twee weken voor de eerste, tweede, derde, vierde, vijfde of zelfs zesde keer met dezelfde vragen geconfronteerd wordt. Vier voorbeelden maken dit duidelijk:

1. Wanneer bereikte Columbus Amerika?
.....
2. Wanneer bereikte Columbus Amerika?
 - a. 1942
 - b. 1592
 - c. 1492
 - d. 1249
3. Bereikte Columbus Amerika in 1492?
 - a. Ja.
 - b. Nee.
4. In bereikte Columbus Amerika.

Uit dit voorbeeld zal duidelijk zijn, dat de leerling anders tegen de vierde vraag aankijkt dan tegen de eerste. Slechts in één studie wordt dit aspect aangeroerd, doch ook hier wordt niet gespeculeerd omtrent de mogelijke invloeden ten gevolge van deze toetsingsprocedure (Traub & Fisher, 1977). Zo zouden geheugen-effecten een belangrijke rol kunnen gaan spelen, wanneer een leerling in een kort tijdsbestek verschillende malen dezelfde vragen moet beantwoorden (vergelijk bij voorbeeld: Hogan, 1981).

Ten tweede moet een voor de toetsenbetrouwbaarheid gecorrigeerde correlatie-coëfficiënt beschouwd worden als een overschatting van de werkelijke samenhang tussen twee toetsen (Lord & Novick, 1968, p. 338). Immers, de betrouwbaarheid (α) is vrijwel altijd een onderschatting van de werkelijke betrouwbaarheid van een toets; de voor toetsenbetrouwbaarheid gecorrigeerde correlatie-coëfficiënt (p_{xy}) zal dus vrijwel altijd een overschatting, tot op een onbekende hoogte, van de werkelijke samenhang tussen de twee toetsen zijn.

Ten derde trekken verschillende auteurs een andere conclusie bij een even hoge correlatie-coëfficiënt. Mellenbergh (1971) bij voorbeeld toetst of de gevonden (gecorrigeerde) coëfficiënt significant afwijkt van 1.00; andere auteurs concluderen dat twee toetsen hetzelfde meten wanneer de gecorrigeerde correlatie-coëfficiënt groter is dan .70 (Van Westrhenen, 1973) of groter is dan .80 (Peterson, 1926; Hurd, 1932).

Binnen de studies met een gerandomizeerde opzet, waarin de vergelijkbaarheid van open en meerkeuze-vragen onderzocht werd, onderscheiden we studies met een gerandomizeerde-groepen-posttoets design en studies met een factoriële opzet van gegevensverzameling. Voorbeelden van studies met een gerandomizeerde-groepen-posttoets design zijn studies waarin de gevoeligheid van toetsen met verschillende vraagtypen om leerresultaten te evalueren bestudeerd werd. Shulson & Crawford (1928) lieten twee groepen studenten een tekst bestuderen. De eerste groep mocht de tekst een uur bestuderen, de tweede groep een half uur. Een derde groep studenten kreeg geen gelegenheid om de tekst te lezen. Daarna maakten de leerlingen een toets met hetzij open, hetzij gesloten vragen. De resultaten van dit type onderzoek kunnen het beste weergegeven worden met de woorden van de onderzoekers:

'The most important outcome... was (that) of the total of eight experiments four were in favor to the true-false and four to the completion tests. In other words... the two tests are equally good to distinguish the students who have studied the lessons from whom have not.' (Shulson & Crawford, 1928, p. 538).

Shohamy (1984) onderzocht de invloed van vraagvorm met een (2 x 2 x 2) factoriële design. Acht gerandomizeerde groepen Israëlische studenten ($n = 655$) lasen of één van de twee Hebreeuwse teksten, of een Engelse vertaling hiervan waarbij of vierkeuze-vragen of open vragen gesteld werden. Zij concludeerde dat zowel de taal van de tekst als de vraagvorm van invloed was op de prestaties van de studenten. Op de toetsen met meerkeuze-vragen bleken de studenten significant hogere scores te behalen dan op de toetsen met open vragen. Voorts bleek dat hoe lager de begrijpend lezen-score van de studenten, hoe sterker de invloed van de vraagvorm en de taal van de tekst. Dat wil zeggen: studenten die over een lage leesvaardigheid beschikten scoorden op de toetsen met meerkeuze-vragen relatief hoger dan de studenten met een grotere leesvaardigheid (zie ook Murphy (1982) voor een vergelijkbare conclusie).

Samenvattend kunnen we stellen dat de onderzoeksresultaten een tegenstrijdig beeld te zien geven: in studies met twee herhaalde metingen wordt weliswaar geconcludeerd dat open en meerkeuze-vragen equivalent zijn, maar gezien de beperkingen die aan deze onderzoeksopzet kleven kan getwijfeld worden aan de houdbaarheid van deze conclusie. Op grond van studies met meer herhaalde metingen lijkt voorlopig de conclusie gerechtvaardigd dat meerkeuze-vragen niet exact hetzelfde meten als open vragen; en wat betreft de studies met een gerandomiseerd design is nog geen beslissing mogelijk. Met name het geringe aantal studies met gerandomizeerde opzet speelt hierbij een rol, terwijl dit design zich juist bijzonder goed leent voor onderzoek naar de validiteit van de meerkeuze-vragen.

In het onderhavige onderzoek wordt de relatie tussen open en meerkeuze-vragen tekstbegrip met behulp van een gerandomiseerd design onderzocht. Elke leerling maakt één van de vier, voor dit doel geconstrueerde tekstbegriptoetsen. Bij twee van de toetsen worden open vragen gesteld, en bij twee toetsen worden meerkeuze-vragen gesteld. De meerkeuze-vragen zijn, op de vraagvorm na, identiek aan de open vragen. De toewijzing van leerlingen aan één van de vier teksten is gerandomiseerd.

De relatie tussen open en meerkeuze-tekstbegripvragen wordt op verschillende manieren onderzocht. Allereerst wordt nagegaan of meerkeuze-toetsen inderdaad een hoger gemiddelde en een lagere betrouwbaarheid hebben dan toetsen met open vragen, zoals verschillende auteurs rapporteren. Vervolgens wordt onderzocht of de eventuele effecten van vraagvorm van leerlingen op twee schooltypen (MAVO vs. LBO) in even sterke mate gelden. Immers, wanneer meerkeuze-tekstbegripvragen qua betrokken cognitie-vaardigheden vergelijkbaar zijn met open tekstbegripvragen, dan mogen geen differentiële effecten van vraagvorm op de prestaties

van LBO- en MAVO-leerlingen aangetoond kunnen worden. Bovendien is deze vraagstelling van belang voor de 'nieuwe' examenpraktijk. Wanneer meerkeuze-vragen een differentiële invloed hebben op de prestaties van LBO- en MAVO-leerlingen, dan zijn de nieuwe examens tekstbegrip niet valide ten opzichte van de oude examens.

2. AFNAME VAN TEKSTBEGRIPTOESTEN

Er zijn vier tekstbegriptoetsen afgenomen aan 590 leerlingen uit de derde klas van het LBO ($n = 326$) en MAVO ($n = 264$). De leerlingen zaten op twaalf verschillende scholen, uit vijf provincies. Elke leerling maakte één tekstbegriptoets. Bij twee van de teksten werden 25 open vragen gesteld, bij de andere twee toetsen werden dezelfde vragen in vierkeuze-vorm gesteld. De afleiders van de vierkeuze-vragen waren gebaseerd op foute antwoorden van derde klassers LBO en MAVO in een beproeving van de instrumenten in hun open vraagvorm (Van den Bergh, 1987).

De beproeving van de tekstbegriptoetsen diende naast de constructie van de afleiders nog een tweede doel. Er kon een uitgebreid classificatie-schema van correcte en foute antwoorden opgesteld worden. Met behulp van dit schema zijn de open tekstbegripvragen betrouwbaar beoordeeld; één docent beoordeelde alle vragen en 100 open vraag-toetsen werden door een tweede corrector beoordeeld. De stabiliteit van de eerste docent was hoog ($r = .91$; $n = 50$), en ook de correlatie tussen beide docenten bleek bevredigend ($r = .86$; $n = 100$).

3. RESULTATEN

Allereerst zijn enkele psychometrische gegevens voor de vier tekstbegriptoetsen berekend. Deze gegevens zijn in Tabel 1 samengevat.

Uit Tabel 1 blijkt, zoals verwacht, dat de leerlingen op de beide tekstbegriptoetsen met gesloten vragen hoger presteren dan op de equivalente toetsen met open vragen. Meerkeuze-vragen hebben gemiddeld een hogere p-waarde dan vergelijkbare open vragen tekstbegrip. Voorts lijken de toetsen in hun open vraagvorm beduidend betrouwbaarder dan in hun gesloten vraagvorm.

Tabel 1. De resultaten op twee tekstbegriptoetsen met open vragen en dezelfde toetsen in meerkeuze-vorm (N = aantal leerlingen; gem. = gemiddelde; sd. = standaarddeviatie; KR-20 = betrouwbaarheid).

Tekst	Vraagvorm	N	gem.	sd.	KR-20
A	open	167	15.11	4.09	.84
	vierkeuze	129	16.26	3.15	.63
B	open	149	14.02	4.08	.83
	vierkeuze	145	17.37	3.57	.63

Beide effecten bleken zowel op het LBO als op het MAVO: op de toetsen in hun open vraagvorm behalen de leerlingen een lagere score dan op de toetsen in hun meerkeuze-vorm, en de betrouwbaarheidschatting van de toetsen in hun open vraagvorm is hoger dan van de equivalente meerkeuze-toetsen (tabel 2).

Ter verdere analyse van de uit Tabel 1 en 2 gebleken verschillen tussen de open en vierkeuze-vragen voor tekstbegrip is per tekst een kruistabel geconstrueerd. De beide kruistabellen zijn driedimensionaal met als dimensies: vraagvorm (V) met twee niveaus: open vs. vierkeuze-vragen; schooltype (S) met twee niveaus: LBO en MAVO; en item (I) met 25 niveaus: vraag 1 tot en met 25. De afhankelijke variabele is het antwoord (P) met twee niveaus: goed (1) en fout (2).

Tabel 2. De resultaten op twee tekstbegriptoetsen met open vragen, en dezelfde vragen in meerkeuze-vorm uitgesplitst naar schooltype (N = aantal leerlingen; gem. = gemiddelde; sd. = standaarddeviatie; KR-20 = betrouwbaarheidschatting).

Tekst	Vraagvorm	Schooltype	N	gem.	sd.	KR-20
A	open	LBO	83	13.40	3.67	.70
	open	MAVO	84	16.81	3.79	.72
	vierkeuze	LBO	75	15.61	3.08	.52
	vierkeuze	MAVO	54	17.15	3.06	.54
B	open	LBO	87	13.30	3.83	.70
	open	MAVO	62	14.54	4.61	.79
	vierkeuze	LBO	81	16.98	3.79	.68
	vierkeuze	MAVO	64	17.88	2.99	.50

De geobserveerde proportie van correcte antwoorden (P) kan volledig beschreven worden met het logit-model:

$$\ln(P_{ijk}/P_{2ijk}) = C + V_i + S_j + I_k + VS_{ij} + VI_{ik} + SI_{jk} + VSI_{ijk}. \quad (\text{model 1})$$

De termen rechts van het gelijkteken kunnen opgevat worden als effecten, zoals het in variantie-analytisch model: C is een algemeen effect, V_i het effect van vraagvorm (i), S_j het effect van schooltype (j), I_k het effect van item (k), VS_{ij} , VI_{ik} , SI_{jk} zijn interactie-effecten van de eerste orde, en VSI_{ijk} is een tweede-orde-interactie-effect.

Het verzadigde model (1) bevat vier effecten van de variabele 'vraagvorm' (V_i , VS_{ij} , VI_{ik} en VSI_{ijk}) op de afhankelijke variabele: proportie correcte antwoorden (P). Door deze effecten successievelijk uit het model te verwijderen, ontstaat een model waarin de proportie correcte antwoorden onafhankelijk is van de variabele 'vraagvorm':

$$\ln(P_{ijk}/P_{2ijk}) = C + S_j + I_k + SI_{jk} \quad (\text{model 8})$$

Daar elk volgend model minder parameters bevat dan het verzadigd model (1), zal het de gegevens minder nauwkeurig beschrijven. Of een volgend model de gegevens niettemin adequaat beschrijft, wordt nagegaan door eerst de parameters te schatten, en vervolgens op grond van deze schattingen de verwachte proporties goede antwoorden te berekenen en daarna te vergelijken met de geobserveerde proporties goede antwoorden. Verschillen de geobserveerde en verwachte proporties weinig, dan is het model een adequate beschrijving. De steekproefgrootte die gebruikt wordt om de passing van het model bij de feitelijke gegevens weer te geven is een 'likelihood-ratio'. Daarbij heeft $-2LN \lambda$ bij benadering een chi-kwadraat steekproevenverdeling.

In Tabel 3 zijn de gegevens van de passing van verschillende logitmodellen voor beide teksten weergegeven.

Uit Tabel 3 kunnen we de conclusie trekken dat model (8) zonder een effect van vraagvorm voor geen van beide tekstbegriptoetsen past. Voor tekst A zijn model (2), model (3) en model (5) de meest spaarzame modellen die de gegevens nog adequaat beschrijven. Model (5) past niet voor tekst B ($P < .05$). Model (2) past zowel voor tekst A als tekst B, maar in dit model is een effect meer gespecificeerd (SI_{jk}) dan in model (3). Daar er voor geen van de twee teksten qua passing een significant verschil is tussen model (2) en model (3), geven wij de voorkeur aan het meer spaarzame model (3).

Model (3), een passend model voor tekst A en B, omvat hoofdeffecten van vraagvorm (V_i), schooltype (S_j) en item (I_k), en interactie-effecten van vraagvorm x schooltype (VS_{ij}) en vraagvorm x item (VI_{ik}). Deze effecten kunnen we als volgt omschrijven: de leerlingen presteren beter op de tekstbegripvragen met gesloten dan op dezelfde vragen in open vorm (V_i). LBO-

Tabel 3. De passing van diverse logitmodellen voor beide tekstbegriptoetsen (C = constante; V_i = effect van vraagvorm; S_j = effect van schooltype; I_k = effect van item).

Model	Tekst A			Tekst B		
	X2	df	p	X2	df	p
1: $C+V_i+S_j+I_k+VS_{ij}+VI_{ik}+SI_{jk}+VSI_{ijk}$						
2: $C+V_i+S_j+I_k+VS_{ij}+VI_{ik}+SI_{jk}$	11.48	24	< .5	13.79	24	< .5
3: $C+V_i+S_j+I_k+VS_{ij}+VI_{ik}$	53.26	48	.28	32.69	48	< .5
4: $C+V_i+S_j+I_k+VS_{ij}+SI_{jk}$	112.27	48	.00	209.13	48	.00
5: $C+V_i+S_j+I_k+VI_{ik}+SI_{jk}$	34.31	25	.10	44.67	15	.01
6: $C+V_i+S_j+I_k+VI_{ik}$	77.81	49	.01	78.51	49	.01
7: $C+V_i+S_j+I_k$	180.53	73	.00	369.80	73	.00
8: $C+S_j+I_k+S_{jk}$	77.15	50	.00	335.96	50	.00

Tabel 4. Het verschil tussen de p-waarde van open en meerkeuze-vragen voor vijf ranges van de p-waarde van de open vragen. Weergegeven zijn: per range de gemiddelde p-waarde van de open vragen (p_{ov}); het aantal open vragen in deze desbetreffende range (I_{ov}); de gemiddelde p-waarde van de equivalente meerkeuze-vragen (p_{gv}), en het gemiddelde verschil tussen beide p-waarden ($p_{ov}-p_{gv}$).

range	p_{ov}	I_{ov}	p	$p_{ov}-p_{gv}$
0.00 tot 0.20	.15	2	.30	— .15
0.20 tot 0.40	.32	7	.56	— .24
0.40 tot 0.60	.48	20	.61	— .13
0.60 tot 0.80	.69	10	.75	— .06
0.80 tot 1.00	.90	11	.84	+ .07

leerlingen behalen gemiddeld lagere scores dan MAVO-leerlingen (S_j). De ene vraag is moeilijker dan de andere vraag (I_k). De LBO-leerlingen bleken relatief meer voordeel te ondervinden van vierkeuze-vragen tekstbegrip dan MAVO-leerlingen; LBO-leerlingen presteren in verhouding beter op de meerkeuze-toets dan op dezelfde toets in open vraagvorm dan MAVO-leerlingen (VS_{ij}) (zie Tabel 5).

Voor de interpretatie van het interactie-effect tussen vraagvorm en item (VI_{ik}) is Tabel 4 geconstrueerd. In deze tabel wordt het gemiddelde verschil in p-waarde van de open en meerkeuze-vragen weergegeven. Hierbij is de range in p-waarde van de open vragen opgesplitst in vijf categorieën.

Op grond van de in Tabel 4 tot uitdrukking komende verschillen in p-waarde tussen open en meerkeuze-vragen tekstbegrip concluderen we dat het verschil bij vragen met een lagere p-waarde groter is dan bij vragen met een hogere p-waarde. Enkel de eerste categorie wijkt af van deze regel, doch in deze categorie vallen slechts twee vragen.

Nu de verschillende effecten geïdentificeerd zijn, rijst de vraag naar de relatieve grootte van de verschillende effecten in model (3). Immers, kleine effecten kunnen wel significant zijn, maar zij zijn tevens minder relevant voor de onderwijs- en toetspraktijk. In Tabel 5 wordt voor beide teksten de relatieve grootte van de in model (3) onderscheiden effecten gespecificeerd (de hier gerapporteerde gestandaardiseerde effect-grootten zijn berekend door de effecten te delen door hun respectievelijke standaardfouten). Voor de overzichtelijkheid is voor effecten waarin de variabele item (I_k) een rol speelt de gemiddelde effect-grootte berekend.

Uit Tabel 5 kunnen we concluderen dat de grootte van alle gespecificeerde effecten aanzienlijk is, ook voor de interactie-effecten van vraagvorm x schooltype (VS_{ij}) en vraagvorm x

Tabel 5. De gestandaardiseerde grootte van de in model (3) gespecificeerde effecten voor beide tekstbegrip-toetsen (onder de kopjes min, gem, max en sd wordt respectievelijk de minimale, gemiddelde, maximale effect-grootte en de standaarddeviatie van effecten waarin de variabele item (I_k) een rol speelt weergegeven).

Tekst A: gestandaardiseerd effect				
Effect	min	gem	max	sd
Algemeen gemiddelde		5.24		
Vraagvorm (V_i)		4.22		
Schooltype (S_j)	-	2.68		
Item (I_k)	-7.43	3.37	16.10	7.09
Vraagvorm x School-type (VS_{ij})	-	2.82		
Vraagvorm x Item (VI_{ik})	-4.59	.06	4.89	1.92

Tekst B: gestandaardiseerd effect				
	min	gem	max	sd
		3.34		
		11.24		
		- 5.24		
	- 5.79	1.01	8.51	4.23
		- 5.61		
	- 5.30	.20	6.57	3.12

item (VI_{ik}). Om de grootte van de effecten te interpreteren is het handig de verschillende effect-grootten (binnen een tekst) te vergelijken met de grootte van het effect voor schooltype (S_j). We zien dan dat de resultaten op beide tekstbegrip-toetsen opmerkelijk convergeren. Het blijkt dat het hoofdeffect van vraagvorm (V_i) ongeveer twee keer zo groot is als het effect van schooltype (S_j), en dat de effecten van de verschillende items een aanzienlijke range vertonen. Het interactie-effect van vraagvorm x schooltype (VS_{ij}) is ongeveer even groot als het effect van schooltype (S_j), en het interactie-effect van vraagvorm x item (VI_{ik}) ligt in een range van nul $\pm$ het schooltype-effect. Op grond van deze resultaten kunnen we concluderen dat in dit onderzoek aanzienlijke verschillen tussen tekstbegrip-toetsen met open en meerkeuze-vragen aangetoond zijn.

4. DISCUSSIE

De resultaten van het hier gerapporteerde onderzoek zijn aanleiding om de volgende conclusies te trekken: tekstbegrip-toetsen met open vragen zijn gemiddeld moeilijker, maar betrouwbaarder, dan equivalente toetsen met vierkeuze-vragen.

MAVO-leerlingen presteren gemiddeld beter op tekstbegrip-toetsen dan LBO-leerlingen. Echter, het verschil tussen de prestaties van LBO- en MAVO-leerlingen is beduidend kleiner bij meerkeuze-toetsen, dan bij tekstbegrip-toetsen met open vragen. Dat wil zeggen: het verschil in prestaties tussen toetsen met open vragen en toetsen met gesloten vragen is groter bij LBO-leerlingen, dan bij MAVO-leerlingen. Dit zou opgevat kunnen worden als steun voor de in de inleiding geopperde hypothese dat minder taalvaardige leerlingen relatief meer voordeel onder-vinden van meerkeuze-vragen.

Voorts bleek het verschil in p-waarde van de open en gesloten vragen afhankelijk van de moeilijkheid van de vraag. Bij moeilijke vragen (vragen met een lage p-waarde) was het verschil

groter dan bij gemakkelijke vragen (vragen met een lage p-waarde). Dit is logisch, omdat bij moeilijke meerkeuze-vragen de proportie leerlingen die naar het goede antwoord 'raadt' groter is dan bij makkelijke vragen. Daar een deel van deze leerlingen het antwoord goed 'raadt', mogen grotere verschillen in p-waarde tussen moeilijke open en gesloten vragen verwacht worden dan bij relatief makkelijke vragen. Dit is tevens een goede verklaring voor het verschijnsel dat minder taalvaardige leerlingen meer voordeel ondervinden (hogere scores behalen op) van gesloten toetsen.

Het is nog de vraag of we dit 'raad-effect' moeten interpreteren als 'raad-effect' of een 'partiële-kennis-effect'. Immers, dit is afhankelijk van de andere strategie van de leerlingen: kiezen zij, wanneer zij het antwoord op een meerkeuze-vraag niet weten, at random een van de alternatieven, of hanteren zij een meer geavanceerde antwoord-strategie door bij voorbeeld eerst de minst waarschijnlijke alternatieven weg te strepen?

Door de gerapporteerde verschillen tussen open en gesloten tekstbegripvragen mogen we twijfelen aan de vergelijkbaarheid van beide vraagvormen. Op grond van deze studie moeten we concluderen dat open en gesloten vragen niet exact hetzelfde meten; meerkeuze-vragen tekstbegrip zijn qua oplossingsgedrag niet geheel vergelijkbaar met open tekstbegripvragen. We kunnen echter vooralsnog geen uitspraak doen over de processen die bij de beantwoording van beide vraagvormen betrokken zijn. Momenteel wordt aan de Stichting Centrum voor Onderwijsonderzoek (SCO) een studie verricht, waarin antwoord gegeven moet worden op de vraag welke cognitieve (deel-)vaardigheden differentieel betrokken zijn bij de beantwoording van open en gesloten tekstbegripvragen. Op grond van de resultaten van dit onderzoek kan deze vraag beantwoord worden.

GERAADPLEEGDE LITERATUUR

- Bergh, H. van den (1987), *Oude en nieuwe examens vergeleken: instrumentconstructie* (interim-rapport SVO: 305.4075), Amsterdam: SCO.
- Benson, J. (1981), A redefinition of content validity. *Journal of Educational and Psychological Measurement*, 741, 3, 793-802.
- Carcelli, H. (1980), The effect of item-format on phonics subtest scores on standardized reading achievement tests. *Educational Documents*.
- Carter, H.D., & Crone, A.P. (1940), The reliability of new-type or objective tests in a normal classroom situation. *Journal of Applied Psychology*, 24, 353-368.
- Choppin, B.M. (1968), *The report of crossover study in literature*. Ongepubliceerd rapport: IEA/TR/6, London.
- Cook, D.C., geciteerd in Ebel R.L. (1965), *Measuring Educational Achievement*. Englewood Cliffs, N.J.: Prentice-Hall Inc.
- Correy, S.M. (1930), The correlation between new type and essay examination scores and the relationship between them and intelligence as measured by army alpha. *School and Society*, 32, 849-850.
- Frery, R.B. (1985), Multiple-choice versus free-response: a simulation study, *Journal of Educational Measurement*, 22, 1, 21-31.
- Heim, A.W. & Watts, R.B. (1967), An experiment on multiple choice versus open ended answering in a vocabulary test. *British Journal of Educational Psychology*, 37, 339-346.
- Hogan, T.P. (1981), *Relationships between free-response and choice-type tests of achievement: a review of literature*. Washington: National Institute of Education.
- Hurd, A.W. (1930), Comparison of short-answers and multiple-choice tests covering identical subject content. *Journal of Educational Research*, 26, 28-30.
- Hurlburt, D. (1954), The relative value of recall and negotiation techniques for measuring precise knowledge of word meaning nouns, verbs, adjectives. *Journal of Educational Research*, 47, 561-576.

- Kempa, R.F. & L'Odiaga, J. (1984), Criterion-referenced interpretation of examination grades. *Educational Research*, 26, 1, 56-64.
- Lord, F.M. & Novick, H.R. (1968), *Statistical theories of mental test scores*. Reading, Mass.: Addison - Wesley.
- Mellenbergh, G.J. (1971), *Studies in studietoetsen*, Amsterdam, Psychologisch lab van de Universiteit van Amsterdam.
- Murphy, R.J. (1982), Sex differences in objective test performance. *British Journal of Educational Psychology*, 52, 213-219.
- Peterson, D.C. (1925) Do old and new type examinations measure different mental functions? *School and Society*, 24, 246-248.
- Ruch, C.M. & Stoddard, G.D. (1927), *Test and measurement in high school instruction*. Yonkers: World Book.
- Shohamy, E. (1984), Does the testing method make difference? The case of reading comprehension. *Language Testing*, 1, 2, 147-169.
- Shulson, V. & Crawford, C.C. (1928), Experimental comparisons of true-false and completion tests. *Journal of Educational Psychology*, 19, 580-583.
- Toops, H.A. (1921), *Trade tests in education* (Teachers college contributions to education, no. 115). New York: Teachers College, Columbia University.
- Traub, R.E. & Fisher, C.W. (1977), On the equivalence of constructed response and multiple-choice tests. *Applied Psychological Measurement*, 1, 3, 335-369.
- Vernon, P.E. (1962), The determinants of reading comprehension. *Educational and Psychological Measurement*, 22, 269-286.
- Ward, W.C. (1982) A comparison of free-response and multiple-choice forms of verbal optitude test, *Applied Psychological Measurement*, 6, 1-11.
- Westrhenen, J. van (1973), *Eindrapport proefexamens 1973*, Arnhem: CITO.

Manuscript ontvangen 25-6-1987.

Definitieve versie ontvangen 6-8-1987.

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie

Kees Beintema (RUL), Peter van den Dool (SCO), Dirk de Jong (LICOR), Ewoud Roede (SCO) en Peter van der Sijde (UT, TO, secretaris).

Redactie-adres

Peter van der Sijde, Faculteit der Toegepaste Onderwijskunde, Vakgroep Instructietechnologie, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893560.

Jrg. 11, nr. 5

October 1987

Inhoudsopgave

Redactioneel	314
VOR nieuws	314
Mededelingen	314
Verslag VOR ledenvergadering 26 mei 1987	315
In memoriam Hildo Wesdorp	316
Over de kwaliteit van de arbeid in de wetenschap <i>Floris van den Berg van Saparoea</i>	317
Proefschriften	322
Congressen	323
Ontvangen publicaties	324

Redactioneel

De reeks artikelen over de programmering van onderwijsonderzoek is onlangs definitief afgesloten met het boek 'Programmering van Onderwijsonderzoek'. Dit boek bevat naast de reeds eerder gepubliceerde bijdragen de voordrachten die tijdens de ORD werden gehouden. Dit verslag is bij de redactiesecretaris te bestellen en kost f 17,50.

In dit nummer starten wij met een nieuw thema over het 'werkklimaat' en de bijdrage van Van den Berg van Saparoea opent de reeks. Uw reacties hierop wachten wij gaarne in.

Verder willen wij u van tevoren op de hoogte brengen van promoties en oraties. Hiervoor is het noodzakelijk dat de promovendi en hoogleraren ons tijdig hiervan op de hoogte stellen. Uw medewerking in deze wordt op prijs gesteld.

VOR nieuws

Agenda voor de Algemene Ledenvergadering

datum: 12 november 1987

plaats: Radboudzaal, Stiltecentrum in Hoog Catharijnen te Utrecht

tijd: 14.00-16.00 uur

- 1 Opening
- 2 Verslag van de vorige ALV
- 3 Begroting 1988 (wordt de leden toegestuurd)
- 4 W.v.t.t.k.
- 5 Rondvraag
- 6 Sluiting

Mededelingen

Een nieuw tijdschrift

De professoren R.W. Webb en R.S. Sherman, beiden verbonden aan de School of Education van de University of Florida (Gainesville), hebben een nieuw tijdschrift opgericht: *International Journal of Qualitative Studies in Education*. Het wordt uitgegeven door Taylor & Francis (London-New York-Philadelphia). De bedoeling is het kwalitatief onderzoek op het terrein van de pedagogische wetenschappen te bevorderen. Men richt zich op onderzoeksbijdragen waarin een verscheidenheid van kwalitatieve methoden en technieken worden beschreven. Met het oog op het aan bod laten komen van Europese bijdragen werd een netwerk opgericht waarvan ondergetekende de coördinator is. Bijdragen moeten gestuurd worden naar onderstaand adres, waar men ook bijkomende informatie kan ontvangen.

Prof. R. Vandenberghe, K.U.-Leuven, Centrum voor Onderwijsbeleid en -vernieuwing, Vesaliusstraat 2, 30000 Leuven (Tel. 016/23.39.41).

Verslag VOR-ledenvergadering 26 mei 1987

Aanwezig: Claessen, Gerritsen v.d. Hoop, Hoebe, Roede, Beintema, Van Hout Wolters, Span, v.d. Werf.

Afwezig met kennisgeving: Scheerens

1. Opening:

Claessen opent de vergadering en heet de leden hartelijk welkom.

2. Verslag ledenvergadering november 1986:

Goedgekeurd, geen opmerkingen.

3. Jaarrekening '86:

Het negatief vermogen is omgezet in een positief saldo. Het ledenaantal is iets toegenomen. Door een actieve wervingscampagne wordt verwacht dat het ledenaantal dit jaar nog zal toenemen. De integratie TOR-VOR-bulletin is financieel gunstig. Met het positieve saldo zullen internationale activiteiten, service aan de leden en de uitgave van een jaarboek kunnen worden gefinancierd.

De kascontrolecommissie zou schriftelijk verslag uitbrengen van haar bevindingen. Helaas is het verslag nog niet binnen. Bij ontstentenis van dit verslag nu, krijgt het bestuur van de ledenvergadering de bevoegdheid dit in de volgende bestuursvergadering te doen.

4. Jaarverslag '86

Geen opmerkingen.

5. Samenstelling kascontrolecommissie:

Voorgesteld wordt om de commissie '86 te laten bestaan gezien haar expertise.

6. Agenda en frequentie ledenvergadering:

De opkomst van de leden is altijd weer mager. Statutair moet in ieder geval één keer per jaar een ledenvergadering zijn. Dit moet aan het eind van het jaar wegens begroting en bestuurswisseling.

Misschien komen er meer leden als deze vergadering wordt gekoppeld aan een andere activiteit. Voorgesteld wordt twee keer per jaar te handhaven met de mogelijkheid dat één keer vervalt. Bij de planning van de volgende ledenvergadering zal rekening worden gehouden met een andere bijeenkomst.

7. W.v.t.t.k.

- Van Hout Wolters vraagt aandacht te hebben voor evaluerende opmerkingen over de samenvoeging TOR-VOR-bulletin.
- Hoebe vraagt, naar aanleiding van de ledenwervingsfolder, of het beleid van de VOR gericht is op een Belgisch (Vlaamse)/Nederlandse vereniging. Dit is geen actief beleid, maar een onderhuidse trend. In VOR-bestuur zal aan de orde worden gesteld of dit actief beleid moet worden. Hoebe zou dit toejuichen.

8. Rondvraag:

Geen punten.

9. Sluiting.

Greetje v.d. Werf,
secretaris

In memoriam Hildo Wesdorp

Op 31 juli is collega Hildo Wesdorp gestorven. Senioronderzoeker en coördinator van het SCO, bijzonder hoogleraar aan de Universiteit van Amsterdam, en oud-bestuurder van de Vereniging voor Onderwijsresearch VOR. Al geruime tijd ziek was hij afwisselend thuis, op zijn instituut en in het ziekenhuis. Begin juli nog had ik hem thuis opgezocht. Net als vroeger was het een geanimeerd gesprek geworden. Over ziekte en pijn sprak hij op een manier die zijn artsen volgens zijn vrouw Betty 'dissimulerend' noemden. Toen ik buiten stond wist ik wat er ontbroken had. Hij had deze keer niets gezegd over zijn plannen.

Over plannen had ik Hildo de afgelopen twintig jaar zo dikwijls met vuur horen praten. In de begintijd van het RITP over toetsen en tekstbegrip. Later over betere metingen van productief taalgebruik. In de tijd van het VOR-bestuur over verenigingsplannen. De laatste vijf jaar spraken we over ons aandeel in het internationaal onderzoek van de IEA en over opstelbeoordelingen. Alle plannen waarover ik hem had horen spreken in die jaren waren ook echt uitgevoerd. Dat kon omdat hij een echte onderzoeker was: nieuwsgierig, precies en vooral vasthoudend.

Bij de collega's in het buitenland was Hildo geliefd, maar zijn onverbiddelijke vasthoudendheid werd lastig gevonden. Hildo kon zich publiekelijk opwinden als er iets verkeerd ging in opzet of uitvoering van het project. Hij nam dan geen blad voor de mond en spaarde niemand. Zijn kritiek op personen en zaken was gevreesd en werd zelden onmiddellijk aanvaard. Misschien was hij lastig, maar toch hield iedereen van hem. Dat kwam door de onbaatzuchtigheid en de deskundigheid van zijn inzet, maar ook door zijn oprechtheid en warmte. Dankzij die kenmerken kon hij uiteindelijk anderen wel overtuigen en zijn eigen werk ook telkens weer tot een goed einde brengen.

Dat alles is nu voorbij. Er zullen anderen komen die zijn werk voortzetten. Maar zelf zal hij er niet meer bij zijn. Dat laatste is voor velen en ook voor mij een erg droeve gedachte.

Egbert Warries
oud-voorzitter VOR.

Over de kwaliteit van de arbeid in de wetenschap

Floris van den Berg van Saparoea

SCO

Grote Bickersstraat 72

1013 KS Amsterdam

Samenvatting

Dit artikel geeft een beschrijving van het begrip kwaliteit van de arbeid (Quality of working life, Humanisering der Arbeit). Het begrip heeft twee zijden. Enerzijds is het de subjectieve kwaliteit van het werkend leven, zoals een werkende die ervaart. Anderzijds is het een ontwerpprincipe dat gebruikt kan worden bij het ontwerp van arbeidsplaatsen/functies. Het is dan een aanvulling op ontwerpprincipes als productietechniek, efficiëntie, effectiviteit, ergonomie, arbeidsbelasting, arbeidsomstandigheden. Door rekening te houden met werkende mensen kunnen werk, techniek en organisatie wederzijds op elkaar worden afgestemd.

Er is de laatste jaren veel veranderd in het werkend leven van wetenschappers. De aanvangssalarissen zijn verlaagd. De AIO en AO-regelingen zijn ingevoerd waardoor beginners niet langer worden gezien als volwaardige onderzoekers maar als leerlingen. Op de universiteiten is bovendien een nieuw rangstelsel ingevoerd, waardoor een bureaucratische hiërarchisering op gang dreigt te komen. De kans bestaat dat onder invloed van deze ontwikkelingen een arbeidsdeling op de universiteiten plaatsvindt, waarbij de beginners met alleen uitgeholde, uitvoerende taken blijven zitten, terwijl onderzoeksplannen en planningen gemaakt worden door al gevestigde wetenschappers. Dit betekent een verarming van het werk en een verslechtering van de positie van pas afgestudeerden.

De vraag is of met deze veranderingen in het werk ook de kwaliteit van de arbeid voor hen is veranderd. Deze is niet zondermeer te beantwoorden. Buiten de universiteiten bestaat echter een lange traditie van onderzoek naar de kwaliteit van het werkend leven. Het kan zijn dat de begrippen die daar ontwikkeld zijn door arbeidssociologen en -psychologen van dienst kunnen zijn bij het onderzoeken en beschrijven van het werk van wetenschappers en de kwaliteit die het voor hen heeft. Over die onderzoekstraditie gaat dit artikel. Informatie over de dingen die onderzoekers belangrijk vinden in hun werk is van belang bij het inrichten van researchorganisaties.

1. Kwaliteit van de arbeid is maakbaar

Kwaliteit van de arbeid is een centraal begrip in de arbeidswetenschap. Het duidt op de betekenis die het werk voor de werkenden heeft. Aan het begrip zitten twee zijden. Aan de ene kant relateert men de inhoud van het werk of de arbeidssituatie aan de betrokken medewerkers. In de tweede zoekt men naar de aspecten van werk (-beleving) die het management binnen een bedrijf kan regelen en beschouwt men kwaliteit van de arbeid als een ontwerpprincipe. Een voorbeeld van de eerste zijde vinden we in de definitie van Huygen, zoals die is geaccepteerd in de SISWO werkgroep Kwaliteit van de arbeid: "De term kwaliteit van de arbeid duidt op de relatie tussen kenmerken van de arbeidssituatie en kenmerken van de werker (werknemer, arbeider, employé). Het gaat daarbij om de mate van overeenstemming tussen enerzijds de handelingsmogelijkheden die de werksituatie biedt en de eisen die deze aan de werker stelt en anderzijds om het arbeidsvermogen (de capaciteiten) en de belangen van de werker en diens wensen en verwachtingen met betrekking tot de arbeidssituatie. Aldus opgevat heeft het begrip kwaliteit van de arbeid zowel objectieve als subjectieve aspecten. Het is een relatief dynamisch en normatief begrip" (Huygen, 1983, p. 17). De subjectiviteit van de ervaring maakt dat kwaliteit van de arbeid alleen te onderzoeken is in overleg met en samen met de betreffende medewerkers.

De Sitter benadrukt de andere zijde, de maakbaarheid van werksituaties. Hij beschrijft kwaliteit

van de arbeid minder als relatie mens-werk, en sterker als structureigenschappen in het arbeidsproces: "... hoe groter de regelcapaciteit in een arbeidstaak, hoe groter het vermogen het netwerk van sociale (dus ook economische) relaties te ontwikkelen, en zich (de medewerkers F.B.S.) in deze zin sociaal economisch te ontplooiën" (De Sitter, 1981, p. 115). De mate van kwaliteit van de arbeid is de verhouding tussen regelcapaciteit en werkdruk. Kennis, vaardigheid, beschikbare technische middelen en overlegmogelijkheden verschaffen regelcapaciteit, i.e. mogelijkheden om problemen op te lossen die zich in het werk voordoen. De werkdruk bestaat uit de hoeveelheid werk per tijdseenheid, de variatie in de opdrachten en tegelijk de precisie van de ontvangen opdrachten. Door de werkdruk te verlagen en de regelcapaciteiten in een functie op te voeren laat het management de relatieve autonomie van een medewerker in een functie, of die van een hele afdeling, toenemen. Daarmee neemt ook de kwaliteit van de arbeid voor die medewerker(s) toe (De Sitter, 1981; Van Eynatten, 1986). Als aan deze voorwaarden is voldaan, zal een medewerker, mits hij aan de functie-eisen aangepaste kwalificaties heeft, zichzelf kunnen verwerkelijken.

Vanuit het standpunt van mensen als De Sitter en Van Eynatten geredeneerd, is er een direct verband tussen de kwaliteit van de arbeid en de eisen van het werk aan de kwalificaties van de werkenden. Het verhogen van de in een functie ingebouwde kwaliteit van de arbeid verzwaart die functie. De grotere gegeven autonomie motiveert de functiedrager tot leren van zijn werk.

En daar in onze arbeidsverhoudingen vooral de vereiste kennis en vaardigheden beloond worden stijgt met de autonomie het aanzien van de functie in het bedrijf. In termen van functiewaardering kunnen we stellen: "Hoe meer verantwoordelijkheid, ervaring en scholing vereist is, hoe hoger de kwaliteit van een arbeidstaak, maar ook hoe hoger het loon" (De Sitter, 1981; Baitsch & Frei, 1980; In 't Veld, 1978). De twee definities vullen elkaar aan. De eerste benadrukt dat alleen betrokkenen kunnen zeggen of het werk kwaliteit heeft of niet. De tweede benadrukt dat de kwaliteit bewust maakbaar is en als een van de ontwerpprincipes gehanteerd kan worden bij de inrichting van organisaties.

In dit artikel leggen wij de nadruk op de definiëring van de kwaliteit van de arbeid als een ontwerpprincipe. Dit opent de ogen voor het feit dat deze kwaliteit maakbaar is en dat bestaande functies ook uitingen zijn van voltooid beleid op dit gebied, zij het impliciet, zij het expliciet. De benadering van Huygen kiezen we als aanvulling hierop.

2. Maar...

Voor we ingaan op mogelijkheden om het begrip toe te passen op het werk van onderzoekers, moeten we er twee kanttekeningen bij plaatsen. Het gaat hierbij om de figuratie waarbinnen de terminologie en het onderzoek ernaar zijn ontwikkeld en om de voorwaarden waaronder kwaliteit van de arbeid te verbeteren is.

Het begrip "kwaliteit van de arbeid" komt uit de Engels/Amerikaanse sociologische discussie over hoe managers lagere werknemers kunnen motiveren tot werken. De discussie ontstond omstreeks 1950. De achtergrond ervan was een push-pull beweging van enerzijds emanciperende arbeiders en anderzijds sociaal bewogen wetenschappers en vanuit concurrentie-overwegingen automatiserende ondernemers. Het proces is het beste te schetsen als de ontwikkeling van de sociale verhoudingen binnen bedrijven van een bevelshuishouding naar een onderhandelingshuishouding (De Swaan, 1982).

Het geven van meer speelruimte aan de werknemers zou een dubbel effect hebben. Zo zou een synergie ontstaan waarbij de kwaliteit van het produkt, de produktiviteit en de kwaliteit van de arbeid tegelijkertijd stijgen. Het scheppen van nieuwe verhoudingen in bedrijven was nodig om de productie mogelijk of efficiënter te maken. Binnen die nieuwe verhoudingen worden de werkenden minder direct gestuurd door hun chefs. Ze moeten zichzelf leren beheersen. Kwaliteit van de arbeid wordt dikwijls besproken alsof het een inzet is bij de strijd tussen arbeiders en bazen of tussen arbeid en kapitaal om de zeggenschap om het arbeidsproces. Hoewel dit meespeelt is het zeker niet het enige punt.

Kortom, de figuratie waarin het begrip ontwikkeld werd was een dubbele. Enerzijds was er het netwerk van om winstkansen concurrerende ondernemers. Anderzijds was er het bedrijfsinterne

netwerk van gevestigden en relatieve buitenstaanders. Die buitenstaanders waren veelal laaggeschoolde handarbeiders.

De voorwaarden waaronder kwaliteit van de arbeid op de werkplek te verwezenlijken is hangen samen met de figuratie waarbinnen dat werk gedaan wordt. De kwaliteit is een kenmerk van de arbeidsplek en van de verhouding daarvan met een medewerker. De werkplek is echter niet los te zien van haar omgeving. Ze is een subsysteem van een bedrijf. De bedrijfsorganisatie als geheel oefent er dan ook een grote invloed op uit. Tot de genoemde variabelen van het bedrijfsniveau horen:

- arbeidsomstandigheden (lawaai, stank, ruziesfeer)
- het gevoerde bedrijfsbeleid ten aanzien van zaken als personeelsbeleid
- manier van leidinggeven door chefs
- relaties met collega's, ondergeschikten, bovengeschikten
- werkzekerheid
- loopbaanmogelijkheden en
- een voldoende hoog inkomen voor de betrokkenen (Hill, 1971).

Als deze punten niet goed geregeld zijn, dan kan er geen sprake zijn van kwaliteit van de arbeid. Dan heeft het vergroten van de speelruimte binnen functies geen effect. Werknemers zullen dan niet gemotiveerd zijn om te werken. Omgekeerd is het zo dat als deze punten wel in orde zijn dan is dat geen garantie voor een goede motivatie. Daar moet de verbetering van de kwaliteit van de arbeid, ook wel humanisering van het werk, voor zorgen.

3. Geldt dit ook voor wetenschappers?

Het begrip kwaliteit van de arbeid is vruchtbaar gebleken bij het onderzoeken van de betekenis die werk voor mensen kan hebben. Het heeft de ogen geopend voor het feit dat er niet één beste manier is om specifieke arbeidstaken in te richten. Daarbij is erop gewezen dat techniek en organisatie twee zijden zijn van een medaille. Als de techniek verandert, als de doelen veranderen, als de mensen die met die techniek werken veranderen, dan zal men ook de organisatie van het werk moeten veranderen om een goede aansluiting te bereiken tussen de eisen die het werk aan de werkenden stelt en hun kwalificaties en behoeften.

Is het begrip bruikbaar bij het bekijken van onderzoekswerk? We denken van wel. Wetenschappers zullen andere eisen aan hun werk stellen dan industrie-arbeiders of kantoorpersoneel. Daarmee moet rekening gehouden worden bij het onderzoeken van de kwaliteit van de arbeid. Dit betekent dat we, als we het begrip willen toepassen op wetenschappelijk werk, rekening zullen moeten houden met de waarden die gelden binnen het wetenschappelijk systeem. Uit eerste onderzoeken blijkt dat het dan een vruchtbaar uitgangspunt biedt voor het analyseren van arbeidsplaatsen.

Als we kijken naar de punten die vanuit de instelling te regelen zijn aan de kwaliteit van de arbeid, moet gelet worden op het volgende. Daar de samenstelling van een functie steeds de resultante is van de gekozen techniek en de gekozen organisatievorm, kunnen we verwachten dat verschillende wetenschapsgebieden een verschillende arbeidsorganisatie zullen behoeven. Dit zal ook het geval zijn bij de organisatie van vakgroepen en op toegepast onderzoek gerichte instituten. Fisser (1986) kwam dergelijke verschillen tegen in zijn onderzoek naar verschillen in de organisatie van onderzoekswerk in verschillende settings.

Er zal dus niet één ideale arbeidsorganisatie voor wetenschappelijk werk gevonden kunnen worden. Diverse soorten werk en onderscheiden wetenschapsgebieden zullen ieder hun eigen eisen stellen. De wetenschappers zelf zullen in verschillende omstandigheden verschillende manieren van werken en samenwerken prefereren. De volgende punten worden in de meeste artikelen genoemd als van belang:

1. De mate waarin collega's en senioronderzoekers hen met respect behandelen en vertrouwen stellen in hun capaciteiten,
2. De mate van gevarieerdheid in de dagelijkse werkzaamheden,
3. De mate waarin werken een uitdaging betekent, de mate waarin men zijn of haar capaciteiten c.q. kwalificaties kan gebruiken en ontwikkelen in het werk,

4. De mate waarin het bedrijf of het management de werkenden promotiekansen of andere kansen op goede toekomstmogelijkheden biedt,
5. Kun je een gevoel van eigenwaarde aan je werk ontleen? Dit hangt ondermeer samen met het krijgen van feed back op wat je produceert,
6. De mate waarin het werk een zinvolle bijdrage levert aan de maatschappij.

De discussie rond "kwaliteit van de arbeid" komt voort uit het economisch systeem. Daar geldt dat iemands aanzien binnen en buiten het bedrijf voor een groot deel bepaald wordt door zijn inkomen. Voor wetenschappers gaat dit niet zonder meer op. Bij hen wordt veelal het inkomen bepaald door het aanzien dat ze binnen de groep wetenschappers hebben opgebouwd (Elias, 1982).

Dit roept de vraag op of het mogelijk is dat bij wetenschappers andere voorwaarden voor kwaliteit van de arbeid gelden dan bij andere beroepsgroepen. Wetenschappelijke erkenning door collega's kan voor AIO's bijvoorbeeld een belangrijker voorwaarde voor kwaliteit van de arbeid zijn dan inkomen. En het verwerven van erkenning is voor nieuwkomers en relatieve buitenstaanders nu juist een probleem. Ze kunnen die erkenning alleen verwerven als ze werken als volledige onderzoekers.

Bij de voorwaarden op instituutsniveau voor verbetering of handhaving van de kwaliteit kan gelet worden op de al genoemde punten:

- arbeidsomstandigheden: zijn er voldoende bureau's of werkplekken, is er lawaai of overlast door apparatuur?
- het gevoerde bedrijfsbeleid ten aanzien van zaken als personeel: zijn er scholingsmogelijkheden, is congresbezoek mogelijk, is er voldoende geld voor materialen en middelen?
- Hoe is de manier van leidinggeven door bovengeschikten: Zijn de opdrachten precies omschreven of ruim, worden resultaten teruggekoppeld naar de onderzoeker?
- Hoe zijn de relaties met collega's, ondergeschikten, bovengeschikten: Zijn de verhoudingen collegiaal of hiërarchisch?
- Is er werkzekerheid: Werkt men met een arbeidscontract voor onbepaalde tijd of is men in tijdelijke dienst?
- Zijn er loopbaanmogelijkheden en
- publicatiemogelijkheden?
- is er een voldoende hoog inkomen voor de betrokkenen te verdienen? (Hill, 1971).

Nu komen we tot een conclusie. Het begrip kwaliteit van de arbeid is vruchtbaar gebleken bij het onderzoeken van de betekenis die werk voor mensen kan hebben. Het heeft de ogen geopend voor het feit dat er niet één beste manier is om specifieke arbeidstaken in te richten. Daarbij is erop gewezen dat subjectiviteit en maakbaarheid twee zijden zijn van een medaille. Wetenschappers zullen andere eisen aan hun werk stellen dan industrie-arbeiders of kantoorpersoneel en wetenschappelijk werk zal andere eisen aan de medewerkers stellen dan kantoor- of fabriekswerk. Daarmee moet rekening gehouden worden bij het onderzoeken van de kwaliteit van de arbeid. Dan zal het mogelijk blijken om te zoeken naar methoden om wetenschappelijk werk zo te organiseren dat instituut en onderzoekers beiden profiteren van verbetering van de kwaliteit van zowel het produkt als de arbeid.

In een volgend artikel zal ingegaan worden op gebleken veranderingen in de kwaliteit van de arbeid van onderzoekers.

LITERATUUR

Baitsch, C., Frei, F., *Qualifizierung in der Arbeitstätigkeit*. Bern, Stuttgart, Wien, Verlag Hans Huber, 1980.

Blauner, Robert, *Alienation and Freedom*. The Factory Worker and his Industry. London, 1964.

Elias, N., Scientific Establishments. in: Elias, N., Martins, and Whitley, R., (eds), *Scientific Establishments and Hierarchies*. Dordrecht, Boston, London, D. Reidel Publishing Co., 1982. pp. 3-69.

Eynatten, Frans M. van, *S.T.T.A., Naar een nieuw werkstrukturingsparadigma*. Proefschrift Sociale Wetenschappen K.U.N. Eindhoven, Nederlandse Philips Bedrijven BV, 1985.

Fisscher, O.A.M., *Wetenschap bedrijven*. Onderzoek in drie researchinstellingen naar de sturing en motivatie in samenhang met de gerichtheid van het researchwerk. Groningen, BV Erven V. van der Kamp, 1986.

Fricke, W., *Arbeitsorganisation und Qualifikation*. Bonn, 1975.

Hill, P., *Towards a new philosophy of management*. London, Gower Press, 1971.

Huijgen, F., Werkne(e)m(st)ers, Management en kwaliteit van de Arbeid. In: Doorewaard, H., Klaveren, M. van, Pot, F., Woude, M. van der (red.), *Stratego op bedrijfsnivo*. Nijmegen, ITS, 1983.

Levine, Mark F., Taylor, James C., Davis, Louis E., Defining Quality of Working Life. In: *Human Relations*, Vol. 37, nr. 1, 1984. pp. 81-104.

Sitter, L.U. de, *Op weg naar nieuwe fabrieken kantoren*. Deventer, Kluwer, 1981.

Swaan, Abram de, *Uitgaansbeperking en uitgaansangst; over de verschuiving van bevelshuishouding naar onderhandelingshuishouding*. In: idem, *De mens is de mens een zorg*. Opstellen 1971-1981. Amsterdam, Meulenhoff, 1984. 1e editie 1982.

Veld, J. in 't, *Analyse van organisatieproblemen, een toepassing van het denken in systemen der processen*. Amsterdam, Elsevier, 1978.

Proefschriften

Bernadette van Hout Wolters: Markeren van kerngedeelten in studieteksten, een proces-product benadering (Swets & Zeitlinger, 1986).

Promotores: prof. dr. L.F.W. de Klerk en prof. dr. A.M.P. Knoers.

Promotie: Katholieke Universiteit Brabant, 21 november 1986 (cum laude).

Korte samenvatting

Het proefschrift heeft betrekking op het bestuderen van schriftelijk studiemateriaal. Het doel is een bijdrage te leveren aan de verbetering van de onderzoeksmethoden en opzet van het onderzoek op dit terrein. Er worden enkele onderzoeken beschreven die de auteur uitvoerde ten behoeve van de constructie van onderzoeksteksten en meetinstrumenten. Speciale aandacht krijgen de methoden voor het verzamelen van informatie over de leerprocessen en de mentale inspanning van leerlingen tijdens het bestuderen van een studietekst.

De geselecteerde methoden werden toegepast in een onderzoek naar het markeren (onderstrepen) van kerngedeelten in studieteksten door de auteur of docent en door de leerling. Onderzocht werd wat het effect was van drie experimentele condities (docent-onderstreepte kerngedeelten, leerling-onderstreepte kerngedeelten, geen onderstreepte kerngedeelten) op:

- de herkenning en herinnering van de onderstreepte gedeelten door derde klas havo-leerlingen (produktgegevens);
- hun selectieve aandacht, herhaling en mentale inspanning tijdens het bestuderen van de studietekst (procesgegevens).

Iedere leerling kreeg alle experimentele condities (Graeco-Latin Square design). De procesgegevens werden tijdens het bestuderen van de tekst verzameld met de hiervoor ontwikkelde 'XY-methode' en de 'neventaakmethode'. Na de tekstbestudering vulden de leerlingen een proces- en evaluatievragenlijst in.

Zowel de onderstreping van de kerngedeelten door de docent als de onderstreping door de leerling bleken positieve effecten te hebben op de herinnering van die kerngedeelten, in vergelijking met de controleconditie. Onderstreping door de leerling had echter significant méér effect dan onderstreping door de docent. Er waren enkele interacties met leerlingkenmerken (bijv. cognitieve stijl).

De positieve herinneringseffecten bij de leerling-onderstreueringen konden verklaard worden uit een grotere selectieve aandacht voor de onderstreepte kerngedeelten, frequentere herhaling van deze gedeelten en een toename van mentale inspanning bij het selecteren en lezen ervan.

Het proefschrift gaat in op de theoretische en praktische implicaties van deze resultaten en eindigt met een evaluatie van de gehanteerde onderzoeksmethoden.

Congressen

Studieconferentie VERLANGD ONDERWIJS EN SCHOLENPLANNING

20 november 1987

Op vrijdag 20 november 1987 organiseren de Vereniging voor Onderwijsrecht (VOR) en het Instituut voor Toegepaste Sociale Wetenschappen te Nijmegen (ITS) een studieconferentie over verlangd onderwijs en scholenplanning. Aanleiding hiervoor is de discussie over de juridische mogelijkheden om de voorkeuren van ouders, te meten door middel van enquêtes, te betrekken bij de scholenplanning. Dergelijke onderzoeken zijn de laatste paar jaar in verschillende regio's en gemeenten gehouden (o.a. Bollenstreek, Roermond, Tilburg, Maastricht).

Tijdens de conferentie komen in het ochtendgedeelte de sociaal-wetenschappelijke aspecten van deze 'directe meting' aan de orde; 's middags zullen de juridische implicaties daarvan ter discussie worden gesteld.

De conferentie zal plaatsvinden in een van de gebouwen van de Katholieke Universiteit te Nijmegen.

De deelnameprijs bedraagt fl. 125,-; voor leden van de Vereniging voor Onderwijsrecht fl. 95,-; voor studenten geldt op aanvraag een gereduceerd tarief. De deelnameprijs omvat ook de lunch en de ontvangst van het conferentieboek na afloop van de conferentie.

Inlichtingen kunnen worden verkregen bij het secretariaat, telefoon 080-780111, mevrouw Chérie Bocken.

COLLOQUIAREEKS SCO

Evaluatie van de implementatie van de Tweefasenstructuur. Drs. J.K. Koppen 25-11-87

Hoger Onderwijs en Arbeidsmarkt. Drs. F. van den Berg van Saparoea 27-01-88

Wisselwerking tussen onderwijs en arbeid. 24-02-88

Analyse op drie niveaus: systemen, scholen en bedrijven, leren en werken.

Drs. P. van den Dool

Onderwijsonderzoek en onderwijspraktijk: hoe groot is de kloof? 23-03-88

Een gevalsbeschrijving aan de hand van onderzoek naar het vak Nederlandse taal.

Dr. H. Blok.

Inlichtingen: SCO, 020-264547.

LANDELIJKE DAG STUDIEVAARDIGHEDEN 1988

29 januari 1988, Katholieke Universiteit Brabant

Oproep voor bijdragen

Ondermeer door eens per jaar een landelijke dag te organiseren, bevordert het Landelijk Overleg Studievaardigheden (LOS) contacten tussen diegenen die een praktische, theoretische en/of onderzoeksmatige interesse hebben op het gebied van studievoordigheden in het hoger onderwijs.

Op vrijdag 29 januari 1988 vindt de 8e Landelijke Dag Studievoordigheden plaats in Tilburg in de gebouwen van de katholieke Universiteit Brabant. De bijdragen aan deze dag zullen alle in het teken staan van een centraal thema:

"Rendementsverbetering via studiebegeleiding?"

Het is de bedoeling dat in alle bijdragen op de één of andere manier aandacht wordt besteed aan de vraag of en zo ja hoe door middel van de diverse vormen van studiebegeleiding het studierendement verbeterd kan worden. Dit kan men doen door onderzoeksresultaten te presenteren of te interpreteren, maar ook door praktijkervaringen te bespreken of visies, theorieën of plannen te bediscussieren.

Inlichtingen: OnderwijsResearchCentrum, secr. LDS, KUB, Postbus 90153, 5000 LE Tilburg.

Ontvangen publicaties

- De Corte, E. (1986). Onderwijsleerprocessen, bevindingen van Leuvens onderwijspsychologisch onderzoek. Amersfoort: Acco (f43,25)
- Earl, T. (1987). Cursusontwikkeling. Almere: Versluys.
- Van Dijk, C. (1986). Handboek stages. Nijmegen: ITS.
- Van Kessel, N. (1987). Van meester naar professional. Een onderzoek naar het functioneren van de OBD-Rijnmond-Zuid. Nijmegen: ITS.
- COA (1987). Nieuwsbrief 5.
- Franssen, H.A.M., Stoverinck, T.W.H.M. & Van der Ploeg, D.A. (1987). Gericht geëvalueerd, eindrapport SVO-project 52704. Utrecht: VOU.
- Everts, H., Golhof, A., Strassen, P. & Teunissen, J. (1986). De cultureel-etnische situatie op OVB-scholen. Utrecht: VOU.
- Kallen, D.B.P. (1987). Arbeidsoriëntatie geëvalueerd. Rotterdam: COA.
- Arbeidsoriëntatie, Informatie voor scholen. Rotterdam: COA.
- Verspreidingsplan van het EG-project 'Arbeidsoriëntatie voor Jongeren' 1987-1991. Rotterdam: COA.
- Van Kroonenburg, J., Corten, H., Van Kessel, N. & Van Kuijk, J. (1986). Op weg naar Pedagogische Academies voor het Basisonderwijs. Nijmegen: ITS.
- Oud-de Glas, M., & Buis, Th. (1987). De toekomst van het voortgezet onderwijs. Nijmegen: ITS.
- Pennings, A.H., Van den Born, J.J.M.L., Fischer, C., Van de Perel, R., Tissink, J. & Vandewall, J. (1987). Letterwerk. Utrecht: VOU + handleiding bij het programma letterwerk (COO aan kinderen met leerstoornissen).
- Wolbert, R.G.M., Schaap-Hummel, I. & Span, P. (1986). Individualisering en differentiatie in de basisschool (+ bijlagen). Utrecht: VOU.
- Van Dijk, W. & Span, P. (1986). De opvang van begaafde leerlingen. Utrecht: VOU.
- Jansen Schoonhoven, A., Van der Linden, J., Posthuma de Boer, M., Span, P. & Van Susante, M. (1987). Hoogbegaafde leerlingen. Een onderzoek naar het onderkennen en onderwijzen van hoogbegaafde leerlingen in het voortgezet onderwijs. Utrecht: VOU (+ bijlage).
- Hamaekers, A., Lkoundi, M., & Teunissen, J.M.F. (1986). Materiaalontwikkeling voor het onderwijs in de arabische taal en cultuur. Utrecht: VOU.

BEKROOND ORD PAPER 1987

Dit artikel is een bewerking van het door de VOR bekroonde ORD-paper.

Overleving in de Bovenstroom van het Voortgezet Onderwijs: Veranderde Milieu-effecten voor Jongens en Meisjes*

Paula Willemse

ABSTRACT

This article focuses on the changes in effect of social background, gender and achievement on the individual school careers in secondary education since 1950. In comparing two Dutch generations of students, the first born around 1940, the second around 1965, this study concentrates especially on the transitions between the schoolyears in the upper-stream of secondary education (vhmo/vwo). This approach yields a clear picture of the changes in inequality of educational opportunity, possibly as a result of educational reforms of the sixties. The data have been analysed by means of the simulation-model of Boudon. The results indicate that the effect of achievement has increased, the position of girls has improved and there is a trend to postpone selection in secondary education.

INLEIDING

In dit artikel wordt verslag gedaan van een onderzoek naar veranderingen in de invloed van milieu, sexe en prestatie op de individuele schoolloopbanen in het voortgezet onderwijs, sinds de Tweede Wereldoorlog.

Theoretisch sluit het onderzoek aan bij een belangrijke discussie binnen de sociologie, die zich toespitst op de vraag welke factoren een rol spelen in het verwerven van een positie in het onderwijs, die vervolgens mede bepalend is voor de positie in de maatschappij. Worden deze posities verworven op basis van vooral toegeschreven kenmerken zoals milieu en sexe, of zijn variabelen die meer te maken hebben met eigen inspanningen zoals leerprestaties hierin bepalend? Daarbij is het interessant na te gaan of in een bepaalde periode de ongelijkheid in de kans op een bepaalde onderwijspositie konstant blijft of verandert en in hoeverre deze verandering toegeschreven kan worden aan bepaalde onderwijsmaatregelen. Eerder onderzoek naar veranderingen in onderwijskansen spitste zich toe op de vergelijking van schoolloopbanen in het lager onderwijs en de aansluiting met het voortgezet onderwijs (zie voor een overzicht: Dronkers, 1986). Inmiddels zijn er data beschikbaar die het mogelijk maken schoolloopbanen in het voortgezet onderwijs van verschillende generaties met elkaar te vergelijken. Faasse, Bakker, Dronkers en Schijf (1986), bij wie dit onderzoek aansluit, hebben onderzocht of de effecten, waarvan verondersteld wordt dat ze de schoolloopbanen in het voortgezet onderwijs beïnvloeden, veranderd zijn door twee generaties leerlingen beide uit de provincie Noord-Brabant, met elkaar te vergelijken. De ene generatie leerlingen is geboren rond 1940, de andere rond 1965. Zij vonden dat de invloed van sexe over een periode van vijftientig jaar vrijwel is verdwenen, de invloed van milieu op onderwijsposities via prestatie (het primaire milieu-effect) is stabiel gebleven en de invloed van milieu waarbij prestatie konstant wordt gehouden (het sekundaire milieu-effect) is afgenomen (zie voor vergelijkbare resultaten: Meester en de Leeuw, 1985). In het onderzoek van Faasse e.a. zit echter een ernstige beperking doordat schoolloopba-

* Dit artikel is een ingekorte versie van het paper zoals gepresenteerd tijdens de ORD. De paper-versie met uitgebreidere analyses en toelichtingen is wel opvraagbaar bij de auteur: Paula Willemse, Duiventoren 10, 5046 ML Tilburg.

nen gekarakteriseerd worden door na te gaan wat de keuze voor een bepaald schooltype op twaalfjarige leeftijd is, vervolgens kijken zij wat het laatst gevolgde type onderwijs is op ongeveer zeventienjarige leeftijd. Door alleen het begin- en eindpunt van het voortgezet onderwijs te gebruiken als omschrijving van de schoolloopbaan, gaat interessante informatie over verandering van ongelijkheid in de doorstroming binnen zo'n schoolloopbaan verloren. In mijn vergelijking van twee cohorten leerlingen, concentreer ik me juist op de overgangen tussen de verschillende schooljaren in het voortgezet onderwijs. Dit maakt het mogelijk uitspraken te doen over eventuele veranderingen per breekpunt of schooljaar en of leerlingen een bepaald onderwijsnivo blijven volgen, afhaken of juist weer gaan volgen. In deze studie concentreer ik me op het al dan niet volgen van het vmbo of vwo.

Niet alleen vanuit theoretisch, maar ook vanuit maatschappelijk oogpunt, is het belangrijk inzicht te verkrijgen in de veranderde invloed van milieu, sexe en prestatie op doorstroming binnen het voortgezet onderwijs. Tussen 1952 en 1977, de jaren waarin de generaties leerlingen begonnen met het voortgezet onderwijs, heeft zich namelijk een ingrijpende onderwijsvernieuwing voltrokken in de vorm van de Mammoetwet. Een van de doelstellingen was het bevorderen van meer gelijke kansen voor gelijk begaafden, door onder andere invoering van brugklassen en scholengemeenschappen. Door vergelijking van de twee generaties kan empirisch aangetoond worden of deze doelstelling ook gerealiseerd is. Dit zou moeten blijken uit een afnemend sekundair effect van milieu op onderwijskansen, in het voortgezet onderwijs.

De probleemstelling die in het onderzoek centraal staat is driedelig:

1. verminderen in een tijdsbestek van vijftientwintig jaar de sekundaire milieu-effecten in schoolloopbanen en zijn deze veranderingen voor jongens en meisjes gelijk?
2. in welk schooljaar of op welk breekpunt zijn die verminderingen in sekundaire effecten te lokaliseren en geven jongens en meisjes een afwijkend beeld te zien?
3. neemt in vijftientwintig jaar het belang van prestatienivo aan het einde van de lagere school in gewicht toe, voor het verder verloop van de schoolloopbaan en geldt dit in gelijke mate voor jongens en meisjes?

DATA

De probleemstelling wordt onderzocht aan de hand van gegevens over twee generaties leerlingen uit Noord-Brabant. In dit onderzoek wordt gebruik gemaakt van gegevens uit beide generaties, zoals die door Faasse e.a. (1986) bewerkt zijn. Het oudste cohort bestaat uit het Sonnemans-Matthijssen-cohort en bevat leerlingen die rond 1940 geboren zijn en in 1952 naar het voortgezet onderwijs gingen. Matthijssen en Sonnemans hebben op basis van de prestatiescores de meest presterende helft geselecteerd en hebben in 1957 en 1958 deze leerlingen opnieuw ondervraagd over het verdere verloop van hun school- en beroepsloopbaan.

De generatie 1965 wordt gevormd door het Brabantse deel van het smvo-cohort. Dit bestand is opgebouwd door het Centraal Bureau voor de Statistiek. Leerlingen uit dit cohort doen in 1977 hun intrede in het voortgezet onderwijs. Na een weging zijn de landelijke gegevens representatief voor de leerlingen uit de tweede klas van het voortgezet onderwijs in Noord-Brabant. Uit dit cohort zijn ook de meest presterende leerlingen geselecteerd, zodat de cohorten te vergelijken zijn (1).

Beide, door Faasse e.a. bewerkte, cohorten bevatten gegevens over de sociale achtergrond van de leerlingen, over hun prestatienivo en advies eind lagere school en over gegevens van hun schoolloopbaan sinds het einde van de lagere school. Alle gegevens die nodig zijn om, met Boudon als analysekader, de probleemstelling te onderzoeken zijn dus beschikbaar (2).

DE ANALYSETECHNIEK

Het analysekader dat in dit onderzoek gebruikt wordt is afkomstig van Boudon (Boudon 1974;

zie ook Dronkers e.a. 1982). Er zijn natuurlijk nog andere analysemethoden mogelijk waarmee men de probleemstelling kan benaderen. Boudon echter levert in het licht van de probleemstelling een geschikt analysekader, omdat met behulp van zijn model de mogelijke variatie in beslissingsprocessen naar sociale achtergrond empirisch geanalyseerd kan worden in elke stap van de schoolloopbaan.

In zijn simulatiemodel van de relaties tussen onderwijs en sociale ongelijkheid staan een aantal axioma's centraal, die bruikbaar zijn voor het schatten van parameters. De axioma's worden hieronder kort toegelicht.

Axioma 1:

De Westerse samenleving is gelaagd: 10% neemt een hoge positie in (C1), 30% een middenpositie (C2) en 60% een lage (C3).

Axioma 2:

Het primaire effect van sociaal milieu op schoolloopbanen. Volgens Boudon hebben kinderen afkomstig uit een hoger sociaal milieu aan het einde van de lagere school een hoger prestatienivo dan kinderen uit een lager sociaal milieu. Een lagere sociale status gaat nl. gepaard met een armere culturele achtergrond, waardoor leerlingen uit dit milieu een lager prestatienivo hebben aan het einde van de lagere school dan leerlingen uit een hoger sociaal milieu.

Axioma 3:

Naast het primaire milieu-effect onderscheidt Boudon een sekundair milieu-effect. Als twee leerlingen gelijk zijn met betrekking tot prestatie, leeftijd, e.d. en niet wat betreft milieu, dan zullen zij op grond van verschillende kosten- en batenanalyses, andere keuzes maken met betrekking tot het mee blijven doen in de stroom die toegang geeft tot het wetenschappelijk onderwijs (het zgn. high curriculum). Met andere woorden: verwachtingspatronen, waarop de kosten- en batenanalyses gemaakt worden, verschillen per positie die iemand inneemt in het stratifikatie-systeem.

De mogelijkheid om op het hoogste nivo door te blijven stromen, hangt af van zowel het prestatienivo als de sociale achtergrond.

Axioma 4:

Een leerling moet tijdens zijn verdere schoolloopbaan in het voortgezet onderwijs kiezen of hij wel of niet op het hoogste nivo mee blijft doen. Boudon veronderstelt dat vanaf het einde van de lagere school een aantal breekpunten te onderscheiden is. Het percentage leerlingen dat bij een bepaald breekpunt het hoogst mogelijke nivo kiest, wordt overlevingspercentage genoemd.

In ons onderzoek gaan we bij de start van het voortgezet onderwijs uit van deze axioma's en de bijbehorende waarden van de parameters. Vervolgens kijken we hoe de overlevingspercentages per overgang in de schoolloopbaan zich ontwikkelen.

Om het model van Boudon zo goed mogelijk na te bootsen, zijn allereerst beide cohorten opgedeeld in drie milieugroepen (zie axioma 1), waarbij het beroep van de vader als indicator is gebruikt (3). De percentages van Boudon zijn als uitgangspunt genomen.

Tabel 1: frequentieverdeling van sociale achtergronden bij beide cohorten en de parameters van Boudon.

sociale achtergrond	generatie '40	generatie '65	Boudon
C1 (hoog)	8% (n = 192)	12% (n = 148)	10%
C2	33% (n = 823)	36% (n = 446)	30%
C3 (laag)	59% (n = 1493)	53% (n = 662)	60%

Vervolgens is een tweede driedeling van prestatienivo gemaakt (zie axioma 2), waarbij weer de percentages vergelijkbaar zijn gemaakt met het model van Boudon, zodat zoveel mogelijk het veronderstelde primaire effect van sociale achtergrond wordt gesimuleerd (4).

Volgens deze procedure worden aan de start van het voortgezet onderwijs negen milieu-prestatiegroepen vastgesteld, voor de beide generaties. Deze vergelijkbare groepen worden vervolgens nog eens uitgesplitst naar sexe.

Tabel 2: verdeling prestatienivo's over de sociale achtergronden in procenten en ter vergelijking de percentage van Boudon.

	generatie '40			generatie '65			Boudon		
	R1	R2	R3	R1	R2	R3	R1	R2	R3*
C1	64	26	10	50	34	16	60	30	10
C2	45	36	19	43	33	24	50	30	20
C3	32	35	33	34	36	30	30	40	30

* C1 = hoog milieu
C2 = midden milieu
C3 = laag milieu

R1 = hoog prestatienivo
R2 = midden prestatienivo
R3 = laag prestatienivo

Om te kunnen onderzoeken waar nu in de schoolloopbaan de overlevingskansen meer of minder milieugebonden zijn in de vergelijking van twee generaties leerlingen, worden een aantal breekpunten onderscheiden. Een leerling moet tijdens de schoolloopbaan op verschillende overgangen of breekpunten kiezen of hij wel of niet op het hoogste nivo mee blijft doen (5). De breekpunten van de stroom die toegang geeft tot het wetenschappelijk onderwijs zijn in dit onderzoek onderscheiden in:

1. advies vhm/vwo
2. keuze begin eerste jaar vhm/vwo of brugklas
3. keuze begin tweede jaar vhm/vwo
4. keuze begin derde jaar vhm/vwo
5. keuze begin vierde jaar vhm/vwo
6. keuze begin vijfde jaar vhm/vwo (6).

Ten aanzien van het tweede breekpunt: het bepalen van de keuze voor het eerste jaar van het vhm (hbs, gymnasium, lyceum of seminarie) levert voor het cohort '40 geen problemen op. Na de invoering van de Mammoetwet is echter in veel gevallen sprake van een uitgestelde keuze. Nadat één of meer brugjaren doorlopen zijn, wordt pas voor een bepaald schooltype gekozen. In dit onderzoek worden brugklassers bij het vwo ingedeeld als bekend is dat leerlingen zich in een brugklas met vwo-mogelijkheid bevinden. Het zo omgaan met de brugklas is in overeenstemming met eerdere studies (zie Faasse e.a.). Een van de doelen van het invoeren van brede brugklassen, een centrale maatregel binnen de Mammoetwet, is het bevorderen van een betere doorstroming. Het effect van de maatregel kunnen we op deze wijze onderzoeken. Nadat volgens de axioma's van Boudon vergelijkbare milieu/prestatiegroepen geconstrueerd zijn, wordt bij elk breekpunt berekend hoeveel procent van de leerlingen van elke groep in de stroom blijft die toegang geeft tot het wetenschappelijk onderwijs *en* hoeveel procent van de leerlingen die een ander breekpunt niet overleefde later weer in de stroom terugkeert (het terugstroompercentage). Boudon betreft deze terugkeermogelijkheid niet in zijn model. In navolging van Dronkers, de Graaf e.a. (1985) worden terugstroompercentages in dit onderzoek wel in de analyse betrokken.

De C2R1-groep van de generatie '65 bestaat uit 192 leerlingen (zie ook tabel 1 en 2). 46% van deze leerlingen krijgt een vhm/vwo-advies. Van degenen die een positief advies hebben gekregen maakt bijna iedereen (93%) ook daadwerkelijk de keuze voor het vwo of brugklas die daar aansluiting op geeft. Van degenen die niet het vwo-advies hebben gekregen maakt 74% de keuze voor het vwo. 77% overleeft vervolgens het tweede jaar van het vwo, het terugstroompercentage bedraagt 9%. Deze procedure wordt herhaald tot en met het laatste breekpunt. De overlevingspercentages van deze groep zijn terug te vinden in tabel 3.

De tabellen die het resultaat zijn van deze analyses zullen hier echter niet alle weergegeven worden. We beperken ons tot de weergave van de analyse aan de hand van de overlevingspercentages.

Een voorbeeld:

Start: C2R1-groep generatie '65 n = 192

	overleving		afstroom
1. advies	89 $89/192 \times 100 = 46\%$		103-
2. keuze begin eerste jaar	83+ $83/89 \times 100 = 93\%$	6- terugstroom 76+ $76/103 \times 100 = 74\%$	27-
3. keuze begin tweede jaar	159 122+ $122/159 \times 100 = 77\%$	37- 3+ $3/33 \times 100 = 9\%$	33- 30-
4. keuze begin derde jaar	125 109+ $109/125 \times 100 = 87\%$	16- 1+ $1/66 = 2\%$	67- 66-
5. keuze begin vierde jaar	110 95+ $95/110 = 86\%$	15- 4+ $4/82 = 5\%$	82- 78-
6. keuze begin vijfde jaar	99 94+ $94/99 = 95\%$	5- 1+ $1/93 = 1\%$	93- 92-

EERSTE RESULTATEN

Uit de analyse, zoals weergegeven in tabel 3 en 4, blijkt dat voor beide generaties het eerste breekpunt 'advies vmo/vwo' in vergelijking met andere breekpunten een zwaar breekpunt is. Voor alle groepen is de uitval hier nogal groot. Ook al heeft een leerling bijvoorbeeld een hoog milieu en een hoog prestatie-nivo, dan nog krijgt maar ongeveer 50% uit deze groepen een vmo/vwo-advies. In de generatie '65 wordt het advies van de onderwijzer relatief vaker opgevolgd dan in de generatie '40. Van degene die een vwo-advies heeft, maakt bijna iedereen ook daadwerkelijk de keuze voor het begin van het eerste jaar vwo, of brugklas die hier aansluiting op geeft. Verder valt op in tabel 3 dat leerlingen uit de generatie '40, die eenmaal het tweede breekpunt gepasseerd zijn, verder een grote kans op overleving hebben in de volgende vier jaren van het vmo. De uitval in deze generatie is vooral in het begin van de schoolloopbaan groot. Dit beeld is voor de '65-generatie anders. De uitval na het eerste breekpunt concentreert zich niet op een bepaald schooljaar, maar is als het ware meer uitgesmeerd over de gehele schoolloopbaan. In eerste instantie worden leerlingen uit dit cohort, door onder andere brugklassen, langer bij elkaar gehouden, echter in het verdere verloop van de schoolloopbaan is de uitval groter dan vijftientig jaar terug.

Als we in tabel 4 de jongens uit de generatie '40 vergelijken met de meisjes uit diezelfde generatie, is duidelijk zichtbaar dat jongens vooral op het eerste breekpunt een voorsprong hebben ten opzichte van meisjes. Vijftientig jaar later hebben de meisjes hun achterstand ten opzichte van jongens ingehaald en vertonen zij op sommige breekpunten een voorsprong. Over meisjes kunnen we verder niet zoveel uitspraken doen, omdat we weinig gegevens hebben van deze categorie. Dit is omdat in de generatie '40 de overleving bij de meisjes gering is, waardoor de aantallen leerlingen in de cellen erg klein worden. Statistisch gezien kan men moeilijk een

Tabel 3: overlevingspercentages voor het '40-cohort en het '65-cohort, totale populatie*

		generatie n40			generatie '65		
		R1	R2	R3	R1	R2	R3
Breekpunt 1.							
Advies vhm/vwo	C1	50	25	–	48	12	8
	C2	28	13	5	46	8	2
	C3	18	6	2	33	4	1
Breekpunt 2.							
Keuze eerste jaar vhm/vwo	C1	75	–	–	100	–	–
	C2	76	61	–	93	–	–
	C3	54	66	–	93	–	–
Breekpunt 3.							
Keuze tweede jaar vhm/vwo	C1	97	–	–	81	47	–
	C2	96	94	–	77	55	18
	C3	86	92	–	64	35	16
Breekpunt 4.							
Keuze derde jaar vhm/vwo	C1	97	–	–	84	–	–
	C2	91	93	–	87	60	–
	C3	85	88	–	84	59	–
Breekpunt 5.							
Keuze vierde jaar vhm/vwo	C1	94	–	–	96	–	–
	C2	95	85	–	86	84	–
	C3	87	92	–	84	75	–
Breekpunt 6.							
Keuze vijfde jaar vhm/vwo	C1	86	–	–	89	–	–
	C2	82	86	–	95	95	–
	C3	93	71	–	91	–	–

* percentages berekend uit een totaal aantal van minder dan twintig zijn weggelaten.

lege cel ('40) met een gevulde cel vergelijken ('65). Wel kan geconcludeerd worden dat de meisjes van het jongste cohort relatief vaker overleven dan de meisjes uit de generatie '40. In de vergelijking jongens '40 met jongens '65 zien we dat ook hier het eerste breekpunt 'advies' een relatief zwaar breekpunt is. In de jongste generatie wordt het advies, in zoverre we daar gegevens over hebben, strak opgevolgd. Op het derde breekpunt, dus na het eerste jaar van het voortgezet onderwijs, is voor de generatie '65 de uitval erg groot in vergelijking met het '40-cohort. Deze terugval blijft bovendien nog een paar schooljaren nadreunen. Een mogelijk effect van de brugklas komt hier naar voren. In eerste instantie worden leerlingen langer bij elkaar gehouden, na het verlaten van de brugklas is de overleving geringer in vergelijking met het '40-cohort. Met andere woorden aan het begin van de schoolloopbaan is de doorstroom verbeterd, na de brugklas is deze tendens niet meer zo duidelijk waar te nemen.

Opvallend in de terugstroompercentages voor de totale populatie (hier niet in tabelvorm weergegeven) is dat in de generatie '65 deze percentages op het tweede breekpunt veel hoger zijn dan die van de generatie '40. Relatief veel leerlingen die niet het advies voor het vwo kregen, maken toch de keuze voor dit type onderwijs of voor de brugklas die hier aansluiting op geeft. Een andere conclusie die we uit deze eerste gegevens kunnen trekken is dat in de generatie '40 vooral leerlingen met een hoog milieu kans hadden alsnog terug te keren in de stroom die toegang geeft tot het wetenschappelijk onderwijs.

Vijfentwintig jaar later zien we een tendens dat een hoog prestatienivo die positie heeft overgenomen.

Verder kan opgemerkt worden dat in de generatie '40 de jongens relatief vaker terugstroomden dan de meisjes, in het '65-cohort is dat niet zo duidelijk te stellen. Wel is, voor zowel jongens

Tabel 4: Overlevingspercentages voor het '40-cohort en het '65-cohort, apart voor jongens en meisjes.

		'40-jongens			'40-meisjes			'65-jongens			'65-meisjes		
		R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
Breekpunt 1.													
Advies vmo/vwo	C1	70	40	-	26	8	-	46	8	-	50	15	-
	C2	39	21	9	12	4	1	45	11	1	48	6	3
	C3	26	10	4	9	1	1	35	7	2	30	0	0
Breekpunt 2.													
Keuze begin eerste jaar vmo/vwo	C1	78	-	-	-	-	-	-	-	-	-	-	-
	C2	80	68	-	-	-	-	89	-	-	98	-	-
	C3	65	69	-	20	-	-	90	-	-	97	0	0
Breekpunt 3.													
Keuze begin tweede jaar vmo/vwo	C1	95	-	-	-	-	-	81	-	-	80	-	-
	C2	96	100	-	96	-	-	73	60	15	81	50	-
	C3	87	96	-	-	-	-	57	28	14	71	42	20
Breekpunt 4.													
Keuze begin derde jaar vmo/vwo	C1	98	-	-	95	-	-	85	-	-	83	-	-
	C2	94	91	-	79	-	-	89	65	-	86	-	-
	C3	84	88	-	-	-	-	83	-	-	85	60	-
Breekpunt 5.													
Keuze begin vierde jaar vmo/vwo	C1	96	-	-	90	-	-	91	-	-	100	-	-
	C2	95	85	-	95	-	-	89	-	-	84	-	-
	C3	84	91	-	-	-	-	89	-	-	80	-	-
Breekpunt 6.													
Keuze begin vijfde jaar vmo/vwo	C1	89	-	-	-	-	-	80	-	-	96	-	-
	C2	83	83	-	-	-	-	94	-	-	96	-	-
	C3	94	74	-	-	-	-	89	-	-	93	-	-

als meisjes, de positie van de leerlingen met een lage sociale achtergrond maar met een hoog prestatienivo aanmerkelijk verbeterd is in vergelijking met de generatie '40.

HET VERGELIJKEN VAN OVERLEVINGSPERCENTAGES

Tabel 3 en 4 kunnen beschouwd worden als basisgegevens, aan de hand waarvan de probleemstelling benaderd wordt. Alvorens dieper ingegaan wordt op de vragen zoals geformuleerd in de probleemstelling, geef ik in het onderstaande kort het probleem aan dat zich voordoet op het moment dat overlevingspercentages vergeleken gaan worden. Het volgende voorbeeld, ontleend aan Jean-Claude Combessie, zal dit probleem verduidelijken.

Tabel 5: deelnamecijfers van twee generaties, ontleend aan Boudon.

sociaal milieu	generatie geboren voor 1910	generatie geboren tussen 1935 en 1940
klasse 1 (hoog)	37%	62%
klasse 2 (laag)	1%	10%

Deze percentages kunnen op verschillende wijzen met elkaar vergeleken worden, bijvoorbeeld:

- | | | |
|-------------------|-------------------------------|----------------------------|
| a. $37/1 = 37$ | vergelijken met $62/10 = 6,2$ | ongelijkheid is afgenomen |
| b. $62/37 = 1,68$ | vergelijken met $10/1 = 10$ | ongelijkheid is afgenomen |
| c. $37-1 = 36$ | vergelijken met $62-10 = 52$ | ongelijkheid is toegenomen |
| d. $62-37 = 25$ | vergelijken met $10-1 = 9$ | ongelijkheid is toegenomen |

Duidelijk wordt dat afhankelijk van de berekeningswijze die toegepast wordt, men tot tegenstrijdige conclusies kan komen over de ontwikkeling van ongelijkheid in onderwijskansen. Jean-Claude Combessie (1984) gaat in zijn artikel dieper in op dit probleem. Aan de hand van bovenstaande vier gegevens past hij meer dan 24 bewerkingen toe. Boudon betreft zowel verhoudingen van als verschillen tussen percentages in zijn analyse. Hij spreekt geen duidelijke voorkeur uit voor één van deze procedures. Uit literatuur wordt niet duidelijk welke procedure de beste is. Het zou te ver voeren om in het kader van dit onderzoek 24 verschillende procedures toe te passen. In mijn vergelijking van overlevings- en terugstroompercentages van twee generaties leerlingen heb ik me beperkt tot vier vergelijkingsprocedures (zie voorbeeld Combessie), die illustratief kunnen zijn voor het door Combessie geschetste probleem.

De probleemstelling wordt steeds onderzocht vanuit vier verschillende invalshoeken. Bij twee rekenprocedures worden verhoudingen van percentages met elkaar vergeleken, bij de andere twee procedures worden verschillen tussen percentages met elkaar vergeleken. De overlevings- en terugstroompercentages van de totale populatie en van jongens en meisjes van beide generaties worden steeds als uitgangspunt genomen.

DE AFNAME VAN HET SEKUNDAIRE MILIEU-EFFEKT IN VIJFENTWINTIG JAAR

Volgens de vier vergelijkingsprocedures is berekend of het sekundaire milieu-effekt in overlevings- en terugstroomkansen in de vergelijking van twee cohorten, kleiner wordt, stabiel blijft of groter wordt. De vergelijkingen zijn alleen uitgevoerd indien er betrouwbare gegevens waren. Uitgaande van tabel 3 kunnen we bijvoorbeeld zien dat op het eerste breekpunt voor C1 - C3 twee vergelijkingen gemaakt kunnen worden volgens de relatieve milieuverschillenprocedure. Eén vergelijking binnen R1 en één vergelijking binnen R2. Op de andere breekpunten kan steeds slechts binnen R1 deze vergelijking gemaakt worden.

1. eerste breekpunt	$50/18=2,78$	$48/33=1,45$	kleiner
2. eerste breekpunt	$25/6=4,17$	$12/4=3,00$	kleiner
3. tweede breekpunt	$75/54=1,39$	$100/93=1,08$	kleiner
4. derde breekpunt	$97/86=1,13$	$81/64=1,27$	groter
5. vierde breekpunt	$97/85=1,14$	$84/84=1,00$	kleiner
6. vijfde breekpunt	$94/87=1,08$	$96/84=1,14$	stabiel
7. zesde breekpunt	$86/93=0,92$	$89/91=0,98$	stabiel

Tabel 6 is het eindresultaat van deze procedures. De resultaten van de analyse van terugstroompercentages zullen hier niet weergegeven worden. Het sekundair milieu-effekt is berekend voor de totale populatie en nog eens apart voor jongens en meisjes, op elk breekpunt. In de analyse is niet alleen de hoge milieugroep met de laagste milieugroep vergeleken. Voor een completer inzicht in de veranderde milieu-effecten is ook de vergelijking van de tussenmilieugroep met de hoogste en laagste groep in de analyse betrokken.

De eerste vraag van de probleemstelling, verminderen sekundaire milieu-effecten en geldt dit voor zowel jongens als meisjes, kan ten dele positief beantwoord worden. Uit de vergelijking van twee generaties leerlingen blijkt dat de invloed van sociaal milieu op het al of niet blijven doorstromen in de bovenstroom van het voortgezet onderwijs kleiner is geworden. Deze conclusie is in overeenstemming met hetgeen Faasse e.a. (1986) constateerden. Echter Faasse e.a. gebruikten onder meer correlaties die weinig of niet gevoelig zijn voor het hier gevonden type veranderingen. Het voordeel van de techniek die in dit onderzoek gebruikt is, is dat men op deze manier inzicht krijgt in welke milieu-kategorieën de sekundaire milieu-effecten veranderen. Uit de analyse komt namelijk naar voren dat niet alle milieugroepen evenredig van de afname in ongelijke onderwijskansen profiteren. De vermindering in sekundaire milieu-effecten gaat wel op in de vergelijking van het hoogste met het laagste milieu en ook tussen het hoogste en het middelste nivo. Maar voor de laagste sociale groepen is hier beduidend minder sprake van. In de kansongelijkheid tussen de twee laagste milieugroepen, gecontroleerd voor

Tabel 6: verandering van sekundaire milieu-effecten, gebaseerd op metingen van verhoudingen van en verschillen tussen alle betrouwbare overlevingspercentages van beide cohorten.**

	Ongelijkheidsverandering tussen C1 en C3			Ongelijkheidsverandering tussen C1 en C2			Ongelijkheidsverandering tussen C2 en C3		
	kleiner	stabiël	groter	kleiner	stabiël	groter	kleiner	stabiël	groter
a*. totale populatie	4	2	1	3	3	1	4	4	3
b*. totale populatie	4	2	1	3	2	2	2	6	3
a. jongens	4	1	1	3	2	1	5	1	3
b. jongens	4	1	1	4	1	1	4	2	3
a. meisjes	1	-	-	2	-	2	-	-	1
b. meisjes	-	1	-	2	1	1	-	2	1
c*. totale populatie	4	2	1	3	2	2	3	5	3
d*. totale populatie	4	2	1	4	1	2	2	6	3
c. jongens	4	1	1	3	3	-	5	2	2
d. jongens	4	1	1	4	1	1	4	2	3
c. meisjes	1	-	-	2	-	2	-	-	1
d. meisjes	-	1	-	2	1	1	-	2	1

** De cijfers in de tabel verwijzen naar het aantal betrouwbare metingen van ongelijkheidsveranderingen. De metingen zijn gesommeerd over de drie prestatienivo's en over alle breekpunten.

a*: relatieve milieuverschillen:	C1'40 / C3'40 met C1'65 / C3'65; C1'40 / C2'40 met C1'65 / C2'65; C2'40 met C2'65 / C3'65.
b*: absolute milieuverschillen:	C1'40 - C3'40 met C1'65 - C3'65; C1'40 - C2'40 met C1'65 - C2'65; C2'40 - C3'40 met C2'65 - C3'65.
c*: relatieve generatieverschillen:	C1'65 / C1'40 met C3'65 / C3'40; C1'65 / C1'40 met C2'65 / C2'40; C2'65 / C2'40 met C3'65 / C3'40.
d*: absolute generatieverschillen:	C1'65 - C1'40 met C3'65 - C3'40; C1'65 - C1'40 met C2'65 - C2'40; C2'65 - C2'40 met C3'65 - C3'40.

prestatienivo aan het einde van de lagere school, verandert over twee generaties weinig. Beslissingen ten aanzien van het door blijven gaan in de bovenstroom blijven voor deze groepen in even sterke mate milieugebonden. Het feit dat de afstand tussen de midden- en laagste milieukategorie gelijk blijft kan samen gaan met een over-all afnemend effect van sociaal milieu omdat de twee hoogste sociale categorieën als het ware dichter bij elkaar komen te liggen.

Schematisch zou het er ongeveer zo uit kunnen zien:

generatie '40: hoog-----midden-----laag

generatie '65: hoog-----midden-----laag

De tendensen in de terugstroompercentages zijn niet eenduidig. De conclusies variëren afhankelijk van de toegepaste rekenprocedures. Met betrekking tot veranderingen in sekundaire milieu-effecten voor jongens en meisjes kunnen we het volgende concluderen: het sekundaire milieu-effect neemt voor jongens verhoudingsgewijs in veel meer gevallen af dan voor meisjes. Voor de meisjes is er weinig veranderd in de invloed van milieu onder konstanthouding van prestatie: evenveel metingen laten een verminderd effect zien als een groter wordend effect. Aan de hand van tabel 4 is het overigens ook mogelijk de ongelijkheidsverandering tussen jongens en meisjes uit de ene generatie te vergelijken met de andere generatie. Uit de analyse blijkt dat jongens van de '40-generatie gemiddeld bijna drie keer meer kans hadden te overleven dan meisjes. Het '65-cohort laat een geheel ander beeld zien: de helft van de metingen geeft een voorsprong van meisjes ten opzichte van jongens te zien.

Ook al kunnen we wat betreft de meisjes weinig mathematische vergelijkingen maken, duidelijk is dat de meisjes sterk vooruitgegaan zijn in vergelijking met de generatie '40. In de jongste generatie overleven beduidend meer meisjes dan in de generatie '40. In de vergelijking meisjes met jongens in beide generaties, is gebleken dat meisjes hun achterstand ten opzichte van jongens ingehaald hebben.

LOKALISATIE VAN VERANDERINGEN IN SEKUNDAIRE MILIEU-EFFEKTEN IN DE SCHOOLLOOPBAAN

Om inzicht te krijgen in welk schooljaar of op welk breekpunt de verminderingen in sekundaire milieu-effecten te lokaliseren zijn en of jongens en meisjes hierin hetzelfde beeld te zien geven, wordt in principe dezelfde procedure gevolgd als bij het eerste deel van de probleemstelling. Alleen wordt nu de afname, het stabiel blijven of de toename van sekundaire effecten per breekpunt bekeken.

Tabel 7: lokalisatie van veranderingen in sekundaire milieu-effecten, gemeten aan de hand van overlevingspercentages, voor de totale populatie en uitgesplitst naar jongens en meisjes.*

Totale populatie:	C1 - C3			C1 - C2			C2 - C3		
	kleiner	stabiel	groter	kleiner	stabiel	groter	kleiner	stabiel	groter
Breekpunt 1	8	—	—	8	—	—	5	7	—
Breekpunt 2	4	—	—	—	1	3	4	—	—
Breekpunt 3	—	—	4	—	4	—	—	4	4
Breekpunt 4	4	—	—	2	2	—	—	8	—
Breekpunt 5	—	4	—	—	—	4	2	2	4
Breekpunt 6	—	4	—	3	1	—	—	—	4

	Jongens:									Meisjes:								
	kl	st	gr	kl	st	gr	kl	st	gr	kl	st	gr	kl	st	gr	kl	st	gr
Breekpunt 1	8	—	—	8	—	—	10	2	—	2	2	—	4	2	2	—	4	4
Breekpunt 2	—	—	—	—	—	—	—	4	—	—	—	—	—	—	—	—	—	—
Breekpunt 3	—	—	4	—	1	3	—	1	7	—	—	—	—	—	—	—	—	—
Breekpunt 4	4	—	—	2	2	—	—	4	—	—	—	—	4	—	—	—	—	—
Breekpunt 5	4	—	—	—	4	—	4	—	—	—	—	—	4	—	—	—	—	—
Breekpunt 6	—	4	—	4	—	—	—	—	4	—	—	—	—	—	—	—	—	—

* De cijfers in de tabel verwijzen naar het aantal betrouwbare metingen gesommeerd over de vier meetprocedures, apart voor de verschillende milieukategorieën. Prestatie wordt konstant gehouden.

Uit de gegevens komt naar voren dat vermindering van sekundaire milieu-effecten voor de totale populatie gelokaliseerd is op het eerste breekpunt en in iets mindere mate op het tweede breekpunt. Bij vooral de breekpunten "advies" en "keuze begin eerste jaar vmo/vwo", speelt milieu, onder konstanthouding van prestatie, een minder zware rol voor de generatie 1965 dan voor de generatie 1940. Dit afnemend effect van milieu wordt echter voor een deel teniet gedaan op het moment dat leerlingen, veelal na het verlaten van de brugklas, moeten gaan kiezen voor het tweede jaar van het vwo. Op dat punt aangekomen is er duidelijk sprake van een toenemend sekundair milieu-effect bij bijna alle milieukategorieën. Na het advies en na het verlaten van de brugklas spelen in de generatie '65 milieugebonden beslissingen, met betrekking tot het verder blijven volgen van de stroom die toegang geeft tot het wetenschappelijk onderwijs, een sterkere rol dan voor de generatie '40. Dit onderstreept nogmaals het belang van het duidelijk onderscheiden van de brugklas in de vergelijking van twee cohorten. Bovenstaande uitspraken gaan op voor jongens, voor wat betreft meisjes beschikken we over weinig gegevens. Bekijken

we de tweede helft van de schoolloopbaan voor de totale populatie, dan kunnen we konstaten dat geen enkele meting op het vierde breekpunt een toename van het sekundaire milieu-effekt laat zien. Bij "keuze begin derde jaar vmo/vwo" wordt dit effect dus niet groter. Dit geldt voor alle leerlingen. Vanaf het vierde breekpunt is in de vergelijking van het hoge milieu met het laagste milieu, voor de totale populatie sprake van een stabiel blijvend effect. Deze uitspraak wordt veelzeggender als ook de vergelijking van de tussengroep (C2) met C1 en met C3 in de analyse betrokken wordt. Bij de C1 - C2-vergelijking worden de sekundaire effecten kleiner, bij de C2 - C3-vergelijking daarentegen vooral groter. Het lijkt zinvol onderscheid te maken tussen de verschillende milieukategorieën. Door alleen uiterste milieugroepen te vergelijken gaat informatie verloren over veranderingen in tussenliggende milieukategorieën. De resultaten van de metingen bij de jongens wijzen in dezelfde richting als die van de totale populatie. Om uitspraken te kunnen doen over de meisjes, zijn er te weinig betrouwbare gegevens.

Afname van sekundaire milieu-effecten berekend aan de hand van terugstroompercentages, is voor de totale populatie vooral op het tweede breekpunt "keuze begin eerste jaar vmo/vwo" geconcentreerd. De beslissing om ondanks het ontbreken van een vwo-advies toch naar het vwo te gaan of een brugklas die daar aansluiting op geeft, wordt minder milieu-gebonden. Dit geldt ook als de totale populatie uitgesplitst wordt naar sexe. De laatste drie breekpunten in de terugstroompercentages wijzen voor alle leerlingen bijna eenduidig op een stabiel blijvend sekundair milieu-effekt.

HET TOENEMEND BELANG VAN PRESTATIE

In het laatste deel van de probleemstelling staat de vraag naar het al dan niet toenemend belang van prestatie centraal. Nu worden niet de milieugroepen met elkaar vergeleken, maar de verschillende prestatienivo's.

Tabel 8: veranderingen van prestatie-effecten, gebaseerd op metingen van verhoudingen van en verschillen tussen alle betrouwbare overlevingspercentages van beide cohorten*

		Ongelijkheidsverandering tussen R1 en R3			Ongelijkheidsverandering tussen R1 en R3			Ongelijkheidsverandering tussen R1 en R3		
		kleiner	stabiel	groter	kleiner	stabiel	groter	kleiner	stabiel	groter
a.	totale popula- tie	—	—	2	—	2	8	—	—	2
b.	totale popula- tie	—	—	2	—	2	8	—	2	—
a.	jongens	—	—	2	—	—	6	—	—	2
b.	jongens	—	—	2	—	—	6	—	2	—
a.	meisjes	—	—	2	—	1	1	—	—	—
b.	meisjes	—	—	1	—	—	3	—	—	—
c.	totale popula- tie	—	—	2	—	2	8	—	—	2
d.	totale popula- tie	—	—	2	—	2	8	—	—	2
c.	jongens	—	—	2	—	—	6	—	—	2
d.	jongens	—	—	2	—	1	5	—	2	—
c.	meisjes	—	—	2	—	1	1	—	—	—
d.	meisjes	—	—	1	—	—	3	—	—	—

* De cijfers in de tabel verwijzen naar het aantal betrouwbare metingen van ongelijkheidsveranderingen. De metingen zijn gesommeerd over de drie milieu-kategorieën en over alle breekpunten.

Uit deze tabel komt een duidelijke trend naar voren die wijst op een groter wordend effect van prestatienivo eind lagere school op overlevingskansen in het vwo. Met andere woorden: onder konstanthouding van milieu is het belang van prestatienivo eind lager onderwijs voor de schoolloopbaan in het voortgezet onderwijs tussen beide generaties toegenomen. Bij de selectie van de stroom die toegang geeft tot het wetenschappelijk onderwijs, is prestatienivo voor de generatie '65 een meer doorslaggevend factor in vergelijking met de generatie '40. Dit gaat op voor alle milieukategorieën en zowel voor jongens als voor meisjes. Tabel 9 geeft deze meritokratiseringstendens nog eens duidelijk weer.

Tabel 9: Overlevingspercentages na vijf schooljaren voor de negen milieu- en prestatiegroepen, van beide generaties.

	Generatie '40			Generatie '65		
Totale populatie	R1	R2	R3	R1	R2	R3
C1	48%	22%	37%	53%	8%	8%
C2	22%	11%	3%	49%	14%	0%
C3	8%	3%	2%	37%	5%	0%
Jongens						
C1	66%	42%	50%	44%	17%	15%
C2	33%	17%	5%	49%	16%	0%
C3	14%	7%	4%	35%	6%	0%
Meisjes						
C1	26%	8%	0%	61%	0%	0%
C2	9%	4%	1%	50%	11%	3%
C3	3%	0%	0%	39%	5%	0%

Deze tabel geeft het over-all effect van prestatie te zien, de breekpunten zijn hier weggelaten. Uitgaande van de startaantallen is berekend hoeveel procent uit de negen onderscheiden milieu-prestatiegroepen het einde van het vijfde schooljaar haalt. Zichtbaar is dat weliswaar in beide cohorten de percentages onder R1 hoger zijn dan onder R2 of R3, echter het verschil in percentages tussen de prestatiegroepen is voor het jongste cohort veel groter. In deze generatie heeft een leerling, ongeacht de sociale achtergrond, met een laag prestatienivo nauwelijks kans van overleven. Dit geldt voor jongens en meisjes. In het '40-cohort ging, uitgezonderd voor meisjes, een hoog milieu vooral samen met een hoge overlevingskans. Vijfentwintig jaar later neemt een hoog prestatienivo (R1) deze selectiefunctie als het ware over. Eerst waren de leerlingen met een hoge sociale achtergrond, min of meer ongeacht prestatienivo, redelijk verzekerd van een positie in het vmo. Nu geldt dat vooral voor leerlingen uit de hoge prestatienivogroep, min of meer ongeacht milieu. Dit beeld gaat overigens wel op voor jongens maar niet voor meisjes. Meisjes uit de generatie '40 moesten en uit een hoog sociaal milieu komen en een hoog prestatienivo hebben wilden ze het einde van het vmo halen (26%). Van de generatie '65 overleven vooral die meisjes die een hoog prestatienivo hebben. Overlevingspercentages uit deze categorie zijn hoger dan die van de jongens uit de generatie '65. Daar staat tegenover dat jongens uit een hoge milieugroep meer kans tot overleven hebben dan de meisjes uit C1 van de jongste generatie.

Tot slot is het al of niet toenemend belang van prestatie geanalyseerd aan de hand van de terugstroompercentages. Uit deze gegevens valt op te maken dat in de kans om weer opgenomen te worden in de stroom die toegang geeft tot het wetenschappelijk onderwijs, de rol van prestatie voor de totale populatie overwegend stabiel is gebleven. Jongens en meisjes

vertonen hierin hetzelfde beeld als de totale populatie. Uit inspectie van de achterliggende berekeningen komt naar voren dat na het tweede breekpunt alle berekeningen wijzen op een stabiel blijvend of groter wordend effect van prestatie. Bij het breekpunt "keuze eerste jaar vmo/vwo" verschijnen echter tegenstrijdige conclusies, afhankelijk van de meetprocedure die toegepast wordt.

BESLUIT

Allereerst wordt kort ingegaan op de methodische aspecten van dit onderzoek. Vervolgens worden de belangrijkste bevindingen uit dit onderzoek samengevat, waarbij geprobeerd wordt de resultaten in een breder perspectief te plaatsen. In dit onderzoek zijn veranderingen in onderwijskansen op vier verschillende manieren gemeten. Vanuit vier verschillende invalshoeken is de probleemstelling benaderd: twee invalshoeken waarbij relatieve vergelijkingen voorop stonden, twee invalshoeken waarbij de overlevings- en terugstroompercentages absoluut werden vergeleken. Niet alle rekenprocedures leidden tot tegenstrijdige conclusies, echter verschillende keren is naar voren gekomen dat dit wel het geval was. Het is gebleken dat Combessie gelijk heeft: met verschillende meetinstrumenten kun je verschillende uitkomsten krijgen. We kunnen dit feit niet negeren. Bovendien zal de moeilijkheid nog in omvang toenemen als de berekeningswijzen complexer worden. We kunnen wellicht accepteren dat berekeningswijzen verschillen, het wordt moeilijker als de betekenis van de resultaten verschillend geïnterpreteerd kan worden met betrekking tot de ontwikkeling in ongelijke onderwijskansen. In dit onderzoek zijn vier uiteenlopende rekenprocedures gekozen, er zijn nog vele andere mogelijkheden. Wie of wat garandeert de juistheid van deze keuze? En welke criteria worden in het oordeel hierover gehanteerd? Er zal een discussie moeten komen die de keuze van het statistisch gereedschap fundeert of legitimeert. De concepten of vooronderstellingen die achter de berekeningswijzen schuilen zouden verhelderd moeten worden op grond van hetgeen waarnaar men op zoek is. De relatie tussen enerzijds concepten van ongelijkheid en anderzijds de statistische meetinstrumenten zouden verduidelijkt moeten worden.

De eerste vraag van de probleemstelling luidde: verminderen in een tijdsbestek van vijftientwintig jaar de sekundaire milieu-effecten en zijn deze veranderingen voor jongens en meisjes gelijk? Uit het onderzoek is een afnemend sekundair milieu-effect naar voren gekomen. De afstand tussen de hoogste en laagste milieugroep wordt kleiner, dit gaat gepaard met een stabiel blijvend milieu-effect voor de laagste milieukategorieën. Het is niet onwaarschijnlijk dat de komst van de havo, een typisch uitvloeisel van de Mammoetwet, hier mede debet aan is. In de hoogste milieugroep (met name bij de jongens) is er sprake van een afname in overlevingspercentages. Doordat deze groep relatief minder deelneemt aan het vwo, wordt de afstand tussen C1 en C2 verkleind. Deze afname in overleving leidt vervolgens tot een afname in sekundaire milieu-effecten tussen deze twee groepen. Ongelijkheid in overlevingspercentages, onafhankelijk van prestatienivo eind lagere school, vermindert dus wel voor de hoogste milieugroepen. In de ongelijkheid in overlevingskansen van de laagste sociale milieus verandert er weinig. Deze resultaten vormen een belangrijke aanvulling op de bevindingen van Faasse e.a. We hebben nu immers meer inzicht gekregen waar in de milieuhierarchie sekundaire milieu-effecten optreden. Bovenstaande uitspraken zijn van toepassing op de totale populatie van beide generaties en op jongens uit deze generaties. De veranderingen in sekundaire milieu-effecten zijn niet voor jongens en meisjes gelijk. Voor zover we gegevens hebben kunnen we concluderen dat meisjes hun achterstand ten opzichte van jongens ingehaald hebben. Waarschijnlijk heeft o.a. het feminisme een klimaat geschapen waarin positieve veranderingen in de richting van meer gelijke onderwijskansen versneld werden. Echter de resultaten in dit onderzoek met betrekking tot gelijke onderwijskansen tussen jongens en meisjes kunnen bedrieglijk zijn, in die zin dat jongens bijvoorbeeld wellicht 'zwaardere' vakkenpakketten kiezen dan meisjes en om die reden weer een voorsprong opbouwen. De traditioneel-vrouwelijke vakkenpakketten kunnen juist

weer leiden tot een onrechtvaardige verdeling van onderwijskansen, die zich kan uiten op de arbeidsmarkt. Handl (1986), bijvoorbeeld wijst er in zijn onderzoek op dat meer gelijke onderwijskansen tussen jongens en meisjes niet noodzakelijk gepaard hoeft te gaan met een reducering van geslachtsspecifieke segregatie op de arbeidsmarkt.

Met betrekking tot de lokalisatie van veranderingen in milieu-effecten, kunnen we concluderen dat verminderingen in sekundaire milieu-effecten vooral geconcentreerd zijn op het eerste breekpunt "advies" en in iets mindere mate op het tweede breekpunt "keuze eerste jaar vmo/vwo". Je zou kunnen stellen dat aan het begin van de schoolloopbaan sprake is van meer gelijke kansen voor gelijkbegaafden, één van de doelstellingen van de Mammoetwet. De doorstroom naar het vwo is in vergelijking met het '40-cohort beter geworden, voor alle milieugroepen. Dit sekundair milieu-effect neemt echter niet gedurende de gehele schoolloopbaan in het voortgezet onderwijs in gelijke mate af. Bij het derde breekpunt, het punt waarop de meeste leerlingen een meer definitieve schoolkeuze moeten gaan maken, wordt het sekundair milieu-effect juist groter. Het bovenstaande gaat ook op voor jongens, voor wat betreft meisjes beschikken we over weinig gegevens. Dit wordt wel enigszins gecompenseerd door een kleiner wordend sekundair milieu-effect in de terugstroompercentages.

De precieze invloed van brugjaren is moeilijk vast te stellen, wel komt uit de resultaten naar voren dat ná de brugklas, het moment waarop structurele differentiatie in het onderwijs dwingen tot het maken van keuzes, deze keuzes meer milieugebonden zijn. In het begin van de schoolloopbaan lijkt het brugjaar wel effect te hebben, bij de keuze van het tweede jaar en daarna, wordt dit effect weer afgezwakt. Het milieu-effect neemt dan weer toe of blijft overwegend stabiel. Relatief hoge overlevingspercentages op met name het tweede breekpunt, wijzen er op dat leerlingen in eerste instantie langer bij elkaar gehouden worden. Bij het derde breekpunt is de uitval daarentegen relatief groot. Deze terugval blijft bovendien nog een paar schooljaren als het ware nadreunen. Het lijkt alsof in de generatie '65 meer hindernissen genomen moeten worden. Dit zou een aanduiding kunnen zijn voor interne selectiemechanismen die werkzaam zijn in een meer geïntegreerd onderwijsstelsel. Dit in tegenstelling tot de generatie '40, waarin naar voren komt dat als leerlingen eenmaal het eerste jaar overleefd hebben, zij een grote kans hebben de eindstreep te halen. In deze generatie, zo zou je kunnen stellen, is meer sprake van een expeditiemodel of sponsored model. Leerlingen uit de generatie '65 komen gedurende de schoolloopbaan meer horden tegen, met name op het derde breekpunt. Deze tendens wijst meer in de richting van een hordeloop- of contest model (7).

Tot slot kunnen we concluderen dat het belang van prestatiescores aan het einde van de lagere school in gewicht toeneemt. In vergelijking van twee cohorten worden, zowel voor jongens als voor meisjes, schoolprestaties belangrijker voor het al dan niet succesvol doorlopen van het vwo. Deze resultaten komen overeen met die van Faasse e.a., maar ook andere onderzoeken komen tot deze conclusie. In het verwerven van onderwijsposities nemen kenmerken die meer te maken hebben met eigen inspanningen in belang toe. Ongelijkheid in de generatie '65 is meer dan in de generatie '40 gebaseerd op het prestatienivo dat een leerling heeft op het einde van het lager onderwijs. De tegenstellingen tussen een hoog en een laag prestatienivo worden groter in verband met de kans tot overleven of terugstromen. Prestatie neemt als onderscheidingscriterium in gewicht toe.

NOTEN

1. Gedwongen werken met de meest presterende helft kan konsekventies hebben voor de analyse, met name indien de relatie tussen milieu en prestatie tussen 1940 en 1965 drastisch gewijzigd is. In dat geval weerspiegelen de resultaten niet alleen veranderde overlevingspercentages maar tegelijkertijd de veranderde relaties tussen milieu en prestatie. Hoewel er sprake is van verandering in deze relaties is die niet van zodanige aard dat we geen uitspraken zouden kunnen doen over verschuivingen in overlevings- en terugstroompercentages.
2. De data zijn via het Steinmetz-archief (Herengracht 410-412, Amsterdam) onder bepaalde voorwaarden beschikbaar.

3. Als indicator voor sociale achtergrond is, in navolging van Boudon, voor beide generaties het beroep van de vader gebruikt. Het was ook mogelijk geweest om afwijkend van Boudon naast of in plaats van het beroep van de vader, de opleiding van de ouders als indicator voor milieu te kiezen. Men krijgt dan wellicht een betere invulling van 'milieu', een bezwaar is echter dat de vergelijkbaarheid tussen de cohorten, door de snel veranderde opleidingsnivo's, afneemt.
4. prestatienovo eind lagere school
Leerlingen van beide generaties hebben een toets gemaakt met zowel een rekenkundige als een taalkundige component. Deze toetsen zijn tot op zekere hoogte met elkaar vergelijkbaar (zie Faasse e.a. 1986).
5. Overal waar in de tekst hij, hem, zijn etc. staat, bedoel ik uiteraard ook de vrouwelijke vorm zij en haar.
6. Van de schooltypen: alleen vglo, lbo, ulo, mms, hbs, gymnasium of seminarie, worden alleen hbs, gymnasium of seminarie tot het vmo gerekend. In navolging van Meesters e.a. (1983) en Faasse e.a. (1986) wordt de mms niet tot het vmo gerekend.
Zittenblijvers worden niet in de analyse betrokken.
7. Zie: Turner, R., 'Modes of social ascent through education: sponsored and contest mobility'. In: A.H. Halsey e.a. (red), *Education, Economy, and Society*. The Free Press, New York. 1961. 121-139.

LITERATUUR

- Boudon, R. (1974). *Education, Inequality and Social Opportunity*, New York, Wiley.
- Combessie, J. (1984). L'évolution comparée des inégalités: problèmes statistiques. *Revue française de sociologie*, 20, 233-254.
- Dronkers, J. (1986). Onderwijsongelijkheid in het verleden, in: *Onderwijs: Bestel en beleid 2*, Onderwijs en samenleving a, red. van Kemenade J.A. e.a.
- Dronkers, J., de Graaf, P., de Haan, D., Peek, J. & Swanborn P. (1982). Veranderingen in ongelijke onderwijskansen. *Comenius*, jrg. 2.
- Faasse, J.H., Bakker, B., Dronkers, J. & Schijf, H. (1986). Voor en na de Mammoet. *Mens en Maatschappij*, 61, nr. 1.
- Handl, J. (1986). Führt die Angleichung der Bildungschancen zum Abbau geschlechtsspezifischer beruflicher Segregation? Eine einfache Frage, aber kontroverse Antworten. *Zeitschrift für Soziologie*, 15, 125-132.
- Meester, A.C., & de Leeuw, J. (1985). De regressie van intelligentiescores op schoolloopbaanvariabelen onder constanthouding van sociale achtergrond en geslacht. *Tijdschrift voor Onderwijsresearch*, 10, nr. 1.
- Meesters, M., Dronkers, J., & Schijf, H. (1983). Veranderde onderwijskansen? Een derde voorbeeld en afrondende conclusies. *Mens en Maatschappij*, 58, 5-27.

Manuscript ontvangen 22-7-1987.

Definitieve versie ontvangen 5-10-1987.

Effectief Gebruik van Rekenmethoden: Evaluatieonderzoek in het laatste jaar van de basisschool

C. Suhre
Rion, Groningen

ABSTRACT

In Dutch primary education teachers use either traditional or modern arithmetic curricula. Up till now there has not been any serious research about the effective use of these two types of curricula. This article reports on an evaluation study of the educational factors that, together with the type of curriculum, could be related to arithmetic skills acquired by pupils when they leave primary school. Arithmetic skills were measured with a test for basic arithmetic skills and with a test in problem solving. Hypotheses about the relations between educational factors and learning outcomes were put to a test by means of LISREL VI, a procedure for analyzing structural relationships in correlational data. The results of the analysis confirm our expectations that teachers of the two types of curricula differ in their arithmetic instruction. Yet, independent of differences between the type of curriculum teachers use, the number of lessons given and the content covered in an arithmetic curriculum are positively related to the basic skills and problem-solving abilities of pupils.

INLEIDING

De afgelopen vijftien jaar hebben zich grote veranderingen voorgedaan in de visie op het rekenonderwijs. Als reactie hierop zijn er een aantal moderne reken-wiskundemethoden op de markt verschenen. Begin 1985 was de situatie zo, dat iets meer dan de helft van de basisscholen een traditionele methode had en iets minder dan de helft een moderne rekenmethode. De traditionele rekenmethoden propageren directe instructie van het eindalgoritme van rekenbewerkingen en oefenen leerlingen in de uitvoering ervan door middel van rijen zgn. "kale" oefensommen. Daarnaast zijn er opgaven (vraagstukjes) voor het toepassen van rekenbewerkingen. Moderne methoden zijn op een andere leest geschoeid. Er wordt uitgegaan van anschouwelijke modellen waarmee leerlingen vraagstukjes rond een thema leren oplossen. De modellen zijn hulpmiddelen bij het leren oplossen van rekenbewerkingen. Moderne methoden bevatten daarnaast ook veel gevarieerde oefeningen in rekenvaardigheid. Hoewel traditionele en moderne rekenmethoden verschillen in hun visie op onderwijzen en het leren rekenen zijn de verschillen in leerstofinhoud niet erg groot (zie Harskamp en Suhre, 1986). De Jong (1986) heeft alle op de markt verkrijgbare rekenmethoden geanalyseerd op kenmerken die typerend zijn voor de moderne rekenwiskunde didactiek. Op grond van deze analyse worden moderne methoden als "realistische methoden" bestempeld en traditionele methoden als "mechanistisch".

Richtte de aandacht met betrekking tot het reken-wiskundeonderwijs zich het afgelopen decennium vooral op vernieuwing in de rekendidactiek, de laatste jaren is er een toenemende aandacht te bespeuren voor de bereikte rekenvaardigheid van leerlingen. Een voorbeeld van deze toegenomen aandacht vormt de Periodieke Peiling van het Onderwijsniveau in Nederland (PPON) die door het Cito uitgevoerd wordt. Naar de doelmatigheid van rekenmethoden met betrekking tot de rekenvaardigheid van leerlingen is tot op heden weinig onderzoek verricht. Slechts incidenteel zijn pogingen ondernomen om onderdelen van nieuwe rekenmethoden te vergelijken met onderdelen van traditionele methoden (zie Vink, 1978, Vrolijk, 1981, Rengerink 1983, Bruins en Bloemkolk, 1987). Een probleem van deze onderzoeken is dat ze methodisch technische onvolkomenheden bevatten, zoals zeer kleine steekproeven, het ontbreken van

betrouwbare, valide toetsinstrumenten of het ontbreken van relevante controlevariabelen bij de evaluatie van methode-effecten. In geen van deze onderzoeken wordt bovendien een summatieve evaluatie van effecten van rekenmethoden op de rekenvaardigheid van leerlingen aan het einde van de basisschool nagestreefd.

Het doel van het onderzoek waarover hier wordt gerapporteerd was tweeledig. De belangrijkste doelstelling was objectieve empirische informatie aan te dragen over de relatieve doelmatigheid van rekenmethoden met betrekking tot de na te streven leerdoelen bij leerlingen. Dit is informatie waarover scholen die een nieuwe rekenmethode willen aanschaffen, niet kunnen beschikken. Tot dusver was het zo dat scholen die een nieuwe rekenmethode aan willen schaffen, voor informatie aangewezen zijn op meningen van vakdeskundigen en ervaringen van collega's op andere scholen.

Een tweede doelstelling van het onderzoek was na te gaan welke onderwijsfactoren in samenhang met de gebruikte rekenmethode van invloed zijn op de rekenvaardigheid van leerlingen. Deze onderzoeksvraag is van belang voor methodemakers en voor leerkrachten, omdat hierdoor licht wordt geworpen op omstandigheden waaronder rekenmethoden meer of minder effectief worden toegepast. Het is met name deze tweede onderzoeksvraag die in dit artikel wordt aangesneden.

BEGRIPPENKADER

In de dagelijkse praktijk van het rekenonderwijs neemt de *rekenmethode* een dominante plaats in. De rekenmethode bevat de bouwstenen voor de rekenlessen die leerkrachten geven. Rekenmethoden zijn via de leerkracht van invloed op de rekenvaardigheid die leerlingen bereiken. Onder *rekenvaardigheid* wordt verstaan het complex aan cognitieve (deel-) vaardigheden waarover leerlingen moeten beschikken om rekenproblemen van uiteenlopende aard op te kunnen lossen. Met betrekking tot rekenvaardigheid kan een onderscheid gemaakt worden in basisrekenvaardigheid en toegepaste rekenvaardigheid. Basisrekenvaardigheid omvat die cognitieve vaardigheden, die nodig zijn om cijfersommen en eenvoudige redactieopgaven op te lossen (bijvoorbeeld inzicht hebben in de getallenstructuur, kennis van tafels en het correct uit kunnen voeren van rekenoperaties). Toegepaste rekenvaardigheid moet blijken in situaties waar leerlingen gedwongen zijn de basisrekenvaardigheid toe te passen. De aanwezigheid van basisrekenvaardigheid bij leerlingen biedt op zich nog geen garantie dat ze concrete rekenproblemen zoals die zich ook in het dagelijks leven voordoen (bijvoorbeeld in de supermarkt) op kunnen lossen.

Naast het verschil in traditionele of moderne rekenmethode spelen in de onderwijspraktijk ook andere factoren een rol bij het leren rekenen. Brophy (1986) wijst in een overzichtsartikel over de relatie tussen onderwijzen en het leren van rekenen en wiskunde op nog twee categorieën van factoren, te weten leerlingfactoren en onderwijsfactoren. Bekend is dat met name *leerlingenkenmerken* als intelligentie en sociaal milieu in hoge mate bepalend zijn voor de rekenprestaties van leerlingen. Over onderwijsfactoren die het leren rekenen bevorderen is weinig informatie voorhanden. Ten aanzien van onderwijsfactoren zijn leerkrachtkenmerken en implementatiekenmerken te onderscheiden. *Implementatiekenmerken* hebben betrekking op het gebruik van de methode door de leerkracht. Appelhof (1979) en Harskamp en Hofman (1982) onderscheiden hierin twee soorten: algemene en vakdidactische kenmerken. Algemene kenmerken zijn van belang voor goed gebruik van elke methode. Het gaat hierbij om de mate waarin de leerstof uit een methode wordt behandeld en om de hoeveelheid lestijd die daaraan wordt besteed. Veel onderzoekers wijzen op de positieve invloed die adequaat gebruik van leerstof en lestijd kan hebben op leerprestaties (zie Veenman, 1981, Slavenburg, 1978, Rosenshine, 1979 of Walberg, 1985). Vakdidactische kenmerken betreffen specifieke didactische suggesties voor vormgeving van de lessen, die rekenmethoden leerkrachten aanbieden. Voorbeelden zijn: het toepassen van bepaalde vormen van differentiatie, het samenwerken in groepen en het variëren van soorten leerstof tijdens lessen. Dit laatste implementatiekenmerk

verwijst naar het vakdidactische principe van moderne methoden dat er tijdens de lessen afwisseling plaats dient te hebben van algemene rekenvaardigheden, hoofdrekenen en toepassingsopgaven. Deze afwisseling dient kennis en vaardigheden van leerlingen flexibel te houden. Wanneer leerkrachten bepaalde vakdidactische kenmerken toepassen, kan worden vastgesteld in welke mate zij volgens de bedoelingen van de methode lesgeven. Vakdidactische kenmerken zijn vaak niet expliciet uitgewerkt in de methoden maar moeten worden afgeleid uit de handleidingen bij de methoden.

Behalve methode-, leerling- en implementatiekenmerken kunnen ook leerkrachtkenmerken meespelen bij de wijze waarop leerlingen leren rekenen. Een *leerkrachtkenmerk* dat van belang lijkt is de lesattitude van de leerkracht. Om volgens de bedoelingen met een moderne methode te werken zullen leerkrachten een leerlinggerichte attitude moeten hebben (leerkrachten moeten leerlingen helpen om zelf problemen op te lossen en niet alleen "kant en klare" kennis overdragen). Andere leerkrachtkenmerken zoals jaren ervaring met de rekenmethode, wel of geen wiskunde in de vooropleiding etc. zijn niet in deze beschouwing opgenomen. Over het belang van deze leerkrachtkenmerken voor doelmatig gebruik van een methode is de laatste tijd wel enige discussie maar empirische evidentie ontbreekt (zie Harskamp en Suhre, 1986).

ONTWERP VAN EEN ONDERWIJSMODEL

Voor beschrijving van verbanden tussen aan de ene kant leerkracht- en implementatiekenmerken en aan de andere kant rekenprestaties van leerlingen is een causaal model opgesteld. In dit model zijn verschillen tussen rekenmethoden geoperationaliseerd in het onderscheid in moderne en traditionele rekenmethoden. Dit onderscheid berust op inhoudelijke gronden (zie Inleiding). Voor de vaststelling van de rekenvaardigheid van leerlingen is gebruik gemaakt van twee toetsinstrumenten. Een toets voor basisrekenvaardigheid (het onderdeel Rekenen uit de Cito-eindtoets voor het Basisonderwijs) en een toets voor toegepaste rekenvaardigheid (een speciaal door het Rion ontworpen rekentoets).

Als algemene implementatiekenmerken is in dit onderzoek gekozen voor:

- de behandeling van de leerstof uit de methoden, geconcretiseerd in het percentage leerlingen in de klas waarmee alle basisstof uit de rekenmethode is behandeld en
- de leertijd, afgemeten aan het gemiddelde aantal rekenlessen per week dat de leerlingen in de laatste twee leerjaren (groep 7 en 8) van de basisschool hebben ontvangen. Deze twee implementatiekenmerken verwijzen naar algemene kenmerken voor effectief gebruik van rekenmethoden.

In het model zijn twee vakdidactische implementatiekenmerken opgenomen, namelijk:

- de mate waarin leerstofdifferentiatie wordt toegepast, vastgesteld door middel van een summatieve schaal (ontleend aan Harskamp, Pijl en Prins, 1984) en
- de variatie van leerstof. Dit kenmerk is geoperationaliseerd als het relatieve percentage van de lestijd dat leerkrachten aan toegepaste activiteiten als hoofdrekenen en redactierekenen besteden.

De lesattitude van leerkrachten is in het model vertegenwoordigd door een summatieve schaal waarmee de mate van leerling versus leerstofgerichtheid wordt gemeten.

Het hierna te beschrijven model is te beschouwen als een theoretisch onderbouwde verzameling van hypothesen over de relatie tussen onderwijsfactoren en leereffecten in het rekenwiskundeonderwijs (zie Begrippenkader). Het model beschrijft de relaties op klasniveau. Het lijkt misschien vreemd om de relaties van leerkracht- en implementatiekenmerken met reken- toetsprestaties te beschrijven op klasniveau. Leerkracht- en implementatiekenmerken kunnen immers binnen klassen verschillend op leerlingen uitwerken. Aangezien rekenlessen echter in klasverband gegeven worden, ligt het voor de hand verschillen in relaties tussen onderwijsfactoren enerzijds en rekenprestaties van leerlingen uit de verschillende klassen anderzijds op klasniveau te evalueren (zie bijv. Poyner, 1975). Om op klasniveau de relaties te kunnen beschrijven is gebruik gemaakt van gemiddelde toetsprestaties voor elke klas. Deze ongewogen

klasgemiddelden vormen een indicatie voor het niveau van rekenvaardigheid dat de leerkrachten bij hun leerlingen hebben bereikt. In de relatie tussen leerkracht- en implementatiekenmerken enerzijds en gemiddelde klasprestaties anderzijds ligt, met inachtneming van verschillen tussen klassen in achtergrondkenmerken zoals het intelligentieniveau van leerlingen in de klas, het overall-effect van leerkracht- en implementatiekenmerken op rekentoetsprestaties van leerlingen uit de verschillende klassen vervat (zie bijv. Cooley & Lohnes, 1976, p. 234).

De centrale rol in het onderwijsmodel is weggelegd voor het type rekenmethode waar leerkrachten mee werken, traditioneel of modern. Als klaskenmerk is de gemiddelde score op de ISI-intelligentietoets gebruikt. Overwogen is om naast de gemiddelde ISI-score ook nog een sociaal-milieuscore (gebaseerd op het beroep van de vader) als klaskenmerk in het model op te nemen. Gezien de hoge samenhang op klasniveau tussen de intelligentiescore en de sociaal-milieuscore voor klassen is met de ISI-score als klaskenmerk volstaan. De in het onderwijsmodel gebruikte variabelen zijn hieronder op een rijtje gezet. Predictoren zijn de variabelen 1 tot en met 7 terwijl 8 en 9 de criteriumvariabelen zijn. Bij de opsomming van de variabelen die voor de validering van het model gebruikt zijn, wordt eerst de afkorting waarmee ze in het geoperationaliseerde model (figuur 1) worden aangeduid, gegeven.

(1)	METHODE	Het type rekenmethode traditioneel (code 1) versus modern (code 2);
(2)	IQ	De gemiddelde ISI-intelligentiescore per klas;
(3)	ATT	De lesattitude van een leerkracht (mate van leerling- versus leerstofgerichtheid);
(4)	LESSEN	Het gemiddeld aantal rekenlessen per schoolweek dat de leerlingen de laatste twee jaar gehad hebben;
(5)	PERCBA	Het percentage leerlingen uit de klas dat alle basisstof uit de gebruikte methode heeft gehad;
(6)	VARIATIE	Gemiddelde percentage van de lestijd besteed aan hoofdrekenen en toepassingsopgaven versus de lestijd besteed aan algemene rekenvaardigheden;
(7)	DIFF	De mate waarin de leerkracht leerstofdifferentiatie toepast.
(8)	CITOR	Gemiddelde toetsscore van een klas op de Cito-toets Rekenen;
(9)	RIONT	Gemiddelde toetsscore van een klas op de Rion-toets.

Verwachtingen met betrekking tot de relatie tussen onderwijsfactoren en de rekenprestaties van leerlingen in klassen die op voorhand geformuleerd konden worden zijn de volgende:

a) hypothesen over de rekenmethode:

(1) Ten aanzien van het onderscheid tussen traditionele en moderne methoden kan verwacht worden dat moderne methoden vooral op scholen gebruikt worden waar de leerlingen in doorsnee over een meer dan gemiddeld intelligentieniveau beschikken. Wie de leerboekjes van moderne rekenmethoden openslaat, ziet dat ook in de basisleerstof complexe vraagstukjes zijn opgenomen die een sterk beroep doen op het probleemoplossend vermogen van leerlingen. In traditionele methoden komen moeilijke opgaven bijna uitsluitend voor in de extra stof voor snelle en goede leerlingen.

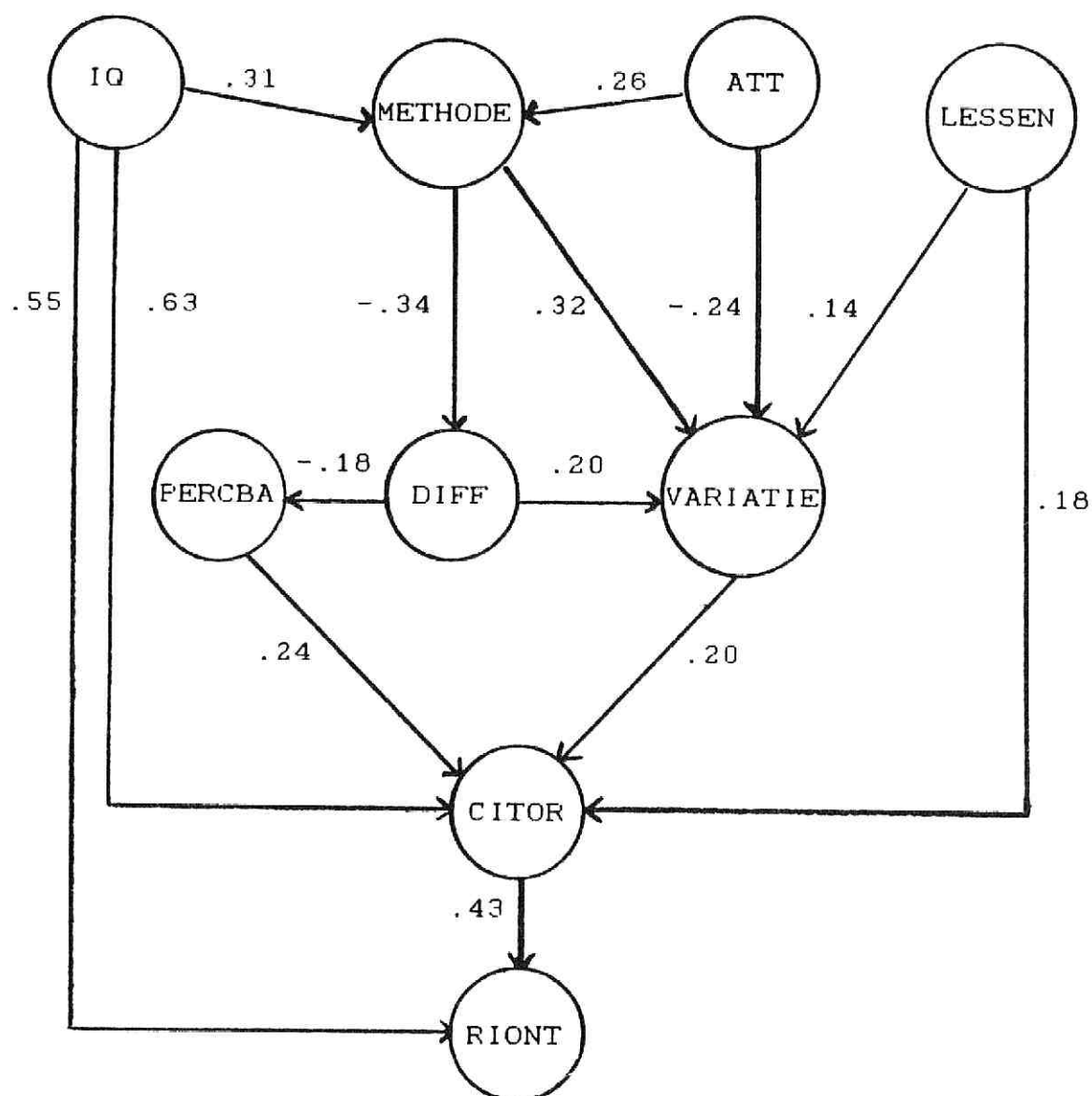
(2) Van de leerkrachten die met een moderne methode lesgeven wordt mede hierom verwacht dat ze meer lestijd besteden aan toegepaste rekenactiviteiten in de lessen.

(3) Leerkrachten met een traditionele methode zullen volgens de bedoeling van de methode meer leerstofdifferentiatie toepassen dan leerkrachten met een moderne methode.

b) hypothesen over de lesattitude van de leerkrachten:

De lesattitude van leerkrachten zal een rol spelen bij de voorkeur voor het type rekenmethode en bij de behandeling van leerstof in de lessen.

(4) Leerkrachten met een meer leerlinggerichte lesattitude zullen vaker met een moderne rekenmethode lesgeven.



Figuur 1: Model van onderwijsfactoren en rekenprestaties.

Tabel 1: Gemiddelde scores op implementatie variabelen voor traditionele en moderne rekenmethoden

	implementatie factoren			leerkrachtfactor	
	perc. leerl. dat de basisstof deed (PERCBA)	variatie in leer stof tijdens lessen (VARIATIE)	gemidd. aantal rekenlessen per week (LESSEN)	leerstof diffe- rentiatie (DIFF)	leerling gerichte onderwijsopvatt. (ATT)
Alle methoden (n = 118)	72 (29)	44 (15)	5.1 (.7)	23.0 (4.5)	18.4 (2.6)
Alle traditionele methoden (n = 76)	69 (29)	41 (15)	5.0 (.7)	24.2 (4.3)	17.9 (2.7)
Alle moderne methoden (n = 42)	78 (27)	48 (14)	5.2 (.5)	21.0 (4.2)	19.4 (2.3)

* Tussen haakjes zijn de standaarddeviaties gegeven

(5) Leerkrachten met een meer leerstofgerichte lesattitude zullen, los van de methode die ze gebruiken, meer moeite doen om variatie in leerstof te waarborgen.

c) Hypothesen over het aantal lessen:

(6) Het aantal rekenlessen dat de leerlingen de laatste twee jaar gehad hebben, zal een positieve relatie hebben met de rekenprestaties van leerlingen.

(7) Tevens mag verwacht worden, dat leerkrachten die meer rekenlessen geven meer tijd vinden om variatie in leerstof in de rekenlessen te brengen.

(d) hypothesen over de leerstof:

(8) Er mag verwacht worden, dat naarmate leerkrachten er in slagen met relatief meer leerlingen uit hun klas alle basistof uit de methode te behandelen, de rekenprestaties beter zijn.

(9) Hetzelfde geldt ten aanzien van de variatie in leerstof. In klassen waar de leerkracht relatief meer variatie in de leerstof aanbrengt zullen de rekenprestaties beter zijn dan in klassen waar de leerkracht minder variatie in de leerstof aanbrengt.

e) hypothesen over leerstofdifferentiatie:

De algemene verwachting ten aanzien van leerstofdifferentiatie is (zie Harskamp en Prins, 1984 of Brophy, 1986) dat naarmate leerkrachten meer leerstofdifferentiatie toepassen, zij minder efficiënt met leerstof en lestijd zullen omgaan.

(10) Meer leerstofdifferentiatie zal leiden tot meer variatie in leerstof, doordat verschillende leerlingen met verschillende taken bezig zijn (hetgeen tot positieve resultaten kan leiden bij goede en snelle leerlingen).

(11) Ver doorgevoerde leerstofdifferentiatie zal (door motivatieverlies bij zwakke leerlingen) ertoe leiden dat met relatief minder leerlingen alle basistof wordt behandeld.

f) hypothese over de rekenvaardigheid:

(12) Tussen prestaties op de rekentoetsen voor basis- en toegepaste rekenvaardigheid zal een positief verband bestaan.

g) hypothesen over verschillen in intelligentieniveau tussen klassen:

Bij de hiervoor genoemde hypothesen moet rekening gehouden worden met het intelligentieniveau in klassen dat in grote mate bepalend zal zijn voor

(13) de prestaties op de Cito-toets Rekenen (CITOR) en voor

(14) de prestaties op de Rion-toets (RIONT).

ONDERZOEKSOPZET

Steekproef.

De volgende rekenmethoden die in 1985 ingevoerd waren tot groep 8 van de basisschool zijn in het onderzoek opgenomen:

Traditionele methoden:

- Op Veilig Spoor (OVS)
- Naar Aanleg en Tempo (NAT)
- Naar Zelfstandig Rekenen (NZR)
- Nieuw Rekenen (NR)
- Niveau Cursus Rekenen (NCR)

Moderne methoden:

- Getal In Beeld (GIB)
- Operator Rekenen (OR)
- Taltaal (TT)

Deze methoden bestrijken 85% van het totale methodenbestand op de basisscholen. Uit het bestand van scholen waar in 1985 de Cito-eindtoets voor het Basisonderwijs is afgenomen, is een gestratificeerde a-selecte steekproef getrokken, met als stratificatiecriteria de rekenmethoden en een globale regio-indeling (Noord-, Oost-, Zuid-, West- en Midden-Nederland). Welke

rekenmethode op de Cito-scholen gebruikt werd is in een vooronderzoek vastgesteld. De effectieve steekproef van klassen voor dit onderzoek bestond uit 133 klassen. Er zijn van in totaal 118 groep 8 klassen datasets van zowel leerkrachten als leerlingen verkregen. In 76 van deze klassen werd een traditionele methode gebruikt en in 42 klassen een moderne methode.

Onderzoeksinstrumenten.

a. leerlingtoetsen/vragenlijst.

Op grond van een inhoudsanalyse van de acht methoden zijn rekentoetsen geselecteerd en ontworpen. Over de selectie en de ontwikkeling van deze toetsen is intensief overleg gevoerd met de uitgevers van de acht rekenmethoden. Een aantal uitgevers, met name de uitgevers van de moderne rekenmethoden, kwamen met het verzoek om extra toetsen te ontwerpen voor opgaven die in de Cito-toets onvoldoende aan bod kwamen. Voor verslag van het overleg met de uitgevers van de acht rekenmethoden wordt verwezen naar Harskamp, Suhre en Wiersma, 1985. De gebruikte leerlingtoetsen zijn:

1. uit de Cito-eindtoets voor het Basisonderwijs het onderdeel Rekenen (voor verantwoording van deze toets wordt verwezen naar Wijnstra, 1984).
2. de speciaal ontwikkelde Rion-toets (Suhre en Harskamp, 1986). Deze toets bestaat uit vier subtoetsen met toepassingsopgaven op het gebied van: Schatten, Verhoudingen, Meten en Combinatoriek. Beide toetsen dekken over het geheel genomen de inhoud van de rekenmethoden. De Rion-toets legt echter op de vier genoemde rekendomeinen, op verzoek van de uitgevers, zwaardere accenten dan het relatieve leerstofaanbod in de methoden aangeeft (zie Wiersma en Willemsen, 1986). De betrouwbaarheid (uitgedrukt in coëfficiënt alpha) van de Cito-toets is .94 en van de Rion-toets .89.
3. De ISI-intelligentietoets. De betrouwbaarheid van de ISI-toets bedraagt .96 (zie voor meer gegevens Boxtel, Snijders en Welten, 1982).
4. Een vragenlijst met achtergrondvragen over beroep en opleiding van de vader en moeder.

b. Docentenvragenlijst.

Bij de leerkrachten is een vragenlijst afgenomen. De vragenlijst bevatte vragen over achtergrondkenmerken van de school, over de leerkracht, en over implementatiekenmerken. Om de lesattitude van leerkrachten vast te kunnen stellen was een summatieve Likertschaal, ontleend aan Meijnen (1982), opgenomen. Een hoge score op deze schaal geeft aan dat een leerkracht leerlinggericht is. Een lage score duidt op meer leerstofgerichtheid. De schaal bestaat uit zeven items en heeft een betrouwbaarheid van .70. De vragen met betrekking tot de implementatie van de rekenmethode waren deels objectief te beantwoorden vragen (vragen over het aantal lessen per week), deels goed in te schatten vragen (bijvoorbeeld vragen over het percentage leerlingen dat alle stof uit bepaalde leerboekjes heeft gedaan). De vragenlijst bevatte verder een summatieve Likert-schaal met 8 vragen over diverse aspecten van leerstofdifferentiatie. Deze differentieschaal heeft een betrouwbaarheid van .70. Indicatie voor validiteit van de schaal is te vinden bij Reezigt e.a (1986).

Statistische analyse.

a) de methodenvergelijking

De acht rekenmethoden zijn vergeleken op de implementatie en op de bereikte rekenprestaties bij leerlingen. De vergelijking van de rekenmethoden ten aanzien van de implementatiekenmerken is gebeurd door middel van univariate variantie-analyses. De vergelijking ten aanzien van de rekenprestaties is uitgevoerd door middel van multivariate covariantie-analyse met als covariaten de ISI-intelligentiescore en een sociaal-milieuscore (beroep van de vader) en als afhankelijke variabelen de gemiddelde toetsscores van de klassen op de Cito- en de Rion-toets.

b) toetsing van het onderwijsmodel

Het onderwijsmodel is op grond van a-priori verwachtingen (zie Ontwerp van een onderwijsmodel) gespecificeerd in een recursief padmodel (figuur 1). Het gespecificeerde model is met

behulp van het programma LISREL VI getoetst op de correlatiematrix van de 9 hiervoor genoemde variabelen. Deze correlatiematrix is weergegeven in tabel 2.

De modelparameters zijn geschat met de maximum likelihood methode. Wat het programma LISREL VI in essentie doet is, rekening houdend met de modelspecificaties, zoeken naar een zo goed mogelijke reproductie van deze correlatiematrix. In hoeverre de uiteindelijke reproductie afwijkt van de oorspronkelijke correlatiematrix kan vastgesteld worden op een aantal manieren. Jöreskog en Sörbom (1985) geven drie indicatoren waaraan globaal kan worden afgemeten of het model bij de data past. Eén indicator is de χ^2 -toets. Twee andere aanwijzingen geven de goodness of fit index (GFI) en de root mean square residual (RMR). Of het model op onderdelen adequaat is kan afgemeten worden aan de genormaliseerde residuen.

RESULTATEN METHODENVERGELIJKING

In het onderzoek zijn de acht methoden met elkaar vergeleken ten aanzien van implementatie en doelmatigheid bij leerlingen. Omdat de methodenvergelijking in dit artikel niet centraal staat wordt hier volstaan met een beknopte samenvatting van de resultaten van deze methodenvergelijking. Voor meer details over de methodenvergelijking verwijs ik naar Harskamp en Suhre (1986).

Vergelijking ten aanzien van de implementatie.

Tussen de acht onderzochte methoden zijn verschillen geconstateerd ten aanzien van implementatiekenmerken. Het percentage leerlingen dat de basisstof uit de methode heeft gedaan verschilt ($F(7, 110) = 5.5$; $p = .00$).

De basisstof heeft in alle methoden betrekking op de leerstofcategorieën die in de Cito-eindtoets Rekenen zijn opgenomen. Hoewel er lichte verschillen zijn, kan door de grote overeenkomst in leerstof tussen de rekenmethoden de basisstof als vergelijkbaar worden beschouwd. De mate van variatie in leerstof tijdens lessen verschilt ook tussen de methoden ($F(7, 110) = 2.8$; $p = .01$). Het gemiddeld aantal rekenlessen dat per schoolweek wordt gegeven verschilt niet tussen de methoden. De meeste leerlingen ontvangen per schoolweek vijf rekenlessen van circa 50 minuten. Ten aanzien van de leerstofdifferentiatie zijn er eveneens verschillen tussen de rekenmethoden ($F(7, 110) = 4.9$; $p = .00$). Deze verschillen komen overeen met de nadruk die in de methoden op leerstofdifferentiatie wordt gelegd. De verschillen tussen de methoden op de implementatiekenmerken zijn vooral groot tussen de groep traditionele versus de groep moderne methoden. In tabel 1 zijn de gemiddelde waarden voor beide groepen rekenmethoden uitgezet.

Vergelijking op rekenprestaties.

De belangrijkste vraag van het onderzoek was of er verschillen zijn tussen rekenmethoden ten aanzien van de bereikte rekenprestaties bij leerlingen. Dit blijkt niet het geval. De multivariate toets voor het methode-effect is niet significant (Pillais-toets = .16; $p = .23$).

Tabel 2: De correlatiematrix van de 9 variabelen in het padmodel.

Variabelen:	1	2	3	4	5	6	7	8	9
(1) METHODE	1.00								
(2) IQ	.32	1.00							
(3) ATT	.26	.03	1.00						
(4) LESSEN	.14	.19	.00	1.00					
(5) PERCBA	.14	.12	-.08	-.02	1.00				
(6) VARIATIE	.21	.06	-.15	.16	.13	1.00			
(7) DIFF	-.34	-.13	.00	-.15	-.18	.07	1.00		
(8) CITOR	.27	.71	-.14	.33	.34	.30	-.20	1.00	
(9) RIONT	.31	.85	.00	.32	.15	.21	-.13	.82	1.00

RESULTATEN TOETSING VAN HET ONDERWIJSMODEL

Bij toetsing van het padmodel blijkt dat dit model de verwachte relaties kan beschrijven. De χ^2 -toets (met 19 vrijheidsgraden) voor het model is 23.9 ($p = .20$). De GFI is .96 en de RMR is .05. De grootste genormaliseerde residu is 1.60 (zie tabel 3), wat acceptabel is. Met betrekking tot de causale paden valt op te merken dat alle T-waarden significant zijn op 5% niveau. Het model is weergegeven in figuur 1. In de figuur zijn alleen de padcoëfficiënten tussen de variabelen weergegeven. Deze padcoëfficiënten zijn maximum likelihood schattingen gebaseerd op de oorspronkelijke correlaties tussen de variabelen. De stoortermen zijn weggelaten. Alleen met betrekking tot de variabelen CITOR en RIONT zijn de waarden van de stoortermen interessant, omdat ze de hoeveelheid niet verklaarde variantie op deze variabelen aangeven. Van CITOR wordt 35% van de variantie niet verklaard. Van RIONT wordt 18% niet verklaard.

Het model geeft een beschrijving van zeven variabelen die van invloed zijn op de rekenprestaties aan het einde van de basisschool. Onderaan in figuur 1 zijn de rekenprestaties weergegeven door twee variabelen: CITOR, een toets voor basisrekenvaardigheid en RIONT, een toets voor toegepast rekenen. Het model laat zien dat een hoge mate van algemene rekenvaardigheid leidt tot een hoge mate van toepassingsvaardigheid ($b = .43$). In het model liggen de volgende relaties vevat:

Ten aanzien van het intelligentieniveau van klassen:

Uit de analyse blijkt een grote directe invloed van de gemiddelde intelligentiescore van klassen (IQ) op CITOR ($b = .64$) en RIONT ($b = .55$). Het totale effect van IQ op RIONT (direct en indirect via CITOR) is .82. Er is geen andere factor in het model met een dergelijk grote invloed op rekenprestaties. Het intelligentieniveau van een klas speelt een belangrijke rol in de keuze voor een moderne of traditionele methode. Moderne methoden worden vooral gebruikt bij klassen met een hoger intelligentieniveau ($b = .31$).

Ten aanzien van de rekenmethode:

Uit de analyse blijkt dat moderne rekenmethoden tot minder leerstofdifferentiatie leiden dan traditionele methoden ($b = -.34$) en tot meer variatie van leerstof ($b = .32$). Dit is in overeenstemming met de vakdidactische bedoelingen van de methoden. Moderne methoden leiden dus tot enigszins ander rekenonderwijs dan traditionele methoden. Deze verschillen zijn het duidelijkst ten aanzien van leerstofdifferentiatie en de variatie van leerstof.

Ten aanzien van de lesattitude van de leerkracht:

Een leerlinggerichte opvatting over onderwijzen en leren is van invloed op (de aanschaf en) het gebruik van een moderne rekenmethode ($b = .26$). Meer leerstofgerichte leerkrachten brengen volgens verwachting meer variatie in leerstof aan ($b = -.24$). Blijkbaar behandelen meer leerlinggerichte leerkrachten de leerstof uit de methode minder volledig.

Tabel 3: De genormaliseerde residuenmatrix bij het padmodel.

Variabelen:	1	2	3	4	5	6	7	8	9
(1) METHODE	.00								
(2) IQ	.00	.00							
(3) ATT	.00	.00	.00						
(4) LESSEN	.89	.00	.00	.00					
(5) PERCBA	.92	1.14	-.98	-.23	.00				
(6) VARIATIE	.13	-.39	.19	.01	1.60	-.06			
(7) DIFF	.00	-.20	.95	-1.36	.00	-.42	.00		
(8) CITOR	.04	.16	-1.41	-.05	.99	.11	-1.22	.27	
(9) RIONT	.28	.06	-.10	.77	.41	.29	-.32	.21	.11

Ten aanzien van het aantal rekenlessen:

Het gemiddeld aantal rekenlessen dat de leerlingen de laatste twee leerjaren gehad hebben heeft een directe invloed op CITOR ($b = .18$). Tevens is er een geringe indirecte invloed via variatie van leerstof in de rekenlessen, .03. Het totaaleffect op CITOR is .21 en op RIONT .09. De hoeveelheid rekenlessen heeft dus een positieve invloed op rekenprestaties. Dit geldt onafhankelijk van het intelligentie-niveau van klassen.

Ten aanzien van de behandeling van de basisstof:

Er is een direct effect van het percentage leerlingen in de klas dat de basisstof heeft gedaan op CITOR ($b = .24$). Naarmate een groter gedeelte van de leerlingen uit de klas de basisstof gehad heeft zijn de rekenprestaties beter. Dit is van kracht ongeacht het intelligentieniveau van klassen.

Ten aanzien van variatie in leerstof tijdens de lessen:

Naarmate er meer variatie in leerstof tijdens de lessen voorkomt zijn de rekenprestaties beter. De ruimte voor variatie van leerstof in de rekenlessen hangt af van het aantal rekenlessen ($b = .14$) en indirect van het intelligentieniveau van klassen (totaal effect .08).

Ten aanzien van leerstofdifferentiatie:

Voor de hypothese dat leerstofdifferentiatie leidt tot daling van rekenprestaties kan enige empirische grond worden gevonden. Leerstofdifferentiatie leidt tot een lager percentage leerlingen in de klas waarmee alle basisstof uit de methode wordt behandeld en daardoor indirect tot lagere klasprestaties (indirect effect via PERCBA -.04). Hier tegenover staat echter een vergelijkbaar positief effect van leerstofdifferentiatie op rekenprestaties door meer variatie in leerstof (indirect effect via VARIATIE .04). Differentiatie heeft dus zowel negatieve als positieve invloed op de prestaties van leerlingen in klassen.

CONCLUSIES EN DISCUSSIE

Uit de inhoudsanalyse van de rekenmethoden (zie Wiersma & Willemsen, 1986) blijkt dat, voor zover er verschillen zijn tussen de methoden, deze verschillen liggen in de bij de methode behorende didactiek. Qua inhoud ontlopen de de methoden elkaar niet zoveel. Uit het onderzoek is ook gebleken dat de prestaties van groepen leerlingen die met een van de acht rekenmethoden les hebben gehad, niet van elkaar verschillen wanneer rekening gehouden wordt met verschillen in achtergrondkenmerken van de leerlingen. De beschrijving van de implementatiekenmerken laat zien dat moderne methoden gemiddeld iets beter geïmplementeerd zijn (zie tabel 1). Traditionele methoden leiden tot een grotere mate van leerstofdifferentiatie en tot een geringere variatie van leerstof in lessen. Moderne methoden leiden tot minder leerstofdifferentiatie, maar tot meer variatie in leerstofaanbod. Voor zover het type rekenmethode van invloed is op de rekenprestaties is dit via deze vakdidactische kenmerken. De leertijd (in het padmodel vertegenwoordigd door de variabele LESSEN) oefent los van de methode positieve invloed uit op leerprestaties. Opmerkelijk zijn de tegenstrijdige effecten van leerstofdifferentiatie in het padmodel. Het lijkt erop dat van leerstofdifferentiatie een positief effect uitgaat, mits er voor gezorgd wordt dat met alle leerlingen de basisstof wordt gedaan.

Voor verder onderzoek liggen hier belangrijk vragen naar de wijze waarop leerkrachten hun rekenmethode hanteren en in hoeverre zij bij de implementatie van vakdidactische kenmerken knelpunten ervaren met betrekking tot leerstof en lestijd.

Praktisch gezien levert dit onderzoek enige aanwijzingen op voor het bijstellen van rekenmethoden en tot vergroting van het doelmatig gebruik van methoden. Van belang is dat leerkrachten zich houden aan de planning van lessen (ca. 5 per schoolweek) en aan de te behandelen leerstof. Ook wanneer zij in klassen werken met een wat lager intelligentieniveau kunnen deze aanbevelingen tot doelmatig gebruik van methoden leiden.

LITERATUUR

- Appelhof, P. N. (1979). *Begeleide onderwijsvernieuwing: Evaluatie van een curriculuminnovatie van het aanvankelijk leesonderwijs*. Tilburg: Zwijzen.
- Boxtel, H. W., Snijders, J. Th., Welten, V. J. (1982). *ISI-handleiding voor de gehele testreeks en verantwoording van het prestatiegedeelte. Publicatievorm III*. Groningen: Wolters-Noordhoff.
- Bruins, R. en Bloemkolk M. (1987, Mei). *Kun je rekenen op je methode*. ORD-paper, Groningen.
- Brophy, J. (1986). Teaching and learning mathematics: where research should be going. *Journal for Research in Mathematics Education*, 17, (5), 323-346.
- Cooley, W. W. & Lohnes, P. R. (1976). *Evaluation Research in Education*. New York: John Wiley & Sons.
- Harskamp, E. G., Hofman R., Lugthart, E., Hoef. A. van der, Bosveld, H. (1982). *Eindverslag van het onderzoeksproject 'Evaluatie van Leermethoden in het Basisonderwijs', een empirische studie naar factoren en effecten van curriculumimplementatie. deel A en B*. RION, Haren.
- Harskamp, E. G. & Prins, J. (1985). De organisatie van basisscholen: stand van zaken en aanbevelingen voor verder onderzoek. In R. Vandenberghen en B. P. M. Creemers (eds.). *Vernieuwing in het Basisonderwijs* (p. 45-57). Amersfoort: ACCO.
- Harskamp, E. G., Pijl, Y. & Prins, J. (1984). *Werken aan de nieuwe basisschool, Eindrapport*. RION, Haren.
- Harskamp, E. G., Suhre C., Wiersma, Y. (1985). *Vergelijking van Rekenmethoden in het Basisonderwijs, Interimrapport I*. RION, Groningen.
- Harskamp, E. G. & Suhre, C. (1986). *Vergelijking van Rekenmethoden in het Basisonderwijs, Eindrapport*. RION, Groningen.
- Jong, R. A. de (1986). *Wiskobas in Methoden*. Dissertatie. Vakgroep OW & OC, Utrecht.
- Jöreskog, K. G. & Sörbom D. (1985). *Analysis of Linear Structural Relationships by the method of Maximum Likelihood, user's guide, Lisrel VI*. Department of Statistics, University of Uppsala, Uppsala.
- Meijnen, G. W. (1982). *Schoolloopbanen in het lager onderwijs*. Sociologisch Instituut, Groningen.
- Poyner, H. (1974). Selecting units of analysis. In G. B. Borich (ed.), *Evaluating educational programs and products* (p. 399-414). Englewood Cliffs, NJ: Prentice-Hall.
- Reezigt, G. J., van Dijk, M. H., Bosveld J. J. F. (1986). *Differentiatie op de basisschool, Eindrapport*. RION, Groningen.
- Rengerink, J. (1983). *De staartdeling. een geïntegreerde aanpak volgens het principe van progressieve schematisering*. Vakgroep OW & OC, Utrecht.
- Rosenshine, B. (1979). Time and direct instruction. In P. L. Peterson & H. J. Walberg (eds.), *Research on Teaching* (p. 28-55). Berkeley: McCutchan.
- Slavenburg, J. (1978). *Het project Onderwijs en Sociaal Milieu: een bundel artikelen opgedragen aan Dr. J. Grandia*. Tilburg: Zwijzen.
- Suhre, C., Harskamp E. G. (1986). *Vergelijking van Rekenmethoden in het Basisonderwijs: Verantwoording van eindtoetsen Toegepast Rekenen in het basisonderwijs. Interimrapport IV*. RION, Groningen.
- Veenman, S. A. M. (1981). Leertijd, directe instructie en de relatie onderzoekspraktijk. In Veenman en Kok (eds.), *Opleiding van onderwijsgeevenden* (p. 23-45). 's Gravenhage: Staatsuitgeverij.
- Vink, H. (1978). Algorithmen met kruispunten en eieren. Leerpsychologische verkenningen en een onderzoek naar het leerresultaat en het hanteren van een algoritme leergang vermenigvuldigen in het derde leerjaar van de basisschool. In H. Dekker, H. ter Heege & A. Treffers, *Cijferend vermenigvuldigen en delen volgens wiskobas* (p. 229). Vakgroep OW & OC, Utrecht.
- Vrolijk, A. (1981). *Algorithmen in het onderwijs*. Doctoraalscriptie. PAW, Utrecht.
- Walberg, H. J. (1984). Improving the Productivity of America's Schools. *Educational Leadership*, 42, 19-27.
- Wiersma, Y. & Willemsen, T. (1986). *Vergelijking van Rekenmethoden, Interimrapport III*. Rion, Groningen.
- Wijnstra, J. M. (1984). *Verantwoording Eindtoets Basisonderwijs 1981*. Cito, Arnhem.

Manuscript ontvangen 30-6-1987.

Definitieve versie ontvangen 7-10-1987.

Het Rasch model: Naar een alternatieve wijze van itemselectie

Diana M.P. Smedts

ABSTRACT

In this article we prove that the Rasch difficulty estimates are not robust when the item discriminations are different and/or when the subjects are guessing. This fact has implications for the relation between dimensionality and the inequality of discriminations. It seems impossible to represent item characteristic curves with different slopes in an unweighted Euclidean space. We propose the weighted Euclidean space as an alternative and suggest INDSCAL as a possible step in the item selection procedure. An illustration is provided by means of simulated data.

In dit artikel wordt, d.m.v. de logitmethode, aangetoond dat de itemparameterschattingen onder het Rasch model niet robuust zijn t.a.v. schendingen van de onderstellingen der gelijke discriminaties en/of afwezigheid van raden. Deze vaststelling heeft implicaties voor de relatie tussen de dimensionaliteit en de ongelijkheid der discriminatieparameters. Hierbij aansluitend wordt INDSCAL voorgesteld als hulpmiddel bij de itemselectie. Onze bevindingen worden geïllustreerd aan de hand van gesimuleerde data.

ONDERZOEK NAAR DE ROBUUSTHEID

In onze uiteenzetting onderscheiden wij twee mogelijke schendingen van het Rasch model, nl. de discriminatiewaarden verschillen en er wordt geraden. Deze schendingen kunnen elk afzonderlijk of samen gevat worden in het twee-parameter logistisch model (2-PLM), het drie-parameter logistisch model met gelijke discriminaties (3'-PLM) respect, het drie-parameter logistisch model met verschillende discriminaties (3-PLM). De vraag is in welke mate de parameterschattingen onder het Rasch model beïnvloed worden door de aanwezige schendingen.

Om enig inzicht te krijgen in dit probleem maken wij gebruik van de logit-methode.

Stel dat de data voldoen aan het 2-PLM, nl.

$$P_{vi} = \frac{e^{a_i(b_v - d_i)}}{1 + e^{a_i(b_v - d_i)}}$$

waarin b_v het eigenschapsniveau is van persoon v , $-\infty < b_v < +\infty$;

d_i de moeilijkheidswaarde is van item i , $-\infty < d_i < +\infty$;

a_i de discriminatiewaarde is van item i , $0 < a_i < +\infty$;

P_{vi} de kans is dat een persoon met eigenschapsniveau b_v een item met moeilijkheidswaarde d_i correct oplost.

De logit bedraagt:

$$\ln \frac{P_{vi}}{1 - P_{vi}} = a_i (b_v - d_i) \quad (2)$$

Indien data gegenereerd volgens het 2-PLM aan een Rasch-analyse onderworpen worden, kunnen we

$$\ln \frac{P_{vi}}{1 - P_{vi}} = \hat{b}_v - \hat{d}_i \quad (3)$$

als schattingsvergelijking naar voor schuiven.

De vergelijking van formule (2) met (3) toont aan dat de schattingen onder het Rasch model systematisch afwijken t.o.v. de werkelijke waarden:

$$\hat{d}_i = a_i d_i$$

$$\hat{b}_v = a_i b_v$$

De scores worden dus gewogen met een factor gelijk aan de discriminatiewaarde van het beschouwde item.

Bestuderen wij nu data gegenereerd volgens het 3'-PLM. De kans op een correct antwoord wordt gedefinieerd als:

$$P_{vi} = c_i + (1 - c_i) \frac{e^{b_v - d_i}}{1 + e^{b_v - d_i}} \quad (4)$$

waarin c_i de kans voorstelt dat het correcte antwoord geraden wordt ($c_i > 0$).

Het is makkelijk af te leiden dat

$$\ln \frac{P_{vi}}{1 - P_{vi}} = \ln \frac{\gamma_i + f_{vi}}{1 - f_{vi}} = \hat{b}_v - \hat{d}_i \quad (5)$$

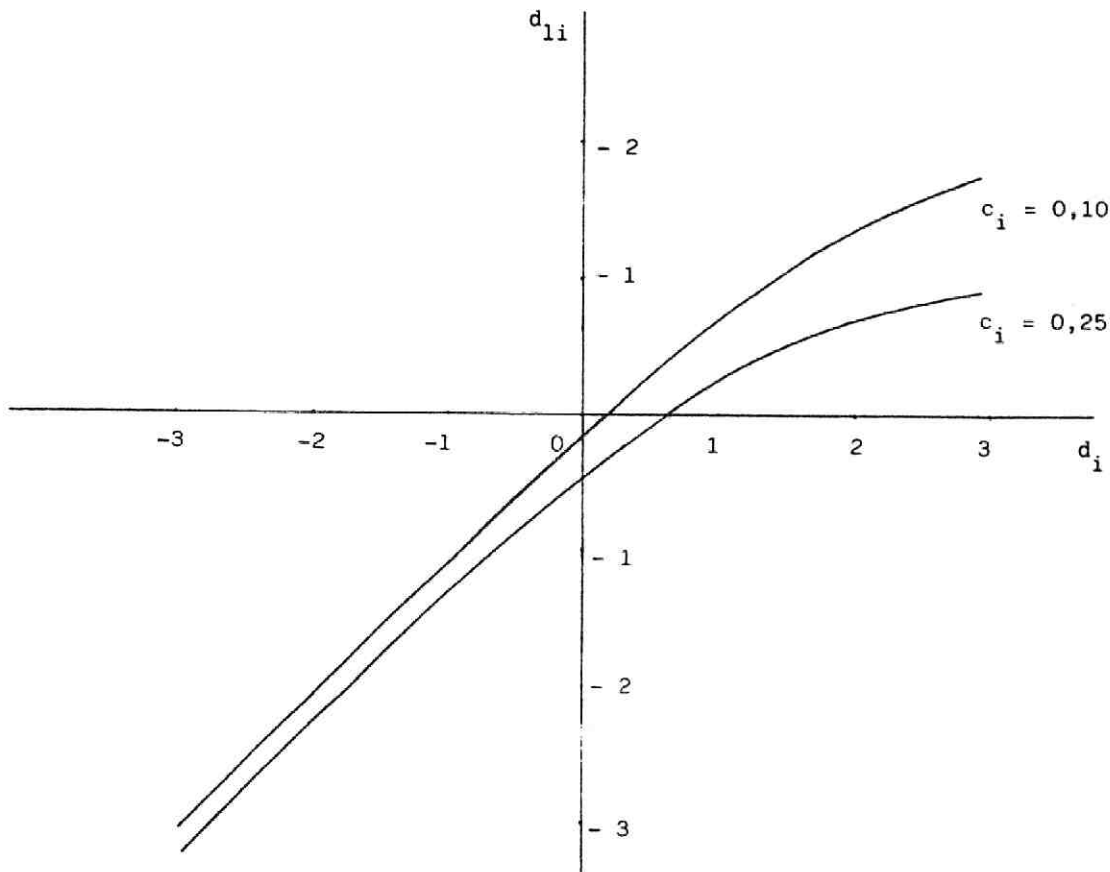
$$\text{met } \gamma_i = \frac{\bar{c}_i}{1 - c_i} \quad (6)$$

$$f_{vi} = \frac{e^{b_v - d_i}}{1 + e^{b_v - d_i}} \quad (7)$$

De formules voor het 3-PLM zijn, mits de introductie van de discriminatieparameter a_i , volledig analoog.

Om een duidelijk beeld te krijgen van de afwijking die de aanwezigheid van raden veroorzaakt in de Rasch-schattingen van d_i , geven wij in figuur 1 het verband tussen de werkelijke moeilijkheid d_i en de op basis van formule (5) geschatte moeilijkheid. In de berekeningen wordt de vaardigheidsparameter gelijkgesteld aan nul.

Hieruit blijkt dat de afwijking, bij constant houden van de eigenschap, des te groter is naarmate het item moeilijker is.



Figuur 1: Voorspelde (d_{li}) t.o.v. ware moeilijkheid (d_i).

TOEPASSING OP GESIMULEERDE DATA

Uit wat voorafgaat mag blijken dat indien de werkelijke itemparameters gekend zijn, de logit-methode kan gebruikt worden om de Rasch-schattingen van de itemmoeilijkheid te voorspellen. In deze paragraaf vergelijken wij de verschillende moeilijkheidswaarden voor gesimuleerde data.

Voor het genereren van data stelden wij een programma op uitgaande van de omschrijving van van de Vijver (1982). Bij wijze van voorbeeld geven wij de werkwijze voor data die voldoen aan het 2-PLM:

- 10 itemparameters d_i en a_i worden opgegeven;
- 500 waarden van persoonparameters b_v worden aselekt uit een standaardnormale verdeling getrokken;
- de kans op een correct antwoord gegeven d_i , a_i en b_v wordt berekend d.m.v. formule (1);
- een (500×10) -matrix U met elementen $0 < u_{vi} < 1$ wordt gegenereerd uit een uniforme verdeling;
- de datamatrix X wordt bepaald als $x_{vi} = 1$ als $p_{vi} \geq u_{vi}$
 $x_{vi} = 0$ anders;
- personen die geen of alle items juist hebben en items die nooit of altijd correct beantwoord worden, worden verwijderd.

Analoog worden gegevens gegenereerd volgens het 3'- en 3-PLM.

Voor elk van de 3 datasets worden de Rasch-schattingen $\hat{d}_i$ bepaald d.m.v. PML (Gustafsson, 1981; Molenaar, 1981). Voorspellingen van deze schattingen, nl. d_{li} , worden bekomen d.m.v. de logit-methode.

In tabel 1 worden de werkelijke itemparameters, d_{ii} , en $\hat{d}_i$ afgedrukt voor elk van de datasets. De schattingsfout van $\hat{d}_i$ varieert van 0,098 tot 0,138.

Uit de resultaten blijkt dat $\hat{d}_i$, rekening houdend met de schattingsfout, d_{ii} dicht benadert. Ter vergelijking met de resultaten die van de Vijver (1982) geeft in zijn onderzoek naar de robuustheid, geven wij in tabel 1 ook de standaardafwijking van de moeilijkheidsparameters en de correlatie tussen werkelijke en geschatte itemmoeilijkheid. Zowel in figuur 1 als uit de vergelijking der standaardafwijkingen blijkt dat de spreiding der itemmoeilijkheid kleiner is in geval van raden. Ook verschillen in discriminatie beïnvloeden de spreiding: 1,002 voor het 2-PLM t.o.v. 1,436 voor het Rasch model. Dit komt in de resultaten van van de Vijver (1982, 1986) niet tot uiting omdat hij werkt met een gemiddelde discriminatie van 0,65 en $\bar{a} \cdot s_d = 0,65 \cdot 1,436 = 0,9334$, dit is een ruwe benadering van de gevonden 1,002.

De correlatie tussen werkelijke en geschatte moeilijkheid is zeer hoog en perfect in te passen in de resultaten van van de Vijver. De grootte van de correlatie wordt medebepaald door de spreiding der itemmoeilijkheid: indien de spreiding van d_i en $\hat{d}_i$ veel groter is dan de verschuivingen van d_i naar $\hat{d}_i$, zal de correlatie toch hoog zijn. Deze grootheid wordt dus beter niet gebruikt als maat voor de robuustheid.

UNIDIMENSIONALITEIT VERSUS VERSCHILLEN IN DISCRIMINATIE.

Twee van de onderstellingen waarop het Rasch model steunt, zijn de unidimensionaliteit en de gelijke discriminaties. Wij stellen ons hier de vraag of de unidimensionaliteit strijdig is met het toelaten van verschillen in discriminatie.

In de literatuur vinden wij geen eenduidig antwoord. Enerzijds toont Lumsden (1978) aan dat bij een unidimensionele test alle itemdiscriminaties gelijk moeten zijn. Anderzijds kan het Rasch model beschouwd worden als een speciaal geval van het 2-PLM, nl. als alle discriminatiewaarden gelijk zijn. Het 2-PLM onderstelt evenzeer een unidimensionele test.

Wij formuleren kort een omschrijving van unidimensionaliteit. Vervolgens lichten wij de notie discriminatiewaarde toe en onderzoeken dan of verschillen in discriminatieparameters te verzoenen zijn met de unidimensionaliteit.

Een test is unidimensioneel als één eigenschap volstaat om de interindividuele verschillen in antwoordgedrag te verklaren. De andere beïnvloedende factoren staan loodrecht op, d.i. correleren niet met de testuitslag.

Tabel 1: Werkelijke itemparameters versus voorspelde (d_{ii}) en geschatte moeilijkheid ($\hat{d}_i$).

Item	2-PLM				3'-PLM			3-PLM			
	d_i	a_i	d_{ii}	$\hat{d}_i$	c_i	d_{ii}	$\hat{d}_i$	a_i	c_i	d_{ii}	$\hat{d}_i$
1	2,25	0,5	1,132	1,297	0,10	1,928	2,095	0,5	0,10	1,150	1,185
2	1,75	0,8	1,408	1,495	0,10	1,641	1,637	0,8	0,10	1,353	1,225
3	1,25	0,8	1,008	1,008	0,25	0,786	0,618	0,8	0,25	0,592	0,534
4	0,75	0,5	0,382	0,513	0,10	0,903	0,853	0,5	0,10	0,532	0,475
5	0,25	0,5	0,132	0,297	0,25	0,134	-0,064	0,5	0,25	-0,014	-0,119
6	-0,25	0,8	-0,192	-0,203	0,25	-0,265	-0,313	0,8	0,25	-0,275	-0,146
7	-0,75	0,8	-0,592	-0,621	0,10	-0,451	-0,392	0,8	0,10	-0,360	-0,300
8	-1,25	0,5	-0,618	-0,640	0,10	-0,933	-0,896	0,5	0,10	-0,384	-0,338
9	-1,75	0,5	-0,868	-1,024	0,25	-1,630	-1,712	0,5	0,25	-0,863	-0,801
10	-2,25	0,8	-1,792	-1,833	0,25	-2,113	-1,826	0,8	0,25	-1,730	-1,716
s_d	1,436		0,958	1,002		1,270	1,245			0,887	0,851
$r_{d_i \hat{d}_i}$				0,982			0,983				0,962

De discriminatiewaarde a_i wordt in het 2- en 3-PLM omschreven als de helling, de richtingscoëfficiënt van de item-antwoordcurve in het punt $b_v = d_i$.
Uit Lord en Novick (1968, p. 377-379) kan afgeleid worden dat voor het 2-PLM

$$a_i = \frac{1,7 \rho_i}{\sqrt{1 - \rho_i}} \quad (8)$$

waarin ρ_i de biseriële correlatie is tussen item i en eigenschap b .

De discriminatiewaarde wordt dus bepaald door de mate waarin het item de eigenschap meet.

Items die voldoen aan het Rasch model vormen een unidimensionele test en verschillen alleen in moeilijkheid. Deze laatste kan opgevat worden als de vaardigheid die vereist is om het item met een kans van 0,5 correct op te lossen. Het is derhalve mogelijk om de itemmoeilijkheid te situeren op diezelfde eigenschapsdimensie. Een gewone Euclidische ruimte (van dimensionaliteit één) volstaat om de eigenschapsdimensie voor te stellen.

In het 2-parameter logistisch model onderscheiden de items zich t.a.v. twee aspecten, nl. de moeilijkheids- en de discriminatiewaarde. In de voorgaande paragraaf toonden wij aan dat voor data gegenereerd volgens het 2-parameter logistisch model de Rasch-schatting van de moeilijkheid $\hat{d}_i$ gelijk is aan $a_i d_i$. Dit betekent dat de eigenschapsschaal, waarop $\hat{d}_i$ kan gesitueerd worden, gewogen wordt volgens de item-discriminatie. Dit is niet te verzoenen met de gewone, wel met de gewogen Euclidische ruimte.

Bij een gewogen Euclidische ruimte wordt onderscheid gemaakt tussen twee ruimtes. Voor de toepassing die wij op het oog hebben, noemen wij ze de subject- en de itemruimte. Onder het 2-parameter logistisch model zijn beide ruimtes unidimensioneel.

De subjectruimte is een Euclidische ruimte waarin de meetuitslagen als punten gelokaliseerd worden.

De itemruimte is dan eveneens een unidimensionele Euclidische ruimte. De items worden voorgesteld als vectoren waarvan de lengte overeenkomt met het gedeelte van de variantie dat verklaard wordt door die ene dimensie.

De afstand $d_{vw}^{(i)}$ tussen twee personen v en w wordt in de gewogen Euclidische ruimte gedefinieerd als:

$$d_{vw}^{(i)} = a_i (b_v - b_w) \quad (9)$$

Dit betekent dat de eigenschapsdimensie afhankelijk van het item een ander gewicht krijgt.

Het is intuïtief aanvaardbaar dat een hoog discriminerend item de verschillende vaardigheidswaarden beter onderscheidt dan een laag discriminerend. In de gewogen Euclidische ruimte wordt dit weerspiegeld door de afstand tussen de eigenschapswaarden te wegen volgens de discriminatie. De parameter a_i in formule (9) kan dus beschouwd worden als de discriminatiewaarde. Bovendien stemt deze voorstelling in de gewogen Euclidische ruimte perfect overeen met onze bevindingen in de voorgaande paragraaf, nl. dat $\hat{d}_i = a_i d_i$ en $\hat{b}_v = a_i b_v$.

Terloops merken wij op dat het bij de introductie van het twee-parameter logistisch model gebruikelijk is de itemkarakteristieke curves met verschillende helling op dezelfde eigenschapsschaal voor te stellen. Dit gebruik dient verworpen te worden: de itemkarakteristieke curves hebben dezelfde helling in de gemeenschappelijke itemruimte. Bij dergelijke figuur dienen de discriminatieparameters aangegeven te worden.

Onder het alternatief van een gewogen Euclidische ruimte sluiten de onderstelling der unidimensionaliteit en het toelaten van verschillen in discriminatie elkaar niet uit. Het toelaten van verschillen in discriminatie impliceert wel dat men aanvaardt dat de hoeveelheid verklaarde variantie en bijgevolg de foutenvariantie mag verschillen van item tot item (zie de interpretatie van a_i^4). De foutenvariantie bepalen wij als de niet-verklaarde variantie. Die foutenvariantie kan zowel de variantie tengevolge van itemspecifieke factoren als de toevallige foutenvariantie

omvatten. Gezien onze omschrijving van de dimensionaliteit van een test is het onmogelijk om deze twee aspecten van de foutenvariantie te onderscheiden.

ITEMSELECTIE

Aansluitend bij de logit-methode en de resulterende conclusies kan de $(k-1) \times k$ -matrix K met per scoregroep ($v = 1, \dots, k-1$) en per item ($i = 1, \dots, k$) het element in $[P_{vi}/(1 - P_{vi})] = b_v - d_i$ beschouwd worden.

$$k = |k_{vi}| = \begin{bmatrix} b_{k-1}-d_1 & b_{k-1}-d_2 & \dots & b_{k-1}-d_i & \dots & b_{k-1}-d_k \\ b_{k-2}-d_1 & b_{k-2}-d_2 & \dots & b_{k-2}-d_i & \dots & b_{k-2}-d_k \\ \vdots & \vdots & & \vdots & & \vdots \\ b_v-d_1 & b_v-d_2 & \dots & b_v-d_i & \dots & b_v-d_k \\ \vdots & \vdots & & \vdots & & \vdots \\ b_1-d_1 & b_1-d_2 & \dots & b_1-d_i & \dots & b_1-d_k \end{bmatrix}$$

Als de gegevens werkelijk aan het Rasch model voldoen en niet met toevallige fouten behept zijn, dan kunnen de persoonparameters op een additieve constante na bepaald worden d.m.v.

$$k_{wi} - k_{vi} = (b_w - d_i) - (b_v - d_i) = b_w - b_v \quad \begin{matrix} w, v = 1, \dots, k-1 \\ w \neq v \end{matrix} \quad (10)$$

Alle items i zullen tot dezelfde persoonparameters leiden.

Uit

$$k_{vj} - k_{vi} = (b_v - d_j) - (b_v - d_i) = d_i - d_j \quad \begin{matrix} i, j = 1, \dots, k \\ i \neq j \end{matrix} \quad (11)$$

worden de itemparameters berekend. Elke scoregroep v ($v = 1, \dots, k-1$) levert dezelfde waarden op.

In de praktijk komt dergelijk ideaal geval niet voor tengevolge van toevallige fouten. Hoe kunnen wij dan te werk gaan?

Onderstel dat de absolute waarde van formule (10) ons per item de afstand oplevert tussen elke twee persoonparameters. Wij moeten trachten de persoonparameters te lokaliseren in een multidimensionele ruimte zodanig dat de uit (10) bekomen afstanden gereproduceerd worden. Wanneer de itemdiscriminaties verschillen, zullen ook de afstanden tussen de persoonparameters verschillen naargelang de items die als referentie voor de berekening gebruikt worden. Wij moeten dus toelaten dat de items een differentieel gewicht toekennen aan de dimensies.

Voor het situeren van de personen in een r -dimensionele gewogen Euclidische ruimte stellen wij de volgende formule voor:

$$|k_{wi} - k_{vi}| = \sqrt{\sum_{t=1}^r a_{it}^2 (z_{wt} - z_{vt})^2} \quad (12)$$

met a_{it} = het gewicht van item i op dimensie t ($t = 1, \dots, r$)

z_{vt} = de schaalwaarde van persoon v op dimensie t .

Gegeven de afstanden $|k_{wi} - k_{vi}|$ ($w, v = 1, \dots, k-1$; $v \neq w$; $i = 1, \dots, k$) moeten de itemgewichten a_{it} en de persoonparameters z_{vt} geschat worden. Dit kan gebeuren d.m.v. het ALSCAL-programma (Young & Lewyckij, 1980) met als optie INDSCAL (INDividual Differences

SCALing). Deze laatste is een multidimensionele schaalmethode, ontwikkeld door Carroll en Chang (1970).

Wij lichten dit algoritme hier enkel toe voor zover dit nodig is voor de interpretatie van de parameters en voor het afleiden van een antwoord op onze vragen m.b.t. het Rasch model. Voor meer technische details verwijzen wij naar Young en Lewyckij (1980).

Uitgaande van de afstanden

$$d_{wv}^{(i)} = |k_{wi} - k_{vi}| \quad \begin{matrix} w, v = 1, \dots, k-1 \\ i = 1, \dots, k \end{matrix} \quad (13)$$

worden de B_i^* -matrices met scalaire produkten bepaald.

In ALSCAL zijn twee opties mogelijk m.b.t. de matrices B_i^* . De matrix-conditionele optie houdt in dat de matrices B_i^* genormaliseerd worden zodanig dat de te verklaren variantie van $b_{wv}^{(i)} (w \neq v)$ gelijk is aan één. Dit impliceert dat de verschillen in variantie, die een gevolg zijn van de verschillen in discriminatiewaarde, weggewerkt worden. Indien wij INDSCAL wensen te gebruiken als selectiemiddel voor Rasch-homogene items dient deze optie verworpen te worden.

De tweede optie is de niet-conditionele, d.w.z. dat de B_i^* -matrices het uitgangspunt van de analyse vormen.

De output omvat de $(k-1) \times r$ -matrix $\hat{Z}$ met elementen $\hat{z}_{vi}$. De matrix beschrijft de gemeenschappelijke subjectruimte. Hij wordt genormaliseerd zodanig dat de schaalwaarden per dimensie uitgedrukt zijn in standaardwaarden.

Verder omvat de output voor de matrix-conditionele optie de $k \times r$ -matrix $\hat{A}$ met itemgewichten $\hat{a}_{it}^2$. Deze worden genormaliseerd zodanig dat:

$$\sum_t \hat{a}_{it}^4 = R_i^2$$

R_i^2 is het kwadraat van de correlatie tussen dispariteiten (optimaal geschatte afstanden) en geschatte afstanden

$$\hat{d}_{wv}^{(i)} = \sqrt{\sum_t \hat{a}_{it}^2 (\hat{z}_{vi} - \hat{z}_{wt})^2}$$

en geeft als zodanig de proportie van de variantie van de dispariteiten aan die door de schattingen verklaard wordt.

Als globale maat wordt o.a.:

$$R^2 = \sum_i \frac{R_i^2}{k} \quad (15)$$

bepaald.

Indien de dimensies van de gemeenschappelijke subjectruimte onafhankelijk zijn, d.i. als $Z'Z$ een diagonale matrix is, dan kan $\hat{a}_{it}^4$ opgevat worden als proportie van de variantie van de dispariteiten m.b.t. item i verklaard door dimensie t . $\hat{a}_{it}^2$ geeft dan de correlatie tussen item i en dimensie t of, in geometrische termen uitgedrukt, de cosinus van de hoek tussen itemvector en dimensie t .

Zoals wij eerder in deze paragraaf reeds aangaven, is de matrix-conditionele optie niet interessant met het oog op de selectie van Rasch-homogene items. De reden waarom wij de output ervan bespreken is de relatie met de factoranalyse (Smedts, 1986).

Bij de niet-conditionele optie wordt de matrix $\hat{A}$ genormaliseerd zodanig dat:

$$\frac{\sum_i \sum_t \hat{a}_{it}^4}{k} = R^2 \quad (16)$$

R^2 wordt bepaald als:

$$R^2 = \frac{\left[\sum_i \sum_w \sum_v (d_{wv}^{(i)} - d_{..}^{(.,)}) (d_{wv}^{(i)} - d_{..}^{(.,)}) \right]^2}{\left[\sum_i \sum_w \sum_v (d_{wv}^{(i)} - d_{..}^{(.,)})^2 \right] \left[\sum_i \sum_w \sum_v (d_{wv}^{(i)} - d_{..}^{(.,)})^2 \right]} \quad (17)$$

R^2 geeft de globale, d.i. over alle items, proportie van de variantie der dispariteiten die verklaard wordt door de schattingen. De gewichten $\hat{a}_{it}^2$ kunnen groter zijn dan één en weerspiegelen de verschillen in discriminatiewaarde. Nadeel is dat de grootte van $\hat{a}_{it}^2$ ook beïnvloed wordt door de mate waarin de data voor item i door de INDSCAL-oplossing kunnen verklaard worden. Bovendien biedt de niet-conditionele optie geen mogelijkheid om de grootte van deze invloed na te gaan. R^2 berekenen zou nl. zinloos zijn, omdat deze groter kan zijn dan één. Dit noopt ons ertoe te werken met de matrix-conditionele oplossing. De verschillen in discriminatie zullen tot uiting komen in de standaardafwijking per item van $k_{vi} = b_v - d_i$. Voldoen de data aan het Rasch model dan zullen de standaardafwijkingen gelijk zijn. Indien de discriminatiewaarden verschillen, dan zullen de standaardafwijkingen verschillen met een factor gelijk aan de discriminatiewaarde.

Tot slot merken wij op dat de schattingen van de subjectparameters $\hat{z}_{vi}$ op de vermenigvuldiging met de standaardafwijking van $\hat{b}_v$ na moeten overeenstemmen met de Rasch persoonparameters $\hat{b}_v$.

De datasets gegenereerd volgens het Rasch respect. twee-parameter logistisch model worden aan een INDSCAL-analyse onderworpen. Bij de extreme scoregroepen doen zich kansen van 0 of 1 voor. In dit geval werd k_{vi} gelijkgesteld aan -5 respect. +5.

Tabel 2: De INDSCAL-itemruimte (1)

Item	Model			
	R		2-PLM	
	I	II	I	II
1	0,944	0,135	0,928	0,249
2	0,945	0,192	0,975	0,041
3	0,951	0,0	0,870	0,395
4	0,976	0,070	0,773	0,533
5	0,969	0,008	0,777	0,613
6	0,968	0,209	0,847	0,514
7	0,952	0,272	0,889	0,145
8	0,949	0,160	0,976	0,0
9	0,981	0,0	0,798	0,521
10	0,978	0,0	0,996	0,010
R^2	0,945		0,929	

(1) Gezien negatieve itemgewichten niet toegelaten zijn, worden zij behandeld als nullen en afgedrukt als 0,0.

Uit tabel 2 blijkt dat de data gegenereerd volgens het Rasch model quasi-perfect door de INDSCAL-oplossing kunnen verklaard worden. Bij het 2-PLM voldoen de meeste items minder goed aan de verwachtingen. Oorzaak hiervan zijn, enerzijds, toevallige afwijkingen in de volgorde van de pseudo-afstanden k_{vi} en, anderzijds, de ruwe, ingevulde schattingen van + of -5 voor een aantal extreme scoregroepen.

Hier verwijzen wij naar onze bemerking m.b.t. de correlatie gemaakt in de eerste paragraaf.

Dong (1985) stelde een regressievergelijking op waardoor voor random data, de verwachte waarde van de logaritme van R kan voorspeld worden op basis van de logaritmen van het aantal

Tabel 3: De standaardafwijking per item van de pseudo-afstanden k_{vi}

Item	Model			
	R		2-PLM	
	(2)	(3)	(2)	(3)
1	1,603	0,993	1,966	1,264
2	2,296	1,777	1,964	1,617
3	1,875	1,431	1,895	1,230
4	1,624	1,315	1,807	0,937
5	2,333	1,634	2,010	0,932
6	2,339	1,543	2,646	1,357
7	1,892	1,228	1,922	1,345
8	1,977	1,148	1,376	1,198
9	2,104	1,522	1,801	0,823
10	1,876	1,610	1,906	1,581

(2) In deze kolom worden de standaardafwijkingen berekend op de 9 scoregroepen gegeven.

(3) De vermelde standaardafwijkingen werden berekend op de scoregroepen 2 tot en met 8.

scoregroepen, het aantal items en de dimensionaliteit. Passen wij deze formule toe dan komen wij aan een voorspelde waarde van $R = 0,504$ of $R^2 = 0,254$.

Uit de standaardafwijkingen in tabel 3 blijkt dat deze nogal wat verschillen zelfs indien de discriminatiewaarden gelijk zijn. Het weglaten van de extreme scoregroepen schept wel een duidelijker beeld voor het 2-PLM, maar niet voor het Rasch model. Waarschijnlijk moet de steekproef van personen veel groter zijn om tot goede schattingen van P_{vi} en bijgevolg van k_{vi} te komen.

Bekijken wij de scoregroepruimte voor de Rasch-data, dan stellen wij vast dat de coördinaten op de eerste dimensie, zoals voorspeld, na vermenigvuldiging met de standaardafwijking van $\hat{b}_v$ praktisch equivalent zijn met de PML-vaardigheidsschattingen (zie tabel 4).

Tabel 4: Vergelijking van de INDSCAL-scoregroep coördinaten met $\hat{b}_v$.

r	ALSCAL-dimensie			
	I	II	$\hat{b}_v$	$I \times s_b$
9	1,428	2,594	2,929	2,535
8	1,247	0,182	1,945	2,214
7	0,681	-0,082	1,222	1,210
6	0,418	-0,168	0,591	0,742
5	0,057	0,845	-0,006	0,101
4	-0,289	-1,001	-0,601	-0,514
3	-0,743	-0,232	-1,229	-1,318
2	-1,076	-0,184	-1,944	-1,910
1	-1,724	-0,632	-2,918	-3,060
s	1	1	1,775	1,775

CONCLUSIES EN BEDENKINGEN

Door middel van de logit-methode toonden wij aan dat de itemmoeilijkheidsschattingen onder het Rasch model vertekend worden door verschillen in discriminatieparameters en/of aanwezigheid van raden ($c_i > 0$).

Raden heeft voor gevolg dat de itemmoeilijkheid daalt en de spreiding der moeilijkheidsparameters kleiner wordt.

De discriminatiewaarden werken als gewicht voor de eigenschapsschaal.

In het licht van deze bevinding onderzochten wij of de onderstelling der unidimensionaliteit strijdig is met het toelaten van verschillen in discriminatie. Dit blijkt niet het geval op voorwaarde dat men verschillen in foutenvariantie toelaat en dat men werkt met een gewogen Euclidische ruimte.

Van de gewogen Euclidische ruimte naar het INDSCAL-algoritme als methode van onderzoek naar de gelijkheid der itemdiscriminaties is maar een kleine stap. Wij laten, in afwachting van een alternatief, de kritiek op de correlatie als maat van overeenkomst tussen data en oplossing buiten beschouwing. Het voorstel dat wij formuleren is dat INDSCAL als criterium kan gehanteerd worden voor het voldoen van data aan het twee-parameter logistisch model en het Rasch model. Zijn alle itemgewichten op de eerste dimensie en R^2 groter dan 0,9 dan is het twee-parameter logistisch model van toepassing. Indien daarenboven de standaardafwijkingen per item van b_i - d_i gelijk zijn, dan is voldaan aan het Rasch model.

Nadeel van deze procedure is wel dat een zeer groot aantal subjecten vereist is in verhouding tot het aantal items om voor de extreme scoregroepen tot betrouwbare schattingen van P_{vi} te komen.

Om misverstanden te vermijden, wijzen wij erop dat INDSCAL slechts een laatste stap is in een mogelijke selectieprocedure. De voorgaande stappen worden toegelicht in Smedts (1986).

BIBLIOGRAFIE

- Carroll, J.D. & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalisation of Eckart Young decomposition. *Psychometrika*, 35, 289-319.
- Dong, H.K. (1985). Chance baselines for INDSCAL's goodness-of-fit index. *Applied psychological measurement*, 9, 27-30.
- Gustafsson, J.E., (1981). *PML: a computer program for conditional estimation and testing in the Rasch model for dichotomous items*. Institute of Education, University of Göteborg, Mölndal.
- Lord, F.M. & Novick, M.R. (Eds.) (1968). *Statistical theories of mental test scores*. Reading, Addison-Wesley.
- Lumsden, J. (1978). Tests are perfectly reliable. *British journal of mathematical and statistical psychology*, 31, 19-26.
- Molenaar, I.W. (1981). Programmabeschrijving van PML (versie 3.1.) voor het Rasch-model. *Heymans Bulletin*, HB-81-538-RP, 188-213.
- Smedts, Diana (1986). *Het Rasch model. Onderstellingen - itemselectie - aanpassing voor raden*. Katholieke Universiteit Leuven (Doctoral dissertation).
- van de Vijver, F.J.R. (1982). *The robustness of Rasch estimates*. Department of psychology, Tilburg University.
- van de Vijver, F.J.R. (1986). The robustness of Rasch estimates. *Applied psychological measurement*, 10, 45-57.
- Young, F.W., & Lewyckij, R. (1980). *ALSCAL-4 user's guide*. University of North Carolina.

Notities en Commentaren

DE BETROUWBAARHEID VAN DRIE DISCRIMINATIE-INDICES

Henk J.M. van Berkel

Rijksuniversiteit Limburg, Vakgroep Onderwijsontwikkeling en Onderwijsresearch

ABSTRACT

In test practice many substitutes of Rit have been proposed. Studies of Oosterhof (1976) and Beuchert and Mendoza (1979) suggest that substitutes have a strong empirical resemblance with each other and with Rit. These studies compared indices within one and the same group. Item evaluating, however, is usually based on comparing values computed within at least two groups. So, one should be interested in the reliability of the indices. The simulation study reported here compares the reliabilities of three often used item-discrimination indices. It is concluded that the D-index (Johnson, 1951) has the highest reliability in almost every test situation.

INLEIDING

De item-test correlatie is een maat voor de overeenkomst tussen een enkele toetsvraag en de hele toetsscore. Deze correlatie kan ook gezien worden als een indicatie voor het discriminerend vermogen van de vraag: tot welke hoogte maakt deze toetsvraag onderscheid tussen 'masters' en 'nonmasters'? In het verleden zijn vanwege berekeningsproblemen vele vervangers van de product-momentcorrelatie voorgesteld; er waren immers nog geen handzame computers. Zie Van Berkel (1984) voor een overzicht van deze vervangers. In die tijd was de validiteitsvraag aan de orde: welke van de vervangers komt het best overeen met de Rit. De Rit werd dan gezien als het criterium. Hiernaar is veel onderzoek verricht. Engelhart (1965), Oosterhof (1976), Hales (1972) en Aleamoni & Spencer (1969) concludeerden dat verschillen tussen de vervangers klein waren en dat de overeenkomst met de Rit hoog was. Anderen (zie Beuchert & Mendoza, 1979) vonden aanzienlijke verschillen. Beuchert & Mendoza veronderstelden dat deze elkaar tegensprekende bevindingen te wijten zijn aan het gebruik van onvergelykbare vraagverzamelingen. Om de resultaten te mogen generaliseren dient een groot aantal verzamelingen in het onderzoek te worden betrokken. Een simulatiestudie ligt daarom voor de hand. Beuchert & Mendoza verrichtten een dergelijk onderzoek bij tien item-indices. De verschillen tussen de indices waren zeer klein en zij concludeerden dat "selection of an index should be based solely on ease of computation or the need for statistical tests of significance".

In de aangehaalde studies werden de correlaties tussen de verschillende indices berekend binnen een en dezelfde groep. Een toetsvraag maakt deel uit van een vragenvoorraad teneinde de vraag meer keren te gebruiken. De veronderstelling is dan dat de eigenschappen van de vraag, waaronder de moeilijkheid, min of meer constant zijn. Met andere woorden, men dient geïnteresseerd te zijn in de betrouwbaarheid van de index. Analooq aan de definitie van betrouwbaarheid als de correlatie tussen de rangordening van toetsdeelnemers op twee parallelle toetsen, kan de betrouwbaarheid van een vraagindex worden geoperationaliseerd als de correlatie van dezelfde index over de vragen in twee vergelijkbare groepen.

In deze notitie wordt verslag gedaan van een simulatiestudie die deze betrouwbaarheid berekent.

DE SIMULATIESTUDIE

In de simulatie werden de betrouwbaarheden van drie indices, die in praktijk vaak worden gehanteerd, onderzocht: 1. D33% (D), deze is gebaseerd op verschillen in p-waarden van toetsvragen in twee groepen (Johnson, 1951); 2. Rit (R), de product-momentcorrelatie tussen de vraag en de totaalscore, en 3. Rc, de Rit gecorrigeerd voor attenuatie. De vraag die beantwoord gaat worden, is: Welke index heeft de hoogste betrouwbaarheid en in welke testsituatie is dit het geval? Er worden verschillende testsituaties gesimuleerd. Aangenomen wordt dat Birnbaum's model (Lord & Novick, 1968, hfd. 17) van toepassing is. Dat wil zeggen, er is slechts een onder de toetsvragen liggende trek en de vraagkarakteristiek, die de kans aangeeft om de vraag correct te beantwoorden als functie van de trekwaarde, is een logistische kromme. De volgende invoergegevens worden gecontroleerd: het aantal alternatieven, het gemiddelde en de range van de discriminatie parameter 'a', het gemiddelde en de range van de moeilijkheidsparameter 'b', het aantal toetsdeelnemers en het aantal toetsvragen. De berekeningen worden allereerst verricht bij een zogenaamde standaardsituatie. De parameters krijgen dan die waarden die doorgaans in praktijk worden aangetroffen. Vervolgens wordt steeds een parameter systematisch veranderd en onderzocht wat de gevolgen hiervan zijn voor de betrouwbaarheid. Het aantal replicaties is twintig, dat wil zeggen, de hierna gepresenteerde resultaten zijn gemiddelde uitkomsten van twintig maal uitgevoerde simulaties. Voor meer details van het computerprogramma wordt verwezen naar Van Berkel (1984).

RESULTATEN

In tabel 1 staan de resultaten samengevat. Enkele uitkomsten zijn triviaal: de betrouwbaarheden van de indices stijgen naarmate het gemiddeld onderscheidingsvermogen toeneemt, eveneens stijgen ze wanneer het aantal toetsdeelnemers toeneemt. Andere uitkomsten kunnen ook worden verklaard. Bijvoorbeeld, het dalen van de betrouwbaarheden naarmate de toets moeilijker wordt. De raadfactor gaat dan een ongewisse rol spelen. Over het algemeen stijgen de betrouwbaarheden als de range van de discriminatie parameter toeneemt. Een verklaring hiervoor is dat de betrouwbaarheid in deze simulatiestudie een correlatie is. Als de range van het onderscheidingsvermogen van de afzonderlijke vragen stijgt, zal de spreiding van de indices, waarover de correlatie is berekend, toenemen en dus de correlatie verhogen. Eveneens is te zien dat het verminderen van het aantal alternatieven gepaard gaat met een vermindering van de betrouwbaarheden. De kansfactor om een vraag correct te beantwoorden wordt immers hoger bij het minderen van het aantal alternatieven.

CONCLUSIE

Eerder onderzoek (Beuchert & Mendoza, 1979) toonde aan dat vele vervangers van de product-momentcorrelatie, waaronder de D-index, zeer hoog positief correleren met de Rit. Hieruit kan worden geconcludeerd dat ze hetzelfde meten (hetgeen overigens ook blijkt uit de (in deze notitie niet gerapporteerde) hoge intercorrelaties tussen de indices) en daarom tot op zekere hoogte dezelfde validiteit hebben.

Bezien we de betrouwbaarheden van de drie onderzochte indices ten opzichte van elkaar – de vraag luidde immers: "Welke discriminatie-index bezit de hoogste betrouwbaarheid en in welke toetssituatie is dit het geval?" – dan is de conclusie dat in praktisch alle onderzochte toetssituaties de D-index de hoogste betrouwbaarheid bezit. Uit tabel 1 is af te lezen in welke specifieke toetssituatie de betrouwbaarheid van de D-index hoger is dan die van de Rit of de Rc.

Elders (Van Berkel, 1984) is geconcludeerd dat de D-index een zeer geschikte maat is om het discriminerend vermogen van een toetsvraag aan te duiden. De index is niet alleen inzichtelijk maar vertoont ook een hoge empirische overeenkomst met de meer geëigende maten. Uit deze

Tabel 1. Effecten van het systematisch veranderen van het gemiddelde en de range van de moeilijkheid en het onderscheidingsvermogen, het aantal toetsdeelnemers, het aantal toetsvragen en het aantal alternatieven in een 60- en 120-vragentoets.

betrouwbaarheid				betrouwbaarheid			
	D	R	Rc		D	R	Rc
MOEILIKHEID							
gemiddelde				range			
-2.5	.78	.69	.64	1	.49	.55	.56
-1.5	.68	.53	.48	2	.57	.63	.65
-0.5	.56	.45	.43	3	.62	.60	.63
0.5	.57	.58	.60	4	.60	.58	.59
1.5	.42	.52	.52	5	.64	.63	.62
2.5	.39	.43	.42	6	.56	.45	.43
ONDERSCHEIDINGSVERMOGEN							
gemiddelde				range			
0.4	.52	.49	.51	0.2	.60	.28	.19
0.6	.56	.45	.43	0.6	.61	.36	.28
0.8	.72	.58	.49	1.0	.46	.38	.37
1.0	.83	.60	.55	1.2	.56	.45	.43
				1.4	.61	.47	.46
AANTAL TOETSDEELNEMERS				AANTAL TOETSVRAGEN			
50	.44	.32	.32	20	.76	.63	.50
100	.56	.45	.43	30	.67	.50	.43
200	.75	.58	.60	40	.56	.45	.43
1000	.94	.91	.90	60	.58	.49	.48
				80	.57	.51	.52
				120	.60	.56	.56
AANTAL ALTERNATIEVEN				120-vragentoets			
60-vragentoets				3	.55	.51	.51
4	.58	.49	.48	2	.43	.37	.38
3	.54	.45	.44				

simulatiestudie blijkt nu bovendien dat de D-index ook nog betrouwbaar(der) is in tal van toetssituaties. Deze conclusie is merkwaardig. Hoe kan een "substituut" (D) van de vraag-totaal correlatie beter zijn dan het "origineel" (R)? Wordt deze conclusie veroorzaakt door de verschillende berekeningswijzen van de indices? Het lijkt er op dat de D-index minder wordt beïnvloed door eigenschappen van de toets als gemiddelde moeilijkheid en onderscheidingsvermogen, het aantal vragen, alternatieven en toetsdeelnemers. De onderhavige studie betrof echter een simulatiestudie. Een simulatiestudie kan niet in de plaats komen van empirisch onderzoek op echte gegevens. De vraag is immers of in de toetspraktijk voldaan wordt aan de assumpties met betrekking tot het in deze studie gehanteerde model.

LITERATUUR

- Aleamoni, L.M., & Spencer, R.E. (1969). A comparison of biserial discrimination, point biserial discrimination, and difficulty indices in item analysis data. *Educational and Psychological Measurement*, 29, 353-358.
- Berkel, H.J.M. van (1984). *De diagnose van toetsvragen*. Amsterdam: Universiteit van Amsterdam/COWO (academisch proefschrift).
- Beuchert, A.K., & Mendoza, J.L. (1979). A monte carlo comparison of ten item discrimination indices. *Journal of Educational Measurement*, 16, 109-117.
- Engelhart, M.D. (1965). A comparison of several item discrimination indices. *Journal of Educational Measurement*, 2, 69-76.

- Hales, L.W. (1972). Method of obtaining the index of discrimination for item selection and selected test characteristics: A comparative study. *Educational and Psychological Measurement*, 32, 929-937.
- Johnson, A.P. (1951). Notes on a suggested index of item validity: the U-L index. *Journal of Educational Psychology*, 42, 499-504.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.
- Oosterhof, A.C. (1976). Similarity of various item discrimination indices. *Journal of Educational Measurement*, 13, 145-150.

Ontvangen 16-3-1987

Boekbespreking

Brutsaert, H., *Gelijke kansen en leerlinggerichtheid in het secundair onderwijs*. Leuven/Amersfoort: ACCO, 1986.

Hoe en onder welke voorwaarden dragen structurele en pedagogisch-didactische schooleigenschappen bij tot variaties in het gedrag van leerlingen? Aldus luidt de onderzoeksvraag waarmee de Gentse socioloog Brutsaert een vergelijkende effectstudie heeft uitgevoerd in het secundair onderwijs in Nederlandstalig België. Brutsaert ziet zijn studie niet als evaluatie-onderzoek, maar als bijdrage aan het schooleffectenonderzoek. In kort bestek behandelt hij vier invalshoeken, met verschillende theoretische en onderzoekstechnische accenten: de context-analyse (referentieprocessen, geaggregeerde groepskenmerken), de sociaal-normatieve benadering (schoolklimaat, aggregatie van individuele percepties), het input-outputmodel (materiële hulpbronnen van scholen) en de institutionele analyse (globale schoolkenmerken). Op drie punten kritiseert hij het schooleffectenonderzoek: (a) het besteedt te weinig aandacht aan de 'intramuros-schoolvariabelen' die toelaten de band tussen schoolkenmerken en leerlinggedrag te interpreteren, (b) het is eenzijdig gericht op cognitieve effectmaten en (c) cross-sectioneel van opzet. De institutionele analyse acht hij het meest veelbelovend en hij situeert zijn onderzoek in deze traditie.

Sinds begin jaren zeventig is in het Vrije Onderwijsnet (bijzonder, merendeels katholiek onderwijs) een tweesporig stelsel ontstaan, waarin geïntegreerde scholen (Vernieuwd Secundair Onderwijs: VSO) bestaan naast categoriale scholen (door Brutsaert Traditioneel Secundair Onderwijs, TradSO, gedoopt); in het Rijksonderwijsnet (openbaar onderwijs) is het VSO sinds 1975 verplicht. Brutsaert heeft uit het Vrije Onderwijsnet vier vergelijkbare scholen geselecteerd, uit beide stelsels een jongens- en een meisjesschool. Hiervan is de instroompopulatie van 1979/1980 longitudinaal onderzocht, met meetmomenten in het eerste, derde en vijfde leerjaar. De effectvariabelen zijn ontleend aan de twee VSO-doelstellingen: (1) het uitschakelen van milieugeboden selectie (criterium: relatie tussen milieu en curriculumpositie in leerjaar vijf) en (2) het verbeteren van het sociaal-emotioneel functioneren van leerlingen (criteria: zes attitudevariabelen). Brutsaert voert twee typen analyse uit. Eerst gaat hij na of de factor 'schoolstelsel' effect heeft op de afhankelijke variabelen (d.m.v. covariantie-analyse) en vervolgens of de eindkenmerken van leerlingen in beide systemen op dezelfde wijze tot stand komen (d.m.v. regressie-analyses per systeem). Met het laatste beoogt hij de hypothese te toetsen, dat referentiegroepsprocessen binnen een selectief onderwijsstelsel (met een cultuur van 'sponsored mobility') een ondergeschikte en binnen een niet-selectief stelsel ('contest mobility') een cruciale rol spelen bij het beïnvloeden van leerlinggedrag.

De onderzoeksresultaten zijn weinig uitgesproken. De geringe systeemeffecten die Brutsaert vindt zijn licht in het voordeel van het VSO. De relatie tussen sociaal milieu en curriculumpositie is er vlakker dan in het TradSO, wat een gevolg is van het verdwijnen van een zelfstandig milieu-effect in het VSO. Wat de attitudes betreft is er geen systeemeffect met betrekking tot studiebetrokkenheid, houding tegenover de leerkrachten, zelfwaardering en studie-aspiraties; wel neemt de schoolbetrokkenheid in het VSO minder af dan in het TradSO en verschuift de 'locus of control' er iets minder van interne naar externe beheersing. Ook wat betreft de processen binnen de schoolsystemen zijn er enkele verschillen: in het TradSO speelt de mate van 'aanmoediging door de ouders' een rol bij diverse attitudes en heeft de SES een sterk effect op de studie-aspiraties; in het VSO daarentegen spelen de voorafgaande schoolprestaties een sterkere rol; Brutsaert acht hierdoor zijn hypothese gesteund.

Het door Brutsaert gekozen onderzoeksontwerp, een quasi-experimenteel longitudinaal design met metingen over een periode van vijf jaar, is positief te waarderen. Er wordt goed gebruik

gemaakt van de variatie in de Belgische veldsituatie - iets wat in Nederland nog veel te weinig gebeurt. Niet alleen cognitieve, maar ook sociaal-emotionele effecten zijn gemeten. Er is echter ook een aantal tekortkomingen te signaleren.

Zoals helaas usance in onderwijsonderzoek, wordt nergens een opmerking gemaakt over welke effectgrootte nu eigenlijk interessant is. Laten we veronderstellen dat het onderscheidingsvermogen van de te hanteren toetsen in balans moet zijn met het meermalen ghanterde significantieniveau (de alpha-fout) van 5%. In dat geval stellen we de beta-fout ook op 5%. Voorts is de steekproefomvang en het ontwerp bekend, zodat we de effectgrootte kunnen terugrekenen. De kennelijk aangenomen relevante effectgrootte blijkt dan een vierde standaarddeviatie te zijn. Wat is hiervan de inhoudelijke betekenis? Nemen we als voorbeeld de studie-aspiraties (gem. = 2.42, s.d. = 0.60). Een punt hoger scoren op deze schaal betekent: de keuze tussen wel of niet doorleren. Een kwart standaarddeviatie verschil in de populatie houdt in, dat één op de zeven leerlingen in het VSO een niveau hoger zou kiezen dan een vergelijkbare groep in het TradSO. We zouden ons kunnen voorstellen dat kleinere verschillen wel degelijk interessant zijn voor beleidsmakers en wetenschappers. Dat niet gekozen is voor een herhaaldemetingenanalyse is in dit verband te betreuren. De variantie is in zo'n ontwerp beter gebonden en de kans op het opsporen van zelfs kleinere verschillen in de populatie neemt aanmerkelijk toe.

De grootste effecten vindt Brutsaert bij de variabele curriculumpositie. Het VSO scoort hier aanmerkelijk beter: hogere niveaus en een minder milieuspecifieke verdeling. De variabele is echter noodzakelijkerwijze nauw verbonden met wat scholen voor curriculum-opties ('tracks') aanbieden. Dat nu verschilt sterk tussen het TradSO en het VSO. Met grote nieuwsgierigheid zien we daarom een eventuele follow-up, als de leerlingen van beide systemen in de éénsporige reguliere hogere trap tesamen komen, tegemoet. Pas dan weten we of er echt verschillen zijn.

Een derde bezwaar betreft het feit dat Brutsaert bij nader inzien scholen met een smalle instroom vergelijkt met scholen met een brede instroom. Dat verneemt de lezer bijna terloops als toelichting bij de paneluitval, die op de VSO-scholen veel groter is dan op de TradSO-scholen (61% en 60% versus 41% en 35%). Dat komt doordat de VSO-scholen in de eerste twee leerjaren ook de leerlingen bevatten die vervolgens uitstromen naar schooltypen buiten het Algemeen Secundair Onderwijs (ASO). Brutsaert vergelijkt, in Nederlandse termen, een middenschool met AVO-bovenbouw met een AVO-scholengemeenschap en rekent alle middenschoolleerlingen die na de onderbouw doorstromen naar het LBO tot het experimenteel verlies. De generaliseerbaarheid van zijn resultaten wordt daardoor in twee opzichten beperkt. Ten eerste zijn de onderzochte VSO-scholen niet meer representatief voor alle VSO-scholen, waaronder ook scholen met een brede bovenbouw (en daarmee een nog breder instroom) voorkomen (zoals de VSO-school die wordt geanalyseerd in: N. Pype, *Follow-up van een VSO-generatie met gematigd heterogene start. Persoon en Gemeenschap*. 1982, 35, 102-116). Bovendien heeft het onderzoek in feite alleen betrekking op de bestpresterende helft van de relevante leeftijdsgroep. Die beperking zelf is minder merkwaardig dan het gegeven dat Brutsaert deze nergens problematiseert. Geïntegreerd onderwijs wordt doorgaans juist ingevoerd om de positie van de zwakkere leerlingen te verbeteren en de meeste hypothesen en onderzoeksresultaten wijzen eerder in de richting van positieve effecten bij deze groep dan bij de sterkere leerlingen. Dat het VSO ook voor de laatsten gunstig lijkt te zijn is een veelbelovend resultaat, maar laat onbelicht of de effecten bij de hele leeftijdsgroep niet veel gunstiger zijn.

De 'uitgevallen' niet-ASO-leerlingen hebben de eerste twee jaar natuurlijk wel op de VSO-scholen gezeten en daar mede het onderwijsklimaat bepaald - via contextueel te analyseren mechanismen, zoals de samenstelling van de klassen. Brutsaert laat meting van enig schoolkenmerk, geaggregeerd of globaal, achterwege. Nu ware dat in dit geval ook weinig zinvol geweest, aangezien met vier scholen en twee dichotome selectiecriteria de vrijheidsgraden voor de analyse al opgebruikt zijn. Het gevolg is echter, dat VSO en TradSO beide een 'black box' blijven waarvan grotendeels ondoorzichtig is welke onderwijskundige en sociale 'treatment' leerlingen er krijgen. Het VSO wordt hier gelijkgesteld met het ontbreken van interscholair differentiatie in de eerste twee jaar secundair onderwijs. De interklassikale groepeeringsvorm binnen de school (streaming, setting, dakpanconstructies, heterogeen groeperen) blijft buiten

beschouwing, terwijl op dit punt zeker variatie bestaat binnen de VSO-sector. Brutsaert kan zo geen inzicht geven in de rol die de groeperingsvorm en de bijbehorende klascompositie spelen bij het tot stand brengen van de output van het VSO en van de verschillen met het TradSO. Dat is mager, gezien zijn eigen heldere en overtuigende betoog dat 'intra-muros-schoolvariabelen' hiervoor essentieel zijn. Alleen met een steekproef van een voldoende groot aantal scholen en meting en analyse van de relevante schoolkenmerken kan deze belofte worden ingelost.

Alhoewel opzet en uitvoering van het onderzoek volgens bovenstaande kritiek voor verbetering vatbaar zijn, is het boek in meerdere opzichten leerzaam. Het toont de meerwaarde van een theoretisch goed doordachte en een redelijk consequente onderzoeksopzet.

Roel Bosker en Peter van Laarhoven
RION, Groningen

Boekbespreking

H. Blok: 'Taal voor alledag. Feiten en meningen over het taalgebruik van lbo- en mavo-leerlingen in alledaagse situaties'.

PROEFSCHRIFT

In zijn inleidende hoofdstuk stelt Blok dat het object van de studie bestaat uit *doelstellingen* van het Nederlandse moedertaalonderwijs op scholen voor LBO en MAVO. Hij onderscheidt drie functies van het onderzoek van doelstellingen: helpen verduidelijken, de haalbaarheid onderzoeken en het antwoord geven op de vraag of de doelstellingen gehaald zijn. De vraag die in de studie van Blok centraal staat wordt geformuleerd als: 'Wat is het peil van de functionele taalvaardigheid van leerlingen aan het einde van een LBO- of MAVO-opleiding en wat kunnen we daarvan leren?' (p.13)

In hoofdstuk twee wordt de *begripsterm* functionele taalvaardigheid geanalyseerd, waarbij Blok zich laat inspireren door de signifi sche begripsanalyse van De Groot en Medendorp. Bij de context van taalvaardigheid is dan aan de orde de vraag waarnaar taalvaardigheid geacht wordt te verwijzen, door wie en wanneer? In het gewone taalgebruik is dit anders dan in het onderwijs Nederlands, en ook weer anders dan in de wetenschap. Vervolgens wordt de context van functionele taalvaardigheid begripmatig afgepaald en verhelderd. Het uiteindelijk definitievoorstel over functionele taalvaardigheid luidt: 'Als ik het over iemands functionele taalvaardigheid heb, dan heb ik het over de mate waarin iemands taalprogramma effectief is in alledaagse gebruikssituaties' (p.41).

In hoofdstuk drie worden *alledaagse taalgebruikssituaties* aan de orde gesteld. Alledaagse situaties zijn situaties waar iedereen voor kan komen te staan (p.46). Voor wat betreft de indeling naar taalgebruikssituaties kiest Blok voor een eigen uitwerking in plaats van gebruikmaking van een bestaand systeem. De eigen uitwerking gebeurde in de vorm van een vragenlijstonderzoek onder oud-leerlingen van het LBO en MAVO. Een proefonderzoek onder 33 jonge volwassenen resulteerde in 113 taalgebruikssituaties. Het hoofdonderzoek vond plaats bij een steekproef van 800 leerlingen die één tot vijf jaar tevoren het LBO of MAVO verlieten om te gaan werken. Doel was het inventariseren van de frequentie, de beheersing en het belang van de beheersing van 69 verschillende taalgebruikssituaties. De respons bedroeg 64% (p.61).

De resultaten tonen onder andere dat, van de 15 meest voorkomende situaties, er 9 betrekking hebben op mondeling (informeel) taalgebruik. Er zijn 5 lees-taalgebruikssituaties (kranten, tijdschriften, strips); de minst frequente onder deze 15 is een schrijf-taalgebruikssituatie (maken van notities voor eigen gebruik). Dit beeld wordt omgekeerd bevestigd in de 15 situaties met de grootste relatieve tekorten in beheersing, waarvan de oud-leerlingen tevens zeggen dat zij een goede beheersing wel belangrijk vinden. Bij het schrijven gaat het hierbij om meer formele situaties; bij het lezen om nauwkeurigheid van handleidingen en officiële stukken. Het blijkt dat de situaties die relatief vaak voorkomen, ook de situaties zijn die relatief goed beheerst worden. Situaties die moeilijk zijn worden mogelijk vermeden en hierdoor blijven ze moeilijk.

In hoofdstuk vier wordt het instrument voor het *meten van functionele taalvaardigheid* ontwikkeld. Hiertoe worden eerst zeven typen van taalgebruikssituaties geselecteerd: een brief schrijven aan een officiële instantie waarin je gevraagde informatie verschaft of zelf om inlichtingen vraagt, voor jezelf korte notities of aantekeningen maken, een lange/dikke handleiding lezen van een formulier dat je moet invullen, het algemene nieuws in de krant lezen, informatie opzoeken in gidsen of in naslagwerken, mondeling informatie vragen, luisteren (en kijken) naar informatieve programma's. Deze situaties werden in twee thema's (vacantie: 'Tienertoer' en werken: 'Uitzendwerk') door twee groepen docenten uitgewerkt. De

zeven taalgebruikssituaties van een thema werden in één dag afgenomen bij een groep van 12 leerlingen in buitenschoolse locaties. Bij de afname waren beschikbaar: een proefleider en twee assistenten; video- en geluidsopname-apparatuur; telefoontoestellen en -verbindingen; één centrale testzaal plus twee kleinere ruimten voor individuele afname. De twee thema's zijn afgenomen bij leerlingen uit de vierde klas van een LBO- of MAVO-school (thema Uitzendwerk $n=208$; thema Tienertoer $n=239$). De leerlingen waren afkomstig uit vijf in Nederland verspreide plaatsen. Elke leerling ontving voor de deelname aan een dag f 25,- plus vergoeding van de reiskosten.

Bij de scoring van de leerlingresultaten op de taalproeven wordt gelet op verschillende criteria zoals uitvoering (bijv. ontvanger- en afzendergegevens, kantlijn en alinea-indeling), inhoud en taalgebruik. De scoringen worden onderworpen aan Mokkenanalyse, waarbij niet-passende items uit een schaal worden weggelaten. Tenslotte bespreekt Blok de inhouds- en begripsvaliditeit, de nauwkeurigheid en de efficiëntie van de taalproeven. De conclusie (p. 115) is dat de proeven nog niet rijp zijn voor grootschalige afname, maar dat zij hiertoe wel verder ontwikkeld kunnen worden.

In hoofdstuk vijf wordt aandacht besteed aan *verwachtingen van beoordelaars* over het taalgebruik van de leerlingen. Ongeveer 50 beoordelaars (niet-representatief gekozen docenten LHNO, LTO, MAVO; moedertaaldidactici, ouders, beleidsvertegenwoordigers, en vertegenwoordigers van vervolgoopleidingen, overheidsdiensten en bedrijfsleven) hebben getracht te voorspellen hoe moeilijk vijf taalproeven voor de leerlingen zijn. De beoordelaars hebben geschat hoe groot het percentage leerlingen is dat een item juist zou kunnen beantwoorden (voorspelde score) én hebben bij het item aangegeven hoeveel procent van de leerlingen het item goed zou moeten kunnen maken (normscore).

Tussen de beoordelaars bestaan grote verschillen in wat zij verwachten aan moeilijkheidsgraad voor de leerlingen. Gemiddeld valt de voorspelde score (slechts) vier percentielpunten hoger uit dan de feitelijke score van de leerlingen. De conclusie is dat: 'De juistheid van de gegeven voorspellingen varieert per item, per taalproef en met de persoon van de beoordelaar; echter niet met de functie of signatuur van de beoordelaar noch met het door de leerlingen gevolgde schooltype.' (p. 138).

In hoofdstuk zes worden vijf taalproeven en de bijbehorende *schaalscores van leerlingen en beoordelaars* nader bekeken. De proeven betreffen het lezen van de krant, het naslaan van informatie over de arbeidsvoorwaarden bij een uitzendbureau, het invullen van een Europees Schadeformulier (twee formulier-schalen), het maken van aantekeningen tijdens een telefoongesprek, en het schrijven van een brief. De vergelijking tussen schaa scores volgens leerlingresultaat (vgl. hoofdstuk 4) en volgens voorspelde score en gewenste of normscore (vgl. hoofdstuk 5) leert dat de schaalgemiddelden van krant lezen, formulier 1 invullen, aantekeningen en envelop (proef 5) in grote trekken overeenstemmen. Opzoeken, percentage fout gespelde woorden en aantal interpunctiefouten doen de leerlingen beter dan voorspeld. Slechter dan voorspeld doen de leerlingen het op brief inhoud, formulier 2, en taalgebruik brief. Wat betreft de relatie tussen feitelijke leerlingsscore en door de beoordelaars gewenste score blijkt dat de beoordelaars in het algemeen een hogere score wensen; de interpunctie bevindt zich op het gewenste niveau, terwijl de spelling van de leerlingen beter is dan de beoordelaars wensen (en ook beter is dan zij voorspellen). Tussen de schooltypen (lto, lhno en mavo) bestaan geen grote verschillen.

In de *discussie* in hoofdstuk zeven worden de resultaten nog eens op een rij gezet en de in de studie optredende beperkingen vermeld (geen volledige operationalisatie van functionele taalvaardigheid, geen steekproef van leerlingen, geen steekproef van beoordelaars). Vervolgens wordt aandacht besteed aan de relevantie van het onderzoek en de resultaten ervan, en aan de consequenties voor onderwijspraktijk en volgend onderzoek.

Het proefschrift van Blok maakt een goede indruk. De verschillende aspecten van het onderzoek worden helder beschreven en verantwoord, waarbij ook de te maken keuzen en de beperkingen aan de orde gesteld worden. Door de concentratie op functionele taalvaardigheid

waagt Blok zich aan een niet eenvoudig maar wel uitermate relevant onderwerp, en hij doet dit op een creatieve wijze. Zijn aanpak en zijn resultaten zullen de discussie over en hopelijk ook de praktijk van het onderwijs in het vak Nederlands beïnvloeden. In het bijzonder actualiseert Blok de vraag *waarop* de alledaagse onderwijspraktijk in het vak Nederlands gericht dient te zijn, waarbij tevens de vraag rijst naar de middelen waarmee de gekozen doelstellingen het best gerealiseerd kunnen worden. Waar het bestaande onderwijs veelal de indruk wekt dat de doelstellingen van het taalonderwijs *in* het onderwijs zelf liggen, poneert Blok het belang van de buitenschoolse taalgebruikssituaties. In feite worden de bestaande doelstellingen daarmee tot middelen om de beheersing van deze buitenschoolse situaties te beïnvloeden.

Ik heb enkele kritische kanttekeningen bij het gebruik van de termen 'peilingsonderzoek' en 'leerprestaties' in het proefschrift. Blok stelt dat het in peilingsonderzoek gaat om de leerprestaties van leerlingen én om het onderwijs dat de leerlingen krijgen: 'Het gaat dus uitdrukkelijk om twee zaken: de leerprestaties en het onderwijsaanbod.' (p.9). In zijn onderzoek gaat Blok wel de prestaties van leerlingen na, maar hij 'vergeet' dat het onderwijsaanbod óók essentieel is (zie bijv. p. 111, 112, 190). Op p. 143 stelt hij: 'Over de lessen Nederlands die de leerlingen gevolgd hebben, is weinig bekend. Evenmin is zeker dat eventuele verschillen in gemiddelde prestaties te wijten zijn aan verschillen in het gevolgde onderwijs.'. Blok onderzoekt dus slechts (veronderstelde) leerprestaties en niet het onderwijsaanbod. Omdat hij het onderwijsaanbod niet betrokken heeft in zijn onderzoek, verricht hij volgens zijn eigen definitie dus géén peilingsonderzoek. Het is dan onjuist én prematuur om te pleiten voor breder opgezet of grootschaliger peilingsonderzoek (p. 190). De prioriteit zou dienen te liggen bij onderzoek van zowel het onderwijsaanbod als de leerprestaties.

Wat betref de leerprestaties trekt Blok niet de consequentie uit de door hem op p. 143 (zie boven) geconstateerde onhelderheid over de mate waarin de prestaties van de leerlingen op de taalproeven het effect zijn van het gevolgde onderwijs. Het is nu bijvoorbeeld niet duidelijk óf en zo ja in welke mate de door leerlingen geleverde prestaties op de ontwikkelde taalproeven beïnvloedbaar zijn met behulp van een bepaald onderwijsaanbod. Zolang dit onzeker is, is het onjuist om te spreken over leerprestaties: het gaat om prestaties van leerlingen op bepaalde taalproeven. Onderzoek naar het onderwijsaanbod en (longitudinaal) onderzoek naar effecten van dit onderwijsaanbod op leerlingen kan duidelijk maken of leerlingprestaties óók leerprestaties zijn. Het is in principe natuurlijk mogelijk om op basis van prestaties van leerlingen te concluderen dat het onderwijsaanbod op een bepaalde wijze ingericht zou moeten worden (hoewel Blok deze stap niet uitvoert met betrekking tot de leerlingprestaties op zijn taalproeven). Het blijft evenwel nodig de relatie tussen onderwijsaanbod én leerlingprestaties in onderzoek te betrekken voor het verkrijgen van afdoende inzicht in het optreden van leerprestaties.

Tenslotte nog een kleine verschrijving op p. 134: in het midden van de pagina wordt onder a het algemene niveau geoperationaliseerd als de absolute waarde van de discrepantie (f-v); dit moet zijn de relatieve waarde van de discrepantie (f-v).

Ton Mooij
Katholieke Universiteit Nijmegen
Instituut voor Toegepaste Sociale wetenschappen
Graafseweg 274
6532 ZV Nijmegen

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie

Kees Beintema (RUL), Peter van den Dool (SCO), Dirk de Jong (LICOR), Ewoud Roede (SCO) en Peter van der Sijde (UT, TO, secretaris).

Redactie-adres

Peter van der Sijde, Toegepaste Onderwijskunde, OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 11, nr. 6

December 1987

Inhoudsopgave

De ORD in het buitenland	376
Symposia	380
De kwaliteit van de arbeid van assistenten-in-opleiding <i>Ellen van Bommen en Ellen Hasselerharm</i>	381
Proefschriften	384
Nieuwe Publikaties	387

De ORD in het buitenland!

Het is voor de eerste maal dat de Vereniging voor Onderwijsresearch buiten de Nederlandse grenzen treedt om de jaarlijkse Onderwijsresearchdagen te laten plaats hebben. De K.U. Leuven vindt het een eer deze dagen te organiseren. Van bij de aanvang van de besprekingen werd benadrukt dat we aandacht moesten besteden aan de 'lokale kleur'. De stad Leuven en de eeuwenoude traditie van onze Alma Mater bieden daartoe heel wat mogelijkheden.

Maar ook kleine accenten in het programma geven kleur aan de 'Leuvense' ORD. Er is de kortere openingssessie, er zijn symposia met medewerking van de 'invited speakers' en er is het ORD-diner op maandagavond.

De ORD bieden nog steeds een unieke gelegenheid om een verkenning te maken van een aantal onderwijskundige problemen. De organisatoren hebben gemeend op de vier thema's die door de V.O.R. zelf worden voorgesteld, te moeten ingaan. Wel werd ervoor gezorgd om vanuit de organiserende universiteit duidelijke accenten te leggen. Dit vindt men terug in de wijze waarop de thema's worden omschreven. De twee open thema's zijn gekozen vanuit de onderzoeksbelangstelling die in het Departement Pedagogische Wetenschappen van de K.U. Leuven momenteel aanwezig is.

Op een ogenblik waar het beleid geneigd is de geldstroom voor onderzoek binnen de Humane Wetenschappen in te perken, is het belangrijk goed onderzoek binnen eigen kring naar voren te brengen, maar ook op een breder maatschappelijk forum uitdrukkelijk aanwezig te zijn.

We hopen dan ook dat velen zullen ingaan op de uitnodiging deel te nemen aan de ORD '88.

R. Vandenbergh

THEMA'S OP DE ONDERWIJSRESEARCHDAGEN 1988

1. Onderwijsleerprocessen

Coördinatoren: E. De Corte (K.U. Leuven) en P. Span (R.U. Utrecht)

Het thema valt uiteen in twee gedeelten: een symposium over het verwerven van algemene cognitieve vaardigheden en een serie paperpresentaties over leren denken binnen vakdomeinen.

Het symposium over algemene cognitieve vaardigheden wordt georganiseerd rond de aanwezigheid op de ORD van Gavriel Salomon (University of Arizona, Tucson). In het subthema leren denken binnen vakgebieden zal aandacht besteed worden aan de rol van voorkennis, het belang van meer denkvaardigheden en het effect van diverse interventie-strategieën.

2. Schoolorganisatie, beleid en onderwijskwaliteit

Coördinatoren: J. Verhoeven (K.U. Leuven) en J. Scheerens (Univ. Twente)

In dit thema worden twee subthema's onderscheiden: instellingsbeleid in relatie tot het regionale en het nationale onderwijsbeleid en schoolorganisatorische condities voor effectief onderwijs.

Aandachtspunten in het eerste subthema zijn:

- onderzoek over de betekenis van de nationale en regionale beleidsorganen voor het instellingsbeleid;
- onderzoek over de structuur en het functioneren van het instellingsbeleid, vooral met betrekking tot het onderwijskundig beleid, het curriculum, het personeelsbeleid en het materiële en financiële beleid.

Bij het tweede subthema dienen de papers betrekking te hebben op de relaties tussen

schoolorganisatiekenmerken, onderwijsprocessen en -effecten. Hierbij gaat het om onderzoek waarin de organisatie en vormgevingskenmerken van het onderwijs in verband worden gebracht met outputindicatoren. Ook onderzoek waarin alleen gegevens verzameld zijn over organisatorische condities en vormgevingskenmerken van het onderwijs en waarbij een gefundeerde aanname bestaat dat deze vormkenmerken indicatief zijn voor doelmatig onderwijs, wordt aanvaardbaar geacht.

3. Methodologie/technologie

Coördinatoren: J. Lowyck (K.U. Leuven) en L.J.Th. van der Kamp (R.U. Leiden)

Methodologie

Een blijvend knelpunt in de totstandkoming van waardevolle uitkomsten van empirisch onderzoek is de deugdelijkheid van de gehanteerde methodologie. Dit wezenlijk aspect van het wetenschappelijk bedrijf komt dan ook recentelijk vaak aan de orde. Voorgesteld wordt om bijdragen aan te trekken die specifiek op methodologische kwesties zijn gericht, ook al kunnen hierbij de gevonden uitkomsten mede aan bod komen. Gedacht is o.m. aan de problematiek van het longitudinaal onderzoek, van de kwaliteit en functie van meta-analyses, van het z.g. multi-level onderzoek, e.d.m.

Technologie

Geopteerd is om deze thematiek breed open te stellen voor alle onderzoeksbijdragen die media m.i.v. computers in het onderwijs bestuderen. Mogelijke onderwerpen zijn: de functie van media in het afstandsonderwijs, ontwikkeling en evaluatie van programmatuur, toepassingsmogelijkheden van informatietechnologie in onderwijs, opleiding en vorming, rentabiliteit van de diverse media ten aanzien van economische en pedagogisch-didactische criteria, alsmede kritische reflectie op grond van empirisch vastgestelde resultaten van media-onderzoek.

4. Jongeren doorheen het onderwijs en op de arbeidsmarkt

Coördinatoren: J. Van Damme (K.U. Leuven) en J. Dronkers (K.U. Brabant)

Alle bijdragen i.v.m. school- en beroepsloopbaanonderzoek (m.i.v. onderzoek betreffende het studie- en/of beroepskeuzeproces) kunnen ingediend worden.

Bij de selectie zal een zekere prioriteit gegeven worden aan volgende subthema's:

- de overgang van school naar werk en in het bijzonder de samenhang tussen de schoolloopbaan en de beroepsloopbaan;
- het effect van schoolkenmerken op de schoolloopbaan en op de loopbaan in het hoger onderwijs en/of in het beroep;
- de betekenis van het aanvankelijk basisonderwijs voor de verdere loopbaan;
- het studie- en/of beroepskeuzeproces bij jongeren en de begeleiding van dit proces.

5. Leraar als beroep

Coördinatoren: G. Tistaert (K.U. Leuven), S. Janssens (K.U. Leuven) en J. Peters (R.U. Groningen)

Dit thema richt zich op de problematiek van het beroep van leraar. Volgende subthema's kunnen hierbij in aanmerking komen:

- de persoon van de leraar: wie wordt leraar, welke motieven liggen aan de grondslag van de beroepskeuze, welke arbeidssatisfactie, c.q. dissatisfactie ervaren zij?
- de rol van de opleiding in het ontwikkelingsverloop bij leraar worden. Gevraagd wordt naar papers die bijdragen tot inzichten over bekwaamheidsverwervingsprocessen, opleidingsmethoden en opleidingsdidactiek;
- problematiek van beginnende leraren en aanvangsbegeleiding;
- problematiek van nascholing en begeleiding van leraren, zowel op het vlak van de vakinhoudelijke als de algemene didactische bekwaamheidsontplooiing en evolutie.

6. Onderwijsvernieuwing

Coördinatoren: R. Vandenbergh (K.U. Leuven) en R. van der Vegt (K.U. Nijmegen)

In het thema onderwijsvernieuwing worden twee subthema's aan de orde gesteld. Enerzijds wil men aandacht besteden aan 'de school-als-arena voor vernieuwing'. Centrale vragen zijn: hoe organiseert een school zich n.a.v. deelname aan een vernieuwingsproject, welke invloed hebben allerlei vormen van interne begeleiding, welke rol speelt de schoolleider enz... In dit verband wordt een symposium georganiseerd in samenwerking met Michael Fullan (O.I.S.E., Toronto). Anderzijds verwacht men papers in verband met onderzoek naar netwerkvorming tussen scholen onderling en tussen en allerlei varianten van externe begeleiding.

WETENSCHAPPELIJK PROGRAMMA

Het wetenschappelijk programma is de kern van de Onderwijsresearchdagen. De nadruk ligt op het presenteren van papers waarin verslag van recent empirisch onderzoek wordt gedaan. Er zijn drie verschillende mogelijkheden voor het presenteren van papers:

Themapapers

Onderzoekers die een paper willen presenteren over empirisch onderzoek dat binnen een van de zes thema's valt kunnen zich voor een thema aanmelden. Het verdient aanbeveling voor het indienen van de paper contact op te nemen met een van de betreffende themacoördinatoren.

Symposia

Onderzoekers kunnen zich als groep presenteren door papers rond eenzelfde onderwerp. Doorgaans omvat een symposium 3 tot 5 papers, die inhoudelijk verwant zijn. Voorstellen voor symposia moeten door een symposiumvoorzitter worden ingediend.

Vrije papers

Onderwerpen voor vrije papers kunnen in principe het hele gebied van het onderwijs bestrijken dat buiten de zes thema's valt. Ook voor de vrije papers geldt dat zij over empirisch onderzoek dienen te gaan.

INDIENEN VAN PAPERS

Onderzoekers die een paper willen presenteren op de Onderwijsresearchdagen 1988 worden uitgenodigd een voorstel hiertoe te sturen naar het congressecretariaat.

Alle voorstellen (themapapers, symposia, vrije papers) moeten *vóór 20 december 1987* toekomen op het volgend adres:

ORD '88

Pedagogisch Instituut K.U. Leuven

Vesaliusstraat 2

B-3000 Leuven

Ontvankelijk zijn alleen voorstellen

- die geschreven zijn volgens de richtlijnen voor papervoorstellen
- die gericht zijn aan het congressecretariaat te Leuven,
- die ontvangen zijn *vóór 20 december 1987*.

Alle inzenders krijgen *vóór 1 maart 1988* bericht of hun voorstel geaccepteerd is.

RICHTLIJNEN VOOR PAPERVOORSTELLEN

Voorstellen voor themapapers, vrije papers en symposia moeten in 5-voud worden ingediend bij het congressecretariaat.

Een voorstel mag niet meer beslaan dan 3 pagina's tekst op A-4 formaat, getypt op regelafstand 1 1/2.

Ieder *papervoorstel* bevat achtereenvolgens minimaal

- titel van het paper, echter *niet* de auteursnaam in verband met anonieme beoordeling van het voorstel,
- samenvatting van *maximaal 100 woorden* voor het programmaboek,
- beschrijving van het theoretisch kader van het onderzoek,
- beschrijving van de onderzoeksofzet en de gebruikte methoden en technieken,
- overzicht van resultaten en conclusies.

Elk van de 5 kopieën moet voorzien zijn van een afzonderlijk titelblad, waarop vermeld zijn:

- nogmaals de titel van het paper,
- naam, volledig adres en telefoonnummer van de auteur(s),
- instelling waar men werkzaam is.

Een voorstel voor een *symposium* bevat naast de afzonderlijke papervoorstellen, die aan bovengenoemde richtlijnen moeten voldoen:

- titel van het symposium, echter *niet* de namen van de auteurs in verband met de anonieme beoordeling van het voorstel,
- samenvatting van *maximaal 100 woorden* voor het programmaboek.

Elk van de 5 kopieën moet voorzien zijn van een afzonderlijk titelblad, waarop vermeld zijn:

- titel van het symposium,
- naam, volledig adres en telefoonnummer van de symposiumvoorzitter,
- opsomming van papers met namen van de auteurs.

Paper- en symposiumvoorstellen worden voorgelegd aan leden van de beoordelingscommissie. De beoordelaars krijgen niet de beschikking over de namen van de auteurs. De acceptatie van voorstellen en de samenstelling van het programma zijn geheel voorbehouden aan de Organisatiecommissie ORD 1988.

V.O.R.-PRIJS 1988

De Vereniging voor Onderwijsresearch heeft in 1984 een jaarlijks toe te kennen aanmoedigingsprijs ingesteld. Het doel van de prijs is de bevordering van de presentatie van onderzoek van hoge kwaliteit door jonge onderzoekers. De prijs bestaat uit een orkonde en een bedrag van fl. 1000,- te besteden aan verdere professionalisering, terwijl tevens publikatie van het bekroonde paper in het Tijdschrift voor Onderwijsresearch bevorderd wordt. De prijswinnaar wordt traditiegetrouw tijdens het openingsprogramma van de ORD bekend gemaakt.

Indiening van papers staat open voor alle onderwijsonderzoekers die tijdens de ORD '88 niet ouder dan 30 jaar zijn. Bij indiening van gezamenlijke papers dienen alle auteurs aan deze voorwaarde te voldoen.

Deelnemers aan de ORD '88 die voor deze prijs in aanmerking willen komen dienen in te zenden:

- vijf exemplaren van het *volledig uitgeschreven paper* dat geheel identiek is aan het te lezen paper op de ORD '88; deze exemplaren behoren ontdaan te zijn van de mogelijkheid tot identificatie van de auteur(s);
- bijgesloten een afzonderlijke gesloten enveloppe met daarin een opgave van de naam van de auteur(s), de titel van het paper en een kort curriculum vitae waaruit blijkt dat aan bovengenoemde voorwaarde is voldaan.

Alle inzendingen worden anoniem beoordeeld door een jury, bestaande uit: B.P.M.

Creemers, L.W.F. de Klerk, J.M.G. Leune, W.J. van der Linden, voorzitter.

Inzendingen voor de V.O.R.-prijs worden gericht aan: W.J. van der Linden, Faculteit der Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.

Alle inzendingen dienen uiterlijk vrijdag 13 mei 1988 door de voorzitter te zijn ontvangen.

Symposia

Leren en instructie in arbeidsorganisaties: 11 december

Plaats: Nijmegen, Collegezalencomplex

Tijd: 13.00 - 17.00 uur

Programma:

dr. P.R.J. Simons (Katholieke Universiteit Brabant)

Onderzoek van opleidingsleerprocessen

drs. J. Vermunt en drs. B. Corstiaans (Katholieke Universiteit Brabant)

De verhouding tussen extern en intern gestuurd leren: Implicaties voor onderzoek van bedrijfsopleidingen

drs. J. Thijssen (RABO-Bank Nederland)

Het leren van de verouderende werknemer

dr. J.M. Pieters, drs. O. Jelsma en drs. J.P. Bijlstra (Universiteit Twente)

Het trainen van operator-vaardigheden

drs. E. Harenberg (Mediavision, Laren)

Interactieve video

W. Bulthuis (Philips Telecommunicatie & Data Systemen, SSS-On-Line Assistance Group, Apeldoorn)

Computers op de werkplek: hoe leren we ze te gebruiken?

Nadere inlichtingen:

Secretariaat Onderzoeksthema-groep Onderwijsleerprocessen, dr. J.M. Pieters, Universiteit Twente, Faculteit der Toegepaste Onderwijskunde, Postbus 217, 7500 AE Enschede, tel. 053-893594

De kwaliteit van de arbeid van assistenten-in-opleiding

Ellen van Bommel en Ellen Hasselerharm

INLEIDING

In het artikel 'Over de kwaliteit van de arbeid in de wetenschap' in het VOR-bulletin van October 1987 geeft Van den Berg van Saporoa aan dat het begrip 'de kwaliteit van de arbeid' ook bruikbaar is bij het bekijken van onderzoekswerk. Aan het begrip worden twee aspecten onderscheiden. Enerzijds de subjectieve kwaliteit van het werkend leven, zoals de werkende die ervaart. Anderzijds kan het begrip beschouwd worden als een ontwerpprincipe, dat gebruikt kan worden bij het ontwerpen van arbeidsplaatsen/functies. Deze twee definities vullen elkaar aan. De mate waarin de organisatie er in is geslaagd te voldoen aan de eisen die het werk aan de werkenden stelt en aan de eisen en behoeften van de werkenden (= ontwerpprincipe) is bepalend voor de wijze waarop de werkende zijn werk ervaart (= subjectieve kwaliteit).

Indien er om welke redenen dan ook verandering optreedt in de werksituatie van onderzoekers heeft dit ongetwijfeld ook gevolgen voor de kwaliteit van de arbeid. De universiteit zal de inrichting van de organisatie en het beleid moeten afstemmen op die verandering zodat de kwaliteit van de arbeid tenminste wordt gehandhaafd. Hierbij gaat het om de definitie van de kwaliteit van de arbeid als een ontwerpprincipe omdat ervan uitgegaan wordt dat de kwaliteit bewust maakbaar is.

Een van de veranderingen die binnen de universiteit doorgevoerd is, betreft de invoering van de Rechtspositieregeling Assistenten-in-Opleiding (kortweg AIO-stelsel). Dit betekende de introductie van een nieuwe generatie onderzoekers. Hierop aansluitend staat in dit artikel de volgende algemene vraag centraal: 'Betekent de invoering van het AIO-stelsel een verandering in de kwaliteit van het onderzoekswerk?'

Voor de beantwoording van deze vraag wordt gebruik gemaakt van de resultaten van een onlangs door het landelijk AIO-overleg gehouden enquête onder 520 AIO's.¹⁾ Om de positie van de AIO beter te kunnen begrijpen is enige achtergrondinformatie nodig.

2. HET AIO-STELSEL

Het AIO-stelsel vloeit voort uit de instelling van de tweefasenstructuur. De tweede fase is voor zover het de onderzoeksopleiding betreft gereld in de Rechtspositieregeling Assistenten-in-Opleiding. De AIO wordt aangesteld in tijdelijke dienst ter verdere wetenschappelijke opleiding en vorming. Dit laatste is essentieel omdat een goede opleiding en begeleiding er zorg voor moet dragen dat de AIO zich ontwikkelt tot een gekwalificeerd wetenschappelijk onderzoeker. De aanstelling kan verder resulteren in een proefschrift (in welk geval de aanstelling ten hoogste vier jaar bedraagt) of een technologisch proefontwerp (aanstellingsduur ten hoogste twee jaar).

Daarnaast vormt de opleiding en begeleiding van de AIO een legitimering van de aparte regeling voor de salariëring van de AIO. De AIO ontvangt voor het gedeelte van de aanstelling dat hij in verband met zijn opleiding en begeleiding wordt geacht geen productieve arbeid te verrichten, ook geen salaris. In het eerste aanstellingsjaar is hiervoor een percentage vastgesteld van 45%; in het tweede jaar van 40%; in het derde jaar van 30% en in het laatste jaar van 15%.

¹⁾ Rapport 'Een jaar AIO-stelsel, de Rechtspositieregeling Assistenten-in-Opleiding in de praktijk.

De rechtspositieregeling behandelt een aantal zaken betreffende de inhoud en omvang van de taakstelling, de werving en selectie van de assistenten-in-opleiding, de aanstellingsvoorwaarden, de salarisopbouw, de beoordeling en de door het College van Bestuur vast te stellen voorschriften. Verder, dat bij aanstelling van iedere AIO een opleidings- en begeleidingsplan uitgereikt moet worden. In dit plan dient opgenomen te zijn: welke kennis en vaardigheden verworven dienen te worden en op welke wijze dit dient plaats te vinden; wie de toezichthouder is, onder wiens toezicht de AIO werkzaam is; en de omvang, in uren per maand, van door de toezichthouder te geven persoonlijke begeleiding waarop de assistent ten minste recht heeft. Het AIO-stelsel trad op 1 september 1986 in werking.

3. DE AIO'S EN DE KWALITEIT VAN HUN ARBEID

De enquête zoals die onder AIO's is gehouden, is uitgevoerd om antwoord te krijgen op de volgende twee vragen:

- Is er een discrepantie tussen de beoogde regelingen en voorwaarden, zoals neergelegd in de Rechtspositieregeling en de daarbij behorende toelichting en de feitelijke vertaling hiervan in de universitaire praktijk? Zo ja, wat is de aard van de discrepantie?
- Welke knelpunten doen zich voor de assistent-in-opleiding voor bij het uitvoeren van hun taken als wetenschappelijk onderzoeker en promovendus? Hoe tevreden zijn de assistenten-in-opleiding met hun aanstelling en de faciliteiten die daar bij horen en welke factoren zijn daarop van invloed?

Deze twee vragen kunnen beschouwd worden als Subvragen van de in de inleiding gestelde vraag: 'Betekent de invoering van het AIO-stelsel een verandering in de kwaliteit van het onderzoekswerk?'

De hierboven genoemde subvragen komen overeen met de twee door Van den Berg van Saporoa onderscheiden aspecten van het begrip 'kwaliteit van de arbeid'.

De eerste vraag duidt op de definiëring van de kwaliteit van de arbeid als een ontwerp-principe. De tweede vraag duidt op het begrip als de subjectieve kwaliteit van het werkend leven.

Voor de beantwoording van de eerste vraagstelling zijn de volgende resultaten van de enquête van belang.

De gemiddelde leeftijd van de assistent is 27 jaar en het merendeel van de assistenten heeft de oude-stijl-doctoraal-studie gevolgd. 18% van de assistenten had bij indiensttreding relevante werkervaring (was bijvoorbeeld al eerder wetenschappelijk medewerker).

De belangrijkste gegevens omtrent de aard van de opleiding tonen, dat eerste fase onderwijs, post-academisch onderwijs en individueel 2e fase onderwijs zonder begeleiding de belangrijkste onderwijsvormen zijn.

Dat betekent dat de opleiding van de AIO vooral plaats vindt binnen bestaande kaders. Zeker daar waar sprake is van eerste fase onderwijs kan dat nooit de bedoeling zijn geweest van de Rechtspositieregeling. Speciaal onderwijs opgezet voor assistenten (anders dan het leren van algemene vaardigheden) komt niet of nauwelijks van de grond. Dit betekent dat twijfels geuit kunnen worden over de kwaliteit en relevantie van de opleidingscomponent, zeker in het licht van de doelstelling van het a.i.o.-schap, het opleiden tot hoogwaardig gekwalificeerde onderzoekers. Wel in overeenstemming met de Rechtspositieregeling is het volgen van post-doctorale cursussen, seminars e.d., maar opgemerkt dient te worden dat slechts de helft van de respondenten aangeeft aan deze onderwijsvoorzieningen deel te nemen.

De resultaten met betrekking tot de aanwezigheid van een opleidings- en begeleidingsplan geven aan dat 17% niet over een opleidings- en begeleidingsplan beschikt. De overige 83% geeft aan wel over een opleidings- en begeleidingsplan te beschikken, maar typeert deze plannen als (zeer) globaal. Dit betekent dat over een aantal belangrijke zaken geen afspraken worden gemaakt. Opmerkelijk is bijvoorbeeld dat 37% geen afspraak maakt over de te ontvangen begeleiding en 36% geen afspraak over het te volgen onderwijs.

De tijd die de AIO besteedt aan het volgen van onderwijs blijkt in de praktijk rond de 20% van

de totale werktijd te liggen. (De Rechtspositieregeling geeft voor het eerste jaar een percentage van 45). Belangrijkste verklaring voor dit gegeven is dat de universiteiten een percentage hoger dan 20 in strijd achten met het uiteindelijke doel van het AIO-schap, het schrijven van een proefschrift.

Op grond van het voorgaande kan geconcludeerd worden dat de universiteiten ten aanzien van het AIO-stelsel hun beleid onvoldoende afgestemd hebben op de veranderingen die het AIO-stelsel met zich meebrengt. De hierdoor ontstane discrepantie tussen beoogde regelingen en de universitaire praktijk leidt tot ontevredenheid bij de AIO's over de kwaliteit van het onderzoekswerk, zoals die subjectief wordt ervaren. Dit komt tot uitdrukking in de beantwoording van de tweede vraagstelling die betrekking heeft op de taakbeleving van de AIO.

De resultaten die voor de beantwoording van de tweede vraagstelling van belang zijn volgen hieronder.

Het salaris wordt als het grootste knelpunt ervaren door de assistenten. 90% geeft aan veel tot zeer veel moeite te hebben met het salaris. Dit gegeven kan ons inziens verklaard worden uit een drietal factoren.

Ten eerste speelt de aard van het werk een belangrijke rol. Promoveren is een intensieve, veel vergende en complexe taak. De aard van de taak staat in schril contrast met de beloning die daar tegenover staat.

Een tweede factor kan gevonden worden in de sfeer van de opleiding en begeleiding. Assistenten worden, voor het gedeelte van hun werkzaamheden dat zij in verband met opleiding en begeleiding geen produktieve werkzaamheden verrichten, gekort op hun salaris. Veel assistenten ervaren die korting als onrechtvaardig, zeker wanneer men de relevantie en mate van opleiding tegen het licht houdt van deze korting. Het is voor assistenten een wrang gegeven óf heel weinig opleiding te verkrijgen óf seminars te volgen, waarbij de deelname van de assistent als opleiding wordt beschouwd en de deelname van collega's aan hetzelfde seminar als (betaalde) bijscholing.

Een derde factor vormen de kosten die verbonden zijn aan de opleiding en aan de aard van het werk. Bijna geen enkele assistent krijgt een vergoeding voor studiemateriaal en boeken, terwijl de financiële positie van de vakgroep een belangrijke factor voor de vraag of bijvoorbeeld zaken als congresbezoek vergoed worden. Daarbij komen op langere termijn de kosten van een promotie. Duidelijk is, dat deze zaken voor assistenten een belangrijke rol spelen als zij hun salaris in ogenschouw nemen.

Knelpunten zitten niet in de inhoudelijke kanten van het werk. Positief tot zeer positief laat de AIO zich uit over de aantrekkelijkheid van het werk. De aard, inhoud en autonomie van het werk en het werken in een (intellectueel stimulerende) omgeving zijn hiervoor belangrijke overwegingen. Ondanks de positieve waardering met betrekking tot de aantrekkelijkheid van het werk twijfelt 47% van de respondenten eraan of men de juiste baan gekozen heeft. 21% zegt een andere baan te nemen indien de gelegenheid zich voordoet. Een belangrijke reden voor de twijfel is het onzekere toekomstperspectief van een gepromoveerde onderzoeker. Men vermoedt dat er aan het einde van de rit meer gepromoveerden zijn dan beschikbare posities op de universiteiten.

Op grond van de twijfel die de AIO heeft over zijn toekomstperspectief, de onvoldoende invulling van de Rechtspositieregeling en het lage salaris, concluderen wij dat de kwaliteit van het onderzoekswerk met de invoering van het AIO-stelsel verminderd is.

LITERATUUR

- Groenendijk, N. en Witzijs, B. (1987). Een jaar AIO-stelsel, de rechtspositieregeling Assistent-in-Opleiding in de praktijk. Enschede: Landelijk Overleg Assistenten in Opleiding.

Proefschriften

H. Kelderman: Quasi-Loglinear Models for Test and Item Analysis.

Promotores: Prof. Dr. W.J. van der Linden en Prof. Dr. G.J. Mellenbergh

Promotie: Universiteit Twente, 12 juni 1987

KORTE SAMENVATTING

Dit proefschrift handelt over test-en itemanalyse met behulp van loglineaire modellen. Met loglineaire modellen kan een soort variantieanalyse voor discrete gegevens worden uitgevoerd. In dit proefschrift wordt aangetoond dat deze loglineaire modellen ook als itemresponse modellen kunnen worden gebruikt. Itemresponsmodellen verklaren de kans op een juist antwoord op een bepaald test item op grond van een aantal parameters die specifiek zijn voor het item en de persoon.

Het voordeel van dit resultaat is nu dat normale discrete variabelen zoals sexe, en de antwoorden op testitems in één model kunnen worden gemodelleerd en geanalyseerd. Daarnaast komen er allerlei nieuwe mogelijkheden beschikbaar om na te gaan of het itemresponse model, in dit geval het Rasch model, bij de testitems past.

Met de gebruikelijke methoden voor de schatting van de parameters van de loglineaire modellen konden slechts een gering aantal variabelen worden geanalyseerd. Om loglineaire modellen ook bruikbaar te maken voor de meestal grotere aantallen testitems is een speciaal algoritme ontwikkeld, waarvoor het computerprogramma *GELORA* (*G*ENERAL *L*OGlinear *R*ASch modeling) is geschreven. Met de huidige versie van *GELORA* is het mogelijk maximum likelihood schattingen van de parameters van loglineaire modellen tot 60 variabelen te berekenen.

Het model is toegepast op twee problemen uit de praktijk van toetsconstrutie: het opsporen van itemonzuiverheid (item bias) en het equivaleren van verschillende vormen van dezelfde toets (test equating). Item bias wordt onderzocht door statisch te toetsen of één of meer items voor sommige groepen leerlingen, bijvoorbeeld meisjes of ethnische minderheden moeilijker zijn dan voor anderen. Item bias bleek goed op te sporen met het model. Bovendien kon worden nagegaan of verschillende items om dezelfde reden biased zijn.

Voor het equating probleem werd een model geformuleerd waarin de itemparameters voor de verschillende deelverzamelingen items, die afgenomen zijn bij verschillende groepen personen, op eenzelfde schaal geplaatst. Op deze wijze worden de item moeilijkheden en de persoonsvaardigheden over de subgroepen vergelijkbaar, zodat normhandhaving over verschillende examens mogelijk is.

Cees Terlouw: De FUNDES-procedure in onderwijsontwikkeling. Evaluatie van een procedure van onderwijsontwikkeling voor het leren probleemoplossen (Onderwijskundig Centrum Universiteit Twente, 1987).

Promotores: Prof. Dr. Tj. Plomp, Prof. Dr. J.A.M. Carpay en Prof. Mr. D.W.P. Ruiter.

Promotie: Universiteit Twente, 19 juni 1987.

KORTE SAMENVATTING

Het proefschrift heeft betrekking op de generalisatie van de FUNDES-procedure, een procedure van onderwijsontwikkeling voor het leren probleemoplossen (FUNDES is een samentrekking van FUNctional DESign). Eerder leidde de toepassing van de procedure tot effectief onderwijs in natuurwetenschappelijke vakken in het hoger onderwijs. Het doel is een bijdrage te leveren aan het vergroten van de reikwijdte van de FUNDES. Daartoe is de procedure toegepast op sociaalwetenschappelijke vakinhouden (Bestuurskunde) en een daarbij horend probleemtype (ontwerpen van beleid). Het centrale evaluatie-criterium was de bruikbaarheid van de FUNDES en betrof de volgende aspecten: (1) aanpassingen van de FUNDES mogen de uitgangspunten van de procedure niet aantasten; (2) FUNDES moet een adequate handleiding zijn voor de professionele onderwijsontwikkelaar; en (3) FUNDES moet een effectief onderwijsprogramma opleveren. Deze aspecten zijn in het proefschrift nader geoperationaliseerd.

De eerste twee aspecten werden nagegaan met behulp van methoden en technieken voor een case study met speciale aandacht voor de wederzijdse validatie van gegevens en conclusies.

Het derde aspect werd geëvalueerd met behulp van een quasi-experimenteel cohort design met speciale aandacht voor interne-, externe- en constructvaliditeit. Het ontwikkelde onderwijs werd gegeven aan 2e en 3e jaars Bestuurskunde-studenten. Onderzocht werd of het onderwijsprogramma effect had op het gebruik van een systematische ontwerpaanpak voor beleid.

De conclusies van het proefschrift:

- De FUNDES-procedure moest voor de toepassing op een sociaalwetenschappelijk vakgebied op een groot aantal punten worden aangepast om een adequate handleiding te zijn. Deze aanpassingen tastten de uitgangspunten van de procedure echter niet aan.
- De aangepaste FUNDES-procedure leidde tot effectief onderwijs.
- Gelet op de voorgaande conclusies is de reikwijdte van de FUNDES vergroot.

Het proefschrift besluit met een methodische reflectie.

Helge Bonset: Onderwijs in heterogene groepen. Een case-study naar het vak Nederlands in een breed-heterogene brugklas van een reguliere scholengemeenschap, en een literatuurstudie naar heterogeniteit en interne differentiatie (Muusses, 1987).

Promotor: Prof. Dr. J.S. ten Brinke

Promotie: Rijksuniversiteit Utrecht, 18 juni 1987

KORTE SAMENVATTING

Doel van het proefschrift is bij te dragen aan praktijkdidactische theorie omtrent wat in de onderwijswerkelijkheid plaatsvindt bij onderwijs in heterogene groepen in het voortgezet onderwijs in Nederland. Het bestaat uit drie delen: een case-study naar een lessenreeks Nederlands in een breed-heterogene (LBO tot en met VWO-) brugklas van een reguliere scholengemeenschap, een literatuurstudie naar heterogeniteit en interne differentiatie in de eerste fase van het voortgezet onderwijs in Nederland, en een confrontatie tussen de gegevens uit beide. Deze confrontatie heeft plaats aan de hand van de volgende thema's: de leraarstaak (benodigde bekwaamheden, en uitvoerbaarheid van de taak); beoordeling en selectie; werksfeer in de klas (motivatie bij leerlingen, en orde); leereffekten bij leerlingen (in het bijzonder de zwakkere en de sterkere); en groepswork in heterogene subgroepen.

De leraar uit de case-study slaagt er tot grote hoogte in om in de breed-heterogene brugklas vorm te geven aan de kommunikatieve benadering van het vak Nederlands, via interne differentiatie naar niveau bij het onderdeel grammatika en voornamelijk naar belangstelling bij de andere vakonderdelen. Hij acht zijn taak in de bewuste klas uitvoerbaar, en demonstreert in zijn lessen de benodigde bekwaamheden, die hij zich al doende heeft eigen gemaakt. Zijn leerlingen tonen zich goed gemotiveerd, en in de klas heerst goede orde. De leerlingen zijn van mening dat zij voldoende leren bij Nederlands in de heterogene brugklas. Groepswork komt in ruime mate voor, en de leerlingen staan er positief tegenover. Een aanzienlijk aantal leerlingen van de bewuste brugklas (ongeveer 40%) verlaat deze met doorstromingsadvies dat hoger ligt dan het advies van de basisschool, en profiteert zodoende van het geboden uitstel van selectie. Bovenstaande bevindingen zijn gebaseerd op interviews, enquête, learner reports, dokumentstudie, en observatie van 22 lessen. De resultaten van de literatuurstudie wijzen grotendeels in dezelfde richting als die van de case-study, en dan niet alleen voor het vak Nederlands.

Het bovenstaande betekent niet dat er geen knelpunten zouden zijn aan te wijzen in het onderwijs aan heterogene groepen. De voornaamste die uit case- en literatuurstudie naar voren komen, zijn het ontbreken van een doelgericht curriculum voor groepswork, en de beperkende randvoorwaarden waaronder onderwijs in heterogene groepen zich moet afspelen (korte duur van de heterogene periode; continue selectie die zijn schaduwen vooruitwerpt).

NIEUWE PUBLICATIES

- Reijmerink, C.H.L., Thijssen, J.A., Mulder, M. & Pieters, J.M. (1987). *Regionale voorzieningen in de werktuigbouwkunde: een structurele oplossing?* (Toegepaste Onderwijskunde: f 10,-).
- Reijmerink C.H.L., & Pieters, J.M. (1987). *Apparatuur in het MTO: ideaal en werkelijkheid.* (Toegepaste Onderwijskunde: f 15,-).
- Reijmerink, C.H.L., & Mulder, M. (1987). *Het profiel van de MTS-er in een startfunctie.* (Toegepaste Onderwijskunde: f 12,50).
- Thijssen, J.A., & Mulder, M. (1987). *Trends in de produktie automatisering en gevolgen voor het beroepsonderwijs.* (Toegepaste Onderwijskunde: f 15,-).
- Thijssen, J.A., & Mulder, M. (1987). *Aansluiting MTS-werktuigbouwkunde-arbeidsmarkt.* (Toegepaste Onderwijskunde: f 12,50).
- Leerplanontwikkeling tussen onderwijsbeleid en schoolontwikkeling* (SLO, f 25,-).
- Een goed produkt.* (SLO, f 21,-).
- Mooij, T. (1987). *Interactional multi-level investigation into pupil behavior, achievement, competence, and orientation in educational situations.* (Selectareeks: f 40,-).
- Koetsier, C.P., Van Driel, C., Janzen, R.W., & Ten Brink, J.S. (1987). *Gebleken leraarsbekwaamheid.* (Afdeling Didaktiek van de Biologie, RUU).
- Koetsier, C.P., & Ten Brink, J.S. (1987). *Experiment individuele eindstage-1 op de afdeling Didaktiek van de Biologie - Analyse van de interviews met student-leerkrachten.* (Afdeling Didaktiek van de Biologie, RUU).
- Bonset, H. (1987). *Onderwijs in heterogene groepen.* Purmerend: Muusses.
- Vernieuwing. Tijdschrift voor Onderwijs en Opvoeding* (1987) (samengaan van: "Onderwijs & Opvoeding", en "Vernieuwing").
- Genemans, L. (1987). *Autonomie van scholen en deregulering.* (ITS).
- Van Kessel, N., & Petersen, B. (1987). *Verlangd voortgezet onderwijs in het Westland.* (ITS).
- Smeets, E., & Willems, M. (1987). *Het vreemde talen onderwijs in de geïntegreerde brugperiode.* (ITS).
- Lucassen, P., Oude-de Glas, M., Petersen, B., & Willems, M. (1987). *De behoefte aan moderne vreemde talen bij oud-leerlingen van het MEAO, MMO en MTO.* (ITS).
- SLO (1987) Jaarverslag 1986.
- Venselaar, K., Gerritsen van der Hoop, W., & Van Drunen, P. (1987). *Leerervaringen ingedeeld naar kennisgroepen en kennissoorten.* (Groep Onderwijsresearch, TUE).
- Van der Sijde, P.C. (1987). *Training the teaching script.* Almere: Versluys.
- Pronto Nieuwsbrief 1**
- OTG-Onderwijsleerprocessen. *Nieuwsbrief 15.*
- IEA Newsletter, number 12, July 1987.
- Smit, F., Reijmer, G., Vrieze, G. en Ruijs, A. (1987). *Het functioneren van de medezeggenschap in het onderwijs.* Nijmegen: ITS.
- De Graaf, P.M. (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen.* Nijmegen: ITS (dissertatie).
- Van Kessel, N. en Petersen, B. (1987). *Verlangd voortgezet onderwijs in Veenendaal en omstreken.* Nijmegen: ITS.
- Van Kessel, N. en Petersen, B. (1987). *Verlangd voortgezet onderwijs in Tilburg en omstreken.* Nijmegen: ITS.
- Van Kessel, N. en Petersen B. (1987). *Verlangd voortgezet onderwijs in Culemborg en omstreken.* Nijmegen: ITS.
- Pouwels, J.P.A. en Vrieze, G. (1987). *Het cursorisch onderwijs onder contract.* Nijmegen: ITS.