

Tijdschrift voor onderwijsresearch 1988

<https://hdl.handle.net/1874/208866>

VWB620

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591X

Jaargang 13
1988

Swets & Zeitlinger B.V.
Lisse

Redactie:

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

BIBLIOTHEEK DER
RIJKSUNIVERSITEIT
UTRECHT



VERENIGING VOOR ONDERWIJSRESEARCH

Artikelen

Schrijven en schrijven is twee: een onderzoek naar de samenhang tussen prestaties op verschillende schrijftaken. <i>Huub van den Berg</i>	311
Differentiële effecten van vraagvorm bij aardrijkskunde- en natuurkunde-examens. <i>Huub van den Bergh, Mindert Eiting en Martha Otter</i>	270
De grootte van schooleffecten: Hoe verschillend presteren leerlingen van verschillende scholen? <i>H. Blok en M.H. Eiting</i>	16
De ontwikkeling van de leesattitude in het basisonderwijs: een longitudinale benadering. <i>Theo Boland</i>	3
Organisatorische verschillen tussen basisscholen en hun effect op leerlingprestaties. <i>Hennie P. Brandsma en Anja W.M. Knuver</i>	201
Bekroond ORD Paper 1988. De Rijksinspectie en het Voortgezet Onderwijs. <i>J.F.A. Braster</i>	334
De relatie tussen leerlingkenmerken en schoolverloop bij IBO-leerlingen. <i>D.A. van Damme, J.M.T. Eulderink-Mulder en H.E. de Jonge</i>	151
Inter- en criterium- codeer-overeenstemming van beroepenclassificaties. <i>J.A. Gras en G.J. Mellenbergh</i>	77
Evaluating an Item and an Option Statistic Using the Bootstrap Method. <i>Dato N.M. de Gruijter</i>	345
Stress-situaties op school: een onderzoek naar alledaagse schoolsituaties die door leerlingen als belastend worden waargenomen. <i>J. Hendricksen en M. Boekaerts</i>	161
Het lineair Logistisch Test Model en heterogeniteit van cognitieve strategieën bij balansproblemen. <i>L. van Manen, P.H. Been en K. Sijtsma</i>	301
Het meten van affectieve doelen: De validering en normering van de Belevingsschaal voor Wiskunde (BSW). <i>Mark J. Martinot, Hans B. Kuhlemeier en Hans J.M. Feenstra</i>	65
Reflecteren in systematische samenspraak. <i>J.M.C. Nelissen</i>	245
Longitudinale computeronderkenning van lees- en spellingsmoeilijkheden: een toepassing van het Kalman-filter in de onderwijspraktijk. <i>Johan H.L. Oud en Martin J.C. Mommers</i>	31
Het leren programmeren van computergestuurde gereedschapswerktuigen. Een afbeeldingstheorie en een informatieverwerkingsmodel. <i>Georg Rakers, Jules Pieters en Imme de Bruijn</i>	181
Opsporen van misconcepties bij Middelbare Scholieren. <i>Henk Schmidt, Gerard Spaaij en Willem de Grave</i>	129
Effecten van klasseconsultatie op curriculumimplementatie en leerresultaten. <i>J. Snippe</i>	141
The Influence of a Teacher Training on the Student Perception of Classroom Climate. <i>Pieter C. van der Sijde</i>	88
Het spellen van Papiaments door Curaçaose schoolkinderen. <i>P. Vedder en H. Kook</i>	325
Betrouwbaarheid van observatietoetsen voor praktische vaardigheden in het medisch onderwijs. <i>C.P.M. van der Vleuten en S.J. van Luyk</i>	213

Notities en Commentaren

Aantekeningen bij 'Anatomie van een leeromgeving, een onderwijs-economische analyse van universitair onderwijs', K.J.D.M. van der Drift en P. Vos (Dissertatie Leiden, 1987). <i>J.W. Holleman</i>	95
Naschrift. <i>K.J.D.M. van der Drift en P. Vos</i>	97
A Further Note on the Necessity of Including a Retrospective Pretest in Self-report Pretest-Posttest Designs to Detect Training Effectiveness. <i>Mirjam Sprangers</i>	353
Het gebruik van de PMT-K bij Curaçaose schoolkinderen: een validatiestudie. <i>Paul Vedder</i>	50
De wens naar een nauwere samenhang tussen O.T.G.-onderzoek en opdrachtonderzoek. <i>P. Vedder</i>	285

Boekbesprekingen

J.J.H. van den Akker: <i>Ontwerp en implementatie van natuuronderwijs</i> . (Bert P.M. Creemers)	236
H. Bonset: <i>Onderwijs in heterogene groepen</i> . (Ton Mooij)	102
Th.H.L. Compernelle: <i>Stress op de middelbare school</i> . (J. Hendriksen)	233
Henri Dors: <i>Vriendschap en sociale relaties in multi-etnisch samengestelde schoolklassen</i> . (L. van den Berg-Eldering)	231
P.M. De Graaf: <i>De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen</i> . (G.W. Meijnen)	290
B.H.A.M. Van Hout Wolters: <i>Het markeren van kerngedeelten in studieteksten. Een 'proces-product'-benadering</i> . (J.H. Boonman)	227
Mart-Jan de Jong: <i>Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs</i> . (L. van den Berg-Eldering)	229
Rik Min: <i>Computersimulatie als leermiddel: een inleiding in methoden en technieken</i> . (H. de Jong)	358
Ton Mooij: <i>Interactional multi-level investigation into pupil behaviour, achievement, competence and orientation in educational situations</i> . (P. van den Eeden)	99
J. Neuvel, M.E. Otter en D.J. Bos: <i>Functioneel aanvankelijk leesonderwijs. Een beschrijving en een evaluatie</i> . (Dré van Dongen)	356
W. van Os: <i>Evaluatie in het hoger onderwijs</i> . (W.H. Gijselaers)	363
J.L. Peschar: <i>Zo vader - zo zoon, zo moeder - zo dochter? Vergelijkende analyses naar de processen van statusverwerving en onderwijsmobiliteit in Nederland, Hongarije en Polen</i> . (J.G.M. Lammers)	360
F.W. Siero: <i>Feedback en motivatie in de klas</i> . (John van Berken)	291
P.C. van de Sijde: <i>Training the teaching script</i> . (S. Veenman)	172
C. Terkouw: <i>De FUNDUS-procedure in onderwijsontwikkeling. Evaluatie van een procedure van onderwijsontwikkeling voor het leren probleemoplossen</i> . (P. RobertJan Simon)	55
Fred Wester: <i>Strategieën voor kwalitatief onderzoek</i> . (Ger Tillekens)	106

Redactionele medewerkers

In 1988 werd redactionele medewerking verleend door:

C.A.J. Aarnoutse	KU Nijmegen
J.J.H. van den Akker	Universiteit Twente
J. Ax	UvAmsterdam
G. van den Berg	RION Groningen
M.P.F. Berger	Universiteit Twente
H.J.M. van Berkel	RU Limburg
S.E. de Bie	RU Leiden
S. Boef-van der Meulen	SCP Rijswijk
A.G. Bus	RU Groningen
S. Dijkstra	Universiteit Twente
D. van Dongen	IVA Tilburg
G.P. van Galen	KU Nijmegen
W. van de Grift	RU Utrecht
D.N.M. de Gruijter	RU Leiden
H. Guldemon	RION Groningen
E.G. Harskamp	RION Groningen
R.J. Hiemstra	RU Groningen
W.Th.J.G. Hoebe	RION Groningen
T. Houtveen	RU Utrecht
J.J. Hox	UvAmsterdam
G.G.H. Jansen	RU Groningen
R. de Jong	RION Groningen
P. Jungbluth	ITS Nijmegen
G. Kanselaar	RU Utrecht
L.F.W. de Klerk	KUB Tilburg
K.B. Koster	RU Groningen
H. Kuyper	RION Groningen
P. van Laarhoven	RION Groningen
A. van der Leij	VU Amsterdam
J.M.G. Leune	EU Rotterdam
C.F.M. van Lieshout	KU Nijmegen
J. Lowyck	KU Leuven
K. Meijer	ITS Nijmegen
G.J. Mellenbergh	UvAmsterdam
M.J.C. Mommers	KU Nijmegen
T. Mooij	ITS Nijmegen
A.J. Nederhof	RU Leiden
J.L. Peschar	RU Groningen
J.M. Pieters	Universiteit Twente
J. Rispens	RU Leiden
E. Roede	SCO Amsterdam
J. Scheerens	Universiteit Twente
J.H. Slavenburg	SAD Rotterdam
W.G.R. Stoel	OCTO Enschede
H.H. Tillema	RION Groningen
P.H. Vedder	SVO Den Haag
R.K.W. van der Velden	RION Groningen
N.H. Veldhuijzen	CITO Arnhem
L. Verschaffel	KU Leuven
H. Verstralen	CITO Arnhem
J.M. Wijnstra	CITO Arnhem
G.L.H. van den Wittenboer	UvAmsterdam
M. Zwarts	CITO Arnhem

Van de redactie

Met ingang van 1987 is het Tijdschrift voor Onderwijsresearch samengevoegd met het VOR-Bulletin. Deze samenvoeging gaat niet verder dan het verschijnen binnen een en dezelfde kافت. De redactionele verantwoordelijkheden blijven uiteraard strikt gescheiden.

Het Tijdschrift voor Onderwijsresearch is een wetenschappelijk tijdschrift met een redactie die weliswaar benoemd wordt door het bestuur van de Vereniging voor Onderwijsresearch (VOR), maar onafhankelijk haar werkzaamheden verricht. Wel zijn in het afgelopen jaar een aantal afspraken met het VOR-Bulletin gemaakt teneinde de beschikbare ruimte efficiënt en doelmatig te gebruiken.

Behalve de mededelingen zullen in de toekomst ook de ontvangen publikaties uitsluitend in het VOR-Bulletin worden vermeld. Samenvattingen van dissertaties op het terrein van onderwijsonderzoek kunnen spoedig na verschijning in het VOR-Bulletin worden opgenomen, een kritische bespreking ervan die vaak meer tijd vergt zal vervolgens in het Tijdschrift voor Onderwijsresearch worden opgenomen.

Hoewel naar de mening van de redactie het Tijdschrift voor Onderwijsresearch en het VOR-Bulletin duidelijk onderscheiden zijn en er ook begrip is voor de financiële noodzaak om beide uitgaven van de Vereniging voor Onderwijsresearch op deze wijze goedkoper te kunnen exploiteren, blijken niet alle lezers even gerust te zijn op wat zij noemen samenvoeging van een "wetenschappelijk tijdschrift met een omhooggefallen mededelingenblad". Zij vrezen verving van het onderscheid, niveauverlaging van het tijdschrift en een onterechte opwaardering van bijdragen die in het VOR-Bulletin verschijnen. De handhaving van de kwaliteit van het Tijdschrift voor Onderwijsresearch is een zaak van de redactie en er is vooralsnog geen reden om te twijfelen dat hieraan getornd zal worden. Elke inhoudelijke verbetering van zowel het Tijdschrift voor Onderwijsresearch als het VOR-Bulletin zal verder alleen met vreugde begroet worden en onterechte opwaardering is een zaak van de boekhoudende kwaliteitscontroleurs.

De beschikbare ruimte kon dit jaar maximaal benut worden voor het plaatsen van bijdragen op het terrein van onderwijsonderzoek. De redactie blijft artikelen op het terrein van onderwijsonderzoek graag tegemoet zien. Nog steeds draagt een te gering deel van de onderwijsonderzoekers bij aan het tijdschrift. Het tijdschrift is niet alleen een blad voor onderwijsonderzoekers, het dient ook door onderwijsonderzoekers gevuld te worden. Behalve verslagen van onderzoek en methodologisch-technische beschouwingen kunnen dat ook tutorials en reviews zijn. In het verleden is een beperkt aantal reviews op het terrein van onderwijsonderzoek geplaatst. In de toekomst wil de redactie proberen meer van deze overzichtsartikelen te plaatsen. Daarvoor is contact gelegd met de onderzoeksthmagroepen op het terrein van het onderwijsonderzoek, die op dit moment bezig zijn reviews op het door hen bestreken terrein voor te bereiden. Het moge duidelijk zijn dat ook reviews evenals andere bijdragen aan het tijdschrift ter voorbereiding van besluitvorming in de redactie door minstens twee beoordelaars anoniem van commentaar worden voorzien. Bij tutorials gaat het om inleidingen ten aanzien van meestal onderzoekstechnische onderwerpen die niet eenvoudig in studieboeken of andere tijdschriften te vinden zijn. Het gaat dus niet om gemakkelijk te vinden of elementaire onderdelen uit de onderzoeksmethodologie. Mocht men een tutorial willen voorbereiden voor het Tijdschrift voor Onderwijsresearch, dan is contact met de redactiesecretaris aan te bevelen om uiteindelijk teleurstellingen bij de beoordeling te voorkomen.

In het verleden heeft de redactie zich altijd verzet tegen het samenstellen van nummers over bepaalde onderwerpen. Als tijdschrift op het terrein van onderwijsresearch is er de voorkeur aan gegeven vooral het aanbod van artikelen te volgen, deze te beoordelen, te begeleiden bij het herschrijven en uiteindelijk te plaatsen. Wel werden wanneer er aanleiding toe was rondom een bepaald artikel andere bijdragen gevraagd, die vaak commentaar gaven op het in het artikel gepresenteerde. Met ingang van 1988 streeft de redactie ernaar om een keer per jaar rondom een

belangrijk onderwerp, inhoudelijk of methodologisch-technisch, op het terrein van onderwijs-onderzoek een themanummer samen te stellen. Op dit moment is een nummer in voorbereiding over longitudinaal onderwijsonderzoek, dat in het najaar van 1988 zal verschijnen.

Tenslotte een mededeling van meer huishoudelijke aard. De rubrieken Notities & Commentaren en Boekbesprekingen, waarin zoals hierboven is opgemerkt meer en sneller aandacht zal worden besteed aan dissertaties op het terrein van onderwijsresearch, zullen in de toekomst geredigeerd worden door Dr. M. Voeten. Met hem kan men contact opnemen over notities en commentaren, boekbesprekingen en het inzenden van dissertaties. Het adres is Instituut voor Onderwijskunde, Erasmusplein 1, 6525 GG Nijmegen. Manuscripten, ook voor de rubriek Notities & Commentaren, dienen echter ingeleverd te worden bij het redactiesecretariaat.

De Ontwikkeling van de Leesattitude in het Basisonderwijs: een Longitudinale Benadering

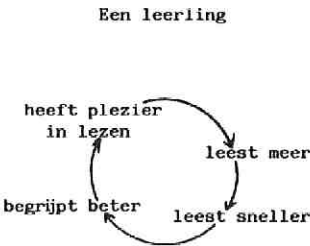
Theo Boland*
Instituut voor Onderwijskunde, K.U.-Nijmegen.

ABSTRACT

This article focusses on the relations between the cognitive and affective aspects of reading in primary education, more specifically on the relations between decoding, reading comprehension and reading attitude. A model is specified to investigate the causal effects of background variables, decoding and reading comprehension on reading attitude. In a longitudinal study data were collected at five time points, viz. Grade 1, Grade 2, Grade 3, Grade 6 (begin) and Grade 6 (end). By using a Lisrel procedure, the data of 310 students of 24 primary schools were analysed. It is shown that a longitudinal model, in which several theoretically based causal relations were assumed, rather well fits to the empirical data. With respect to the cognitive reading components, only reading comprehension influences reading attitude. Contrary to the social economic and reading background variables, the sexe factor seems to be important. Girls and boys differ in their reading attitude development.

INLEIDING

In een recente SLO-publikatie 'Problemen bij tekstbegrip van allochtone leerlingen' (Gerritsma, 1987) is een figuur opgenomen dat als uitgangspunt kan dienen voor het onderwerp van dit artikel. Afgebeeld is wat we zouden kunnen noemen 'de positieve leesspiraal' (figuur 1).



Figuur 1. De positieve leesspiraal.

Vanzelfsprekend bestaat er ook een pendant: de negatieve leesspiraal. (Een leerling leest weinig, leest langzaam, begrijpt het niet, heeft geen plezier in lezen, etc.)

We willen in deze bijdrage onderzoeken hoe de decodeersnelheid ('leest snel') en het begripend lezen ('begrijpt het goed') de leesattitude ('plezier in lezen') beïnvloedt in de loop van het basisonderwijs. Is de in de afbeelding veronderstelde rechtlijnigheid ook inderdaad aanwezig?

1. THEORETISCHE ACHTERGRONDEN

Krathwohl, Bloom en Masia (1964) onderscheiden drie domeinen waarop doelstellingen in het onderwijs betrekking hebben, namelijk het cognitieve, affectieve en het psychomotorische domein. Beperken we ons tot de eerste twee domeinen, dan kunnen we constateren dat er veel meer onderzoek is verricht binnen het cognitieve dan binnen het affectieve domein. Dit geldt ook voor het domein lezen en leesonderwijs in het basisonderwijs. Er is en wordt zeer veel onderzoek verricht naar de cognitieve aspecten van de leesvaardigheid en veel minder naar het affectieve aspect, bijvoorbeeld naar het ontstaan en de ontwikkeling van leesattitudes en leesgewoontes. Toch wordt het belang van de affectieve aspecten van de leesvaardigheid algemeen erkend (cf. ARBO, 1984). Malmquist en Brus (1974) zijn van opvatting dat het bevorderen van een positieve houding ten aanzien van lezen een hoofddoel van het leesonderwijs zou moeten zijn. In de onderwijspraktijk zien we dan ook dat er op verschillende manieren wordt gewerkt aan deze doelstelling, zoals het inrichten van goede klassebibliotheken en het promoten van goede jeugdliteratuur.

Binnen het affectieve aspect van de leesvaardigheid neemt het begrip *leesattitude* een centrale plaats in. Volgens een aan Allport (1935) ontleende, algemeen geaccepteerde definitie, is een attitude een door ervaring gevormde neiging of dispositie om op bepaalde objecten, situaties of gebeurtenissen positief of negatief te reageren. Een *leesattitude* is dan de geleerde of verworven (pre)dispositie om leessituaties en leesmateriaal positief of negatief te evalueren. Volgens Fishbein & Ajzen (1975) kunnen binnen de leesattitude een aantal componenten worden onderscheiden. De cognitieve component is de kennis of ervaring van/met lezen en diverse soorten leesstof, de affectieve component is de positieve of negatieve beoordeling daarvan. De derde, conatieve, component is het leesgedrag, of in Fishbein & Ajzen's termen, de kans dat iemand een bepaald leesgedrag vertoont (cf. Bentler & Speckart, 1979).

Er is weinig bekend over de *relatie* tussen de cognitieve en affectieve component van de leesvaardigheid. De relatie is complexer dan vaak wordt aangenomen (Downing & Leong, 1982). Uit onderzoek is wel gebleken dat de leesprestatie als representant van het cognitieve aspect, geen betrouwbare voorspeller is voor de leesattitude. Het omgekeerde is ook het geval zoals blijkt uit de opmerking van Roettger c.s. '...that contrary to the assumption that a positive attitude is essential to read succesfully, attitudes cannot be used as predictor of academic (reading) achievement' (Roettger, Szymczuk & Millard, 1979, p. 141).

Hiermee zijn we beland bij de centrale onderzoeksvragen van dit artikel: hoe groot is de invloed van resp. technisch lezen (decoderen) en begrijpend lezen op de leesattitude in de loop van het basisonderwijs? Wordt de ontwikkeling van de leesattitude beïnvloed door achtergrondkenmerken als sexe, sociaal milieu en leesmilieu?

Het meeste onderzoek op dit gebied is correlatieel van aard en beperkt zich tot één fase in de ontwikkelingsperiode. De correlaties tussen leesprestatie en leesattitude liggen meestal tussen .10 en .40 (Roettger et al., 1979; Gorman, White, Orchard & Tate, 1981/1982; Bisschop, Aarnoutse & Feenstra, 1985; Boland & Aarnoutse, 1986). In dat laatste onderzoek hebben we vastgesteld dat de variantie in leesattitude voor een deel kan worden verklaard uit de combinatie leesfrequentie en prestatie op begrijpend lezen. Veel variantie bleef echter onverklaard. Bij een onderzoek in leerjaar 6 (Groep 8) stelden we vast dat er geen één-op-één relatie is tussen 'goed-kunnen-lezen' en 'houden-van-lezen'. Ongeveer 5% van de zesde-klassers combineerden hoge leesprestaties met een lage leesattitude (deze groep bestond voor het grootste deel uit jongens); eveneens 5% van de leerlingen, meest meisjes, behaalde hoge scores op de leesattitudemetingen en tegelijkertijd lage scores op toetsen voor begrijpend lezen.

Wat de factor 'sexe' betreft: meestal wordt een lage correlatie gevonden. Meisjes scoren gemiddeld iets hoger op leesattitudeschalen dan jongens (Hansen, 1969; Kennedy & Halinsky, 1975; Gorman et al., 1981/1982; Boland & Mommers, 1985). De invloed van het milieu op de ontwikkeling van de leesattitude is eveneens nog weinig onderzocht. Wigfield & Asher (1984) constateerden een geringe invloed van het sociaal milieu. Een mogelijke bijdrage in de verklaring van verschillen tussen kinderen wat hun houding ten opzichte van lezen betreft, zou

volgens Hansen (1973) geleverd kunnen worden door de leesomgeving thuis: de interactie tussen ouders en kinderen in hun opvatting over lezen en leesmateriaal.

Onderzoeksvragen:

Door middel van het opstellen en toetsen van een longitudinaal model, waarin causale relaties worden verondersteld tussen technisch lezen, begrijpend lezen en leesattitude, gekoppeld aan enkele achtergrondvariabelen: sexe, sociaal milieu en leesmiliieu, zal getracht worden een antwoord te vinden op de volgende vragen.

- 1) Hoe groot is de invloed van technisch, resp. begrijpend lezen op de leesattitude? Treden er in de loop van het basisonderwijs veranderingen op in de mate van beïnvloeding?
- 2) Welk belang moet gehecht worden aan de achtergrondkenmerken: sexe, sociaal milieu en leesmiliieu in hun invloed op de leesprestaties en (vooral) op de leesattitude-ontwikkeling?

2. BESCHRIJVING VAN DE BESCHIKBARE VARIABELEN

In het nu volgende gedeelte geven we een korte beschrijving van de in de analyses opgenomen variabelen. Een samenvattend overzicht met de te gebruiken afkortingen vindt U in Schema 1. In een bijlage zijn de psychometrische gegevens van de geobserveerde variabelen vermeld.

2.1 De affectieve component

Voor het meten van de affectieve component zijn vier meetinstrumenten gebruikt.

De LAS (Rensen, 1981) is een leesattitudeschaal bestaande uit 18 items van het Likert-type bestemd voor de middengroep van het basisonderwijs. De items hebben twee antwoordcategorieën. Een voorbeeld:

'Je krijgt op je verjaardag een leesboek.
Vind je het *fijn* of *niet fijn* om een leesboek te krijgen?'

Antwoorden die wijzen op een positieve instelling ten aanzien van boeken, lezen, leessituaties, etc. zijn met (2) gescoord, de overige antwoorden met (1). De somscore wordt geacht een maat te zijn voor de leesattitude. De LAS is afgenomen in leerjaar twee en drie¹⁾.

De LAS-A is een door Aarnoutse bewerkte leesattitudeschaal van Estes (1971): Scale to Measure Attitudes toward Reading. De schaal heeft 20 items van het Likert-type met 5 antwoordcategorieën. Een voorbeeld:

'De meeste boeken zijn te saai om helemaal uit te lezen.'
 – helemaal mee eens
 – mee eens
 – weet ik niet
 – niet mee eens
 – helemaal niet mee eens

De scoring loopt van 5 naar 1 voor de 8 positief geformuleerde items en van 1 naar 5 voor de 12 negatief geformuleerde items (zoals het voorbeeld hierboven). De somscore wordt gezien als een indicatie voor de mate waarin leerlingen een positieve houding hebben ten aanzien van recreatief of vrij lezen. De LAS-A is twee maal afgenomen: in het begin en eind van leerjaar zes.

¹⁾ In dit verslag gebruiken we de oude terminologie.
Leerjaar 2 = Groep 4; leerjaar 6 = Groep 8.

De *LAS-B* (Aarnoutse & Boland, 1985) is een leesattitudeschaal die qua vorm identiek is aan de *LAS-A*. De *LAS-B* heeft 22 items. Een voorbeeld van een uitspraak waarop gereageerd moet worden: 'In leeslessen wordt teveel tijd besteed aan hardop lezen.' De somscore is een maat voor de houding ten aanzien van lezen en leeslessen. De *LAS-B* is afgenomen aan het eind van leerjaar 6.

Op hetzelfde tijdstip is afgenomen een Vrijtijdspreferentieschaal (Boland, 1985). Een subschaal hieruit is de *Leespreferentieschaal*. Leerlingen ordenen zeven vormen van vrijetijdsbesteding naar mate van voorkeur in een paarsgewijze aanbieding van stimuli in een 'complete balanced design'. Leerlingen met de score (7) prefereren lezen boven alle andere bezigheden (winkelen, spelletjes doen, huiswerk maken, tv-kijken, knutselen, sport). Leerlingen met de score (1) prefereren alle andere vormen van vrijetijdsbesteding boven lezen.

2.2 De cognitieve component

De vaardigheid technisch lezen is gemeten met de Eén-Minuut-Test (Brus & Voeten, 1979). De score op de EMT is een maat voor de decodeervaardigheid. Ze wordt bepaald door na te gaan hoeveel woorden uit een lijst van 116 niet-samenhangende, in moeilijkheidsgraad opklimmende woorden, in één minuut foutloos hardop worden gelezen. Deze toets is afgenomen in de leerjaren 2, 3 en 6; in leerjaar 6 zowel de A-vorm als de (parallel)vorm B, in leerjaar 2 en 3 alleen de A-vorm.

Begrijpend lezen is gemeten met een reeks van toetsen. In het begin van leerjaar 2 is afgenomen de toets *Lees en Begrijp 1b* (CITO, 1979). Dit instrument richt zich op het begrijpen van woorden, zinnen en relaties tussen zinnen.

De toets *Schriftelijke Opdrachten 3* (Brus & Van Bergen, 1973), die afgenomen is in leerjaar 3, gaat uit van de gedachte dat wanneer kinderen een geschreven opdracht goed uitvoeren, aangenomen mag worden dat de tekst begrepen is. De toets bestaat uit een reeks opdrachten die in moeilijkheid opklimmen. In het toetsboekje moet bijvoorbeeld een streep of een kruisje worden gezet of een eenvoudige tekening worden gemaakt. De volgende zes toetsen zijn afgenomen in leerjaar zes; de eerste vier aan het begin, de resterende twee aan het eind van dit leerjaar.

De *Hoofdgedachtetoets* (Aarnoutse, 1984) meet de vaardigheid uit een tekst de hoofdgedachte, het centrale thema, te kunnen afleiden.

De *Woordenschattest* (Stijnen, 1975) meet de receptief-schriftelijke woordenschat.

De subtest *Stillezen* uit de ISI-Vorm 3 (Van Boxtel, Snijders & Welten, 1982) onderzoekt kennis en inzicht in woorden, uitdrukkingen en idiomatische eenheden op zins- en tekstniveau.

De *Cloze Test* voor begrijpend lezen (Boland, 1984) geeft de leerlingen de opdracht een tekst waaruit systematisch woorden zijn weggelaten, te reconstrueren.

Tenslotte twee subtesten uit de Eindtoets Basisonderwijs 1985 (CITO, 1985), namelijk *Lezen van teksten/reproductie* en *Lezen van teksten/conclusies*. Leerlingen moeten woordelijke of geparafraseerde herhalingen van tekstinformatie herkennen en/of expliciteren door middel van het trekken van conclusies.

2.3 Achtergrondkenmerken

De milieufactoren zijn onderverdeeld in twee variabelen: het *sociale milieu* en het *leesmilieu*. De index voor het sociale milieu is geconstrueerd met behulp van de variabelen beroepsniveau van de ouders, opleidingsniveau van de ouders, aantal kinderen in het gezin en nationaliteit. Voor het vaststellen van het leesmilieu is een maat geconstrueerd die berust op de variabelen abonnement op kranten en tijdschriften, leesfrequentie van de ouders, lidmaatschap van een bibliotheek en de frequentie waarmee ouders voor hun kinderen boeken kopen. De gegevens omtrent het sociaal milieu en het leesmilieu zijn verzameld in leerjaar 1. De data zijn via een HOMALS-procedure (Gifi, 1981) omgezet in continue variabelen.

Schema 1. Overzicht van de variabelen.

(Latente) variabelen	Geobserveerde variabelen in leerjaar				
	1	2	3	6 (begin)	6 (eind)
Sociaal milieu Leesmilieu Sexe	SM LM SEXE				
Technisch lezen (TL)		EMT/A	EMT/A	EMT/A EMT/B	
Begrijpend lezen (BL)		L&B1b	SO-3	HGT WST STL CLOZE	LvT/r LvT/c
Leesattitude (LA)		LAS	LAS	LAS-A	LAS-A LAS-B LEESPREF

Toclichting:

SM = index voor het sociale milieu

LM = index voor het leesmilieu

EMT/A resp. EMT/B = de A- en B-versie van de Eén-Minuut-Test (Brus & Voeten)

L&B1b = Lees & Begrijp 1b (CITO/Verhoeven)

SO-3 = Schriftelijke Opdrachten 3 (Brus & v. Bergen)

HGT = Hoofdgedachtetest (Aarnoutse)

WST = Woordenschattest (Stijnen)

STL = subtest Stillezen uit de ISI-Vorm 3 (Van Boxtel, Snijders & Welten)

CLOZE = Cloze-toets voor begrijpend lezen (Boland)

LvT/r = subtest Lezen van teksten/reproductie (CITO-Eindtoets 1985)

LvT/c = subtest Lezen van teksten/conclusies (CITO/Eindtoets 1985)

LAS = leesattitudeschaal (Rensen)

LAS-A = leesattitudeschaal (Estes/Aarnoutse)

LAS-B = leesattitudeschaal (Aarnoutse/Boland)

LEESPREF = leespreferentieschaal (Boland)

3. MODELSPECIFICATIE

In Figuur 2 wordt een uitgangsmodel geïntroduceerd. Hierin worden de latente variabelen aan elkaar gerelateerd op basis van de volgende veronderstellingen.

1) De analyse heeft betrekking op gegevens die op vijf verschillende meetmomenten zijn verzameld, namelijk in het begin van leerjaar 1 (de achtergrondgegevens), in het begin van leerjaar 2, begin leerjaar 3, begin leerjaar 6 en eind leerjaar 6. Om het uitgangsmodel nietodeloos gecompliceerd te maken, beschouwen we voorlopig de metingen in leerjaar 2 en 3 als één meting; hetzelfde geldt voor de twee metingen in leerjaar 6: ook hier voegen we de twee metingen samen. Dit kunnen we doen omdat de metingen in de tijd gezien relatief dicht bij

elkaar liggen. Wel zal getoetst moeten worden of een dergelijke samenvoeging toegestaan is.

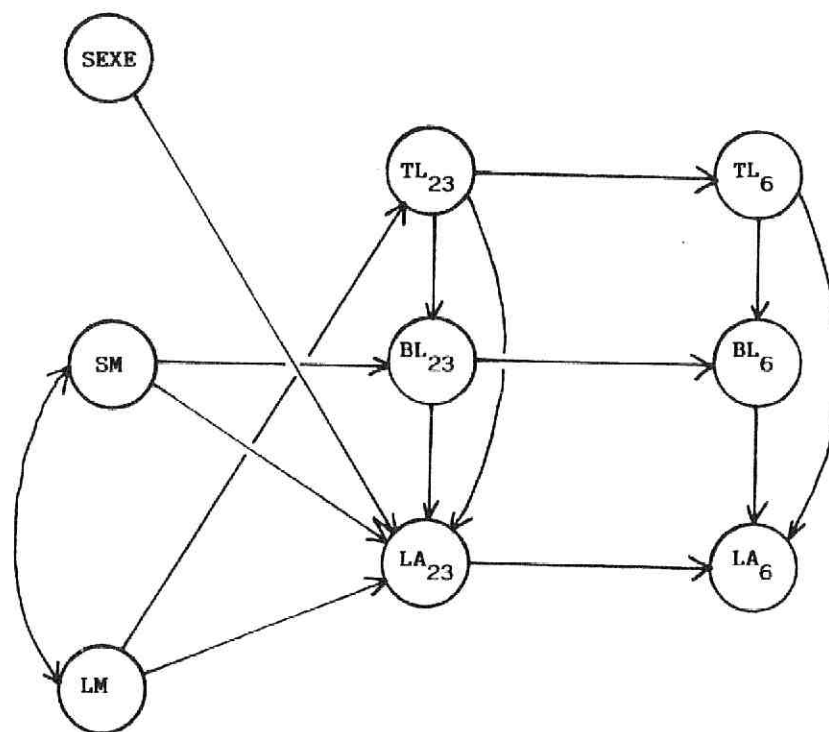
2) We gaan er van uit dat zich binnen één bepaalde latente variabele sterke longitudinale invloeden ontwikkelen. Bijvoorbeeld sterke invloeden van BL_{23} op BL_6 . Deze longitudinale invloeden zullen voor een deel de synchrone invloeden tussen latente variabelen op één meetmoment opheffen. Eerdere analyses hebben in dit verband aangetoond, dat bijv. de invloed van TL op BL in leerjaar 6 sterk afneemt, indien ook de metingen in voorgaande leerjaren in de beschouwing worden opgenomen (Mommers et al., 1986; Boland & Mommers, 1986).

3. Het ligt voor de hand een invloed van zowel TL als BL op de leesattitude te veronderstellen. Het is ook plausibel te stellen dat om van lezen en leesmateriaal te kunnen genieten, er sprake moet zijn van een zekere basisvaardigheid in technisch en begrijpend lezen. Een alternatief model, waarin verwacht wordt dat er een invloed uitgaat van de leesattitude op de leesprestatie, kan niet bij voorbaat worden uitgesloten. Daarom zal eveneens getoetst worden of een alternatief model waarin de pijlen lopen van de leesattitude naar technisch en begrijpend lezen, mogelijk en misschien wel beter is.

4) Uit de onderzoeksliteratuur blijkt niet dat de factor 'sexe' van grote invloed is op de cognitieve aspecten van de leesvaardigheid, althans niet in de beginfase van het onderwijs. Verondersteld wordt alleen een invloed op de leesattitude.

5) De achtergrondkenmerken SM (sociaal milieu) en LM (leesmilieu) zijn waarschijnlijk gecorreleerd. Het leesmilieu wordt geacht van invloed te zijn op de technische leesvaardigheid, dat in het begin van het leesonderwijs een belangrijke plaats inneemt, en op de leesattitude. We verwachten dat er directe lijnen te trekken zijn tussen het sociale milieu en resp. begrijpend lezen en leesattitude.

In het uitgangsmodel zijn dus opgenomen drie exogene variabelen (sexe, sociaal milieu en leesmilieu) en zes endogene latente variabelen, te weten technisch lezen, begrijpend lezen en leesattitude op twee meetmomenten. Gezien de aard van dit longitudinale model zijn alleen relaties verondersteld die 'meegaan met de tijd'; invloeden van een variabele op een variabele eerder in de tijd zijn dus per definitie uitgesloten.



Figuur 2. Uitgangsmodel 1 met 3 exogene en 6 endogene latente variabelen.

4. METHODE

4.1 Proefpersonen

In de analyses zijn opgenomen de gegevens van 310 leerlingen met scorebare data op alle geobserveerde variabelen (164 jongens en 146 meisjes). Het zijn allen doorstromers: in augustus 1979 begonnen aan hun basisschoolperiode, zonder doublure doorgestroomd en in juni 1985 van het basisonderwijs overgestapt naar het voortgezet onderwijs.

De data zijn verzameld binnen de SVO-projecten 492 'Preventie van Leesmoeilijkheden' en 1126. De leerlingen zijn afkomstig van 24 traditionele basisscholen uit de driehoek Amersfoort/ Enschede/Eindhoven.

4.2. Dataverzameling

In de leerjaren 1 t/m begin 6 zijn de toetsen afgenomen door een aantal speciaal getrainde testleiders. De toetsen aan het eind van het leerjaar 6 zijn afgenomen door de leerkrachten. Alle gebruikte instrumenten bleken te voldoen aan de te stellen criteria.

4.3 Analyseprocedure

De gegevens zijn geanalyseerd met behulp van de LISREL-VI procedure (Jöreskog & Sörbom, 1985²). Deze procedure toetst zowel de globale als locale passing (fit) van modellen waarin zowel latente als geobserveerde variabelen zijn opgenomen. Zowel synchrone als longitudinale invloeden kunnen in een dergelijke procedure verdisconteerd worden (cf. Saris & Stronkhorst, 1984). Het *meetmodel* specificeert hoe de latente variabelen (de theoretische constructen) gemeten worden in termen van de geobserveerde variabelen (of: hoe goed de geobserveerde variabelen een latente variabele representeren). Het *model van de structurele vergelijkingen* specificeert de veronderstelde causale relaties tussen latente variabelen.

De correlatiematrix van de geobserveerde variabelen is gebruikt als input voor de LISREL-procedure.

Een indicator voor de mate waarin een gepostuleerd model overeenkomt met de feitelijke data is de proportie door het model verklaarde variantie, de goodness-of-fit index en de verhouding tussen de chi-kwadraat en het aantal bijbehorende vrijheidsgraden. Deze laatste maat is echter afhankelijk van de steekproefgrootte.

De analyse verloopt als volgt. Eerst wordt het uitgangsmodel getoetst aan de empirische werkelijkheid. Mocht het uitgangsmodel wezenlijk afwijken van de data, dan zal geprobeerd worden door het vrijlaten van bepaalde parameters het model te verbeteren. De keuze van de vrij te laten parameter wordt bepaald door de mate waarin de fit verhoogd wordt, mits, en dat is het overheerstende criterium, de verandering niet indruist tegen de theoretische overwegingen die aan het model ten grondslag liggen. De procedure wordt stapsgewijze toegepast. Na het vrijlaten van een bepaalde parameter worden de overige opnieuw geschat. Aan de hand van de modificatieindices en de genormaliseerde residuele varianties wordt de keuze bepaald. Gestreefd wordt naar een acceptabele oplossing met een beperkt aantal relaties tussen variabelen. Gezien de omvang van de steekproef ($N = 310$) moet verwacht worden dat de gevonden modellen significant zullen afwijken. We zullen echter streven naar (gecorrigeerde) goodness-of-fit indices die liggen boven .900 en een verhouding χ^2/df van 2:1 of minder; pas dan kan gesproken worden van min of meer acceptabele modellen.

5. RESULTATEN

5-1) Voordat de structurele modellen getoetst zijn, is eerst onderzocht of de 18 geobserveerde variabelen die betrekking hebben op de leesvaardigheidsaspecten onder te brengen zijn in 2x3 endogene latente variabelen. Met andere woorden, mogen we om het model te vereenvoudigen de metingen in leerjaar twee en drie samennemen? En kunnen we de twee metingen in leerjaar zes (begin en eind) beschouwen als één meting?

In Tabel 1 is te zien welke relatieve bijdrage de diverse geobserveerde variabelen aan hun factor leveren. In enkele gevallen is toegestaan dat geobserveerde variabelen op meerdere factoren laden.

Tabel 1. Resultaat van de toetsende factoranalyse met 6 factoren.

	afgenomen in leerj.	TL ₂₃	BL ₂₃	LA ₂₃	Factor TL ₆	BL ₆	LA ₆
EMT/A	2	.86					
EMT/A	3	.99					
L&B1b	2		.77				
SO-3	3		.86				
LAS	2			.55			
LAS	3			.84			
EMT-A	6a				.97		
EMT-B	6a				.93		
HGT	6a					.81	
WST	6a					.86	
CLOZE	6a		.27			.61	
STL	6a					.84	
LvT/r	6b					.84	
LvT/c	6b					.82	
LAS-A	6a						.77
LAS-A	6b						.99
LAS-B	6b					— .18	.88
LEESPREF	6b			.20			.65

Een toetsende factoranalyse levert een acceptabel resultaat op: met 117 vrijheidsgraden een χ^2 van 222.13 ($p = 0.000$); Goodness-of-Fit = .926.

5-2) Nu vastgesteld is dat het mogelijk is het aantal latente variabelen te beperken door het combineren van de meetmomenten, gaan we over tot de toetsing van het uitgangsmodel 1. Hierin zijn causale relaties verondersteld tussen de cognitieve leesvariabelen TL en BL en de leesattitude (LA). We nemen aan dat er een invloed uitgaat van enerzijds TL en BL op anderzijds LA. De toetsing van het uitgangsmodel levert de volgende oplossing op: $\chi^2/175 = 354.27$ ($p = .000$); GoF = .902; AG (gecorrigeerde Goodness-of-fit) = .871.

Uit de modificatie-indices blijkt dat het model verbeterd kan worden door enkele directe invloeden toe te voegen:

—een invloed van SEXE op resp. BL₂₃ en BL₆;

—een invloed van SM op BL₆.

Verbeteringen zijn dan nog mogelijk door het weglaten van niet-significante relaties:

—van LM naar LA₂₃;

—van SM naar LA₂₃;

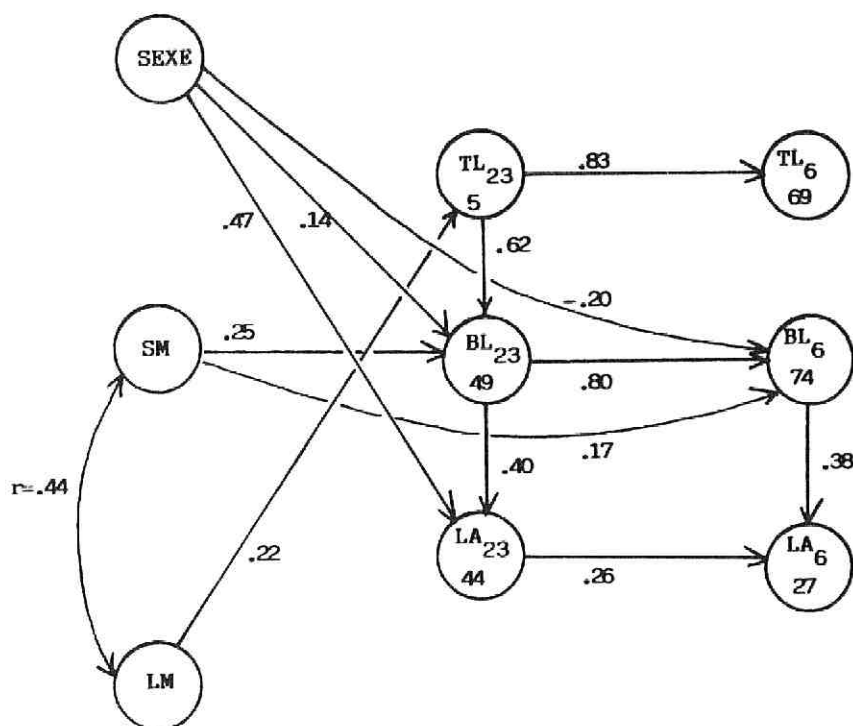
—van TL₂₃ naar LA₂₃;

—van TL₆ naar LA₆;

—van TL₆ naar BL₆.

In Figuur 3 is het uiteindelijke model afgebeeld. De toetsingsgegevens zijn: $\chi^2/174 = 315.55$ ($p = .000$); GoF = .912; AG = .883.

5-3) Op dezelfde manier als hierboven beschreven is, onderzoeken we vervolgens of een alternatief model, waarin verondersteld wordt een invloed van LA op TL en BL, beter of slechter past bij de geobserveerde data. We gebruiken hetzelfde uitgangsmodel, alleen lopen de pijlen van LA₂₃ naar



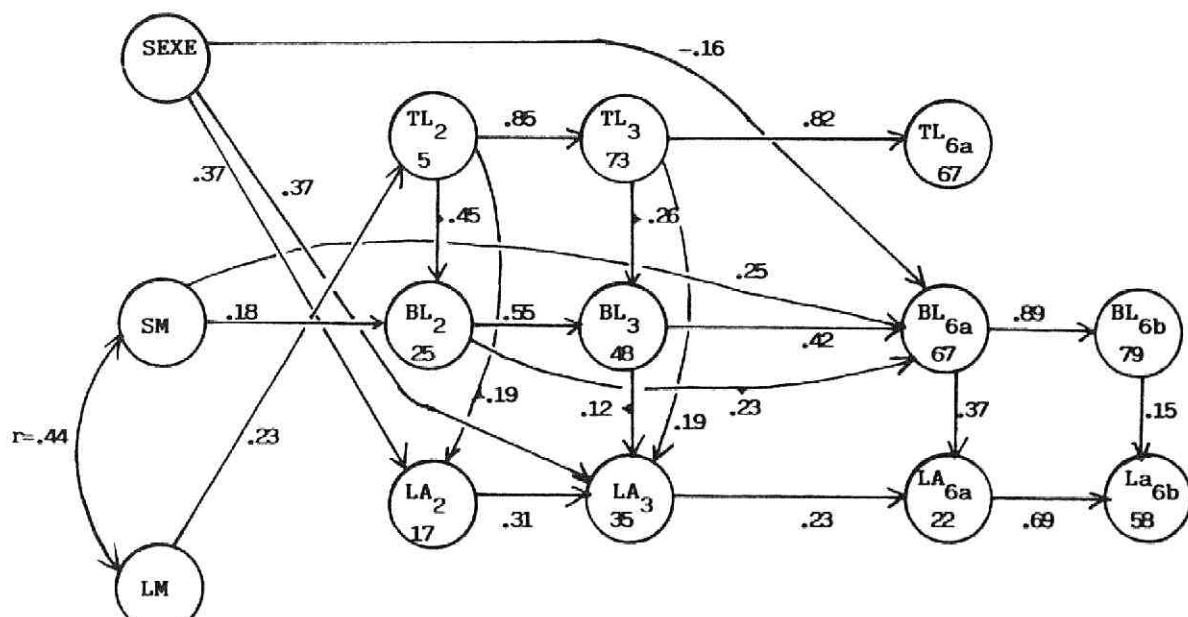
Figuur 3. Het uiteindelijke model met 6 endogene latente variabelen.

Getallen bij de pijlen: gestandaardiseerde coëfficiënten.

Getallen in de cirkels: proportie door het model verklaarde variantie.

TL₂₃ en BL₂₃ en in leerjaar zes loopt de pijl van LA₆ naar BL₆. Dit model geeft het volgende resultaat: $\chi^2/174 = 392.74$ ($p = .000$); GoF = .892; AG = .857. Dit alternatieve model blijkt dus minder goed overeen te stemmen met de data.

5-4) Tenslotte laten we in Figuur 4 zien wat het resultaat is indien we de twee meetmomenten in het begin (leerjaar 2 en 3) en eind (begin en eind leerjaar 6) van het basisonderwijs niet samenvoegen maar als aparte meetmomenten beschouwen. Het nadeel van deze werkwijze is wel dat de diverse latente variabelen in leerjaar 2 en 3 nu slechts vertegenwoordigd zijn door één geobserveerde variabele, zodat met meetfouten geen rekening kan worden gehouden.



Figuur 4. Het uiteindelijke model met 11 endogene latente variabelen.

Getallen bij de pijlen: gestandaardiseerde coëfficiënten.

Getallen in de cirkels: proportie door het model verklaarde variantie.

De toetsingsgegevens van dit uitgebreide model zijn, zoals verwacht kan worden, bijna gelijk aan die uit 5-2): $\chi^2/174 = 307.12$ ($p = .000$); $GoF = .915$; $AG = .887$.

CONCLUSIE EN DISCUSSIE

Keren we nu terug naar de uitgangsvraag. Deze luidde: is het mogelijk via een modelmatige benadering de complexe relatie tussen leesprestatie en leesattitude te verhelderen? Zonder voorbij te gaan aan de beperkingen van dit onderzoek, waarop we aanstonds zullen ingaan, kan gesteld worden dat de poging geslaagd is; het is mogelijk gebleken longitudinale verbanden op te sporen en in kaart te brengen.

Sprekend over de beperkingen die verbonden zijn aan dit onderzoek, maken we een onderscheid tussen meer formele en inhoudelijke aspecten die een juiste interpretatie in de weg kunnen staan. Als eerste moet opgemerkt worden dat een dergelijke Lisrel-procedure een model oplevert dat slechts één uit meerdere mogelijkheden is. De hier gevonden oplossing behoeft niet de enig juiste te zijn. We kunnen hoogstens zeggen dat de hier gevonden modeloplossingen acceptabel zijn omdat ze zich niet verzetten tegen de empirische werkelijkheid. Ze kunnen echter te allen tijde worden ingeruild voor betere. Vervolgens is de vraag aan de orde of de theoretische constructen goed, dat wil zeggen betrouwbaar en valide gemeten zijn. Is het construct 'leesattitude' wel te vangen in het beperkte net van deze geobserveerde variabelen? Is dit net niet te klein of zijn de mazen niet te wijd zodat wezenlijke aspecten van de leesattitude buiten beschouwing blijven?

Vatten we de resultaten samen:

- 1) De invloed van technisch lezen, resp. begrijpend lezen op de leesattitude verschuift in de loop van het basisonderwijs. De invloed van technisch lezen is gering in de eerste leerjaren en neemt in leerjaar zes verder af tot een niet langer significant niveau. Bij begrijpend lezen zien we het tegenovergestelde. In leerjaar twee is de invloed van begrijpend lezen op de leesattitude te verwaarlozen; ze neemt echter in de loop van de tijd aanmerkelijk toe. In leerjaar zes is de directe invloed van begrijpend lezen groter dan die van de leesattitude op eerdere meetmomenten (figuur 3). De indruk dat de specifieke invloed van begrijpend lezen op de leesattitude aan het eind van leerjaar zes weer afneemt (figuur 4) is niet juist: er is een relatief grote invloed van BL_{6a} op LA_{6a} en indirect ook op LA_{6b} .
- 2) Dit onderzoek bevestigt de opvatting dat milieufactoren weinig directe invloed uitoefenen op de ontwikkeling van de leesattitude. Dit zou kunnen betekenen dat de basisschoolperiode en de onderwijsleerprocessen die zich daar afspelen, bepalend zijn voor de ontwikkeling van een positieve danwel negatieve leesattitude. Voor de goede orde, dit resultaat doet geen uitspraak over de milieu-invloeden *tijdens* de periode van het basisonderwijs. We kunnen hoogstens constateren dat vóórschoolse milieu-invloeden alleen (en dan nog in beperkte mate) een invloed uitoefenen op de cognitieve leesprestaties.
- 3) Inderdaad blijken de cognitieve aspecten van de leesvaardigheid (technisch en begrijpend lezen) op termijn gezien een grote mate van stabiliteit te bezitten. Er zijn sterke longitudinale invloeden aangetoond binnen resp. technisch lezen en begrijpend lezen over de vier meetmomenten. De invloed van de leesattitude in het begin van het basisonderwijs op de leesattitude later is veel geringer.
- 4) Dit onderzoek heeft geen antwoord kunnen geven op de vraag waaraan de leesattitude-variantie in leerjaar zes nu moet worden toegeschreven. De proportie door het model verklaarde variantie van LA_6 (figuur 3) is relatief klein. Dit houdt natuurlijk ook verband met de vraag of de leesattitude wel adequaat gemeten is.
- 5) De factor 'sexe' in het geheel verdient nadere aandacht. We wisten al dat in leerjaar zes meisjes gemiddeld een iets hogere score behalen op de leesattitudemetingen, maar er blijken zich ook in de ontwikkeling van de leesattitude verschillen voor te doen. Wanneer we de data van meisjes en jongens apart analyseren, valt op dat bij meisjes de coëfficiënten tussen LA_2 - LA_3 en tussen LA_3 - LA_6 twee keer zo klein zijn als die bij de jongens (tussen leerjaar twee en drie .18 bij de meisjes, .40

bij de jongens; tussen leerjaar drie en zes .15 bij de meisjes en .31 bij de jongens). Dit zou er op kunnen wijzen dat de leesattitude zich bij jongens eerder stabiliseert.

Zoals ieder onderzoek werpt ook dit vragen op die zich lenen voor een verdere overdenking. Een vraag uit de onderwijspraktijk zou kunnen zijn: wat zegt dit resultaat nu over onze pogingen via boekpromotie etc. de leesattitude van onze leerlingen te verhogen. Daarover zegt dit onderzoek niets; we zouden hoogstens kunnen zeggen dat de rechtlijnigheid in de positieve leesspiraal (figuur 1) wat voorbarig is. Het is zeker niet zo dat bij alle leerlingen een toename in de leesprestatie een meer positieve leesattitude tot gevolg heeft.

Een ander punt van aandacht is de vraag of de gemiddelde trends en effecten die we hier vastgesteld hebben, geen afbreuk doen aan de noodzaak aandacht te besteden aan verschillen tussen categorieën leerlingen. We hebben hier te maken met doorstromers, leerlingen zonder een breuk in hun schoolloopbaan. Maar zijn er misschien speciale groepen leerlingen aan te wijzen bij wie dit gevonden model volstrekt niet of juist bijzonder goed opgaat? Eén aspect, het sexe-verschil, is hier ter sprake gekomen, maar hoe staat het met verschillen in leeftijd of schoolloopbaan (doorstromers versus zittenblijvers)?

Zoals in de inleiding al is gezegd, is onderzoek naar de *ontwikkeling* in de relatie tussen leesprestatie en leesattitude zeldzaam. Naar ons weten is het onderzoek waarover hier is gerapporteerd het eerste dat systematisch longitudinale verschillen in de relaties tussen de diverse componenten van de leesvaardigheid heeft opgespoord. Daarmee hebben we geprobeerd een kleine bijdrage te leveren aan de theoretische en praktische uitbouw van inzichten rond de ontwikkeling van een complex begrip als de leesattitude en attitudevorming in het algemeen.

REFERENTIES

- Aarnoutse, C.A.J. Buitenhuis, A., Lucas, H., & Poos-Akerboom, M. (1976). *Doelstellingen van het leesonderwijs*. Deelrapport 2 SVO-0298. Nijmegen: Instituut voor Onderwijskunde.
- Aarnoutse, C.A.J. (1984). *Hoofdgedachtetest*. Test voor begrijpend lezen bestemd voor het vierde en vijfde leerjaar van het basisonderwijs. Verantwoording en Handleiding. Nijmegen: Berkhout BV.
- Aarnoutse, C.A.J., & Boland, Th. (1987). Schalen voor het meten van de leesattitude en de leesvoorkeur. Nijmegen: Berkhout, Testmateriaal.
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations. A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888-918.
- Allport, G.W. (1935). Attitudes. In: C.M. Murchison (ed.), *Handbook of social psychology*. Worcester (Mass.): Clark University Press.
- ARBO (1984). *Het moet ons een zorg zijn*. Advies over zorgverbreding in het basisonderwijs. Zeist: Onderwijscentrum.
- Bentler, P.M., & Speckart, G. (1979). Models of Attitude-Behavior Relations. *Psychological Review*, 86, 5, 425-464.
- Bisschop, P., Aarnoutse, C.A.J., & Feenstra, H. (1985). *Leesattitudeschalen*. Bestemd voor het derde, vierde en vijfde leerjaar van het basisonderwijs. Verantwoording en Handleiding. Nijmegen: Berkhout BV.
- Boland, Th. (1984). *Cloze test voor begrijpend lezen; experimentele versie*. Nijmegen: Instituut voor Onderwijskunde.
- Boland, Th., & Mommers, M.J.C. (1985). *Lees- en leerprestaties in het zesde leerjaar van het lager onderwijs*. Deelrapport 2, SVO-1126. Nijmegen: Instituut voor Onderwijskunde.
- Boland, Th., & Mommers, M.J.C. (1987). *De ontwikkeling van de leesprestatie en de leesattitude in longitudinaal perspectief*. Deelrapport 2, SVO-6564. Nijmegen: Instituut voor Onderwijskunde.
- Boland, Th., & Mommers, M.J.C. (1986). Development of Reading Skills in Dutch Primary Education: Outcomes and Prospects. Paper presented on the AERA-conference San Francisco.
- Boland, Th., & Aarnoutse, C.A.J. (1986). De relatie tussen leesattitude en leesprestatie. ORD-paper 1985. Nijmegen: Instituut voor Onderwijskunde.
- Boxtel, H. van, Snijders, J.Th. & Welten, V.J. (1982). *ISI-Reeks: Interesse, Schoolvorderingen en Intelligentie; Vorm III*. Publicatie 7: Verantwoording en handleiding. Groningen: Wolters/Noordhoff.
- Brus, B.Th., & Voeten, M.J.M. (1973/1979). *Eén-Minut-Test; Vorm A en B*. Schoolvorderingentest voor de technische leesvaardigheid, bestemd voor het tweede t/m zesde leerjaar van het basisonderwijs. Nijmegen: Berkhout BV.

De grootte van schooleffecten: Hoe verschillend presteren leerlingen van verschillende scholen?*

H. Blok en M. H. Eiting

Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam.

ABSTRACT

THE SIZE OF SCHOOL EFFECTS ASSESSED BY THE INTRACLAS CORRELATION COEFFICIENT.

It is a well-established fact that pupils of different schools perform differently. The size of such school effects may be evaluated by the intraclass correlation coefficient (icc), as will be demonstrated in this study. The data consist of a representative sample of 113 Dutch primary schools which took part in a national assessment of mothertongue education. In the grand population of all schools the school effects appeared to be more modest (the icc's for the 33 different language tests varied from $-.03$ to $.23$) than was expected. In one subpopulation (consisting of the 25 percent of schools which receive additional governmental funds, however, the school effects were much larger (icc's between $.07$ and $.39$). This result might be the consequence of the variation in the intake percentage of immigrant children. The amount of instruction time and grouping practices proved to be unrelated to the size of school effects.

1. INLEIDING

Niet alleen presteert de ene leerling meer dan de andere. Ook leveren leerlingen van de ene school gemiddeld betere prestaties dan leerlingen van de andere. Het eerste verschijnsel wordt geïndiceerd door de variantie in de scores die leerlingen op studietoetsen behalen. Het tweede verschijnsel blijkt wanneer men de scores van leerlingen op studietoetsen per school middelt en vervolgens de schoolgemiddelden met elkaar vergelijkt. Het Cito bij voorbeeld berekent jaarlijks schoolgemiddelden op basis van de prestaties van leerlingen op de Eindtoets Basisonderwijs.

Verschillen in de schoolgemiddelden kunnen een uiteenlopende ontstaansgrond hebben. We onderscheiden er vier, die zowel afzonderlijk als in combinatie werkzaam kunnen zijn (Kish, 1965, p. 163):

- Scholen kunnen een ongelijke instroom hebben: sommige scholen trekken relatief veel leerlingen die het leren gemakkelijk aangaat, die thuis een stimulerende leeromgeving hebben of die al veel voorkennis hebben.
- De doeltreffendheid van het door scholen gegeven onderwijs kan variëren: sommige scholen weten met meer succes onderwijs te geven dan andere.
- Van de leerlingen die de school bevolken kan een invloed uitgaan op de prestaties van individuele leerlingen ('mutual interaction within schools').
- Toevallige factoren kunnen werkzaam zijn, bij voorbeeld als de betrouwbaarheid van de studietoets te wensen overlaat of als het schoolgemiddelde berekend wordt op basis van een steekproef van leerlingen (en niet op basis van alle leerlingen).

* Dit onderzoek werd mogelijk gemaakt door een subsidie van het Instituut voor Onderzoek van het Onderwijs (S.V.O.) te Den Haag.

kan worden (op denominatie, schoolgrootte, aantal verwijzingen naar het Speciaal Onderwijs) is de responsgroep van 158 scholen na terugweging representatief voor de populatie van lagere scholen in het schooljaar 1984/85.

Op elk van de 158 scholen hebben ongeveer 15 zesdeklassers een taakboek met lees- en schrijfoopdrachten gemaakt. Er zijn drie taakboeken (aangeduid met A, B, en C) met in totaal 16 leestaken en 10 schrijftaken. Omdat enkele schrijftaken op verscheidene aspecten gescoord zijn, bedraagt het aantal prestatievariabelen 33. Elk taakboek is in principe door vijf aselekt aangevozen leerlingen per school gemaakt. De afnames werden door de eigen onderwijzers verricht. Tabel 1 bevat enkele gegevens over de afgenomen taken en de prestatiepunten.

De gegevens over het *onderwijsaanbod* zijn verzameld bij een omvangrijker steekproef, die overigens de steekproef voor de meting van de leerlingprestaties insloot. De opzet van de aanbodsteekproef was vrijwel gelijk aan die van de leerlingprestatiesteekproef. Van de 1325 scholen wier medewerking was verzocht hebben er 495 (i.e. 37 procent) daadwerkelijk meegewerkt. Het hoofd van de school heeft een lijst met vragen, verdeeld over 48 rubrieken beantwoord. De leerkracht van de zesde klas heeft een lijst met vragen, verdeeld over 79 rubrieken beantwoord. Voorbeelden van rubrieken zijn opleiding en ervaring als leerkracht, algemene taalaanpak (methode, planning, toetsing, leermiddelen, differentiatie), leerdoelen, taalsituaties, grammatica, spelling en woordenschat. De twee vragenlijsten over het onderwijsaanbod zijn opgenomen in Hoeksma, Wesdorp, Van den Bergh & Bos (1985), evenals de lees- en schrijfoopdrachten voor de leerlingen.

Uitgangspunt voor de hier te rapporteren analyses van de relaties tussen aanbod en leerlingprestaties was de gezamenlijke steekproef van stimulerings- en niet-stimuleringscholen waarvoor zowel aanbod- als prestatiegegevens beschikbaar waren ($n = 158$). Uit deze steekproef zijn de scholen verwijderd met relatief veel ontbrekende gegevens over het onderwijsaanbod ($n = 17$). Van de resterende scholen zijn er 56 een stimulerings- en 85 een niet-stimuleringschool. Om de populatie-verhouding te bereiken, werden voor sommige analyses nog 28 stimuleringscholen aselekt uit de steekproef verwijderd. De uiteindelijke steekproef omvat 113 scholen, 85 (75 procent) niet-stimuleringscholen en 28 (25 procent) stimuleringscholen.

3. VOORSPELLINGEN

In de inleiding hebben we vier verklaringsgronden onderscheiden voor het verschijnsel dat schoolgemiddelden variëren. Bij drie van de vier hebben we een specifieke verwachting geformuleerd die getoetst kan worden aan de gegevens uit de Voorstudie P.P.O.N.

a. Het effect van een selectieve instroom

Een school met een selectieve instroom is een school waar vooral leerlingen naar toe gaan die op elkaar lijken, bij voorbeeld wat betreft de reeds verworven kennis en vaardigheden (de voorkennis), maar ook wat betreft hun eigenschappen om te leren (de leergeschiktheid). Een school die veel allochtone leerlingen trekt, is een voorbeeld van een school met een selectieve instroom, omdat zulke leerlingen van hun anders-taligheid duidelijk nadeel ondervinden. Hoe selectiever de instroom van scholen in een bepaalde populatie, des te groter zullen de verschillen tussen scholen zijn en des te hoger ook coëfficiënt rho.

Dit effect kan geëvalueerd worden door rho afzonderlijk te schatten voor de volledige strata van stimuleringscholen ($n=56$) en niet-stimuleringscholen ($n = 85$). De populatie van stimuleringscholen wordt namelijk gekenmerkt door een selectievere instroom dan de populatie van niet-stimuleringscholen. Dat dit zo is, blijkt uit gegevens van de Voorstudie P.P.O.N., die op het schooljaar 1984/85 betrekking hebben. Het valt ook af te leiden uit de door Wijnstra (1987) gepresenteerde gegevens, die betrekking hebben op het daaropvolgende schooljaar 1985/86, het jaar waarin het stimuleringsbeleid opgevolgd is door het onderwijsvoorrangsbeleid. De selectieve instroom heeft niet zozeer betrekking op het percentage kinderen, waarvan de ouders een laag scholingsniveau hebben¹, als wel op het percentage allochtone leerlingen op

Tabel 1: Overzicht van de taalopdrachten uit de Voorstudie P.P.O.N. (aantal items, Cronbach alpha, aantal leerlingen, gemiddelde score, standaardafwijking en intra-klassecorrelatiecoëfficiënt).

taak	omschrijving	k-items	alpha	N-11n.	$\bar{x}$	s (x)	$\hat{\rho}$
A. 1	aanmeldingsformulier invullen	12	.55	715	10.6	1.48	.051
A. 2	opstel: Wandelende boom	n.v.t.	.83 ^a	712	285.7	28.65	.061
A. 3	krante-artikel met meerkeuzevragen	8	.39	711	5.5	1.42	.078
A. 4	museum- pictogrammen	7	.85	709	5.1	1.97	.086
A. 5	telefoonboek hanteren	7	.54	711	4.3	1.56	.232
A. 6	korte mededeling schrijven	7	.81	704	3.6	2.30	.098
A. 7	ongevalsverslag met open vragen	6	.42	711	2.5	1.26	.094
A. 8	Waddenoogse bos: samenvatting maken	9	.65	694	3.9	2.17	.134
A. 9	kook- en baknotities: aanwijzingen lezen	11	.88	651	7.3	3.45	.128
B. 1	aanmeldingsformulier invullen	12	.46	710	10.5	1.40	.160
B. 2	briefje aan je ouders schrijven	4	.29	715	3.0	.97	.198
B. 3a	argumentatief opstel: inhoud	n.v.t.	.87 ^a	713	286.3	42.92	— .003
B. 3b	argumentatief opstel: taalgebruik	n.v.t.	.87 ^b	713	192.0	23.16	.110
B. 4	tekst haas/konijn: meerkeuze- en open vragen	16	.68	712	11.1	2.93	.069
B. 5	radio/tv-gids opzoeken	11	.76	715	9.1	2.11	.053
B. 6a	bestelbriefje schrijven: inhoudselementen	12	.80	706	6.6	3.01	.234
B. 6b	bestelbriefje schrijven: gramm. fouten	n.v.t.	.77 ^c	703	.05	.04	.141
B. 6c	bestelbriefje schrijven: spelfouten	n.v.t.	.92 ^c	703	.04	.05	.160
B. 7	paddestoelentekst met open vragen	8	.74	712	6.2	1.95	.060
B. 8a	instructie schrijven doolhof: inhoud	n.v.t.	.93 ^a	703	299.5	46.26	.111
B. 8b	instructie schrijven doolhof: taalgebruik	n.v.t.	.91 ^b	705	197.4	30.76	.159
B. 9	onderschriften bij konijnenstrip	6	.90	701	5.3	1.59	— .033
C. 1	briefje aan V.V.V. schrijven (inhoudselementen)	9	.64	706	4.2	2.15	.220
C. 2	grafisch materiaal lezen (open vragen)	5	.64	712	2.7	1.45	.109
C. 3	inschrijfformulier invullen	13	.79	714	9.5	2.72	.071
C. 4	boot/busdienst regeling lezen (open vragen)	6	.78	703	3.4	2.02	.174
C. 5	duinentekst met meerkeuzevragen	6	.50	707	2.9	1.48	.060
C. 6a	beschrijven kampeerboerderij: inhoud	n.v.t.	.92 ^a	701	259.3	47.07	.232
C. 6b	beschrijven kampeerboerderij: taalgebruik	n.v.t.	.88 ^b	702	172.6	32.44	.144
C. 7	een route beschrijven (inhoudselementen)	5	.70	708	2.8	1.64	.091
C. 8	museumtekst met meerkeuzevragen	15	.66	705	10.6	2.79	.131
C. 9a	opstel.stemming: inhoud	n.v.t.	.79	693	248.8	37.53	.020
C. 9b	opstel.stemming: taalgebruik	n.v.t.	.85	693	174.3	29.41	.085

^a homogeniteit van een jury van drie onafhankelijke beoordelaars
^b produkt-momentcorrelatie tussen twee onafhankelijke beoordelaars
^c p.m.c. op basis van een steekproef van ongeveer 100 herbeoordeelde schrijfproducten.

school. Dat *varieert* op stimuleringscholen van 0 tot 86 procent (geschat op basis van de gegevens uit de Voorstudie P.P.O.N.) en op de daarmee vergelijkbare scholen² een jaar later van 6 tot 96 procent (volgens Wijnstra, o.c., p. 101). Op niet-stimuleringscholen is de *variatie* in het percentage allochtone leerlingen veel geringer, namelijk van 0 tot 27 procent (volgens de Voorstudie P.P.O.N.) en een jaar later van 1 tot 6 procent (volgens Wijnstra, o.c.). Het feit dat op stimuleringscholen de *variatie* in het percentage instromende allochtone leerlingen groot is, zorgt er in combinatie met de achterstand in schoolprestaties van allochtone leerlingen voor dat de schoolgemiddelden van stimuleringscholen meer van elkaar zullen verschillen dan van niet-stimuleringscholen. Dit voert ons naar de eerste voorspelling.

Voorspelling 1: Coëfficiënt rho zal binnen het stratum van stimuleringscholen hoger zijn dan binnen het stratum van niet-stimuleringscholen.

b. Het effect van het gezamenlijke onderwijs

In het denkbeeldige geval dat geen enkele school aandacht zou schenken aan de kennis en vaardigheden die door een bepaalde studietoets bestreken worden, zouden verschillen in schoolgemiddelden niet groter of kleiner worden dan ze al zijn op grond van andere effecten. Neemt echter in een populatie van scholen de hoeveelheid bestede aandacht toe, dan kan de doeltreffendheid van het onderwijs een rol gaan spelen. We veronderstellen dat het onderwijs niet op alle scholen even doeltreffend gegeven wordt, hetgeen zich dan kan uiten in verschillen in schoolgemiddelden.

In de Voorstudie P.P.O.N. zijn niet alleen uiteenlopende lees- en schrijftaken aan leerlingen voorgelegd. Ook is aan leerkrachten gevraagd, hoe zij hun tijd verdelen over de diverse onderdelen van het leerplan. Als operationalisatie voor de factor 'hoeveelheid onderwijs' is het aantal minuten genomen dat de school - volgens opgave van de leerkracht van de zesde klas - wekelijks besteedt aan gerichte lees- en schrijfoefeningen. Het zogenaamde recreatieve lezen en schrijven blijft hierbij buiten beschouwing. De steekproef van 113 scholen is gedichotomiseerd op de mediaan (150 minuten), behorende bij de verdeling over alle ruim 400 scholen waarvan dit gegeven bekend was. Zodoende is een groep van 'veel taaltijd-scholen' ($n = 40$) gevormd en een groep van 'weinig taaltijd-scholen' ($n = 68$), samen 108 scholen. Van vijf scholen ontbrak het taaltijd-gegeven. In de 'veel taaltijd-scholen' wordt wekelijks gemiddeld 211 minuten aan lezen en schrijven besteed (de standaardafwijking bedraagt 41 minuten), in de 'weinig taaltijd-scholen' bedraagt het gemiddelde ongeveer de helft daarvan, namelijk 112 minuten bij een standaardafwijking van 29 minuten.

Nagegaan is of er contaminatie zou kunnen optreden met het effect van stimulering/niet-stimulering. Gezien het zwakke verband tussen veel/weinig-taaltijd en wel/niet-stimulering dat echter uitgesloten (contingentie C bedraagt .13).

Voorspelling 2: Coëfficiënt rho zal binnen de groep van 'veel taaltijd-scholen' hoger zijn dan binnen de groep van 'weinig taaltijd-scholen'.

c. Het effect van interactie binnen scholen

Er zijn aanwijzingen dat de prestaties van leerlingen afhangen van de samenstelling van de klas of de school waartoe zij behoren. Leerlingen zouden er gemiddeld genomen van profiteren als hun medeleerlingen goede prestaties leveren. Mensen & Guldemond (1987) vonden dit effect in de door Meijnen (1982) verzamelde gegevens, die betrekking hebben op het Lager Onderwijs in de stad Groningen. Het door hen gerapporteerde effect lijkt sterker voor rekenen te gelden dan voor het schoolvak Nederlandse taal, hetgeen zou kunnen samenhangen met het feit dat de beheersing van de Nederlandse taal veel sterker dan bij rekenen het geval is, een invloed ondervindt van de buitenschoolse omstandigheden: taal wordt ook thuis en op straat geleerd. Andere aanwijzingen voor een verband tussen samenstelling van de klas en individuele presta-

ties komen uit onderzoek van Beckerman & Good (1981) en Leiter (1983). Recent is door De Vos (1986) een model voorgesteld ter verklaring van dit verband. Over dit model, een toepassing van de nutstheorie, is een discussie gaande waarvan de uitkomst nog onbekend is (Winkels, 1987; Kalmijn, 1987; De Vos, 1987).

Als we aannemen dat de samenstelling van de groep medeleerlingen inderdaad van invloed is op de individuele prestaties, dan vormt de manier waarop scholen hun leerlingen groeperen een relevant gegeven. In scholen waarin géén groepering naar niveau ('streaming, setting') plaatsvindt, blijft de variantie in prestaties beperkt, doordat de snelle en de trage leerlingen ongeveer hetzelfde aanbod aan onderwijs ontvangen. Naarmate echter scholen meer of vaker niveau-groepen vormen, zal het onderwijsaanbod aan de leerlingen meer variatie gaan vertonen. Elke groep ontvangt onderwijs op haar niveau. Dat heeft gevolgen voor de variatie in prestaties binnen de school: die variatie zal eveneens toenemen. Daarom zullen prestaties sterker variëren binnen scholen waar leerlingen naar niveau gegroepeerd worden dan op klassikale scholen, waar leerlingen niet naar niveau gegroepeerd worden. Dit verschijnsel is inderdaad waargenomen (Shavit & Williams, 1985). Groepering naar niveau daarentegen hoeft niet noodzakelijk tot toename van verschillen tussen scholen te leiden. Toenemende variantie binnen scholen bij gelijkblijvende variantie tussen scholen betekent een afname van de waarde van coëfficiënt rho (zie par. 4).

Van de scholen die aan de Voorstudie P.P.O.N. hebben deelgenomen, is vastgesteld hoe zij hun leerlingen bij het taalonderwijs groeperen. Daarbij is gebruik gemaakt van de antwoorden die de leerkrachten zesde klas verstrekt hebben op drie vragen daaromtrent in de aanbodvragenlijst. De drie vragen hadden betrekking op groepering wat betreft drie aspecten: respectievelijk het leerstofaanbod, de verwerking van de leerstof en de beoordeling van de prestaties. De leerkrachten hebben voor elk van de drie aspecten aangegeven of zij hun leerlingen in dat opzicht naar niveau groepeerden (score 1) of niet (score 0). De antwoorden op de drie vragen zijn gecombineerd tot een schaalscore (bereik 0-3) en vervolgens is de steekproef scholen gedichotomiseerd op de mediaan van de schaalscore³. De groep n-scholen bestaat uit 54 scholen die op alle drie de aspecten naar niveau groeperen of op twee daarvan (in dit geval voornamelijk wat betreft aanbod en verwerking). De groep k-scholen bestaat uit 48 scholen die géén niveaugroepering toepassen of die wat betreft één aspect naar niveau groeperen (voornamelijk wat betreft aanbod). Van 11 scholen uit de steekproef van 113 scholen ontbraken de gegevens over de groeperingswijze. Er is statistisch geen verband tussen wel/niet-stimulering en de schaalscore voor niveaudifferentiatie (contingentie *C* bedraagt .19). Ook de samenhang tussen de variabele niveaudifferentiatie en veel/weining taaltijd is zwak (*C* = .09).

Voorspelling 3: Coëfficiënt rho zal hoger zijn in de groep klassikale scholen dan in de groep niveau-differentiërende scholen.

d. Het effect van toeval

Indien de verschillen tussen scholen volledig op toeval berusten, zal coëfficiënt rho gelijk zijn aan nul (zie par. 4). Het toevalseffect wordt hier geëvalueerd door te toetsen of rho in de steekproef statistisch significant van nul afwijkt.

4. DE INTRA-KLASSE CORRELATIECOËFFICIËNT RHO

Volgens Kish (1965, p. 170) is de intra-klasse correlatiecoëfficiënt rho als volgt gedefinieerd voor de populatie:

$$\rho = \frac{\sigma_{\mu}^2 - \sigma_{\text{binnen}}^2 / (B - 1)}{\sigma_{\text{totaal}}^2} \quad (1)$$

waarin σ_μ^2 de variantie is van de klasse-gemiddelden μ in de populatie, σ_{binnen}^2 de variantie van de scores binnen de klassen, σ_{totaal}^2 de variantie van de scores in de gehele populatie en B de grootte der klassen. Aangenomen is dat de variantie binnen de klassen constant is voor alle klassen (homogeniteit van varianties) en dat de grootte der klassen eveneens constant is.

Vergelijking 1 laat zich op een eenvoudiger manier herschrijven. Uit (1) en

$$\sigma_{\text{totaal}}^2 = \sigma_\mu^2 + \sigma_{\text{binnen}}^2 \quad (2)$$

volgt

$$\hat{\rho} = \frac{B}{(B-1)} \cdot \frac{\sigma_\mu^2}{\sigma_{\text{totaal}}^2} - \frac{1}{(B-1)} \quad (3)$$

Een toestand van extreme heterogeniteit (Kish, 1965, p. 171) houdt in dat alle klassegemiddelde precies gelijk aan elkaar zijn ($\sigma_\mu^2 = 0$, $\sigma_{\text{totaal}}^2 = \sigma_{\text{binnen}}^2$): alle scorevariantie is dan binnen de klassen aan te treffen. Een toestand van extreme homogeniteit betekent dat alle scores binnen klassen gelijk aan elkaar zijn ($\sigma_{\text{binnen}}^2 = 0$, $\sigma_{\text{totaal}}^2 = \sigma_\mu^2$): alle scorevariantie is tussen de klassen aan te treffen. Bijgevolg geldt:

$$-\frac{1}{(B-1)} \leq \rho \leq 1 \quad (4)$$

Op basis van Vergelijking 3 kunnen verschillende vergelijkingen voor een zuivere schatter van rho worden afgeleid, afhankelijk van enkele aannamen.

Stel dat men in een eerste trap aselekt een steekproef van a klassen trekt uit een populatie van A klassen en dat men in een tweede trap uit elke klas aselekt een aantal van b leerlingen trekt uit het totale aantal van B leerlingen. Neemt men aan dat de totale steekproefomvang $n = a * b$ voldoende groot is, dan kan bewezen worden (zie Kish, 1965, p. 171-173) dat

$$\hat{\rho} = \frac{b}{(b-1)} \cdot \frac{s_{\bar{x}}^2}{s_{\text{totaal}}^2} - \frac{1}{(b-1)} \quad (5)$$

een zuivere schatter voor rho is. Hierin is $s_{\bar{x}}^2$ de variantie van de gemiddelden $\bar{x}$ in de steekproeven uit de klassen berekend over de steekproef van klassen, s_{totaal}^2 de variantie van scores in de gehele steekproef en b het aantal leerlingen dat per klas getrokken is. Voor de extremen van de zuivere schatter van rho geldt een vergelijking analoog met Vergelijking 4, waarbij men B vervangt door b .

Ter illustratie van het effect van toeval (zie par. 3) wordt hier het aselekt-model voorgesteld. Men zou kunnen veronderstellen dat klassen in de populatie ontstaan als waren zij aselekte steekproeven uit de populatie van alle leerlingen. Aannemend dat het aantal klassen voldoende groot is, kan men bewijzen dat:

$$\sigma_\mu^2 = \sigma_{\text{totaal}}^2 / B \quad (6)$$

Substitutie van Vergelijking 6 in Vergelijking 3 levert dan een intraklasse correlatie op van nul. De verwachting van de schatter voor rho volgens Vergelijking 5 is in dat geval ook nul; ook wanneer we uit elke klas een steekproef van leerlingen trekken.

Voor de statistische evaluatie van de op basis van een steekproef geschatte rho baseren we ons

op Kish (1965, p. 173). Bewezen kan worden dat, onder het zojuist genoemde aselekt-model en onder de extra aanname dat de scores in de populatie normaal zijn verdeeld, de statistische grootheid F in de volgende vergelijking:

$$\hat{\rho} = \frac{F - 1}{F - 1 + b} \quad (7)$$

een F -verdeling bezit met $df_1 = (a-1)$ en $df_2 = a(b-1)$ vrijheidsgraden. Bij het hanteren van een tabel voor de F -verdeling (zie bij voorbeeld Hays, 1977, appendix C), dient men strikt genomen de tweede vrijheidsgraden-term te vervangen door ∞ . Deze aanname is immers gedaan bij het opstellen van Vergelijking 5.

Een toepassing van Vergelijking 7 laat zich als volgt illustreren. We trekken aselekt een steekproef van 61 scholen ($a = 61$) uit de scholenpopulatie en we trekken per school aselekt een steekproef van 5 leerlingen ($b = 5$). Onder het aselekt-model en de normaliteitsaanname is de kans 95 procent dat we een F vinden kleiner dan of gelijk aan 1.32 ($df_1 = 60$, $df_2 = \infty$). Substitutie in Vergelijking 7 geeft dan een kritieke ρ van .06.

Bij de interpretatie van ρ kan men de vuistregel hanteren dat een betrekkelijk kleine waarde van deze coëfficiënt reeds correspondeert met een aanzienlijk effect in termen van verschillen tussen scholen of klassen. In Figuur 1 geven we voor een op leerlingniveau standaardnormaal verdeelde variabele een illustratie bij ρ , geschat zowel op basis van volledig getrokken klassen van 30 leerlingen ($B = 30$, zie Vergelijking 3) als op basis van steekproeven van vijf leerlingen per klas ($b = 5$, zie Vergelijking 5) en wel voor drie waarden, respectievelijk .05, .25 en .50. De curve die de verdeling met de kleinste variantie representeert, geeft steeds de verdeling van de klassemiddelen weer.

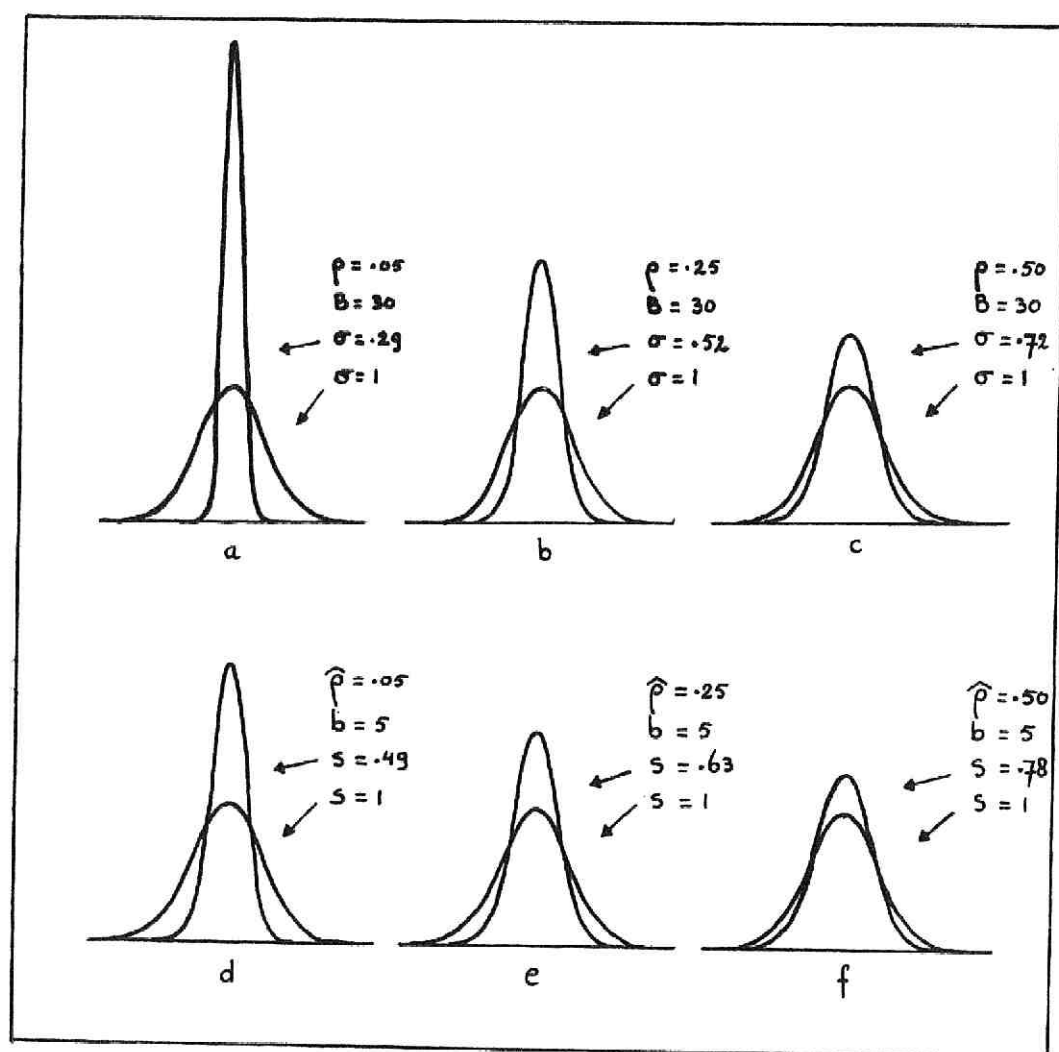
5. RESULTATEN

Voorafgaande aan de resultaten betreffende de in paragraaf 3 ontwikkelde voorspellingen, geven we een overzicht van geschatte ρ 's in de representatieve steekproef (113 scholen, gemiddeld 4.4 leerlingen per school). Deze zijn opgenomen in Tabel 1. Omdat een geschatte ρ van .05 significant is bij een overschrijdingskans van .05, wijken 30 van de 33 ρ 's zeer waarschijnlijk af van nul. De laagste schatting bedraagt -.033, de hoogste .234 en de mediane .109.

Stimulerings- versus niet-stimuleringscholen

De eerste voorspelling die we geuit hebben, is dat stimuleringscholen onderling qua leerling prestaties een grotere verscheidenheid zullen vertonen dan niet-stimuleringscholen. Deze hypothese is geëvalueerd door ρ afzonderlijk te schatten voor het volledige stratum van stimulerings- en niet-stimuleringscholen. Dat is voor elk van de 33 in Tabel 1 genoemde prestatievariabelen gebeurd, waarbij Vergelijking 5 als schatter is gebruikt. Voor grootheid b , het aantal per klas getrokken kinderen, is het empirisch bepaalde (per boekje over alle taalopdrachten) gemiddelde aantal kinderen per klas dat enige taak gemaakt heeft, gesubstitueerd.

Voor de stimuleringscholen blijkt de voor de 33 prestatievariabelen geschatte ρ te variëren van .069 tot .386, terwijl het gemiddelde .241 bedraagt met een standaardafwijking van .077. De 33 ρ 's wijken alle statistisch significant af van nul (bij een overschrijdingskans van .05). De 33 ρ 's voor de niet-stimuleringscholen variëren van -.027 tot .213, waarbij het gemiddelde .065 is met een standaardafwijking van .063. Voor niet-stimuleringscholen wijken maar 15 van de 33 ρ 's statistisch significant van nul af. Uit deze cijfers blijkt reeds dat de gegevens een ondersteuning vormen voor de eerste voorspelling. De ondersteuning is echter nog sterker, omdat voor alle afzonderlijke 33 prestatievariabelen blijkt te gelden dat de geschatte ρ voor stimuleringscholen groter is dan of gelijk aan de geschatte ρ voor niet-stimuleringscholen.



Figuur 1: Enkele illustraties van de verhouding tussen de variantie van leerlingsscores (hier steeds standaardnormaal verdeeld) en de variantie van schoolgemiddelden voor drie gevallen: rho is respectievelijk .05, .25 en .50; zowel indien de gehele klas van $B = 30$ getrokken wordt (a, b en c) als indien per klas een steekproef van $b = 5$ leerlingen getrokken wordt (d, e en f).

Men zie Tabel 2, die in de vorm van een 'stem-and-leaf display' (Tukey, 1977) de frequentieverdeling geeft van de verschillen: de rho's voor de stimuleringscholen minus de rho's voor de niet-stimuleringscholen.

Men dient het gemiddelde verschil van .18 te interpreteren als vrij aanzienlijk (vgl. Figuur 1). Het betekent dat de stimuleringscholen, die in 1984/85 ongeveer 25 procent van de totale populatie van lagere scholen uitmaakten, veel grotere onderlinge verschillen vertonen in gemiddelde schoolprestaties van hun leerlingen, dan de niet-stimuleringscholen. Voor de

Tabel 2: 'Stem-and-leaf display' van de verschillen: rho voor stimuleringscholen minus rho voor niet-stimuleringscholen (voor 33 prestatievariabelen).

.0	0 0 7 8 8 9	gem. = .18
.1	0 1 2 2 3 4 5 5 7 7 7 9 9	s.d. = .09
.2	0 0 1 1 2 3 4 4 5 6 7	
.3	1 7 9	

laatste groep blijkt te gelden dat de onderlinge verschillen vaak verwaarloosbaar zijn en in de helft van de gevallen waarschijnlijk zelfs op toeval berusten.

Meer aanbod versus minder aanbod

Als tweede voorspelling hebben we geformuleerd dat scholen die relatief veel tijd besteden aan het onderwijs in de Nederlandse taal, een grotere onderlinge verscheidenheid zullen vertonen qua bereikte leerresultaten dan scholen die daaraan relatief weinig tijd besteden. Deze veronderstelling is geëvalueerd door coëfficiënt rho afzonderlijk te schatten voor de groep 'veel taaltijd-scholen' en de groep 'weinig taaltijd-scholen'.

Voor de groep 'weinig taaltijd-scholen' blijkt de voor de 33 taalopdrachten geschatte rho te variëren van $-.018$ tot $.310$ met een gemiddelde van $.126$ en een standaarddeviatie van $.080$. Voor de groep 'veel taaltijd-scholen' varieert de geschatte rho van $-.072$ tot en met $.300$ met een gemiddelde van $.101$ en een standaarddeviatie van $.100$. Hieruit valt te concluderen dat de verschillen tussen de beide verdelingen verwaarloosbaar klein zijn. Dezelfde conclusie valt te trekken, wanneer per taalopdracht de beide rho's vergeleken worden. Daartoe is steeds het verschil bepaald (zie Tabel 3 voor een 'stem-and-leaf display'). Het gemiddelde verschil van $-.03$ is te klein om er enige betekenis aan te hechten.

Tabel 3: 'Stem-and-leaf display' van de verschillen: rho voor 'veel taaltijd-scholen' minus rho voor 'weinig taaltijd-scholen' (voor 33 prestatievariabelen).

— .2	0 2 5	gem. = $-.03$
— .1	0 1 1 1 1 4 5 9	s.d. = $.12$
— .0	3 3 4 5 7 7 9	
.0	1 1 4 5 5 6 8 8 8 9	
.1	1 1 1 2	
.2	5	

Alvorens te besluiten dat de tweede hypothese in onze data geen steun vindt, zijn additionele analyses aangewezen. Men zou kunnen veronderstellen dat de gedane analyse te globaal is door het algemene karakter van de vraag naar de tijdsbesteding. Het kan bij voorbeeld zijn dat in de 'veel-taaltijd-groep' scholen opgenomen zijn die aan sommige specifieke taaldoelstellingen zoals die in de taalopdrachten geoperationaliseerd zijn, juist weinig onderwijstijd besteden. Met het oog op deze mogelijkheid is de steekproef scholen per taalopdracht opnieuw gedichotomiseerd. Daarbij is gebruik gemaakt van gegevens die leerkrachten verstrekt hebben over de hoeveelheid aandacht die zij aan een aantal nader omschreven doelstellingen besteden. De vragen hierover, die eveneens opgenomen waren in de aanbodsvragenlijst voor de leerkracht van de zesde klas, waren opzettelijk zo geformuleerd dat er een directe koppeling te maken is tussen de doelstellingen en de taalopdrachten. Per taalopdracht kon zodoende een groep scholen onderscheiden worden die aan de bijbehorende doelstelling veel aandacht schonken en een groep scholen die daaraan weining aandacht schonken. Het resultaat van de op deze wijze verrichte evaluatie is echter gelijk aan de vorige. Ook zo beschouwd vindt de tweede hypothese in onze data geen enkele steun. Er blijkt geen verband te bestaan tussen de hoeveelheid onderwijstijd of aandacht en de grootte van rho.

Klassikale aanpak versus groepering naar niveau

Als derde voorspelling hebben we geformuleerd dat scholen met een klassikale aanpak een grotere onderlinge verscheidenheid zullen vertonen wat betreft de onderlinge prestaties van hun leerlingen dan scholen waar het onderwijs in niveaugroepen gegeven wordt. Deze voorspelling is geëvalueerd door rho afzonderlijk te schatten voor de groep klassikale scholen (k-scholen) en de groep niveaugroeperende scholen (n-scholen). Tegen de derde veronderstelling in blijkt dat de rho's binnen de groep klassikale scholen in het algemeen niet hoger, maar lager zijn dan de

zaakvakken. Door deze omstandigheid kan niet worden uitgesloten dat de verschillen tussen scholen voor andere vakken dan taal groter zijn. In dit verband kunnen de in 1988 te verwachten P.P.O.N.-resultaten, die op rekenen betrekking hebben, belangwekkend zijn.

Ook al is de intra-klasse correlatiecoëfficiënt in het algemeen laag gebleken, er viel toch een systematische variantie van gemiddelden te bespeuren. Van de drie door ons veronderstelde effecten is alleen de eerste, het effect van wel/niet-stimulering, aangetoond. Het effect van bestede tijd aan taalonderwijs is nul en het effect van niveaugroepering is tegengesteld aan onze verwachting.

Het effect van wel/niet-stimulering is hier gebleken het grootste effect te zijn. Dat is op te vatten als een duidelijk steun voor de selectieve instroom-hypothese (zie par. 3). Nederlandse basisscholen verschillen niet zo erg van elkaar in de taalbeheersing van hun leerlingen. Maar voor zover er verschillen zijn, hebben deze waarschijnlijk meer te maken met het effect van een selectieve instroom dan met het effect van het gezamenlijke aanbod. Deze constatering is niet nieuw voor wie zich realiseert, dat de variatie in het onderwijsaanbod door de reglementering van de overheid geringer is dan de variatie in de kwaliteit van de instroom. Ouders zijn vrij in de keuze van de school waar zij hun kind inschrijven. Als gevolg van die vrijheid zijn er scholen waar veel kinderen in achterstandssituaties naar toe gaan naast scholen die door zulke kinderen niet bezocht worden.

De effecten van de selectieve instroom en van het gezamenlijke onderwijsaanbod kunnen ook gedemonstreerd worden door de gemiddelden op de 33 prestatiematen voor de afzonderlijke groepen scholen te berekenen en te vergelijken. Daartoe zijn de prestatiescores getransformeerd tot gestandaardiseerde variabelen en is de effectgrootte geschat, gemiddeld over de 33 prestatievariabelen. Voor de vergelijking wel/niet-stimulering bedraagt de gemiddelde effectgrootte .40 in het voordeel van de niet-stimuleringsscholen. In de terminologie van Cohen (1977) kan men hier van een middelmatig effect spreken. Het voordeel voor de niet-stimuleringsscholen is zeer algemeen, aangezien het zich bij 32 van de 33 prestatiematen voor blijkt te doen.

Voor de vergelijking tussen de veel/weinig taaltijd-scholen bedraagt de gemiddelde effectgrootte .00. De groep 'veel taaltijd-scholen' behaalt op ongeveer evenveel taalopdrachten een hogere gemiddelde score als een lagere gemiddelde score in vergelijking met de groep 'weinig taaltijd-scholen'. Dit resultaat, dat in zeker opzicht verrassend genoemd moet worden, zou tot twijfel kunnen leiden aan de validiteit van de door ons gebruikte variabele onderwijstijd⁴. Ander onderzoek heeft uitgewezen dat er voor zo'n twijfel enige bestaansgrond is. Otter (1987) heeft enkele methoden vergeleken voor de registratie van de tijdbesteding van leerkrachten in het Speciaal Onderwijs. Het blijkt dat totaalschattingen (zoals door ons gebruikt) en meer nauwkeurige tijdregistraties (d.m.v. dagelijks bijgehouden logboeken) betrekkelijk laag correleren. Wij beschouwen daarom het gevonden nuleffect als een dubieus fenomeen.

Het derde door ons veronderstelde effect op ρ - dat van niveaugroepering - bleek tegengesteld aan de verwachting. Binnen de groep scholen waar het minst aan niveaugroepering gedaan wordt, troffen we geen hogere maar juist lagere ρ 's aan. Het is onduidelijk hoe dit verschijnsel verstaan moet worden. In volgend onderzoek moet meer aandacht besteed worden aan de omschrijving van niveaugroepering. Blijkens onze aanbodvragenlijst is er niet sprake van een dichotomie maar van een meervoudige classificatie. In de literatuur blijft vaak onbesproken wat precies het verschil in werkwijze is tussen niveaugroeperende en klassikale scholen (zie bijvoorbeeld Shavit & Williams, 1985). Worden er parallelklassen gevormd of vindt de groepering binnen de klas plaats; op welke aspecten van het onderwijsaanbod wordt de groepering toegepast: instructie, verwerking, beoordeling; of op een combinatie van deze drie aspecten? Zonder een preciezere begripsbepaling is het niet goed mogelijk de resultaten uit verschillend onderzoek met elkaar te combineren. Misschien moet ook de hoedanigheid van de betrokken schoolvakken in de vergelijking betrokken worden. Al eerder is er op gewezen dat de grootste schooleffecten niet bij het vak Nederlandse taal verwacht moeten worden, maar bij vakken die meer expliciet alleen op school geleerd worden: rekenen en in mindere mate de zaakvakken. Van belang is hierbij dat het groeperen van leerlingen naar niveau bij rekenen een veel vaker voorkomend fenomeen is dan bij taal. Kennelijk leent het rekenonderwijs zich veel

beter of gemakkelijker tot niveaugroepering dan het taalonderwijs. Ondanks al deze overwegingen blijft het voorsnog onduidelijk of niveaugroepering nu positief of negatief samenhangt met de omvang van schoolverschillen.

Uit een vergelijking van de gemiddelde effectgrootte (gemiddeld over de 33 prestatievariabelen) blijkt dat beide groepen scholen nauwelijks van elkaar verschillen. Het verschil bedraagt slechts .06 sigma in het voordeel van de groep klassikale scholen⁵. Dit geringe verschil doet vermoeden dat de groeperingswijze geen interessante variabele is in het onderzoek naar schooleffecten.

Tot slot nog een opmerking over de samenhang tussen de 33 prestatievariabelen. Op grond van het feit dat de verschillen tussen scholen betrekkelijk gering zijn, kan men vermoeden dat de verschillen zowel in de tijd gezien weinig stabiel zijn als over de prestatievariabelen weinig consistent. Doordat de Voorstudie P.P.O.N.-gegevens geen longitudinaal karakter hebben, valt de eerste veronderstelling niet te evalueren, de tweede echter wel. Daartoe zijn de schoolgemiddelden per toetsboekje (A, B of C) na standaardisatie gesommeerd, afzonderlijk over de leesvariabelen en de schrijfvariabelen. Uit de correlatiematrix (Tabel 5) blijkt dat de intercorrelaties inderdaad betrekkelijk laag zijn. Scholen waarvan de leerlingen op de opgaven in boekje A een hoge leesscore behaalden, doen dat niet op de leesopgaven in boekje B of C. Voor schrijven valt hetzelfde verschijnsel waar te nemen. Ook tussen lezen en schrijven zijn de correlaties niet zo hoog. Uit het verschijnsel dat de diverse taalprestatievariabelen betrekkelijk laag intercorreleren, kan afgeleid worden dat de onderscheiding tussen hoog en laag presterende scholen zelfs binnen één vakgebied als de Nederlandse taal niet gemakkelijk te maken is. Conclusies uit schoolvakgebonden schooleffectenonderzoek behoeven derhalve beslist niet geldig te zijn binnen andere schoolvakken. Men kan hieraan een pleidooi ontleenen voor het gebruik van een brede toetsbatterij als afhankelijke variabele.

Tabel 5: Produkt-moment correlaties tussen de over de afzonderlijke prestatievariabelen berekende schalen; op de diagonaal de Cronbach alpha; het aantal scholen is minstens 99.

schaal (aantal prestatievariabelen)	1	2	3	4	5	6
1. lezen, boekje A (7)	.84					
2. lezen, boekje B (5)	.19	.74				
3. lezen, boekje C (5)	.52	.46	.79			
4. schrijven, boekje A (2)	.69	.18	.57	.41		
5. schrijven, boekje B (8)	.21	.60	.35	.25	.87	
6. schrijven, boekje C (6)	.41	.46	.67	.61	.35	.82

De hoogste correlaties in Tabel 5 betreffen die tussen lezen en schrijven uit hetzelfde boekje. Deze bedragen .69, .60 en .67 voor respectievelijk boekje A, B en C. Ze zijn echter berekend over scores van dezelfde leerlingen, omdat de taalopdrachten per boekje geroteerd zijn over leerlingen. Ze geven derhalve een geflatteerd beeld. Het verschijnsel dat metingen per school die bij verschillende leerlingen verricht worden veel lager correleren – de mediane correlatie bedraagt .44 – dan metingen per school die bij dezelfde leerlingen verricht worden, doet vermoeden dat de stabiliteit van schooleffecten in de tijd gezien niet hoog zal zijn. Er is longitudinaal onderzoek nodig om over deze belangrijke veronderstelling uitsluitsel te verkrijgen.

NOTEN

1. Ter vergelijking vermelden we dat volgens de Voorstudie P.P.O.N. het percentage leerlingen waarvan de ouders volgens de leerkracht een laag scholingsniveau hebben (glo, lbo, een paar jaar avo) op stimuleringscholen varieert van 0 tot 100 procent (gemiddeld 67 procent met een standaardafwijking van 24.6 procent) en op niet-stimuleringscholen van 0 tot 92 procent (gemiddeld 37 procent met een standaardafwijking van 22.2 procent).

2. Door de verschillen tussen het stimulerings- en het onderwijsvoorrrangsbeleid is een rechtstreekse vergelijking onmogelijk. We vergelijken hier de stimuleringscholen uit 1984/85 (in dat jaar was ongeveer 25 procent van de lagere scholen een stimuleringschool) met de 18 procent scholen die in 1985/86 de hoogste schoolscore hadden (vanaf schoolscore 111). Verder dient men te weten dat Wijnstra's gegevens betrekking hebben op gemiddelde percentages voor groepen van scholen (gegroepeerd naar schoolscore), terwijl de Voorstudie P.P.O.N.-gegevens percentages voor afzonderlijk scholen betreffen. Daardoor is de variatie in Wijnstra's percentages geringer. Zij nog vermeld dat volgens onze gegevens het gemiddelde percentage allochtone leerlingen op stimuleringscholen 9 bedraagt met een standaardafwijking van 17 procent en op niet-stimuleringscholen 3 procent met een standaardafwijking van 7.7 procent.
3. Uit het antwoordpatroon blijkt dat de drie vragen over niveaugroepering uitstekend passen binnen het door Mokken (1970) ontwikkelde stochastische cumulatieve schaalmodel (schaalcoëfficiënt H bedraagt .82; betrouwbaarheidscoëfficiënt ρ bedraagt .87). Ook Cronbachs alpha voor de schaal waarin deze drie vragen gecombineerd worden, is hoog ($\alpha = .82$).
4. Nagegaan is nog of het nul-effect van de variabele 'hoeveelheid onderricht' een artefact kan zijn van onze dichotomisering. Dit blijkt echter niet het geval. Ook de correlaties tussen de als een continue variabele behandelde 'hoeveelheid onderwijstijd' en de 33 prestatiematen zijn uiterst laag. De mediane correlatie bedraagt .013 en slechts één ervan wijkt statistisch significant af van nul (bij een α van .05).
5. Nagegaan is nog of het geringe effect van de groeperingsvorm een artefact kan zijn van onze dichotomisering. Dat blijkt echter niet het geval. Ook de correlaties tussen de variabele 'niveaugroepering' (met vier niveaus; zie par. 3 onder c.) en de 33 prestatiematen zijn uiterst laag. De mediane correlatie bedraagt .012 en 13 ervan wijken statistisch significant af van nul (bij een α van .05).

LITERATUUR

- Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J. & Pincust, J. (1974). *How effective is schooling? A critical review of research*. Englewood Cliffs, N.J.: Prentice-Hall.
- Beckerman, T.M. & Good, T. L. (1981). The classroom ratio of high- and low-aptitude students and its effects on achievement. *American Educational Research Journal*, 18, 317-327.
- Brookover, W., Beady, C., Flood, P. & Wisenbaker, J. (1979). *School social systems and students achievement. Schools can make difference*. New York: Praeger Publishers.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Wood, A. M., Winfeld, F. & York, R. L. (1966). *Equality of educational opportunity*. Washington D.C.: U.S. Govt. Printing Office.
- Glasman, N.S. & Biniaminov, J. (1981). Input-output analysis of schools. *Review of Educational Research*, 51, 509-539.
- Glopper, K. de (1987). *Schrijffprestaties, schrijffonderwijs en achtergrondvariabelen*. Paper gepresenteerd op de Onderwijsresearchdagen, Groningen.
- Good, T. L. & Brophy, J. E. (1986). School effects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (third edition), pp. 570-602. New York: MacMillan Publishing Company.
- Grift, W. van de (1985). Onderwijsleerklimaat en leerlingprestaties. *Pedagogische Studiën*, 62, 401-414.
- Hays, W. L. (1977). *Statistics for the social sciences* (2nd edition). London: Holt, Rinehart & Winston.
- Hoeksma, J. B., Wesdorp, H., Bergh, H. van den & Bos, D. J. (1985). *Voorstudie Periodieke Peiling van het Onderwijsniveau. Deel IV: Technische verantwoording* (SCO-rapport no. 54). Amsterdam: S.C.O.
- Kalmijn, M. (1987). Commentaar op De Vos' model van het interactiesysteem van de schoolklas. *Mens en Maatschappij*, 62, 75-79.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.
- Leiter, J. (1983). Classroom composition and achievement gains. *Sociology of Education*, 56, 126-132.
- Mensen, T. & Guldmond, H. (1987). Groepseffecten in individuele leerprestaties. *Mens en Maatschappij*, 62, 176-199.
- Meijnen, G. W. (1982). *Schoolloopbanen in het lager onderwijs*. Groningen: Sociologisch Instituut.
- Meijnen, G. W. (1984). *Van zes tot twaalf. Een longitudinaal onderzoek naar de milieu- en schooleffecten van loopbanen in het lager onderwijs* (SVO-reeks, no. 79). Harlingen: Flevodruk BV.
- Mokken, R. J. (1970). A theory and procedure of scale analysis. 's-Gravenhage: Mouton.
- Otter, M. E. (1987). *Tijdbesteding aan moedertaalonderwijs op vier manieren gemeten*. Amsterdam: S.C.O.
- Purkey, S. C., & Smith, M. S. (1983). Effective schools: a review. *The Elementary School Journal*, 83, 427-452.
- Ross, K. N. (1978). Sample design for educational survey research. *Evaluation in Education*, 2, 105-195.

- Rutter, M., Maugham, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours. Secondary schools and their effects on children*. London: Open Books.
- Shavit, Y. & Williams, R. A. (1985). Ability grouping and contextual determinants of educational expectations in Israel. *American Sociological Review*, 50, 62-73.
- Singer, J. D. (1987). An intraclass correlation model for analyzing multilevel data. *The Journal of Experimental Education*, 55, 219-228.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley Publ. Comp.
- Vos, H. de (1986). Een model van het interactiesysteem van de schoolklas. *Mens en Maatschappij*, 61, 147-168.
- Vos, H. de (1987). Repliek aan Winkels en Kalmijn, *Mens en Maatschappij*, 62, 80-82.
- Wesdorp, H., Bergh, H. van den, Bos, D. J., Hoeksma, J. B., Oostdam, R. J., Scheerens, J. & Triesscheijn, B. (1986). *De haalbaarheid van periodiek peilingsonderzoek. Een voorstudie op het gebied van het taalonderwijs in de lagere school*. Lisse: Swets & Zeitlinger.
- Winkels, J. (1987). Een alternatief model van het interactiesysteem van de schoolklas. *Mens en Maatschappij*, 62, 72-74.
- Wolf, J. C. van der (1984). *Schooluitval; een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het regulier onderwijs*. Lisse: Swets & Zeitlinger.
- Wijnstra, J. M. (1987). De samenstelling van de schoolbevolking in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 12, 99-103.

Manuscript ontvangen 30-6-1987.

Definitieve versie ontvangen 20-10-1987.

Longitudinale Computerondersteunde Onderkenning van Lees- en Spellingsmoeilijkheden: Een Toepassing van het Kalman-Filter in de Onderwijspraktijk*

Johan H. L. Oud en Martin J. C. Mommers

Instituut voor Orthopedagogiek en Instituut voor Onderwijskunde, Katholieke Universiteit Nijmegen.

ABSTRACT

In this article the use of the Kalman filter is recommended for the assessment of reading and spelling disabilities during beginning reading. In contrast with cross-sectional factorscore estimators, the Kalman filter combines present and past information for optimally estimating the values of imperfectly measurable variables. The Kalman filter presupposes a longitudinal model which in the present application is taken from a previously published Lisrel-model, based on data from 225 pupils who attended the reading program "Veilig leren lezen". The Kalman-filter results are compared with those of Bartlett's cross-sectional factorscore estimator. The proposed procedure is implemented in a microcomputer program which gives results in tabular as well as graphical form and can be used by the teacher in the classroom. The use of the program in assessing reading and spelling achievements is extensively illustrated for three pupils.

1. INLEIDING

In ons land zijn maar weinig longitudinale gegevens over lezen en spellen in het basisonderwijs beschikbaar. Het verzamelen van dergelijke gegevens is tijdrovend en kostbaar. In het kader van het project "Preventie van leesmoeilijkheden" (S.V.O.-project 0492, zie Van Dongen, 1984) zijn longitudinale gegevens beschikbaar gekomen die zich lenen voor nadere secundaire analyses. Het doel van het genoemde project was meer kennis te verwerven over de samenhang tussen technisch lezen, begrijpend lezen en spelling, in de hoop dat daar aanwijzingen uit af te leiden zouden zijn om in de praktijk het aantal kinderen met lees- en spellingsmoeilijkheden wat te kunnen reduceren. Men kan het ook anders formuleren. Door de Advies Raad voor het Basis Onderwijs, het Speciaal Onderwijs en de Opleidingen (ARBO) is aangedrongen op maatregelen die de uitstroom van leerlingen uit het regulier onderwijs naar het speciale onderwijs moeten afremmen. Leerkrachten in het reguliere onderwijs zullen hun zorg breder moeten maken, dat wil zeggen zij zullen zwakke leerlingen met lees- en spellingsmoeilijkheden niet zonder meer moeten verwijzen naar een of andere vorm van een speciaal onderwijs, maar zij moeten proberen deze kinderen tijdig te ontdekken en op te vangen. In dit verband spreekt men wel over diagnostiserend onderwijzen. Het onderwijs moet zo veel mogelijk worden afgestemd op de specifieke instructiebehoeften van de leerlingen.

In dit artikel zal worden aangetoond hoe de onderkenning van lees- en spellingsmoeilijkheden kan worden verbeterd door gebruik te maken van een empirisch ondersteund, longitudinaal verklaringsmodel. Omdat technisch lezen, begrijpend lezen en spellen onderling samenhangen, ligt het voor de hand om deze *drie aspecten* van de taalontwikkeling *gelijktijdig* te onderzoeken. Oud (1978, 1982) heeft erop gewezen dat cross-sectioneel gevonden "causale effecten" vaak geheel of gedeeltelijk worden wegverklaard indien men gegevens verzamelt met

* De auteurs danken R. Sinkeldam die op basis van het Fortran-programma LISKAL het gepresenteerde microcomputer-programma LISKAL-M1 construeerde.

betrekking tot dezelfde variabelen op voorgaande tijdstippen en deze als additionele verklarende variabelen in de analyse betreft. Om deze reden is het van belang de drie aspecten over een langere periode, *longitudinaal* bij leerlingen te volgen. De laatste jaren zijn analysetechnieken ontwikkeld die het mogelijk maken complexe longitudinale verklaringsmodellen te schatten voor groepen leerlingen en hiervan gebruik te maken bij de diagnostiek van individuele leerlingen. Voor de groepsbenadering gebruiken we de Lisrelprocedure (Jöreskog & Sörbom, 1981), voor de individuele benadering het Kalman-filter (Oud, Van den Bercken en Essers, 1986).

De voorgestelde Kalman-filter procedure verschaft voor individuele leerlingen schattingen van de niveaus van technisch lezen, begrijpend lezen en spellen die aanmerkelijk nauwkeuriger zijn dan in een traditionele screeningsprocedure. Het Kalman-filter is optimaal: Geen andere schattingsprocedure kan schattingen met kleinere standaardfouten opleveren. De procedure geeft tevens de mogelijkheid op basis van actuele en verleden informatie voorspellingen van toekomstige prestaties te doen zodat deze in de beschouwing kunnen worden betrokken. Aan de hand van een vergelijking van de voorspellingen met de later feitelijk gerealiseerde ontwikkeling kunnen tenslotte genomen maatregelen ter verbetering van de prestaties worden geëvalueerd.

De voorgestelde procedure is bedoeld als een instrument in de handen van de leerkracht om de vorderingen van de leerlingen op globaal niveau te volgen. Met globaal bedoelen we het signaleren van vorderingen over langere perioden (3 à 6 maanden). Volgen betekent in dit verband vaststellen, of er sprake is van een bevredigende ontwikkeling, gezien dat wat verwacht mag worden, dan wel of er sprake is van relatieve stilstand of mogelijk zelfs achteruitgang. Het instrument is niet geschikt voor *gedetailleerde*, dagelijkse voortgangscontrole en ook niet voor een *verfijnde* diagnostiek naar oorzaken van achterblijven. Het moet worden gezien als een belangrijke eerste stap in een diagnostisch proces.

In paragraaf 2 zullen we ingaan op de uitgevoerde Lisrel-analyses. Paragraaf 3 zal gewijd zijn aan het Kalman-filter met behulp waarvan voor individuele leerlingen het niveau van technisch lezen, begrijpend lezen en spelling optimaal geschat wordt. In het bijzonder zal aandacht worden besteed aan de optimaliteit. De Kalman-filter procedure zal worden vergeleken met de traditionele, cross-sectionele benadering. In paragraaf 4 zal de Kalman-filter procedure worden toegepast op de gegevens van drie concrete leerlingen met behulp van het microcomputerprogramma LISKAL-M1. Aangegeven zal worden hoe LISKAL-M1 door de leerkracht zelf in de klassituatie kan worden toegepast en welke diagnostisch relevante informatie daarbij over de leerlingen ter beschikking komt. Tenslotte zullen we in de conclusies van paragraaf 5 aandacht besteden aan vervolgonderzoek.

2. HET LONGITUDINALE LISREL-MODEL

Het zou te ver voeren hier uitvoerig verslag te doen van de Lisrel-analyses, die op de gegevens van het longitudinale onderzoek zijn uitgevoerd. Deze zijn reeds elders beschreven (Mommers, 1987; Mommers, Van Leeuwe, Oud & Janssens, 1986; Mommers & Oud, 1984). Wij beperken ons hier tot enkele hoofdlijnen aan de hand van het Lisrel-model dat is weergegeven in Figuur 1 en waarvan de geobserveerde variabelen (vierkantjes) zijn opgenomen in Tabel 1.

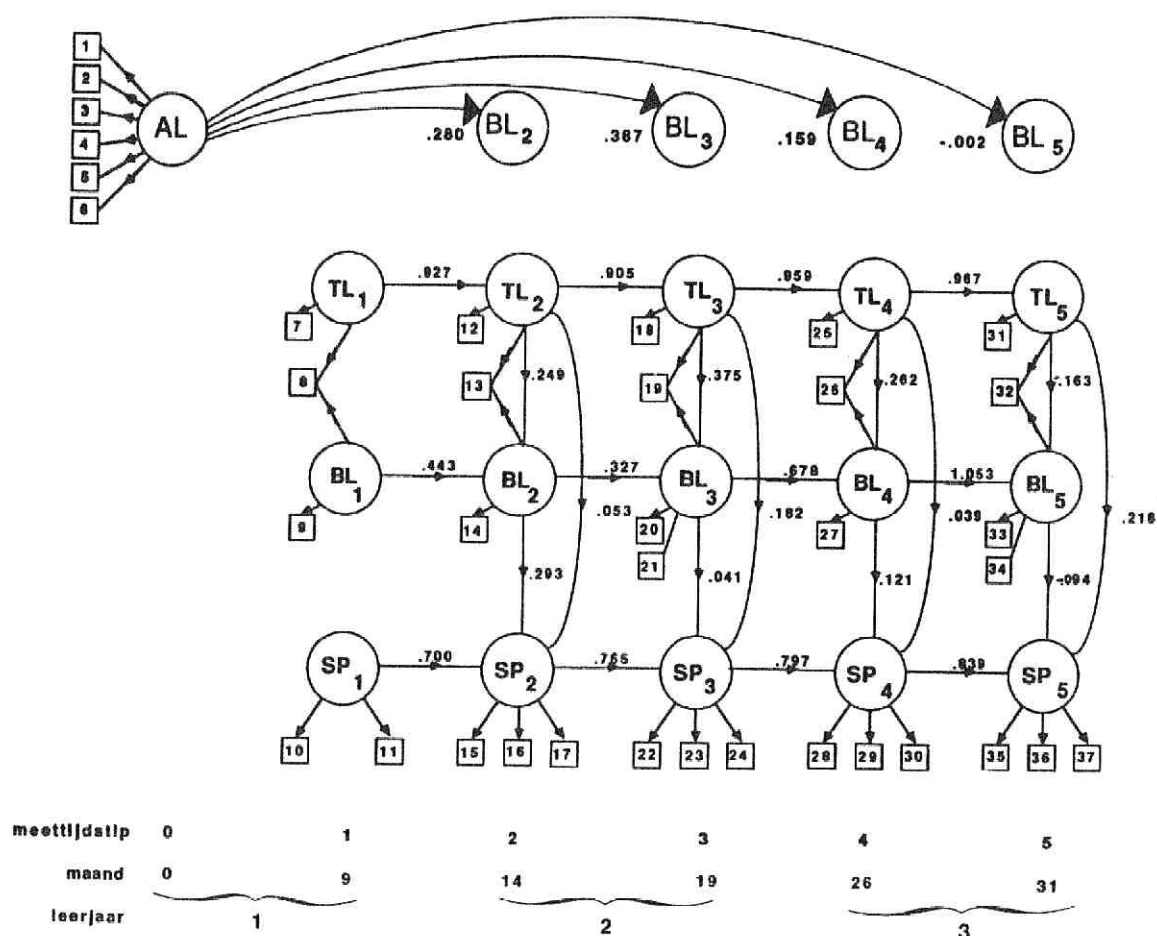
Opgemerkt moet worden, dat het onderzoek Preventie van Leesmoeilijkheden betrekking had op twee steekproeven: de Veilig Leren Lezen-groep en de Letterstad-groep. Het hier gepresenteerde Lisrel-model heeft alleen betrekking op de Veilig Leren Lezen-groep (N=225). Het is een vereenvoudigde vorm van het model 16 voor deze groep in Mommers et al. (1986, p. 110) en daaruit rechtstreeks afleidbaar. De belangrijkste vereenvoudiging is dat van de latente variabelen (cirkels) voorafgaand aan 9 maanden leesonderwijs alleen de Algemene Leesvoorwaardenfactor (AL) is behouden. De eliminatie van latente variabelen en effecten – met name die voor 5 maanden leesonderwijs waar Technisch Lezen (TL), Begrijpend Lezen (BL) en Spellende (SP) nog niet als duidelijk te onderscheiden variabelen naar voren komen – heeft verder geen consequenties voor het overblijvende gedeelte van het model 16 dat is

Tabel 1. Lijst van geobserveerde variabelen.

Meettijdstip 0	Voor het begin van het systematisch leesonderwijs 1 ELI, subtest matrices 2 ELI, subtest natekenen 3 Oordeel van de kleuterleidster over schoolrijpheid 4 Lettercluster identificatie test 5 Auditieve synthese test 6 Fonemische analyse test
Meettijdstip 1	Maand 9 Leesonderwijs 7 Caesar Een Minuut Test 8 AVI-toetskaarten 9 CITO-toets Lees en Begrijp 1A 10 Woorddictee 11 Zinsdictee B1
Meettijdstip 2	Maand 14 Leesonderwijs 12 Een Minuut Test van Brus-Voeten, vorm A 13 AVI-toetskaarten 14 CITO-toets Lees en Begrijp 1B 15 Woorddictee OBCE-1 16 Zinsdictee B2 17 Zinsdictee C1
Meettijdstip 3	Maand 19 Leesonderwijs 18 Een Minuut Test van Brus-Voeten, vorm A 19 AVI-toetskaarten 20 CITO-toets Lees en Begrijp 2 21 Schiftelijke opdrachten 2 22 Woorddictee OBCE-2 23 Zinsdictee C2 24 Zinsdictee D1
Meettijdstip 4	Maand 26 Leesonderwijs 25 Een Minuut Test van Brus-Voeten, vorm A 26 AVI-toetskaarten 27 Schiftelijke opdrachten 3 28 Woorddictee OBCE-2 29 Zinsdictee D2 30 Zinsdictee E1
Meettijdstip 5	Maand 31 Leesonderwijs 31 Een Minuut Test van Brus-Voeten, vorm A 32 AVI-toetskaarten 33 Schiftelijke opdrachten 4 34 CITO-toets Begrijpend Lezen M3 35 Woorddictee OBCE-3 36 Zinsdictee E2 37 Zinsdictee F1

opgenomen in Figuur 1. De reden voor de vereenvoudiging is dat een zinvolle diagnostiek van technisch lezen, begrijpend lezen en spellen en de toepassing van het Kalman-filter voor dat doel, beperkt blijft tot die tijdstippen waarop deze vaardigheden reeds als zodanig zijn te onderscheiden. Dat is, zoals we zullen zien, vanaf ongeveer 9 maanden leesonderwijs het geval.

Voor een goed begrip van Figuur 1 dient nog gewezen te worden op twee bijzonderheden. In de eerste plaats hebben we om de figuur niet te gecompliceerd te maken de BL-factoren



Figuur 1: Longitudinaal Lisrel-model voor lezen en spellen (alle variabelen gestandaardiseerd).

bovenaan de figuur herhaald. De effecten van de Algemene Leesvoorwaardenfactor AL op de opeenvolgende BL-factoren zijn hier aangegeven. In de tweede plaats zijn de geobserveerde variabelen 8, 13, 19, 26 en 32 zowel met de TL-factoren als met de BL-factoren verbonden. Het betreft hier de AVI-toetskaarten waarvan we aanvankelijk veronderstelden dat zij alleen op de TL-factoren zouden laden. Het bleek echter (zie Mommers et al., 1986) dat een verbeterde modelpassing kon worden bereikt door de AVI-toetskaarten tevens op de BL-factoren te laten laden.

Figuur 1 geeft duidelijk aan welke variabelen *direct* van invloed zijn op een score. Bijvoorbeeld de BL3-score wordt direct beïnvloed door TL3, BL2 en AL. De padcoëfficiënten .375, .327 en .387 geven de gestandaardiseerde gewichten van de directe invloeden aan. Uit de figuur kan men bovendien aflezen welke variabelen *indirect* van invloed zijn via andere variabelen. Voorspellingen die op basis van het model zullen worden gedaan, komen via deze directe en indirecte effecten tot stand.

Belangrijke vraagstellingen die aan de Lisrel-analyse ten grondslag hebben gelegen, waren: In hoeverre zijn technische lezen, begrijpend lezen en spellen op de diverse meetmomenten empirisch te onderscheiden en hoe sterk is de invloed van elk van deze factoren op de gelijksoortige factor op een volgend meetmoment? Welke effecten bestaan er tussen de drie factoren op de verschillende meetmomenten? Is er effect van de Algemene Leesvoorwaarden op de drie factoren en op welke meetmomenten?

Mede aan de hand van de hoogte van de (gestandaardiseerde) padcoëfficiënten in Figuur 1 zijn de volgende conclusies getrokken met betrekking tot de periode vanaf 9 maanden lees-

- 1) Technisch lezen, begrijpend lezen en spellen komen als drie duidelijk te onderscheiden factoren naar voren; dit blijkt onder meer uit de sterke longitudinale effecten van TL-, BL- en SP-factoren op zichzelf die verreweg het grootste deel van de verklaarde variantie voor hun rekening nemen en de relatief geringe effecten tussen deze factoren; er bestaat bovendien een duidelijke tendens dat de drie factoren onafhankelijker van elkaar worden naarmate het leesonderwijs voortschrijdt.
- 2) Technisch lezen oefent invloed uit op begrijpend lezen hoewel minder sterk dan Perfetti en Lesgold (1979) in hun cross-sectionele benadering menen te moeten concluderen en bovendien in de tijd afnemend.
- 3) Er zijn betrekkelijk geringe en in de tijd afnemende effecten te constateren van technisch lezen en begrijpend lezen op spelling.
- 4) Begrijpend lezen blijft ook op latere tijdstippen (tot en met 26 maanden leesonderwijs) een niet onaanzienlijke invloed ondergaan van een algemene leesvoorwaardenfactor.

Als consequentie voor de diagnostiek leiden Mommers et al. (1986) uit het Lisrel-model af dat "pupils with reading and spelling difficulties will have to be examined on all three of these factors, because, due to their relative independence, discrepancies may exist between the stage of development of the respective skills". Voor de praktijk van de diagnostiek wordt het probleem hoe men op basis van het Lisrel-model en de daarin gebruikte toetsen (zie Tabel 1) zo goed mogelijk bij individuele leerlingen factorscores voor technisch lezen, begrijpend lezen en spellen kan schatten. Oud et al. (1986) hebben laten zien dat dit probleem het best wordt opgelost door toepassing van het Kalman-filter.

3. HET KALMAN-FILTER

Het is hier niet de plaats uitgebreid in te gaan op het Kalman-filter en de toepassing daarvan in gedragswetenschappelijk onderzoek. We verwijzen hiervoor naar de oorspronkelijke publicaties van Kalman (1960) en Kalman en Bucy (1961) en naar de meer toegankelijke uiteenzettingen bij Kwakernaak en Sivan (1972), Lewis (1986) en Oud et al. (1986). We zullen ons hier beperken tot enkele hoofdzaken die van belang zijn voor de beoogde toepassing op basis van het beschreven longitudinale Lisrel-model.

Het Kalman-filter stamt uit de regeltechniek en wordt daar gebruikt om uit kennis over het verleden van een systeem en uit actuele observaties een zo goed mogelijke schatting van de actuele toestand van het systeem te distilleren. Om bijvoorbeeld een schip te besturen moet men op opeenvolgende momenten zo goed mogelijk de positie van het schip kennen. De positie is de onbekende driedimensionele toestand die men moet zien te achterhalen. Dat kan in principe op twee verschillende manieren die het Kalman-filter op een optimale wijze combineert.

De eerste manier is predictie vanuit de vorige toestandsschatting op basis van een longitudinaal verklaringsmodel. Als men een schatting heeft van de vorige positie en men weet de invloeden die in de tussentijd op het schip hebben ingewerkt, kan men trachten op basis van het verklaringsmodel de actuele positie te voorspellen. De voorspelling zal echter niet perfect zijn vanwege onvolkomenheden in het verklaringsmodel en omdat ook de vorige positie niet precies bekend is. In sociaal-wetenschappelijk jargon heten de onbekende toestandsvariabelen factoren en kan men de voorspelling een predictieve factorscore-schatting noemen. Een tweede manier om de toestand te bepalen is op basis van actuele metingen: een cross-sectionele factorscore-schatting. Duiden we de predictieve factorscore-schatting aan als $\hat{x}_{t-}$, dat wil zeggen de schatting van de onbekende toestand x_t op basis van informatie voorafgaand aan tijdstip t , en de cross-sectionele factorscore-schatting als $H_t y_t$, waarin H_t de gewichtenmatrix is voor het schatten van x_t en waarin y_t de geobserveerde variabelen bevat, dan combineert het Kalman-filter beide soorten schatters als volgt:

$$\hat{x}_t = (I - H_t C_t) \hat{x}_{t-} + H_t y_t \quad (1)$$

waarin

$$\hat{x}_{t-} = A_{t-1}\hat{x}_{t-1} + B_{t-1}u_{t-1}. \quad (2)$$

Voor de toepassing van de Kalman-filter formules (1) en (2), bijvoorbeeld om voor een concreet kind op een bepaald tijdstip t TL, BL en SP te schatten, zijn een vijftal matrices vereist die worden ontleend aan het longitudinale verklaringsmodel, in onze toepassing dus aan het longitudinale Lisrel-model (zie Figuur 1). We zullen deze vijf matrices tezamen met de nog niet genoemde inputvector u_t achtereenvolgens bespreken, voordat we nader ingaan op het Kalman-filter en de daarin zo centrale gewichtenmatrix H_t . Evenals bij traditionele factorscore-schatting (Lawley & Maxwell, 1971, p. 106) worden de coëfficiënten in de matrices bekend verondersteld. De steekproeffluctuaties bij de schatting van het Lisrel-model worden dus verondersteld verwaarloosbaar klein te zijn. Dit houdt in ieder geval in dat de gebruikte steekproef bij de schatting van het Lisrel-model groot moet zijn. Bij aanwezigheid van procesfouten in het model (zie de dadelijk te bespreken matrix Q_t), blijkt het Kalman-filter overigens opvallend robuust voor afwijkingen in de dynamische modelmatrices A_t en B_t (Lewis, 1986, p. 205).

A_t : Deze matrix wordt wel de toestandsovergangsmatrix genoemd en bevat de effecten van de toestand x_t op de nieuwe toestand x_{t+1} , in onze toepassing dus de effecten die de factoren TL, BL en SP op tijdstip t uitoefenen op TL, BL en SP op tijdstip $t+1$. Bijvoorbeeld bij de overgang van tijdstip 1 naar tijdstip 2 worden de volgende effecten gespecificeerd:

$$\begin{array}{c} \text{TL}_1 \quad \text{BL}_1 \quad \text{SP}_1 \\ \text{TL}_2 \quad \begin{bmatrix} .927 & 0 & 0 \\ .231 & .443 & 0 \\ .117 & .130 & .700 \end{bmatrix} \\ \text{BL}_2 \\ \text{SP}_2 \end{array}$$

Blijkens Figuur 1 ondergaat TL_2 alleen een effect .927 van TL_1 . BL_2 ondergaat een effect .443 van BL_1 en een indirect effect $.231 = .927 \times .249$ van TL_1 . SP_2 ondergaat een effect .700 van SP_1 , een indirect effect $.130 = .443 \times .293$ van BL_1 en een samengesteld indirect effect $.117 = .927 \times .053 + .927 \times .249 \times .293$ van TL_1 .

B_t : Hierin zijn de effecten van de input-variabelen in u_t op de nieuwe toestand x_{t+1} opgenomen. Input-variabelen werken van buiten op de toestandsvariabelen in en worden zelf niet door de toestandsvariabelen beïnvloed. In onze toepassing is er slechts één input-variabele, AL, die bovendien over de tijd constant wordt verondersteld. AL heeft een direct effect .280 op BL_2 en een indirect effect $.082 = .280 \times .293$ op SP_2 . Bij de overgang van tijdstip 1 naar tijdstip 2 worden dus de volgende effecten opgenomen in B_t :

$$\begin{array}{c} \text{AL} \\ \text{TL}_2 \quad \begin{bmatrix} 0 \\ .280 \\ .082 \end{bmatrix} \\ \text{BL}_2 \\ \text{SP}_2 \end{array}$$

C_t : Dit is de factormatrix op tijdstip t met de ladingen van de geobserveerde variabelen in y_t op de factoren in x_t . Voor tijdstip 1 in Figuur 1 bevat y_t de vijf geobserveerde variabelen 7, 8, 9, 10 en 11, bevat x_t de drie factoren TL_1 , BL_1 en SP_1 en is de factormatrix C_t dus van de orde 5×3 . De factormatrices C_t op opeenvolgende tijdstippen zijn in de Lisrel-output te vinden als deelmatrices van de Lisrel-factormatrix.

Q_t : Dit is de covariantiematrix van procesfouten w_t in de toestandsovergangsvergelijking $x_{t+1} = A_t x_t + B_t u_t + w_t$ die eveneens gemakkelijk in de Lisrel-output is te vinden of daaruit af te leiden (Oud et al., 1986, p. 115).

R_t : Dit is de covariantiematrix van meetfouten v_t in de factormodel- of uitleesvergelijking $y_t = C_t x_t + v_t$. Ook deze matrix is voor ieder van de tijdstippen gemakkelijk uit de Lisrel-output af te lezen. De diagonaalelementen van R_t worden in de factor-analyse unieke varianties genoemd.

Meer inzicht in het Kalman-filter ontstaat, wanneer we op basis van de voorspelling van de geobserveerde waarden in y_t :

$$\hat{y}_{t-} = C_t \hat{x}_{t-} \quad (3)$$

het Kalman-filter in (1) herschrijven: $\hat{x}_t = \hat{x}_{t-} - H_t \hat{y}_{t-} + H_t y_t$. Het Kalman-filter corrigeert dus de predictieve schatter $\hat{x}_{t-}$ door de cross-sectionele schatter $H_t y_t$ in de plaats te laten komen van de corresponderende predictieve vorm $H_t \hat{y}_{t-}$. Ook kan men schrijven $\hat{x}_t = \hat{x}_{t-} + H_t (y_t - \hat{y}_{t-})$ en dan blijkt de correctie te bestaan uit de weging H_t van de zogenaamde innovaties $y_t - \hat{y}_{t-}$ (afwijking tussen de werkelijke en de voorspelde observaties).

Optimaliteit van het Kalman-filter

De optimaliteit heeft het Kalman-filter te danken aan de specifieke vorm van de gewichtenmatrix of "Kalman gain matrix" H_t :

$$H_t = P_t C_t' R_t^{-1}. \quad (4)$$

H_t ontstaat door de factormatrix C_t' te wegen door postmultiplicatie met de inverse van de meetfout-covariantiematrix R_t en door premultiplicatie met de covariantiematrix P_t van de Kalman-schattingsfouten $e_t = x_t - \hat{x}_t$:

$$\begin{aligned} P_t &= (P_{t-}^{-1} + W_t^{-1})^{-1} \text{ waarin} \\ P_{t-} &= A_{t-1} P_{t-1} A_{t-1}' + Q_{t-1} \text{ en} \\ W_t &= (C_t' R_t^{-1} C_t)^{-1}. \end{aligned} \quad (5)$$

De wortels uit de diagonaalelementen van P_t vormen de standaardfouten van de Kalman-filter schatting. Optimaliteit houdt in dat geen enkele andere schattingsprocedure kleinere standaardfouten oplevert dan het Kalman-filter. Het verdient vermelding dat P_t is samengesteld uit de twee volgende fouten-covariantiematrices: P_{t-} van de predictieve schatter $\hat{x}_{t-}$ en W_t van de in de sociale wetenschap bekende cross-sectionele factorscore-schatter van Bartlett (Lawley & Maxwell, 1971, p. 110). De Bartlett-schatter is vooral bekend vanwege de aantrekkelijke statistische eigenschap van zuiverheid, die eveneens toekomt aan de predictieve schatter en het Kalman-filter.

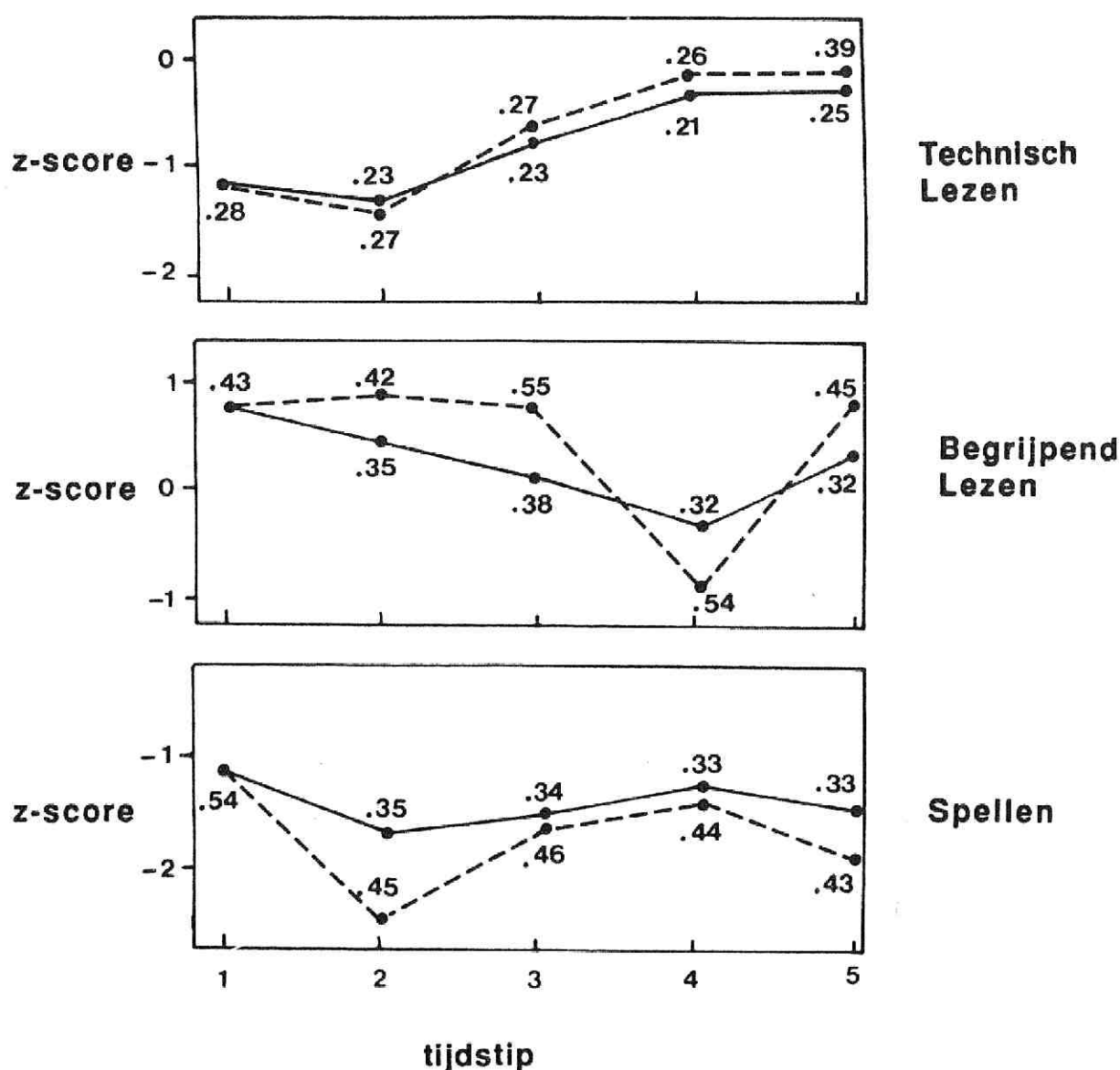
Nu spelen de procesfouten en de meetfouten via respectievelijk de matrices Q_t ($t' < t$) en R_t een essentiële rol bij de totstandkoming van P_{t-} en W_t . Voor een goed inzicht in de optimaliteit van het Kalman-filter is het van belang na te gaan welk effect beide soorten fouten via P_{t-} en W_t hebben op P_t en tenslotte via H_t op de Kalman-schatter $\hat{x}_t$. Hoge procesfouten ($Q_{t'} \rightarrow \infty$, $P_{t-} \rightarrow \infty$ en $P_{t-}^{-1} \rightarrow 0$) laten P_t tenderen naar de Bartlett-covariantiematrix: $P_t \approx (C_t' R_t^{-1} C_t)^{-1}$ en het Kalman-filter naar de Bartlett-schatter: $\hat{x}_t \approx (C_t' R_t^{-1} C_t)^{-1} C_t' R_t^{-1} y_t$ vanwege $(I - H_t C_t) \approx 0$. Met andere woorden: een gebrekkig procesmodel met geringe verklaarde varianties heeft tot gevolg dat de predictieve schatter automatisch aan belang inboet en alleen de actuele metingen nog een rol van betekenis spelen. Van de andere kant hebben hoge meetfouten ($R_t \rightarrow \infty$, $W_t \rightarrow \infty$ en $W_t^{-1} \rightarrow 0$) tot gevolg dat de predictieve schatter gaat domineren en de cross-sectionele schatter wordt uitgeschakeld: $P_t \approx P_{t-}$ en $\hat{x}_t \approx \hat{x}_{t-}$ vanwege $H_t \approx 0$.

In de sociale wetenschap kan men zowel met aanzienlijke procesfouten als met aanzienlijke meetfouten worden geconfronteerd. P_t in (5) laat zien dat het Kalman-filter steeds de voorkeur verdient boven toepassing van ofwel de Bartlett-schatter ofwel de predictieve schatter. Procesfouten maken P_t immers in het ongunstigste geval ($P_{t-} \rightarrow \infty$) gelijk aan maar nooit groter dan de cross-sectionele Bartlett-covariantiematrix W_t en meetfouten maken P_t in het ongunstigste

geval ($W_t \rightarrow \infty$) gelijk aan maar nooit groter dan de predictieve covariantiematrix P_{t-} . Door beide informatie-bronnen (predictief en cross-sectioneel) gelijktijdig te benutten en optimaal te combineren blijft de Kalman-covariantiematrix P_t in de praktijk zowel onder W_t als onder P_{t-} .

Een vergelijking van Kalman-filter, Bartlett- en predictieve schattingen

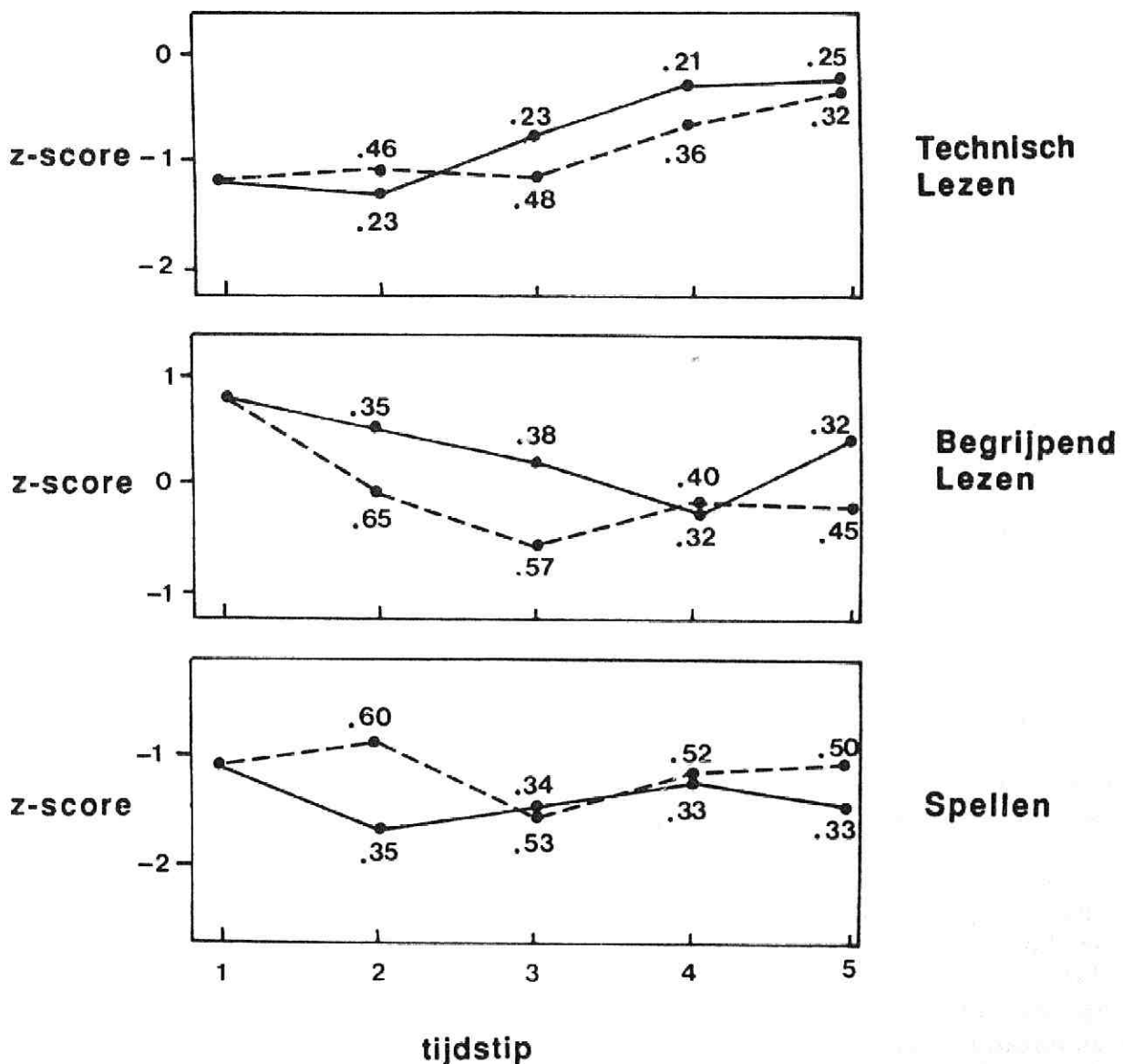
Ter afsluiting van deze paragraaf vergelijken we bij een willekeurig gekozen leerling de Kalman-filter schattingen met respectievelijk de Bartlett-schattingen en de predictieve schattingen. Hoewel de Bartlett-schatter een verfijnder combinatie van scores geeft dan een eenvoudige sommering of middeling van scores, maakt de Bartlett-schatter alleen gebruik van cross-sectionele informatie en illustreert dus de traditionele aanpak in de diagnostiek: combinatie van scores die van hetzelfde tijdstip afkomstig zijn. Ook bij een traditionele, herhaalde screening van leerlingprestaties gebruikt men geen verleden informatie om tot betere actuele schattingen te komen en zijn de resultaten in het algemeen verre van optimaal. Op het begintijdstip 1 in Figuur 2 is uiteraard nog geen Kalman-filter of predictieve schatting mogelijk en hebben we steeds alleen de Bartlett-schatting aangegeven. Zoals is te verwachten op basis van het gebruik



Figuur 2: Kalman-filter schattingen (aaneengesloten lijnen) en Bartlett-schattingen (stippellijnen) van een willekeurig gekozen leerling met vermelding van de bijbehorende standaardfouten en met de Bartlett-schatting als beginschatting.

van informatie uit het verleden, tonen de Kalman-filter schattingen in Figuur 2 een veel stabielere verloop over de tijd dan de Bartlett-schattingen. De laatste laten grotere sprongen zien en de bijbehorende standaardfouten zijn overal groter dan van het Kalman-filter. De grootste reductie in standaardfouten bij toepassing van het Kalman-filter treffen we aan bij Begrijpend Lezen, bijvoorbeeld op tijdstip 4 reduceert standaardfout .54 voor de Bartlett-schatting tot .32 voor de Kalman-filter schatting. Een relatief bescheiden reductie in standaardfouten treffen we aan bij Technisch lezen. Dit wordt veroorzaakt doordat Technisch Lezen voornamelijk wordt gemeten met de zeer betrouwbare Een Minuut Test. De geringe meetfouten van de Een Minuut Test leiden er niet alleen toe dat de Bartlett-schatting zelf goede resultaten geeft maar tevens dat het Kalman-filter de Bartlett-schatting gaat benaderen. Niettemin geeft het Kalman-filter ook hier steeds betere resultaten.

In overeenstemming met formule (2) zijn de predictieve schattingen in Figuur 3 telkens gebaseerd op de Kalman-filter (c.q. Bartlett-)schatting van het voorafgaande tijdstip. De aanzienlijk geringere standaardfouten van de Kalman-filter schattingen in Figuur 3 zijn te danken aan de correctie door actuele metingen op het betreffende tijdstip. In de volgende



Figuur 3: Kalman-filter schattingen (aaneengesloten lijnen) en predictieve schattingen (stippellijnen) van een willekeurig gekozen leerling met vermelding van de bijbehorende standaardfouten en met de Bartlett-schatting als beginschatting; de predictieve schattingen zijn gebaseerd op de voorafgaande Kalman-filter c.q. Bartlett-schatting.

paragraaf zullen we ook gebruik maken van predictieve schattingen over twee en meer tijdsintervallen. De predictieve schattingen worden daarbij gebaseerd op predictieve schattingen van het voorafgaande tijdstip en leiden uiteraard tot grotere en toenemende standaardfouten.

4. HET LONGITUDINALE VERKLARINGSMODEL EN DE KALMAN-FILTER PROCEDURE IN DE PRAKTIJK

Als men in de basisschool de zorg voor de individuele leerlingen wil vergroten, dan is het zeer gewenst, zo niet noodzakelijk, dat de vorderingen van de leerlingen systematisch gevolgd worden (Janssens, 1984). Problemen kunnen dan vroegtijdig worden gesignaleerd. Bovendien kan men nagaan of eerder genomen maatregelen effect hebben gehad. Men heeft daarbij behoefte aan een referentiekader. Normtabellen bieden tot op zekere hoogte een houvast, maar het is beter als men tevens over een longitudinaal verklaringsmodel kan beschikken bij de interpretatie van de scores.

Bij het beoordelen van vorderingen zijn drie verschillende beoordelingscriteria mogelijk: 1. Men vergelijkt de prestaties met die van andere leerlingen, bijvoorbeeld met behulp van normtabellen (sociale vergelijking), 2. De prestaties worden gerelateerd aan de nagestreefde doelstellingen, zoals dat bij mastery learning gebeurt (didactische vergelijking). 3. Men vergelijkt de prestaties met die op een eerder tijdstip en/of met de prestaties die men redelijkerwijze kan verwachten op basis van eerdere gegevens (persoonlijke vergelijking).

Voor de derde vorm van beoordelen levert in de praktijk moeilijkheden op. Het is niet zo eenvoudig de persoonlijke prestatiestandaard van een leerling vast te stellen. Deze kan het meest nauwkeurig worden afgeleid uit schoolvorderingen die over langere tijd geregistreerd zijn (Van Lieshout en Wardenaar, 1985). Hierbij doen zich echter een aantal praktische moeilijkheden voor. In de meeste gevallen kan men over meerdere jaren heen niet dezelfde meetinstrumenten gebruiken en zijn er niet voldoende gegevens beschikbaar voor betrouwbare predicties. Bovendien vereist zo'n predictie nogal wat rekenwerk, vooral als daarbij meerdere samenhangende variabelen worden betrokken. Deze moeilijkheden kunnen voor een deel worden overwonnen door het inschakelen van de computer. Terecht merken Van Lieshout en Wardenaar (1985) op, dat het inzetten van de computer in het onderwijs voor longitudinale gegevensopslag over schoolvorderingen van leerlingen momenteel binnen het bereik liggen.

Voor de toepassing van een empirisch gevalideerd longitudinaal verklaringsmodel, waarbij gebruik wordt gemaakt van geavanceerde technieken zoals het Kalman-filter, is een gebruikersvriendelijk computerprogramma noodzakelijk. Het microcomputer-programma LISKAL-M1 is als zodanig ontworpen. Hoe dit toe te passen is, wordt hier geïllustreerd aan de hand van de gegevens van drie leerlingen, ontleend aan het longitudinale onderzoek van het project "Preventie van Leesmoeilijkheden". Uiteraard is een belangrijke voorwaarde dat de leerkracht de gegevens op het juiste tijdstip met de aangegeven meetinstrumenten verzamelt. Het programma is zo gemaakt dat de leerkracht alleen maar de ruwe scores hoeft in te typen. Het programma verschaft vervolgens drie soorten diagnostische informatie:

- a. De leerkracht krijgt via de Kalman-filter procedure een nauwkeurige beschrijving van het niveau en de ontwikkeling van het technisch lezen, het begrijpend lezen en de spellingsvaardigheid, vergeleken met de ontwikkeling zoals die gemiddeld is verlopen in de groep van 225 leerlingen met Veilig Leren Lezen in het longitudinale onderzoek (Mommers, 1987; Mommers et al., 1986). Het gaat dus niet om een absolute, maar om een inter-individuele ontwikkeling, aangegeven in geschatte standaardscores (z-scores). Bovendien worden de standaardfouten van de schattingen opgegeven. Vanwege het optimale karakter van het Kalman-filter zijn de standaardfouten in het algemeen aanmerkelijk kleiner dan mogelijk zou zijn via een traditionele regelmatige screening van de prestaties. Standaardscores en bijbehorende standaardfouten kunnen zowel in de vorm van een tabel als door middel van een grafische voorstelling worden opgevraagd.

- b. Aan de hand van reeds beschikbare gegevens geeft het computerprogramma *voorspellingen* over het te verwachten niveau op latere tijdstippen met de daarbij behorende standaardfouten, opnieuw op basis van de onderzoeksgroep van 225. Op die manier kan men de vorderingen van de leerlingen niet alleen vergelijken met de gemiddelde ontwikkeling, maar kan men ook de verwachte score van de leerling vergelijken met het resultaat dat op voorafgaande meetmomenten werd bereikt.
- c. Tenslotte kan men zien of de score die voor een bepaald meetmoment werd voorspeld, ook daadwerkelijk is bereikt, wanneer de optimale Kalman-filter schattingen ter beschikking komen. Dit geeft ondermeer de mogelijkheid het effect van genomen maatregelen longitudinaal te evalueren. Met behulp van de standaardfouten van de voorspellingen wordt aangegeven binnen welke schattingsintervallen overeenkomstige leerlingen in de onderzoeksgroep van het Lisrel-model blijven met een bepaalde kans. Effect van genomen maatregelen veronderstelt dat de scores zich voldoende buiten deze schattingsintervallen van longitudinaal verwachte prestaties bewegen.

Dit soort gegevens vormen een belangrijk referentiekader voor het volgen van een leerling en eventueel voor het stellen van een diagnose.

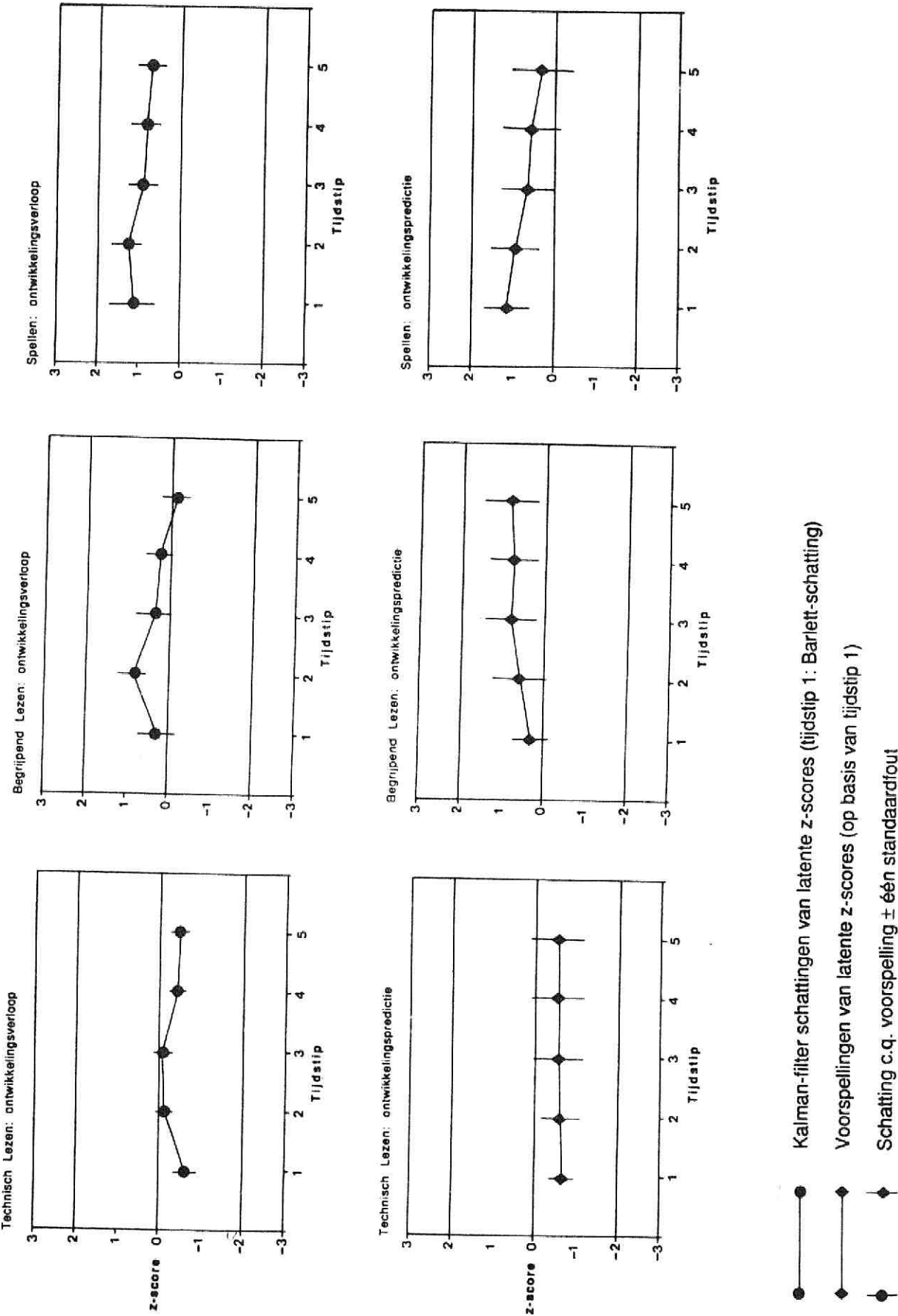
Om dit alles te illustreren volgen hier drie voorbeelden (zie Figuur 4-6). Van elke leerling worden zes plaatjes gegeven: twee voor begrijpend lezen, twee voor spelling en twee voor technisch lezen. Het bovenste plaatje geeft een beschrijving van de ontwikkeling van de betreffende vaardigheden vanaf halfweg leerjaar 1 tot en met leerjaar 3. Het onderste plaatje geeft de voorspelling vanaf het tweede meetmoment, dat is begin leerjaar 2. Vooraf moet worden opgemerkt dat alléén op basis van deze gegevens geen uitputtende diagnostiek mogelijk is. Zoals ook blijkt uit de volgende voorbeelden is aanvulling met andere gegevens, bijvoorbeeld foutenanalyse, noodzakelijk.

Leerling A is een jongen. Zijn vader is een kleine zelfstandige die een lagere beroepsopleiding heeft gehad. Volgens het oordeel van de leerkracht is de mondelinge taalvaardigheid van deze leerling zwak-middelmatig. Hij weet zich in sociaal opzicht goed aan te passen, maar zijn werkhouding en concentratie zijn zwak. Hij is uitgesproken vrijmoedig in zijn gedrag.

Het plaatje waarop de ontwikkeling van het begrijpend lezen wordt *beschreven* laat zien dat er aanvankelijk een stijging optreedt, die gevolgd wordt door een geleidelijke daling. Eind leerjaar 3 behaalt hij een gemiddelde prestatie. Vanaf begin leerjaar 2 (meettijdstip 2) zijn de voorspellingen vrij gunstig. Voor eind leerjaar 3 werd een prestatie verwacht die bijna 1 standaardafwijking boven het gemiddelde ligt. Deze verwachting is echter niet uitgekomen. Wat spelling betreft behaalt deze leerling vrij goede resultaten, maar vanaf meetmoment 2 treedt een lichte daling op. Maar hij blijft behoren tot de goede middelmaat. Kijkt men naar het onderste plaatje, dan ziet men dat er op het meettijdstip 2 een wat sterkere daling werd verwacht dan in feite is opgetreden. De verschillen blijven echter binnen de standaardfout.

In tegenstelling tot begrijpend lezen en spelling, liggen de prestaties bij technische lezen beneden het gemiddelde. Na een lichte stijging treedt aan het begin van het derde leerjaar een kleine daling op. Vanaf begin leerjaar 2 werd een zeer geringe stijging verwacht. Eind leerjaar 3 (meettijdstip 5) komt het daadwerkelijk bereikte niveau overeen met de voorspelling in het begin van het tweede leerjaar.

Als men het geheel overziet, dan baart de ontwikkeling van deze leerling geen ernstige zorgen. Maar toch geven deze plaatjes aanleiding tot een aantal vragen. Als men ervan uitgaat, dat zowel het technisch lezen als het spellen van woorden verband houdt met de kwaliteit en beschikbaarheid van opgeslagen woordrepresentaties, dan is een discrepantie tussen de vorderingen bij technisch lezen en spelling min of meer verrassend. Omdat het juist spellen van woorden hogere eisen stelt aan de kwaliteit van de woordrepresentaties dan de herkenning bij het lezen, mag men aannemen dat deze leerling in staat is woordrepresentaties goed in het geheugen op te slaan. Dat de prestaties bij technisch lezen lager liggen, moet eerder gezocht worden in de *snelheid* van de woordherkenning. Dit bleek ook uit nadere observaties. Deze leerling maakt bij het technisch lezen weinig fouten, maar de woordidentificatie-technieken zijn onvoldoende geautomatiseerd. Melis (1985) spreekt in dit verband over automatiserings-



Figuur 4: LeerlingA: ontwikkelingsverloop en ontwikkelingspredictie (op basis van tijdstip 1) van technisch lezen, begrijpend lezen en spellen.

zwakte. Alle aan tijd gebonden taken worden minder goed gemaakt. In dit geval zijn de lage prestaties waarschijnlijk mede het gevolg van te weinig oefening. Deze jongen leest buiten school maar heel weinig en op school is hij zelden geconcentreerd bezig met lezen.

Onvoldoende automatisch verlopende decodeerprocessen hebben vermoedelijk een negatieve invloed op begrijpend lezen. Daardoor presteert deze leerling waarschijnlijk beneden zijn mogelijkheden. Als belangrijke doelstelling bij deze leerling werd gekozen voor het versterken van de automatisering van de woordherkenningsprocessen. In principe is dit mogelijk met behulp van geëigende computerprogramma's, zoals Roth en Beck (1987) hebben aangetoond. Beschikt men daar niet over dan kan men de leerling aangepaste lectuur aanbieden die zijn belangstelling heeft. Het stimuleren van het leesplezier is vermoedelijk de meest aangewezen weg om toekomstige ontwikkelingen positief te beïnvloeden.

Leerling B is een jongen uit een arbeidersmilieu. Volgens het oordeel van de leerkracht is zijn mondelinge taalvaardigheid aan de zwakke kant. Zijn zelfstandigheid, werkhouding en concentratie zijn middelmatig. De sociale omgang is duidelijk positief, terwijl zijn emotionele aanspreekbaarheid tamelijk gering is.

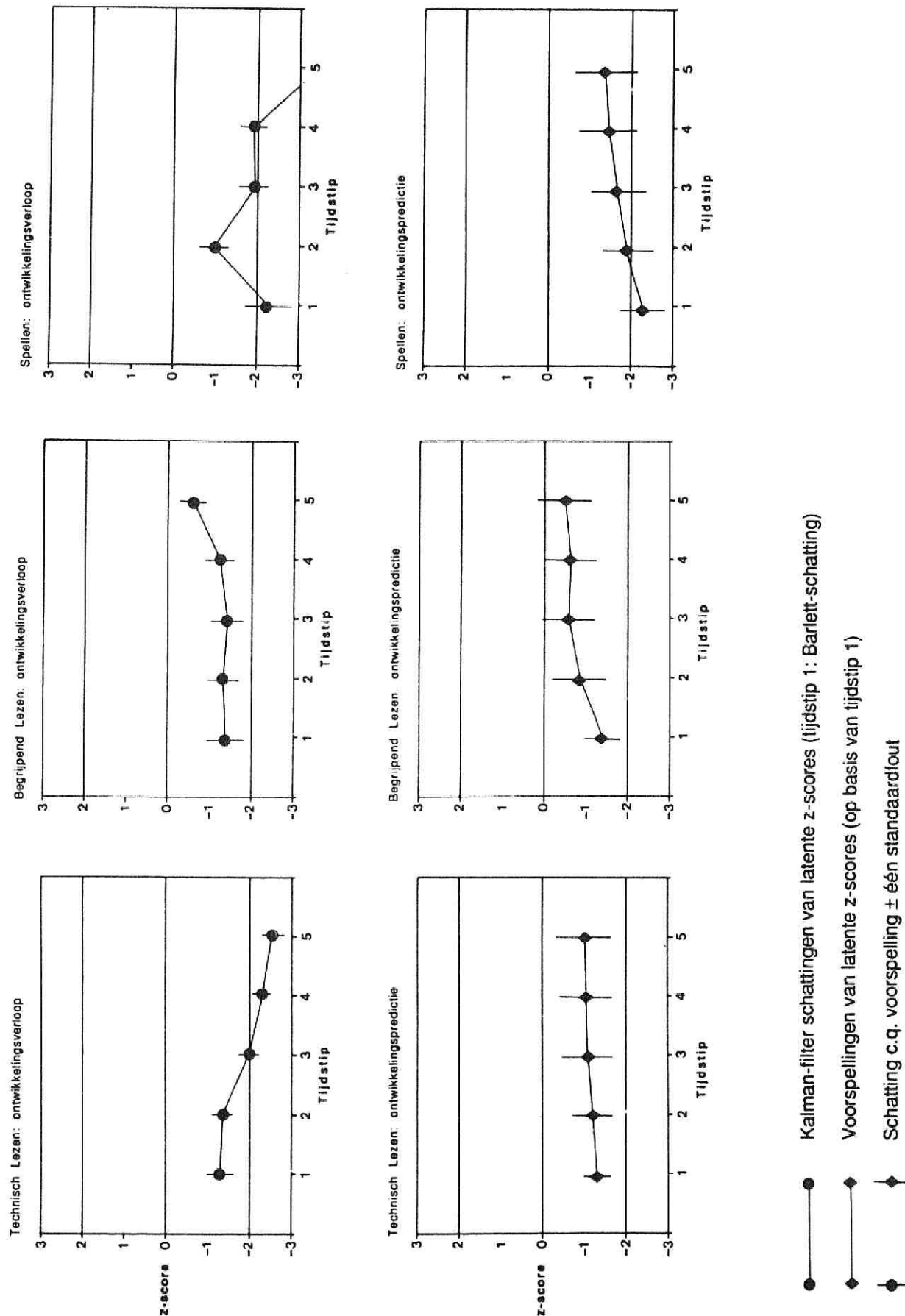
Op de eerste vier meetmomenten liggen zijn prestaties voor begrijpend lezen laag, ongeveer 1,5 standaardafwijking beneden het gemiddelde. Maar na het vierde meetmoment treedt een duidelijke stijging op. Bekijkt men het plaatje met de voorspellingen vanaf het tweede meetmoment, dan wordt daar een geleidelijke stijging verwacht. Het niveau dat daadwerkelijk eind leerjaar 3 wordt bereikt, komt overeen met de voorspelling, maar de stijging is later opgetreden dan in het begin van het tweede leerjaar werd verwacht.

Halfweg het eerste leerjaar behaalt deze leerling voor spelling een lage score, die meer dan twee standaardafwijkingen beneden het gemiddelde ligt. Op het tweede meetmoment treedt een verbetering op, die wordt gevolgd door een daling. Eind leerjaar 3 is het spellingniveau zelfs zo laag, dat het meer dan drie standaardafwijkingen beneden het gemiddelde ligt. Dit wijst op zeer ernstige spellingproblemen. Bij de voorspellingen vanaf het tweede meetmoment werd een geleidelijke verbetering verwacht. Op het derde en vierde meettijdstip lag de feitelijke score nog binnen de grenzen van de standaardfout, maar op het vijfde meetmoment is sprake van een ernstige terugval.

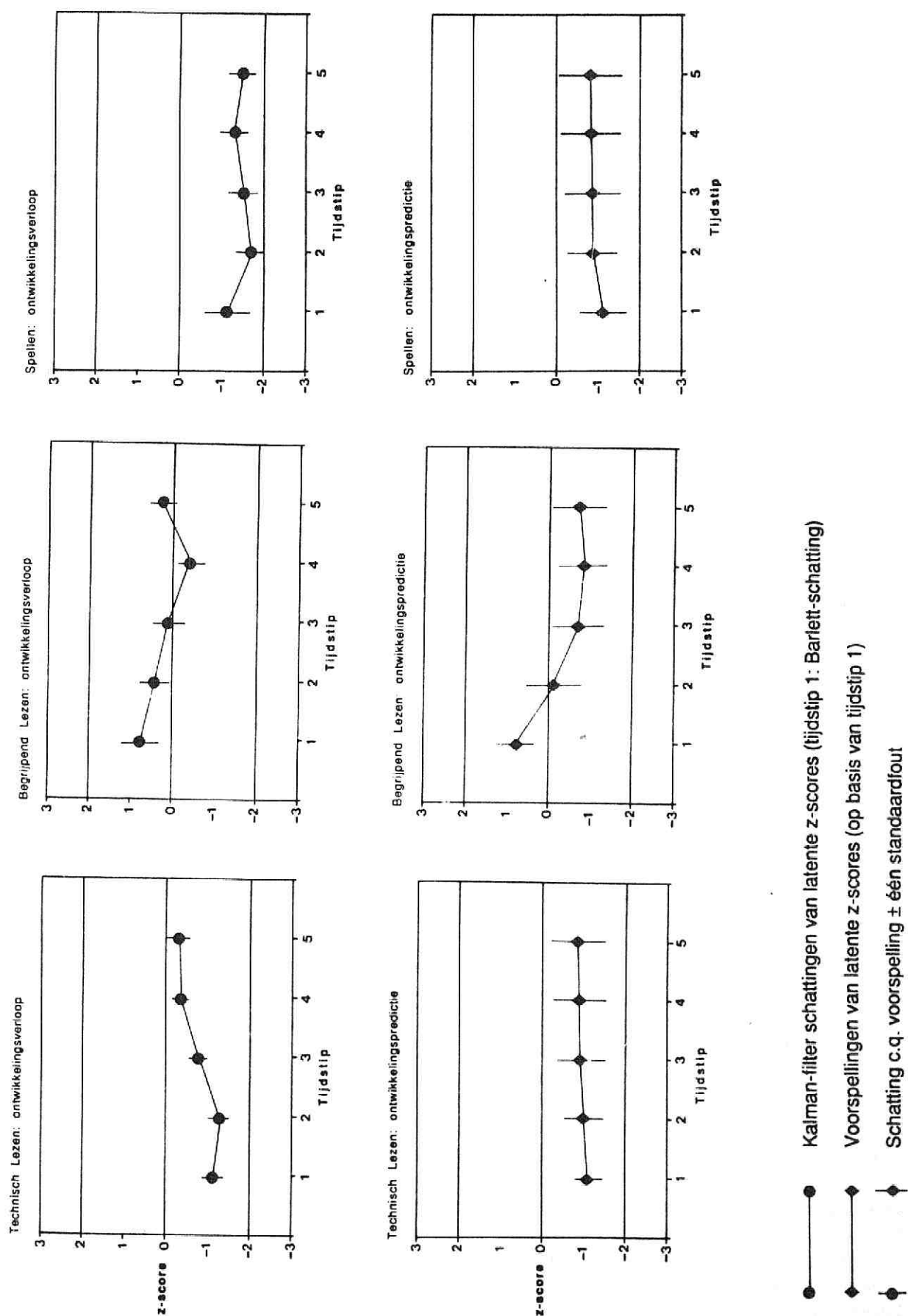
Ook de prestaties voor technisch lezen dalen vrij sterk, naar een niveau dat meer dan twee standaardafwijkingen beneden het gemiddelde ligt. Dit is in tegenstelling met de voorspellingen vanaf begin leerjaar 2, die een lichte stijging deden verwachten.

Deze leerling heeft duidelijk lees- en spellingsmoeilijkheden. Bij nader onderzoek bleek, dat het relatief lang geduurd heeft, voordat de foneem-grafeemkoppeling bij deze leerling geautomatiseerd was. Maar begin leerjaar 2 beheerste hij deze voldoende. Aanvankelijk waren er ook problemen met de auditieve analyse en synthese, maar door extra oefening zijn die reeds in het eerste leerjaar opgelost. Bij minder frequente woorden blijft deze leerling echter spellend lezen. Bij spelling tracht hij de woorden te schrijven zoals hij die hoort. In een vrij groot aantal gevallen slaagt hij daar ook in, maar hij gaat meestal in de fout als de schrijfwijze van een woord fonetisch afwijkt. Hij heeft vooral moeite met woorden die visueel ingeprent moeten worden zoals woorden met ei of ij en ou of au. Ook open lettergrepen en spellingpatronen vormen een bron van talloze spelfouten. Vermoedelijk is hier sprake van structuurzwakte. Orthografische deelstructuren (spellingpatronen en lettergrepen) worden niet herkend en dus ook niet onthouden. Dit wrekt zich vooral bij meer complexe woorden in het derde leerjaar. Dat begrijpend lezen een wat minder zorgelijk beeld te zien geeft, hangt vermoedelijk samen met conceptuele kennis die de leerling langs andere dan schriftelijke weg heeft opgedaan. Het is noodzakelijk voor dit kind speciale maatregelen te treffen, wil het niet tot de groep functionele analfabeten gaan behoren. In elk geval is nader onderzoek nodig.

Leerling C is een meisje. Haar vader is werkzaam als lager employé en heeft een middelbare opleiding genoten. Volgens het oordeel van de leerkracht is het meisje verlegen in de sociale omgang, maar heeft zij een vrij positieve werkhouding. Haar mondelinge taalvaardigheid is zwak-middelmatig.



Figuur 5: Leerling B: ontwikkelingsverloop en ontwikkelingspredictie (op basis van tijdstip 1) van technisch lezen, begrijpend lezen en spellen.



Figuur 6: Leerling C: ontwikkelingsverloop en ontwikkelingspredictie (op basis van tijdstip 1) van technisch lezen, begrijpend lezen en spellen.

Het begrijpend lezen laat op de eerste vier meetmomenten een daling zien, maar daarna treedt een stijging op. Opmerkelijk is, dat vanaf begin leerjaar 2 een veel sterkere daling werd verwacht. Het daadwerkelijk bereikte niveau eind leerjaar 3 ligt duidelijk hoger dan de voorspelling.

Het bereikte resultaat bij spelling ligt aanzienlijk lager dan bij begrijpend lezen. De voorspellingen vanaf het begin van het tweede leerjaar liggen ongeveer op het niveau van 1 standaardafwijking beneden het gemiddelde en zijn gunstiger dan het daadwerkelijk bereikt niveau.

Het technisch lezen vertoont vanaf het tweede meettijdstip een duidelijke stijging en komt eind leerjaar 3 bijna op het gemiddelde peil. Begin leerjaar 2 was er een lichte stijging voorspeld, maar de werkelijke stijging overtreft duidelijk het voorspelde niveau.

Als men bij dit meisje de ontwikkeling van de drie onderscheiden vaardigheden bekijkt, ziet men duidelijk verschillen. Terwijl het technisch lezen verbetert, daalt aanvankelijk het niveau van het begrijpend lezen. Een voor de hand liggende verklaring is, dat het begrijpend lezen werd gehinderd door een onvoldoende automatisering van de woordherkenningsprocessen. Als deze processen op peil zijn zal ook het niveau van begrijpend lezen stijgen. Vaak ziet men bij dit soort leerlingen dat de spelling wat langer achterblijft dan de decodeersnelheid. Gezien de positieve werkhouding van dit meisje kan oefening de aanvankelijke moeilijkheden aanzienlijk hebben verminderd en is de kans aanwezig dat ook de spellingsmoeilijkheden voor een deel overwonnen kunnen worden. Inspectie van de aard van de spelfouten doet vermoeden dat aan twee zaken specifieke aandacht geschonken moet worden. In de eerste plaats is dat de inprenting van weinig frequente woorden die niet klankzuiver zijn. In de tweede plaats zijn dat de spellingregels met betrekking tot de open klinker op het eind van een lettergreep en de verdubbeling van de medeklinker na een beklemtoonde korte klinker. De prestaties bij begrijpend lezen doen vermoeden dat dit meisje voldoende intelligent is om deze regels te leren toepassen.

Om misverstanden te voorkomen is het nuttig erop te wijzen dat hier geen volledige diagnoses werden gegeven. Daartoe zouden meer gegevens in de overwegingen betrokken moeten worden. Maar deze voorbeelden laten wel zien, dat met behulp van het computerprogramma LISKAL-M1 en het longitudinaal referentiekader een veel betrouwbaarder beeld gegeven kan worden van daadwerkelijke en te verwachten ontwikkelingen van instrumentele vaardigheden dan voorheen. Een meer gedetailleerde diagnostiek kan dit als uitgangspunt nemen.

5. CONCLUSIES EN DISCUSSIE

Zoals in de inleiding is aangegeven beoogde het project "Preventie van Leesmoeilijkheden" meer inzicht te verschaffen in de longitudinale samenhang van technische lezen, begrijpend lezen en spellen. Gezien de doelstelling van het project hoopte men daardoor een bijdrage te leveren aan een reductie van het aantal leerlingen met lees- en spellingsmoeilijkheden. Die bijdrage werd vooral gezocht ten aanzien van het tijdig opsporen en begeleiden van leerlingen die dreigen uit te vallen op grond van lees- en spellingsmoeilijkheden. De hier voorgestelde procedure maakt dit beter mogelijk doordat schattingen met aanzienlijk kleinere standaardfouten worden verkregen dan mogelijk is in een traditionele screeningsprocedure. Bovendien geeft de procedure op basis van informatie uit het verleden voorspellingen van toekomstige prestaties en kunnen maatregelen ter verbetering aan de hand van de voorspellingen geëvalueerd worden. Inzicht in de longitudinale samenhang van technisch lezen, begrijpend lezen en spelling is overigens niet alleen van belang voor risico-leerlingen. In het kader van een leerlingvolgsysteem is het voor alle leerlingen wenselijk te kunnen nagaan hoe de ontwikkeling verloopt en in hoeverre deze in overeenstemming is met de verwachtingen.

Paragraaf 2 vat de belangrijkste resultaten van de eerder uitgevoerde Lisrel-analyses en het daarop gebaseerde longitudinale Lisrel-model samen (Figuur 1). Meer dan in het tot nu toe vooral cross-sectioneel gerichte leesonderzoek komen in de longitudinale analyses technisch

lezen, begrijpend lezen en spellen naar voren als relatief autonome vaardigheden. Op grond hiervan kwamen we in een eerdere publikatie (Mommers et al., 1986) tot de nog tamelijk algemeen gestelde conclusie dat leerlingen met lees- en spellingsmoeilijkheden onderzocht moeten worden op ieder van de drie vaardigheden, aangezien aanzienlijke discrepanties in de respectieve ontwikkelingsniveaus kunnen optreden (zie bijvoorbeeld de ontwikkelingscurven van de leerlingen in de Figuur 4-6). Dit artikel laat zien hoe door toepassing van het Kalman-filter de kennis die in het Lisrel-model ligt opgeslagen meer specifiek bruikbaar is te maken voor betrouwbare schattingen van de latente variabelen, die als uitgangspunt kunnen dienen bij de beoordeling van de vorderingen van de leerlingen. Dit is vooral maar niet uitsluitend van belang voor risico-leerlingen. Zo'n leerlingvolgsysteem is overigens niet geschikt voor gedetailleerde, dagelijkse voortgangscontrole en ook niet voor een verfijnde diagnostiek. Daarvoor zijn meer en ook andersoortige gegevens nodig.

Aan de hand van drie concrete leerlingen werd in paragraaf 4 geïllustreerd hoe meer betrouwbare schattingen een bijdrage kunnen leveren aan het opsporen en diagnostiseren van lees- en spellingsmoeilijkheden. Het gepresenteerde microcomputer-programma LISKAL-M1 stelt de leerkracht in staat om zelf in de klassesituatie de longitudinale, op het Kalman-filter gebaseerde schattingen uit te voeren.

Vervolgonderzoek zou zich enerzijds kunnen richten op een vereenvoudiging en anderzijds op een uitbreiding van het gehanteerde Lisrel-model. De vereenvoudiging komt tegemoet aan het probleem dat het huidige Lisrel-model een aantal instrumenten bevat (zie Tabel 1) die niet standaard worden afgenomen in het regulier onderwijs. De toepassing van LISKAL-M1 vereist niettemin dat alle instrumenten op de voorgeschreven tijdstippen worden afgenomen en vergt dus nogal wat tijd en discipline van de leerkracht. Op dit moment wordt een vereenvoudigd Lisrel-model en een daarop aansluitend programma LISKAL-M2 geconstrueerd waarin uitsluitend instrumenten voorkomen die voor leerkrachten goed toegankelijk zijn en gemakkelijk zijn af te nemen. LISKAL-M2 wordt bovendien zo geconstrueerd dat het eerste tijdstip voor afname van instrumenten (in LISKAL-M2 niet 0 maar 9 maanden leesonderwijs: tijdstip 1 in Figuur 1) eventueel ook naar latere tijdstippen kan worden verschoven (tijdstippen 2 t/m 5 in Figuur 1). Hoewel LISKAL-M2 uiteraard meer informatie en betere schattingen geeft, naarmate het begintijdstip vroeger wordt gekozen, blijven de resultaten niettemin geldig wanneer om een of andere reden pas later met de procedure kan worden gestart.

Het tot nu toe gehanteerde Lisrel-model richt zich op de ontwikkeling van de drie vaardigheden: technisch lezen, begrijpend lezen en spellen, daarbij gebruik makend van 37 geobserveerde en 16 latente variabelen (zie Figuur 1). Ondanks de omvangrijkheid van het model is kritiek mogelijk op het feit dat factoren die vaak worden genoemd als determinanten van lees- en spellingsprestaties zoals mondelinge taalvaardigheid, korte duurgeheugen en motivatie niet expliciet zijn opgenomen. Het model leent zich weliswaar voor een meer betrouwbare, longitudinaal gebaseerde onderkenning van leesmoeilijkheden maar laat de gebruiker in de steek wanneer specifieke informatie is gewenst over de oorzaken van het falen en over de ontwikkeling van deze oorzakelijke factoren over de tijd. Op dit punt is de gepresenteerde procedure aan soortgelijke kritiek onderhevig als de discrepantie-formule voor de onderkenning van leerstoornissen (zie bijv. Oud, 1985; Van der Wissel, 1984). De discrepantieformule, waarvan verschillende varianten bestaan, diagnostiseert een leerling als leergestoord, wanneer zijn prestaties duidelijk achterblijven bij zijn intelligentie-niveau (beide uitgedrukt in bijvoorbeeld z-scores) maar geeft evenmin informatie over de specifieke factoren die buiten de intelligentie verantwoordelijk zijn voor de lage prestaties. Nieuw longitudinaal onderzoek waarin tevens determinanten van leerstoornissen worden opgenomen, zou ondersteunend kunnen werken voor de klinische praktijk ten aanzien van lees- en spellingsproblemen. Uiteindelijk hopen we op basis van een uitgebreid Lisrel-model een programma LISKAL-M3 te construeren dat meer specifieke, bij de oorzaken aansluitende aanwijzingen voor de behandeling mogelijk maakt.

Tenslotte willen we nog wijzen op de mogelijkheid van een gecentraliseerde toepassing van LISKAL-M2. Al of niet in samenhang met een leerlingvolgsysteem wordt de laatste tijd van

verschillende kanten aangedrongen op de inrichting van een data-bank waar centraal longitudinale gegevens worden opgeslagen over het onderwijs, in het bijzonder leerlinggegevens. Met betrekking tot lezen ligt het voor de hand gegevens op te nemen van instrumenten die gemakkelijk af te nemen zijn. Alle toetsen die zijn opgenomen in LISKAL-M2, zijn voor iedereen verkrijgbaar, deels bij het CITO en deels bij uitgeverijen. Voorstelbaar is dat de centraal verzamelde gegevens ook centraal worden geanalyseerd en dat de aangesloten scholen voor iedere leerling op ieder gewenst tijdstip de bijgewerkte informatie van LISKAL-M2 kunnen opvragen, zowel op individueel als klas- en schoolniveau. Dit heeft een viertal voordelen. In de eerste plaats wordt de leerkracht ontlast en zal hij gemakkelijker van de diagnostische informatie gebruik maken. In de tweede plaats ontstaat zo ook op centraal niveau een overzicht hoe lees- en spellingsvaardigheden zich ontwikkelen en verkrijgt men tevens indicaties over ernst, frekwentie en spreiding van leesmoeilijkheden over klassen en scholen hetgeen beleidsmatig van groot belang kan zijn. In de derde plaats kan het Lisrel-model dat aan LISKAL-M2 ten grondslag ligt, bij wijze van spreken "on-line" hergeschat worden voor steeds grotere aantallen leerlingen, zodat de steekproeffluctuaties op den duur verwaarloosbaar klein worden. In de vierde plaats ontstaat de mogelijkheid om op gezette tijden na te gaan, of wellicht de parameters van het Lisrel-model aan bepaalde historische wijzigingen onderhevig zijn en bijgesteld moeten worden.

LITERATUUR

- Dongen, D. van (1984). *Leesmoeilijkheden: Naar diagnostiserend onderwijzen bij het leren lezen*. Tilburg: Zwijssen.
- Janssens, F. J. G. (1984). *Vingers aan de pols. Over het registreren van leerlinggegevens*. Onderwijskundige Brochuren Reeks, 305. Tilburg: Zwijssen.
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. New York: Academic Press.
- Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood: User's guide*. Chicago: International Educational Services.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* (Trans. ASME, Ser. D), 82, 35-45.
- Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering* (Trans. ASME, Ser. D), 83, 95-108.
- Kwakernaak, H., & Sivan, R. (1972). *Linear optimal control systems*. New York: Wiley Interscience.
- Lewis, F. L. (1986). *Optimal estimation*. New York: Wiley.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.
- Lieshout, C. F. M. van, & Wardenaar, E. (1985). *Onderwijsdifferentiatie en computergebruik voor beheer en evaluatie van onderwijs*. 's-Gravenhage: Wetenschappelijke Raad voor het Regeringsbeleid.
- Melis, G. N. (1985). *Leerlingenkenmerken en zorgbreedte. Naar onderwijsarrangementen voor kinderen met leerproblemen in het basisonderwijs*. Amsterdam: Elsevier.
- Mommers, M. J. C. (1987). An investigation into the relation between word recognition, reading comprehension and spelling skills in the first two years of primary school. *Journal of Research in Reading*, 10, 122-143.
- Mommers, M. J. C., Leeuwe, J. F. J. van, Oud, J. H. L., & Janssens, J. M. A. M. (1986). Decoding skills, reading comprehension and spelling: A longitudinal investigation. *Tijdschrift voor Onderwijsresearch*, 11, 97-113.
- Mommers, M. J. C., & Oud, J. H. L. (1984). Op zoek naar de samenhang tussen technisch lezen, begrijpend lezen en spellingsvaardigheid in de eerste twee leerjaren van de lagere school. In L. F. W. de Klerk & A. M. P. Knoers (Red.), *Onderwijspsychologisch onderzoek* (pp. 193-202). Lisse: Swets & Zeitlinger.
- Oud, J. H. L. (1978). *Systeem-methodologie in sociaal-wetenschappelijk onderzoek*. Nijmegen: Alfa.
- Oud, J. H. L. (1982). De stap van input-output naar toestand in sociaal-wetenschappelijk onderzoek. In A. F. G. Hanken & J. H. L. Oud (Red.), *Onderneming en overheid in systeemdynamisch perspectief* (pp. 17-33). 's-Gravenhage: Academic Service.
- Oud, J. H. L. (1985). Schooluitval in dialoog. *Tijdschrift voor Orthopedagogiek*, 9, 407-417.
- Oud, J. H. L., Bercken, J. H. L. van den, & Essers, R. J. (1986). Longitudinal factor score estimation using the Kalman filter. *Kwantitatieve Methoden*, 7, (20), 109-129.

- Otter, P. W. (1985). *Dynamic feature space modelling, filtering and self-tuning control of stochastic systems*. New York: Springer Verlag.
- Roth, S. F., & Beck, J. L. (1987). Theoretical and instructional implications of the assessment of two microcomputer word recognition programs. *Reading Research Quarterly*, 27, (2), 197-219.
- Wissel, A. van der (1984). *Schooluitval: Een reeks onderzoeken naar cognitieve aspecten van het falen bij kinderen met ernstige leerproblemen*. Groningen: Proefschrift RU Groningen.

Manuscript ontvangen: 25-3-1987.

Definitieve versie ontvangen: 16-11-1987.

Notities en Commentaren

HET GEBRUIK VAN DE PMT-K BIJ CURAÇAOSE SCHOOLKINDEREN: EEN VALIDATIESTUDIE

Paul Vedder

ABSTRACT

THE USE OF THE PMT-K IN CURAÇAOAN PRIMARY SCHOOLS: A VALIDATION STUDY.

A study in achievement motivation was set up in the fourth, fifth and sixth grade of three Curaçaoan primary schools. A total of 282 children participated. A dutch achievement motivation test, the PMT-K, was translated into Papiamentu, the primary language of most children in Curaçao. Compared to dutch norms Curaçaoan pupils have a relatively strong achievement motivation. The findings further indicate that achievement motivation is related with self-evaluations of learning problems. The relation with more objective criteria for failure and success is less firm. Scores on subscales measuring fear of failure differentiate between children who never have contacts with their fathers and children who once in a while have contacts with their fathers. Children who never see their fathers have better scores.

1. INLEIDING

In de handleiding bij de nieuwste versie van de prestatie-motivatietest voor kinderen van Hermans tot dusver verzamelde validatiegegevens als volgt samen (zie Hermans, 1983, p. 20):

- Correlaties met rapportcijfers worden het meest consistent gevonden bij de prestatie-motivatieschaal en minder consistent bij de schalen m.b.t. faalangst.
- De correlaties van de PMT-K schalen met rapportcijfers liggen over het algemeen hoger dan die met schoolvorderingsindices.
- Ten aanzien van het criterium "advies schoolhoofd" laat de schaal voor prestatie-motivatatie de meest duidelijke verschillen zien. De faalangtschalen laten wel verschillen zien, maar deze zijn moeilijker interpreteerbaar.
- Uit leerkracht beoordelingen komt naar voren dat leerlingen met een hoge prestatie-motivatiescore door de leerkrachten als "meer voorbeeldig" worden gezien dan leerlingen met een lage prestatie-motivatiescore.
- Leerlingen met een hoge score voor negatieve faalangst rapporteren wel meer angst in taaksituaties maar rapporteren niet méér vermijding.
- Leerlingen met een hoge score voor positieve faalangst ervaren zichzelf als meer bekwaam in taaksituaties dan leerlingen met een lage score voor positieve faalangst.

In een onderzoek bij vierde, vijfde en zesde klas leerlingen van Curaçaose lagere scholen (de oude stijl Nederlandse lagere school) werden de bevindingen, zoals samengevat bij de eerste twee punten, nog eens onderstreept (zie Vedder, 1987a). Er zijn evenwel ook interessante aanvullende resultaten gevonden die hierna worden gepresenteerd.

2. METHODE

Proefpersonen

Aan het onderzoek werd deelgenomen door 107 vierde klassers, 91 vijfde klassers en 84 zesde klassers van drie Curaçaose lagere scholen. Iets meer dan 56% van de leerlingen zijn meisjes. De gemiddeld leeftijd bedraagt 12 jaar en 1 maand (s.d. 19.4 maanden). Zevenendertig procent van de leerlingen is één keer blijven zitten en 19% meer dan één keer. Wat betreft de leeftijd en het zittenblijverspercentage zijn de leerlingen representatief voor de Curaçaose populatie lagere schoolkinderen. De kinderen wonen overwegend in buurten met een lage welstand.

Materiaal

Voor het onderzoek is de PMT-K versie 1976 vertaald in het Papiaments en enigszins aangepast aan de Curaçaose situatie. De interne consistentie (Cronbach's alpha) van de diverse schalen varieert in de drie leerjaren van .64 tot .82:

Verder is er gebruik gemaakt van een leerproblemenschaal. Het betreft een vertaling, aanpassing en uitbreiding van een schaal ontwikkeld door De Jong en Batenburg (1985). De items hebben betrekking op de wijze waarop kinderen zelf het leren en leerproblemen ervaren, de beleving van de school en de perceptie van de wijze waarop leerkrachten het leren en leerprestaties beoordelen. De mate van interne consistentie (Cronbach's alpha) varieert in de drie leerjaren van .78 tot .81.

Als derde instrument is een sociometrische vragenlijst gebruikt waarop kinderen m.b.t. zes activiteiten kunnen aangeven met welke twee klasgenoten ze het liefst een betreffende activiteit zouden willen uitvoeren. Op basis van de gegevens kan worden vastgesteld hoe populair een leerling is in zijn klas. Op de zelfde vragenlijst kunnen de leerlingen ook aangeven met welke van de volgende personen ze de betreffende activiteiten het liefst zouden uitvoeren: ouders, broers/zussen, vrienden, andere volwassen personen. Tenslotte kunnen de leerlingen op de vragenlijst aangeven hoe vaak ze hun vader en moeder ontmoeten (elke dag, bijna elke dag, af en toe, of nooit).

Procedure

De toetsen en vragenlijsten zijn klassikaal afgenomen door een speciale toetsleider, die de leerlingen instrueerde in het Papiaments. De items van alle instrumenten werden één voor één behandeld. Deze werkwijze is noodzakelijk, omdat leerlingen niet allemaal goed in staat zijn Papiaments te lezen (zie voor meer informatie over de beheersing van het Papiaments en het Nederlands, Vedder, 1987b). De gegevens zijn verzameld in de periode van januari tot mei 1987.

3. RESULTATEN

De prestatiemotivatie bij Curaçaose en Nederlandse leerlingen

In Tabel 1 worden de gemiddelde scores van Curaçaose jongens en meisjes per leerjaar gepresenteerd. De groepen Nederlandse jongens en meisjes betreffen bijna vijfduizend leerlingen uit de eerste klassen van verschillende vormen van voortgezet onderwijs (zie Hermans, 1983). De standaarddeviaties van de Nederlandse groepen zijn niet bekend.

In Tabel 1 valt op dat er van het vierde naar het zesde leerjaar sprake is van een lichte afname van de prestatiemotivatie, een lichte toename van de negatieve faalangst en een lichte afname van de positieve faalangst. Verder valt er een verschil waar te nemen tussen jongens en meisjes wat betreft de prestatiemotivatie en de sociale wenselijkheid. Meisjes hebben een sterker prestatiemotief en zijn meer geneigd sociaal wenselijke antwoorden te geven dan jongens. Jongens en meisjes lijken niet te verschillen wat betreft de negatieve en positieve faalangst.

Een vergelijking van de Curaçaose gegevens met de Nederlandse resulteert in de vaststelling dat Curaçaose leerlingen over het algemeen een sterker prestatiemotief hebben, vooral de

Tabel 1. De motivatie van de leerlingen per leerjaar en per geslachtsgroep (p.m. is prestatie-motivatie; s.w. is de neiging sociaal wenselijk te antwoorden; f- is de negatieve faalangst en f+ is de positieve faalangst).

groep	p.m. $\bar{X}$	s.d.	s.w. $\bar{X}$	s.d.	f- $\bar{X}$	s.d.	f+ $\bar{X}$	s.d.
Lj. 4	20.3	4.7	9.4	2.9	7.0	3.1	10.1	3.7
5	20.6	5.9	9.1	2.9	7.7	3.6	10.2	3.7
6	18.6	6.0	8.7	3.5	7.9	3.4	9.1	3.7
jongens	17.9	5.5	8.3	3.0	7.3	3.3	10.3	3.3
meisjes	21.5	5.1	9.7	3.1	7.6	3.4	10.0	4.0
Ned. j.	17.2		9.8		7.7		9.8	
Ned. m.	17.4		10.9		9.4		8.0	

meisjes, en dat ze minder geneigd zijn sociaal wenselijk te antwoorden. Wat betreft de faalangst zijn er slechts duidelijke verschillen vast te stellen tussen de Curaçaose en Nederlandse meisjes. De Curaçaose meisjes hebben over het algemeen minder te kampen met negatieve faalangst, terwijl hun positieve faalangst sterker is dan die van de Nederlandse meisjes. Over het algemeen ontstaat er een behoorlijk positief beeld over de motivatie van Curaçaose lagere schoolkinderen.

Prestatiemotivatie en leerproblemen

In Tabel 2 wordt informatie gegeven over de relatie (Pearson p.m. correlaties) tussen scores op de diverse motivatiematen en de scores op de leerproblemenschaal. De correlaties worden gepresenteerd per leerjaar, per geslacht en voor alle leerlingen samen.

Tabel 2. De samenhang tussen motivatie en leerproblemen.

	leerproblemen X		
	p.m.	f-	f+
leerjaar 4	-.55*	.35*	-.14
5	-.53*	.28*	-.39*
6	-.62*	.19*	-.03
jongens	-.49*	.21*	-.24*
meisjes	-.54*	.37*	-.21*
allen	-.56*	.28*	-.20*

* $p < .05$

In Tabel 2 blijkt dat er een redelijk sterke negatieve samenhang bestaat tussen prestatie-motivatie en het eigen oordeel van leerlingen over leerproblemen. Naarmate leerlingen meer problemen menen te hebben, hebben ze een lagere prestatie-motivatie.

De samenhangen zijn minder sterk of soms zelfs afwezig als het om de negatieve en positieve faalangst gaat.

In een stapsgewijze regressie-analyse is de relatieve bijdrage van de scores op diverse maten tot de verklaring van de variantie in prestatie-motivatiescores nagegaan. Het oordeel van leerlingen over hun leerproblemen heeft het grootste aandeel in de verklaring ($R^2 = .29$), daarna komt de score voor sociaal wenselijk antwoordgedrag ($R^2 = .41$), de som van de rapportcijfers voor passen ($R^2 = .45$), de score voor positieve faalangst ($R^2 = .47$) en tenslotte de som van de rapportcijfers voor kerst ($R^2 = .48$).

Prestatiemotivatie en vriendschapsrelaties

In studies over vriendschapsrelaties bij kinderen blijkt dat bij de keuze van vrienden enige congruentie wordt gerealiseerd wat betreft eigen eigenschappen of kenmerken en eigenschappen van de vrienden. Dit is evident wat betreft leeftijd en sexe, maar geldt ook voor persoonlijkheidskenmerken (vgl. Asher, Oden & Gottman, 1977; Byrne & Griffith, 1966). Wij zijn nagegaan of dit streven naar congruentie ook bestaat voor de scores op de PMT-K.

In Tabel 3 wordt informatie gepresenteerd over de samenhang tussen eigen scores en scores van de meest populaire klasgenoot op een aantal maten. Daarbij worden twee groepen onderscheiden: de hele groep (282) en de groep met leerlingen die een sterke voorkeur hebben om diverse activiteiten uit te voeren met vrienden (42).

Tabel 3. De samenhang tussen eigen scores en scores van de meest favoriete klasgenoot.

	allen	voorkeur vrienden
prestatiemotivatie	.04	.02
negatieve faalangst	— .05	— .05
positieve faalangst	.06	.20
populariteit	.33	.54
rapport kerst	.22	.31
rapport pasen	.27	.39

Het blijkt dat kinderen bij de keuze van vrienden over het algemeen congruentie realiseren wat betreft rapportcijfers en wat betreft de populariteit. In de groep kinderen met een sterke voorkeur voor vrienden zijn wat dit betreft de relaties sterker dan in de totale groep. Over het algemeen lijkt de prestatiemotivatie en de faalangst iets dat niet als belangrijk onderscheidend kenmerk bij de keuze van vrienden een rol speelt. Wellicht komt dit doordat er in de omgeving van kinderen (de leerkrachten en ouders) weinig expliciete aandacht aan wordt besteed, en doordat het niet een duidelijk manifeste set van kenmerken betreft.

Motivatie en de regelmaat van contacten met de ouders

Er is nagegaan of er een samenhang bestaat tussen de regelmaat waarmee kinderen hun moeder zien en hun scores op de PMT-K. Er bleken geen zinvolle relaties te bestaan. Hetzelfde is gedaan met betrekking tot de regelmaat van contacten met de vaders. Ook in dit geval is geen zinvolle relatie gevonden. Vervolgens is gekeken naar het verschil in motivatiescores tussen kinderen die hun vader af en toe zien (63) en kinderen die hun vader nooit zien (31). Het blijkt dat kinderen die hun vader af en toe zien een gemiddeld hogere score hebben voor negatieve faalangst ($\bar{X}$ 8.4; s.d. 3.1) en een lagere score voor positieve faalangst ($\bar{X}$ 9.5; s.d. 4.1), dan kinderen die hun vader nooit zien (resp. $\bar{X}$ 7.0; s.d. 3.5 en $\bar{X}$ 11.1; s.d. 4.1). Wat betreft de faalangst van Curaçaose kinderen lijkt het voor hen gunstiger om hun vader helemaal niet te zien dan om hun vader soms te zien. Dit resultaat geeft wellicht iets weer van een complexe en onzekerheids versterkende opvoedingssituatie van kinderen die hun vader slechts soms zien.

4. DISCUSSIE

Zoals gesteld in de inleiding is er veel onderzoek gedaan naar de validiteit van de PMT-K. Het in dit artikel beschreven onderzoek voegt daaraan enige informatie toe.

Het is vreemd dat op Curaçao, waar leerlingen objectief gezien zoveel falen (bijvoorbeeld zittenblijven) de prestatiemotivatie toch relatief hoog is. Wellicht geldt door het vele voorkomen van het zittenblijven en het hebben van leerproblemen de regel "gedeelde smart is halve smart". Het is opmerkelijk dat het persoonlijk oordeel over leerproblemen behoorlijk sterk

samenhangt met de prestatie-motivatie. De resultaten, inclusief die van de regressie-analyse duiden er op dat prestatie-motivatie, zoals gemeten met de PMT-K, meer van doen heeft met de persoonlijke beoordeling van het slagen en falen in leersituaties dan met meer objectieve indices voor slagen en falen.

De vaagheid van het concept prestatie-motivatie (zie Van den Bercken, De Bruyn & Bergen, 1986; De Bruyn, 1979) wordt onderstreept door de bevinding dat prestatie-motivatie geen rol speelt bij de keuze van vrienden.

Het resultaat m.b.t. faalangst in relatie tot de contacten met de vader verdient aandacht in verder onderzoek. Daarbij kan worden nagegaan in hoeverre de gevonden samenhangen specifiek zijn voor de Curaçaose samenleving.

LITERATUUR

- Asher, S. R., Oden, S. L. & Gottman, J. M. (1977). Children's friendships in school settings. In: L. G. Katz (Ed.) *Current topics in early childhood education*. (vol. 1). Norwood, N.J.: Ablex.
- Bercken, J. H. L. van den, Bruyn, E. E. J. de & Bergen, Th. C. M. (red.) (1986). *Achievement and taskmotivation*. Lisse: Swets & Zeitlinger.
- Bruyn, E. E. J. de (red.) (1979). *Ontwikkelingen in het onderzoek naar prestatie-motivatie*. Lisse: Swets & Zeitlinger.
- Byrne, D. & Griffith, W. A developmental investigation of the law of attraction. *Journal of personality and social psychology*, 4 (1966), 699-702.
- Hermans, H. J. M. (1983). *Prestatie-motivatietest voor kinderen - 83*. Lisse: Swets & Zeitlinger.
- Vedder, P. (1987a). *De samenhang tussen schoolse en buitenschoolse socialisatie bij Curaçaose lagere schoolkinderen*. Curaçao: SEDUKAL.
- Vedder, P. (1987b). *Learning in a second language: a study in Curaçaoan primary schools*. Curaçao: KONDESA.

Manuscript ontvangen 30-10-1987

Boekbespreking

Terlouw, C., *De FUNDES-procedure in onderwijsontwikkeling: Evaluatie van een procedure van onderwijsontwikkeling voor het leren probleemoplossen.*

Terlouw heeft het proefschrift, dat hij op 19 juni verdedigde aan de Universiteit Twente (promotoren Plomp, Carpay en Ruiter), een motto meegegeven dat er vrij vertaald op neer komt dat onderwijsontwikkeling moeilijk is en dat je er bedroefder en wijzer van wordt. Naar mijn mening is dit een zeer goed gekozen motto. Het proefschrift maakt op uitgebreide wijze duidelijk dat onderwijsontwikkeling erg moeilijk is. Je kunt je dan ook goed voorstellen dat Terlouw er regelmatig bedroefd van is geworden. Uiteindelijk komt hij er echter duidelijk wijzer uit. Spijtig is wel dat hij voor het beschrijven van deze ervaringen maar liefst 568 pagina's tekst nodig heeft. Ook hier geldt de kwalificatie die Knoers onlangs aan het proefschrift van Van Oers gaf: "In der Beschränkung zeigt sich der Meister".

Hoewel ik van mening ben dat Terlouw zijn boodschap ook en beter in de helft van het door hem nu gebruikte aantal pagina's had kunnen opschrijven, ben ik wel degelijk onder de indruk van het door hem verrichte monnikenwerk, de originaliteit van de gevolgde onderzoeksmethode en de kwaliteit van de ontwikkelde onderwijsmethode in een zeer moeilijk, maar belangrijk domein als het leren ontwerpen van beleid.

De centrale vraagstelling van het proefschrift is: "In welke mate is de gegeneraliseerde handelingstheoretische FUNDES-procedure voor onderwijsontwikkeling in het leren oplossen van problemen in het Hoger Onderwijs bruikbaar voor bestuurskundig probleemoplossen en daarmee onafhankelijk van vakgebied en probleemtype?". De FUNDES-procedure (Functional Design) is de door Mettes en Pilot in hun proefschrift beschreven procedure voor het ontwikkelen van onderwijs in het oplossen van problemen. Deze bestaat uit een zestal stappen: De eerste is de "Probleemanalyse", waarin het probleem (inclusief de context) wordt verkend met inbegrip van een analyse van de voorwaarden voor, alsmede mogelijkheden van een aanpak. In de tweede stap "Cursusontwerp" wordt de opzet van de cursus gemaakt, terwijl in de derde "Cursusconstructie" de vormgeving van deze opzet in een plan van onderwijs- en evaluatiemiddelen plaatsvindt. In de vierde fase wordt de Cursus "uitgevoerd" en formatief geëvalueerd. De vijfde fase is de "summatieve evaluatie" en de laatste de "beslissing" over doorgaan of niet. Deze procedure werd door Mettes en Pilot beschreven voor het ontwikkelen van onderwijs in het probleemoplossen in technische vakken als thermodynamica, maar mogelijk ook voor andere domeinen van probleemoplossen. Het bijzondere van deze procedure is echter niet gelegen in deze zes stappen, want die komen in allerhande procedures voor onderwijsontwikkeling voor. Het gaat vooral om de deelstappen binnen deze hoofdstappen en om het centrale begrip het Gewenst HandelingsVerloop (GHV), dat met name in de ontwerpfasen centraal staat. Afwijkend van andere procedures van onderwijsontwikkeling is ook dat FUNDES zich expliciet en alleen richt op onderwijs in het leren probleemoplossen en wel volgens een handelingstheoretische benadering. Hieronder verstaat Terlouw: "een benadering die bestaat uit het bepalen welke activiteiten men dient te leren voor bepaalde situaties en op welke wijze daarvoor onderwijs vormgegeven kan worden, opdat de oriënteringsmogelijkheden van de studenten in nieuwe situaties vergroot kunnen worden met een verantwoorde leerinspanning". Voor een verdere definiëring van de handelingstheorie verwijst Terlouw naar twee publikaties van Van Parreren uit 1981. De ene ontbreekt echter in de literatuurlijst, de andere is een bijdrage aan het Handboek voor de Onderwijspraktijk(!).

Terlouw is nu nagegaan in welke mate de FUNDES-procedure ook bruikbaar is bij het ontwikkelen van onderwijs op het gebied van het bestuurskundig probleemoplossen, concreet voor cursussen op het gebied van "Stadsvernieuwing" en "EG-landbouwbeleid". Het onderzoek dat hij heeft uitgevoerd noemt hij in navolging van Reigeluth ontwikkelingsonderzoek

(instructional development), dat in tegenstelling tot "instructional design-onderzoek" niet gericht is op het zoeken naar optimale combinaties van onderwijsmethoden en zich niet afspeelt op het niveau van de onderwijsleertheorie, maar gericht is op het verkrijgen van kennis over verschillende ontwikkelprocedures, over optimale combinaties van ontwikkelprocedures en over situaties waarin ontwikkelprocedures optimaal zijn. Het onderzoek richt zich mede op advies van Strike dan ook niet op de "harde kern" van de gehanteerde (handelingstheoretische) onderwijsleertheorie, maar uitsluitend op daaruit voortvloeiende actieplannen en uitkomsten. De onderwijsleertheorie staat niet ter discussie. Getoetst wordt slechts of uit deze harde kern voortvloeiende actieplannen en onderwijsuitkomsten "bruikbaar" zijn. Wanneer blijkt dat dit het geval is, kan het uitgevoerde onderzoeksprogramma "progressief" genoemd worden voorzover in de procedure aangebrachte veranderingen duidelijk verbonden zijn met de theoretisch-normatieve "harde kern" en aan enkele andere eisen is voldaan (eenvoud, consistentie, precisering, uitbreiding van de empirische basis). In zo'n geval, stelt Terlouw, komt de theorie er versterkt uit.

Bruikbaarheid operationaliseert hij in termen van "Houdbaarheid" en "Effectiviteit". Houdbaarheid heeft betrekking op de vraag of de stelling houdbaar is dat de FUNDES-procedure een adequate oriënteringsbasis is voor het ontwerpen van onderwijs in (bestuurskundig) probleemoplossen. De criteria voor houdbaarheid worden ontleend aan de theoretisch-normatieve harde kern, zoals beschreven in het eerder genoemde artikel van Van Parreren uit 1981. Het heeft mij enig zoekwerk gekost om een samenvatting hiervan te vinden. Eerst meende ik te moeten concluderen dat deze in het proefschrift ontbrak. Net toen ik me boos wilde gaan maken, ontdekte ik dat in het hoofdstuk "Operationalisering van het evaluatie-criterium bruikbaarheid" toch een beschrijving van de "harde kern" was opgenomen (p. 62-63). De theoretisch harde kern blijkt te bestaan uit de opvatting dat onderwijs moet bijdragen aan de oriënteringsmogelijkheden van de leerlingen en een zevental voorschriften over de wijze waarop dit moet gebeuren: 1) Onderwijsdoelstellingen moet men formuleren in termen van handelingsstructuren; 2) Elke lerende moet zich kennis eigen maken door actief te handelen; 3) Men moet expliciet kiezen voor cognitieve schema's die als basis kunnen dienen voor bepaalde handelingen; 4) Integratie in het persoonlijk kennisbezit is belangrijk, waarbij aansluiten bij de beginsituatie en zorgdragen voor een positieve leermotivatie cruciaal zijn; 5) Men moet uitgaan van een psychologische analyse van de gewenste eindhandelingsstructuur, waarbij soms uitgegaan kan worden van expertgedrag; 6) Vijf richtlijnen voor de organisatie van het gewenste onderwijsleerproces (tussenfasen opsporen, leerinspanning minimaliseren, foutenanalyse doen, niet-beoogde resultaten in de gaten houden en de stabiliteit van de leermotivatie bevorderen); 7) Een centrale plaats dient te worden ingeruimd voor oriëntering door en in het onderwijs.

Hieruit worden nu criteria voor houdbaarheid en effectiviteit afgeleid: Wat betreft de houdbaarheid wordt nagegaan a) of de FUNDES-toepassing leidt tot een cyclisch verbeteringsproces waarin de ontwikkel(tussen)produkten consistent blijven met de theoretisch-normatieve kern; b) of FUNDES een adequate oriënteringsbasis is voor professionele onderwijsontwikkelaars, uitgesplitst naar de functionaliteit (volledigheid en algemeenheid) van de oriënteringsbasis en praktische hanteerbaarheid ervan. Wat betreft de effectiviteit van de ontwikkelendproducten wordt nagegaan of het ontwikkelde onderwijs zoals bedoeld overeenkomt met gewenste leerresultaten en hoe de relatie is tussen de eindhandelingsstructuur en expert- en beginnersgedrag.

De resultaten van het confronteren van de toepassing van de FUNDES-procedure met de beschreven criteria, afzonderlijk voor vijf van de zes fasen, vormen nu de hoofdmoot van het proefschrift. Over het algemeen bleek de FUNDES-procedure bruikbaar: 69% van de oordelen over de adequaatheid als oriënteringsbasis waren positief. De problemen die werden geconstateerd hadden over het algemeen betrekking op het feit dat de voorschriften te nauw verbonden waren met de natuurkundige vakinhouden en natuurkundige probleemttypen en vooral op een onvolledigheid van de oriënteringsbasis in die zin dat de voorschriften vaak van een te algemeen niveau waren. De meeste problemen deden zich voor in de ontwerpfasen bij de deelhandelingen "het formuleren van een gewenst handelingsverloop" en het "opstellen van

een programma van essentiële kenniselementen". In de fase van de probleemanalyse deden zich vooral problemen voor met de deelhandeling "overleg met docenten en commissies" omdat – zo stelt Terlouw – de innovatiedimensie in de procedure afwezig was. In de constructiefase vormde vooral de opstelling van evaluatieplannen een probleem. Erg positief is Terlouw met name over één van de meest kenmerkende handelingen in de procedure "het afleiden van de uitgebreide versie van de handeling vanuit de onderwijsfuncties".

Over de effectiviteit van de ontwikkelde produkten is Terlouw onomwonden positief: er is een cyclisch verbeteringsproces op gang gezet dat uiteindelijk leidde tot ontwikkelprodukten die zowel intern als extern effectief bleken te zijn.

De FUNDES-procedure behoeft dus nogal wat wijzigingen, uitbreidingen en specificaties voordat hij bruikbaar was bij het ontwikkelen van onderwijs in het bestuurskundig probleemoplossen en was derhalve ten dele niet onafhankelijk van vakgebied en probleemtype. De aangepaste procedure leidde wel tot effectief onderwijs.

Omdat de aangepaste versie van de FUNDES-procedure beter de harde kern van het handelingstheoretische onderzoeksprogramma aangaande onderwijsleerprocessen realiseert dan de oorspronkelijke van Mettes & Pilot en latere versies van het onderwijsplan beter dan eerdere, kan worden geconcludeerd dat het onderzoeksprogramma progressief is en de theoretisch-normatieve harde kern ervan heeft versterkt. In de terug- en vooruitblik formuleert Terlouw, tenslotte, zijn versie van de FUNDES-procedure die in nieuwe gevallen niet klakkeloos kan worden toegepast, maar een "kritische uiteenzetting" vraagt.

Dan volgen nu enkele "jammer-dingen". In de eerste plaats vind ik het jammer dat in het proefschrift zo extreem de nadruk is gelegd op de evaluatie van de ontwikkelprocedure, waar – naar mijn mening – de analyse van en training in vakspecifiek probleemoplossen minstens zo interessant is. Nu komt dit wel in de dissertatie aan bod, maar door de gekozen vormgeving krijgt dit te weinig accent. Dit blijkt bijvoorbeeld ook al uit de gekozen titel. Potentiële lezers die geïnteresseerd zijn in probleemoplossen in complexe terreinen (als het bestuurskundig probleemoplossen) zullen in de titel niet direct aanleiding zien om dit proefschrift te bestuderen. Maar ook in de vele pagina's tekst over de evaluatieresultaten die soms wel een beetje slaapverwekkend zijn raakt de vakinhoudelijke en (cognitief)psychologische analyse van het probleemoplossen wat op de achtergrond.

Jammer vind ik het ook dat Terlouw de theoretisch-normatieve harde kern buiten de discussie stelt. Ten eerste is die kern naar mijn mening niet zo hard als Terlouw beweert. Verschillende van de eerder vermelde handelingstheoretische voorschriften kunnen net zo goed worden afgeleid uit andere (cognitivistische) theorieën (zie Simons, in druk). Ten tweede is ook niet zo erg helder wat hij precies tot de harde kern rekent en wat niet. Eerder wees ik al op het feit dat hij daarvoor zijn toevlucht nam tot een populair-wetenschappelijk betoog uit het Handboek voor de Onderwijspraktijk. Overigens blijkt op pagina 429 dat ook Terlouw zich hierover wel een beetje zorgen heeft gemaakt, waar hij aan het eind van zijn boek, volgende onderzoekers in één van zijn aanbevelingen aanraadt om te systematiseren welke elementen tot de theoretisch-normatieve harde kern behoren of onder actieplannen vallen. Ten derde houdt de theoretisch-normatieve harde kern strikt genomen niet veel meer in dan de opvatting dat onderwijs de leerlingen optimale oriënteringsmogelijkheden moet bieden en een aantal zeer algemene voorschriften (zoals aansluiten bij de beginsituatie) waar niemand het mee oneens zal zijn plus enkele vage specifieke voorschriften. Verder blijken de grootste problemen opgetreden te zijn in het opstellen van een gewenst handelingsverloop: de deelhandeling die het dichtst staat bij wat dan ook de theoretische kern moge zijn. Jammer is tenslotte ook dat Terlouw meent te moeten concluderen dat de theoretisch-normatieve kern versterkt is door de behaalde resultaten. Deze conclusie, die gebaseerd wordt op de redenering dat via een ingrijpend gewijzigde procedure effectief onderwijs is ontworpen dat niet in strijd lijkt te zijn met een vage theoretische kern, gaat mij echt te ver.

Ondanks deze bezwaren, is mijn eindoordeel over dit proefschrift positief. Onderwijsontwikkelaars vinden een aangepaste procedure om vakspecifiek onderwijs in het leren probleemoplossen te ontwerpen en een methodologie om het ontwikkelproces te evalueren. Onderzoekers

die geïnteresseerd zijn in probleemoplossen in complexe domeinen vinden interessante analyses en trainingsmethoden die nog blijken te werken ook. Bestuurskundigen tenslotte vinden een manier om hun studenten te leren om zich adequaat te oriënteren bij het ontwerpen van beleid.

REFERENTIE

Simons, P. R. J. (in druk). Handelingspsychologie versus cognitivistische psychologie. *Pedagogische Studiën*.

P. RobertJan Simons
Katholieke Universiteit Brabant
Sectie Onderwijspsychologie en Bedrijfsopleidingen
Hogeschoollaan 225
5037 GC Tilburg

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 OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 12, nr. 1

Februari 1988

Inhoudsopgave

Outputbepaling onderwijswetenschappen op het juiste spoor?
J.F.M. Claessen

60

Proefschriften

62

Outputbepaling onderwijswetenschappen op juiste spoor?

J. F. M. Claessen, Open Universiteit

Aanleiding

Na jarenlange discussies, vele symposia (o.a. laatste ORD-dagen) en nota's over kwaliteitsbepaling is inmiddels het tijdstip aangebroken dat faculteiten, vakgroepen en onderzoeksinstituten ook daadwerkelijk worden geconfronteerd met kwaliteitsbeoordeling van hun onderzoeksoutput. De voorstellen van de visitatiecommissies voor 'Pedagogiek en Onderwijskunde' en 'Sociologie, Politicologie en Bestuurskunde' waren een eerste proeve van wat in de toekomst regel zal worden. HOOP en het visitatieprotocol van de VSNU laten hier geen twijfel over bestaan.

Het lijkt tijd voor een tussenbalans, mede naar aanleiding van discussies binnen het Landelijk Overleg, vakgroepen Onderwijskunde, de Nederlandse Vereniging van Opvoedkundigen en de VOR. Oogmerk van deze tussenbalans is geenszins de noodzaak van outputbeoordeling in twijfel te trekken. Integendeel, voorop staat dat het om een veelheid van overwegingen legitiem en noodzakelijk is om van tijd tot tijd de maatlat te hanteren ten einde kaf van koren te scheiden. De discussie dient zich veel meer te richten op de keuze van 'meetmiddelen' en eventuele neveneffecten van een wellicht overtrokken outputbepalingscultuur.

Discussiepunten

Beide visitatiecommissies stonden voor de opgave een meetinstrumentarium te ontwikkelen voor het kwantificeren van de diversiteit aan onderzoeksoutput. Het siert beide commissies dat ze gekozen hebben voor een enigermate brede benadering. De visitatiecommissie 'Sociologie, Politicologie en Bestuurskunde' formuleert het als volgt (p. 23): 'Twee wegen stonden de commissie open: beperking van de bepaling van het kwaliteitsoordeel tot publikaties in internationale tijdschriften of bepaling van een kwaliteitsoordeel op basis van een reeks indicatoren, die op output, maar ook op werfkracht, kwaliteit van het kader en mate van erkenning via citaties zijn gebaseerd. De commissie heeft voor dit laatste gekozen'. Ook de visitatiecommissie 'Pedagogiek en Onderwijskunde' heeft voor een brede optiek gekozen. Er worden verschillende wegingsschema's gehanteerd die enerzijds recht doen aan lokale omstandigheden en anderzijds aan verschillen in wetenschappelijke status. 'Voor de toepassingswaarde kunnen echter ook rapporten in het Nederlands van belang worden geacht' (p. 15). De commissie voegt daar aan toe dat de eerste indruk is dat er relatief weinig in buitenlandse tijdschriften of boeken wordt gepubliceerd en relatief veel in rapportages met een onduidelijke status dan wel in een restcategorie.

Een eerste vraag die naar boven komt bij lezing van de rapporten van beide commissies (in het bijzonder 'Pedagogiek en Onderwijskunde') is de mate van realiteit van de verwachting van veel buitenlandse publikaties. Het is een ervaringsgegeven dat het verre van gemakkelijk is entree te krijgen in bijvoorbeeld Engelstalige tijdschriften. Het is niet uitgesloten dat dit voor een deel komt door onroereikende taalbeheersing of door onvoldoende gerichtheid van de onderzoekers op deze media. Het is tegelijkertijd een gegeven dat veel onderzoeksbijdragen specifiek gebonden zijn aan de context van het Nederlandse onderwijs. Nogal wat onderwerpen zijn internationaal nauwelijks overdraagbaar. Het zou een illusie zijn te verwachten dat internationale redacties onder deze omstandigheden veel ruimte zouden geven. Engelstalige publikaties als het hoogste goed aanmerken, lijkt dan ook niet altijd realistisch.

Een tweede aspekt is het toepassingskarakter van de onderwijswetenschappen. Publikaties kunnen een hoge waarde hebben voor praktijk en beleid zonder dat er nu meteen sprake is van excellent wetenschappelijk onderzoek dat tot de internationale verbeelding spreekt. Het gaat dan om de vraag hoe krijgen we deze toepassingswaarde boven tafel. Is hier niet evenzeer een meetinstrumentarium voor nodig? In het verlengde van deze toepassingwaarde ligt het ontwikkelingswerk dat voor een deel ook door de onderwijswetenschappen wordt verricht: bijdragen om curriculumontwikkeling, diagnostisch onderzoek, instrumentontwikkeling e.d. Met gebruikmaking van wetenschappelijke tools wordt in deze sektor waardevolle kennis geproduceerd die in de ratings van de visitatiecommissies nauwelijks wordt verdisconteerd. Hoe kunnen we ervoor zorgdragen dat ook deze wetenschappelijke output recht wordt gedaan?

Een derde aspekt betreft de randvoorwaarden waaronder veel onderwijsonderzoek in Nederland moet worden verricht. Een belangrijk deel van onderwijsonderzoek is contract-research. De ervaring wijst uit dat nogal wat opdrachtgevers vooral geïnteresseerd zijn in kortstondige projecten die snel resultaat moeten opleveren zonder veel wetenschappelijke diepgang. Men wil snel een eindrapport boven tafel hebben en is ook niet genegen middelen ter beschikking te stellen voor publikaties in (internationale) tijdschriften of bijdragen aan conferenties. De financiële marges voor de para-universitaire onderzoeksinstituten zijn daarbij in de afgelopen jaren zodanig versmald dat uit deze hoek geen soulaas meer mag worden verwacht. Nogal wat onderwijsonderzoek dreigt zo, lettend op de criteria van de visitatiecommissies, in een neerwaartse spiraal te belanden. Zijn er mogelijkheden om uit deze impasse te komen?

Tot slot willen we nog een tweetal risico's signaleren die zouden kunnen voortvloeien uit een overtrokken outputbepalingscultuur.

Het getalscriterium dreigt een magische betekenis te gaan aannemen. De publikatiedruk is zo hoog geworden dat eenzelfde publikatie op legio wijzen wordt uitgebracht: een reeks van interimrapporten, een eindrapport, een congresvoorstel, een congrespaper, artikelen in tijdschriften, bijdragen aan readers etc. De kwantiteit neemt zo indrukwekkende afmetingen aan, maar betekent dit ook een evenredige stijging van de kwaliteit?

Een ander niet denkbeeldig gevaar is een wending in het soort onderzoek dat men nog wil entameren. Om snelle publikaties te verzekeren, wordt gekozen voor onderzoek met een beperkte (detail)vraagstelling, onderzoek dat eenvoudig te beïnstrumenteren valt, onderzoek dat snel data oplevert, data die zich lenen voor geavanceerde statistische analyses. Het risico van een verloren publikatiejaar wordt zo geminimaliseerd, maar tegelijkertijd blijven we ons zo bewegen op gebaande wegen.

Proefschriften

Ton Mooij: Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations.

's-Gravenhage: Instituut voor Onderzoek van het Onderwijs (1987).

Promotor: prof. dr. M. Boekaerts; co-referent dr. J. H. L. van den Bercken.

Promotie: Katholieke Universiteit Nijmegen, 11 september 1987.

Korte samenvatting

Het proefschrift is de neerslag van een studie, die op verzoek van de Algemene Bond van Onderwijzend Personeel gesubsidieerd werd door het Instituut voor Onderzoek van het Onderwijs (SVO). De aanleiding tot de studie was het verschijnsel van prestatie- en motivatieproblemen van leerlingen in het voortgezet onderwijs. Gevraagd werd naar de oorzaken en reductiemogelijkheden van de problematiek.

In *theoretisch opzicht* wordt een interactionele multi-niveau theorie ontwikkeld. 'Interactioneel' wil zeggen een concentratie op de longitudinale wederzijdse beïnvloeding van (kenmerken van) een leerling en (kenmerken van) zijn of haar onderwijsomgeving. Deze omgeving kan gezien worden als een multi-niveau systeem. De interactie tussen leerling en onderwijsomgeving wordt dan geconceptualiseerd in termen van verschillende in tijd en ruimte gesitueerde processen. In aansluiting hieraan wordt het *doel van de studie* geformuleerd als: het exploreren van theoretische aspecten van een interactionele multi-niveau benadering van gedrag, prestatie, bekwaamheid en gerichtheid van leerlingen, waarbij drie niveaus worden onderscheiden: een leerlingniveau, een groepsniveau, en een schoolniveau; en het onderzoeken van relevante processen en effecten in de onderwijspraktijk.

Bij de uitwerking van de interactionele multi-niveau theorie wordt de aandacht gericht op verschillende, ten dele complementaire, overlappende of met elkaar rivaliserende processen. Het gaat hierbij om leerling-onderwijsomgeving-processen *tijdens de les* (leerlingniveau en lesgroepniveau), *tijdens een schooljaar* (op het leerlingniveau en tussen het leerlingniveau en het jaargroepniveau), en *tijdens een schooljaar en schoolloopbaan* (op leerlingniveau en tussen het leerlingniveau en het schoolniveau).

Empirisch worden het lbo, mavo en ibo in het onderzoek betrokken. Er worden twee kwantitatieve longitudinale onderzoeken verricht. In het *eerste onderzoek* wordt, in een nonmanipulatieve vergelijkende studie onder 860 leerlingen uit 39 klassen in 10 scholen, getracht een extreme-groups design te realiseren. De relevante variabelen worden zoveel mogelijk op hun 'eigen' niveau gemeten. Bij de meting van de variabelen worden gebruikt: schoolvorderingentests en vragenlijsten in verschillende schoolvakken voor leerlingen, gestructureerde observatie (van leskenmerken in verschillende schoolvakken, en van de onderwijssamenwerking per school), en vragenlijsten voor leerkrachten. Er worden verschillende soorten analyse-technieken toegepast (waaronder multiple regressie-analyse in het kader van de tweestaps procedure). Ten behoeve van een onderlinge vergelijking worden de analyses op drie wijzen uitgevoerd: ongecorrigeerd, gecorrigeerd voor jaargroep, en gecorrigeerd voor jaargroep en school.

Het *tweede onderzoek* is een manipulatief of interventie-onderzoek in een lbo-ibo-school. Hier wordt gepoogd met behulp van door leerkrachten uitgevoerde curriculumontwikkeling bepaalde verbeteringen in kenmerken van onderwijsleersituaties en de school aan te brengen. Volgens de interactionele multi-niveau theorie zouden deze moeten leiden tot bepaalde wijzigingen in de ontwikkeling van leerlingkenmerken.

De *resultaten en conclusies* in beide onderzoeken verlenen steun aan de adequaatheid en meerwaarde van theorie-vorming vanuit de interactionele multi-niveau optiek. De gehanteerde multi-niveau methodologie is essentieel in onderzoek naar schooleffecten (ter vermijding van contaminatie van variabelen van verschillende niveaus in theorie en onderzoek). Met

betrekking tot de onderwijspraktijk wordt geconcludeerd dat (a) de bepaling van vorderingen in de prestatie(s) van een leerling gebaseerd zou moeten zijn op de hoeveelheid beheerste leerstof en dat het gebruik van cijfers betrekking zou moeten hebben op door de leerling zelf te beïnvloeden factoren (zoals inspanning of inzet); (b) leerkrachten zèèr sterk ondersteund zouden moeten worden bij het ontwikkelen van juiste curricula; en (c) het bevorderen van het management van de leerkacht het inhoudelijk taakgericht gedrag van leerlingen zou kunnen vergroten.

Frank van Overwalle: Causes of success and failure of freshmen at university: an attributional approach.

Promotor: prof. dr. H. Rigaux.

Promotie: Vrije Universiteit Brussel, België, 12 juni 1987.

Korte samenvatting

De overgang van het middelbaar onderwijs naar de universiteit verloopt voor vele studenten moeizaam. Grote hoeveelheid leerstof, minder intensieve begeleiding, examens geconcentreerd op het einde van het jaar: dit zijn een aantal factoren die heel wat aanpassing vergen van de eerstejaarsstudenten. Wanneer de cijfers van de eerste tests in het midden van het jaar dan ook nog tegenvallen, laten velen de moed zakken. Dit doctoraat gaat over de problemen die studenten in hun studie ervaren. Tevens wordt een begeleidingsprogramma besproken waarin gepoogd werd de slagingskansen te verhogen. De resultaten van dit programma tonen aan dat studiemoeilijkheden niet onomkeerbaar zijn. Veel probleemstudenten konden door dit programma hun resultaten verbeteren en behaalden in de eerste zittijd gemiddeld 1,3 punten op 20 meer.

Het begeleidingsprogramma beperkt zich niet, zoals gewoonlijk, tot het aanleren van specifieke studietechnieken. In feite nemen die een kleine plaats in. Belangrijker is de psychologische ingesteldheid van de studenten ten aanzien van hun examenprestaties, met name hoe zij hun lage cijfers interpreteren. Dit denkproces wordt beschreven in de *attributietheorie*: de oorzaak die studenten zelf toeschrijven of attribueren aan hun falen bepaalt in sterke mate hun verdere studiegedrag. Bernard Weiner en zijn collega's hebben een vooraanstaande rol gespeeld in de ontwikkeling van de attributietheorie, die model stond voor dit onderzoek naar studiemotivatie en -prestaties.

Eenvoudig geformuleerd, gaat Weiners attributietheorie ervan uit dat studenten die hun slechte cijfers toeschrijven aan *onvoldoende bekwaamheid*, sneller geneigd zullen zijn om de moed te laten zakken en niet meer zullen hopen op betere resultaten. Dit heeft als negatief gevolg dat hun inzet en studierendement daalt. Studenten die daarentegen ervan overtuigd zijn dat hun lage cijfers te wijten zijn aan *onvoldoende of inefficiënte studie*, blijven de hoop koesteren dat zij door meer en juistere inspanning toch nog kunnen slagen. En dit heeft dan ook een gunstige weerslag op hun studieprestaties.

In een drietal studies werden vooraf de theoretische grondslagen van het attributiemodel getoetst in de concrete praktijk aan de universiteit. Uit de resultaten kwamen de volgende punten aan het licht.

(1) Oorzaken van slagen en mislukken die door studenten dikwijls worden vermeld hebben inderdaad een merkbare invloed op examenprestaties in het eerste jaar aan de universiteit, en hebben vooral te maken met factoren die rechtstreeks in verband staan met de studie en de studie-omgeving.

(2) Oorzaken die studenten toeschrijven aan slagen of falen vallen uiteen in drie onderliggende basisdimensies: *locus* (oorzaken binnen of buiten de persoon), *stabiliteit* (permanente of tijdelijke factoren) en *beheersbaarheid* (oorzaken die al dan niet beïnvloedbaar zijn door de persoon of door anderen).

(3) Deze drie causale dimensies hebben een onrechtstreeks, maar aanzienlijk effect op

examenprestaties via studieverwachtingen en gevoelens van tevredenheid en trots (positief zelfbeeld) over vorige resultaten tijdens het jaar.

De resultaten van de drie veldonderzoeken bevestigden dus in grote trekken het attributiemodel voor het concrete studiegedrag aan de universiteit. Op basis van deze bevindingen werd dan een interventieprogramma uitgewerkt. Het principe is als volgt: men poogt de student af te leren om studiemoeilijkheden voorbarig toe te schrijven aan *onvoldoende intelligentie* (een permanent en onbeïnvloedbaar tekort) en tracht hen duidelijk te maken dat er nog andere oorzaken mogelijk zijn: *onvoldoende studie-inspanning* en *ondoeltreffende studiemethode* (tijdelijke en beïnvloedbare factoren). Wanneer een student dit inziet, kan hij zelf op korte termijn deze tekorten verhelpen en komt dit zijn studierendement ten goede.

Uniek aan dit programma is dat het dus gebaseerd is op de subjectieve ervaringen van de studenten, en veel minder op de objectieve principes van de begeleider. Ondermeer worden video's getoond waar ouderejaars vertellen over hun studieproblemen in de eerste kandidatuur en hoe ze die hebben opgelost. Daardoor wint de boodschap aan realiteit en overtuigingskracht. Zoals gezegd, kan door deze training van ongeveer 30 minuten een gemiddelde stijging van iets meer dan 1 op 20 punten bereikt worden over de totale eerste zittijd. Dit experiment werd uitgevoerd bij studenten uit de richtingen economie en handelsingenieur die in januari voor het hoofdvak economie minder dan 9 behaalden.

Dit doctoraat wijst dus op het grote belang van subjectief beleefde oorzaken van slagen en falen voor de motivatie en de prestaties van studenten. Tevens wordt aangetoond dat oorzakelijke attributies in positieve richting kunnen omgebogen worden en daardoor aanleiding geven tot hogere resultaten op de examens.

H. Blok: Feiten en meningen over het taalgebruik van LBO- en MAVO-leerlingen in alledaagse situaties (Stichting voor Onderzoek van het Onderwijs, Selecta reeks, 1987).

Promotores: prof. dr. A. D. de Groot en prof. dr. H. Wesdorp.

Promotie: Universiteit van Amsterdam, 19 juni 1987 (cum laude).

Korte samenvatting

Het onderzoek gaat over doelstellingen voor het vak Nederlands op scholen voor LBO en MAVO. De centraal gestelde vragen zijn: welke doelstellingen weet het moedertaalonderwijs bij leerlingen te realiseren en hoe worden de bereikte leerresultaten in het algemeen gewaardeerd: vallen ze in verhouding tot de heersende verwachtingen mee of juist tegen? De aandacht is met name gericht op taalvaardigheden die van groot praktisch belang zijn in het naschoolse leven. Dergelijke vaardigheden worden tegenwoordig wel aangeduid als functionele taalvaardigheden. Ten behoeve van een inventarisatie van het naschoolse taalgebruik zijn bij jongeren die hun schoolloopbaan inmiddels al enkele jaren hebben afgerond, gegevens verzameld over de typen taalgebruikssituaties waarmee ze geconfronteerd worden en over de problemen die ze daarbij ervaren. Enkele alledaagse taalgebruikssituaties zijn geoperationaliseerd in zogenoemde taalproeven, waarmee bij individuele leerlingen gepeild kan worden hoe doelmatig hun taalgebruik in zulke situaties is. De taalproeven zijn afgenomen bij twee steekproeven van leerlingen aan het einde van een LBO- of MAVO-opleiding. De taalproeven zijn tevens ter beoordeling voorgelegd aan een aantal personen die op het terrein van het moedertaalonderwijs deskundig of belangstellend zijn. Zonder de prestaties van de leerlingen op de taalproeven te kennen, hebben zij hun verwachtingen uitgesproken over de door de leerlingen behaalde scores. Eindresultaat van het onderzoek is een aantal conclusies over de feitelijke doelmatigheid van het taalgebruik van LBO- en MAVO-leerlingen in relatie tot de verwachtingen die daaromtrent bestaan. Deze conclusies zijn van bijzonder belang voor de discussie over een gemeenschappelijke basisvorming voor leerlingen tot 15 à 16 jaar.

Het Meten van Affectieve Doelen: De Validering en Normering van de Belevingsschaal voor Wiskunde (BSW)

Mark J. Martinot, Hans B. Kuhlemeier en Hans J. M. Feenstra¹
Instituut voor Toetsontwikkeling (Cito²) Arnhem

ABSTRACT

This article reports on an empirical study of reliability, internal structure and validity of the Attitude Scale towards Mathematics (BSW). The BSW gives the teacher a class score (mean of pupil results) on four subscales: 'pleasure', 'fear/difficulty', 'interest/devotion', 'profit/relevance'. National norms are reported that have been gathered with this instrument. The findings point towards a satisfying reliability and validity; furthermore, the presented norms do not indicate an extremely negative attitude towards mathematics in the first years of secondary education. The article ends with some conclusions and recommendations.

INLEIDING

In sociaal-wetenschappelijk onderzoek is de toepassing van instrumenten op het affectieve domein reeds lange tijd gebruikelijk. Een overzicht van affectieve instrumenten voor het vakgebied wiskunde wordt gegeven door Korf et al. (1986). In de lespraktijk van het voortgezet onderwijs worden dergelijke instrumenten nog nauwelijks gebruikt.

Het Cito-project Schoolvakbelevingsinstrumenten werkt sinds 1985 aan de constructie en uitgave van de 'Belevingsschaal voor wiskunde' (BSW), een gemodificeerde versie van de destijds door Kremers (1978, 1979, 1981) ontwikkelde wiskunde-attitudeschaal. De schaal zal wiskundeleraars in staat stellen een belangrijk aspect van het affectieve domein te meten, namelijk de mate waarin leerlingen hun wiskunde-onderwijs als plezierig, moeilijk, interessant en nuttig beleven. De aanleiding tot de constructie van een schaal voor wiskunde was mede gelegen in de problemen die zich op dit vakgebied voordoen. Zo zouden veel leerlingen bang zijn voor wiskunde, het vak moeilijk en vervelend vinden, het nut ervan niet inzien, zich er niet voor willen inzetten, etc. (Kremers, 1978). Het is dan ook niet verwonderlijk dat het nastreven van een positieve beleving van het wiskunde-onderwijs in de ogen van veel docenten een belangrijke onderwijsdoelstelling vertegenwoordigt (Kuhlemeier & Martinot, 1987).

Ter voorbereiding op de uitgave van de BSW is de psychometrische kwaliteit van de schaal nader onderzocht en zijn meningen over en ervaringen met de schaal in het onderwijs systematisch geïnventariseerd (Martinot, 1986; Kuhlemeier en Martinot, 1987). Ten behoeve van de interpretatie van de resultaten die met behulp van de BSW worden verkregen, is een normeringsonderzoek uitgevoerd (Martinot & Kuhlemeier, in voorbereiding). In laatstgenoemd onderzoek zijn normen opgesteld voor de eerste drie leerjaren van het voortgezet onderwijs; tevens is het onderzoek naar de psychometrische kwaliteit van de BSW op grotere schaal herhaald. In het onderhavige artikel worden de belangrijkste resultaten van dit normerings- en valideringsonderzoek besproken. Eerst geven we een beknopte beschrijving van de BSW. Daarna gaan we in op de methode van onderzoek (steekproef, respons, instrumentatie).

¹ Met dank aan M. Zwarts, W. van Roosmalen, M. Ouborg

² Adres: Postbus 1034, 6801 MG Arnhem

Vervolgens rapporteren we de belangrijkste onderzoeksresultaten (interne structuur, betrouwbaarheid, validiteit, normen, relatie met externe variabelen). Tenslotte volgen enkele conclusies en aanbevelingen.

DE BELEVINGSSCHAAL VOOR WISKUNDE

De BSW omvat vier subschalen die ieder acht uitspraken over het vak wiskunde bevatten. De subschalen hebben betrekking op de volgende aspecten van de wiskundebeleving.

1 Plezier. Deze schaal betreft het plezier dat leerlingen beleven aan het ' bezig zijn met wiskunde'. De leerlingen geven aan of het vak leuk gevonden wordt, de lestijd snel voorbij gaat, het volgen van wiskunde tot de eigen keuze zou behoren, etc. Een uitspraak uit deze subschaal is bijvoorbeeld: 'Ik vind wiskunde een leuk vak.'

2 Angst en moeilijkheid. Deze schaal heeft betrekking op de ervaren moeilijkheid van wiskunde en daarmee samenhangende angstgevoelens. Gevraagd wordt naar de mate waarin leerlingen zenuwachtig zijn tijdens de wiskundeles, de mate waarin zij bang zijn om fouten te maken, en naar de begrijpelijkheid van de stof. Een voorbeeld van een uitspraak uit deze subschaal luidt: 'Bij wiskunde ben ik banger om fouten te maken dan bij andere vakken.'

3 Inzet en interesse. Hier gaat het erom of de leerlingen zich aangesproken voelen door het vak wiskunde en bereid zijn om er tijd aan te besteden. De uitspraken van deze schaal gaan bijvoorbeeld over de mate waarin de lessen interessant gevonden worden en over de hoeveelheid vrije tijd die aan onderdelen uit de wiskundelessen wordt besteed. Zo wordt aan de leerlingen onder andere de uitspraak voorgelegd: 'Ik maak weleens meer huiswerk voor wiskunde dan we opgekregen hebben.'

4 Nut en relevantie. Deze laatste schaal stelt het praktisch nut van wiskunde centraal. Gevraagd wordt naar de bruikbaarheid van de wiskunde binnen en buiten de school, nu en later. Een van de uitspraken luidt bijvoorbeeld: 'Buiten de school heb je weinig aan wat je in de wiskundelessen leert.'

Voor de bovengenoemde aspecten is gekozen, omdat deze in de beleving van het wiskunde-onderwijs een belangrijke plaats innemen (vgl. Kremers, 1981). Dergelijke belevingsaspecten blijken overigens ook in soortgelijke instrumenten als de BSW geoperationaliseerd te zijn (Crombach, Voeten en Feenstra, 1986; Raat en De Vries, 1986). De uiteindelijke indeling van de BSW-items in subschalen is gebaseerd op de resultaten van het vooronderzoek met een conceptversie van de BSW.

De items van de BSW zijn van het Likert-type. De leerlingen geven op een vijfpuntsschaal aan in hoeverre zij het met elke uitspraak eens zijn. De antwoordalternatieven variëren van 'helemaal mee eens' tot 'helemaal niet mee eens'. Om antwoordtendenties tegen te gaan, bevat iedere subschaal evenveel negatief als positief geformuleerde uitspraken. Voor het verkrijgen van objectieve informatie is het nodig dat de leerlingen de vragen zo eerlijk mogelijk beantwoorden. Om dit te bereiken wordt de BSW anoniem, dus zonder vermelding van persoonsgegevens, door de leerlingen ingevuld. Aan de hand van de antwoorden kan voor iedere leerling per subschaal een score worden berekend. Elke score kan variëren van 0 tot 32, en is altijd hoger naarmate de beleving van het vak wiskunde positiever is.

De leerlingen doen in de wiskundelessen vele gemeenschappelijke ervaringen op. Als gevolg daarvan bestaat er een zekere overeenstemming in de wijze waarop de verschillende leerlingen van een klas het vak wiskunde beleven. Het is het belangrijkste doel van de BSW om deze beleving van de klas als geheel in kaart te brengen. De uitslag van de BSW omvat daarom voor elke subschaal een aparte klassescore die het gemiddelde vormt van de leerlingsscores. Verondersteld wordt dat het onderscheid tussen de verschillende aspecten van de beleving van het vak wiskunde door de resultaten van het onderzoek zowel op leerling- als op klasniveau wordt gerechtvaardigd. Om dit na te gaan is de interne structuur van de BSW zowel op leerling- als op klasniveau nader onderzocht.

METHODE VAN ONDERZOEK

Steekproef en respons

In het onderzoek zijn de subschalen van de BSW voor elk van de eerste drie leerjaren van het voortgezet onderwijs apart genormeerd. Normering kan in het algemeen plaatsvinden op het niveau van de leerling of op dat van de klas. Aangezien de uitslag van de BSW uit klasscores bestaat heeft de normering van de BSW op klasniveau plaatsgevonden.

De omvang van de steekproef is bepaald volgens de vuistregel van Angoff (1984). Deze stelt dat de som van de samplingfout en de meetfout van het instrument maximaal één procent groter mag zijn dan de meetfout van het instrument alleen. Conform deze vuistregel is de wenselijke omvang per normgroep geraamd op minimaal 80-100 klassen (vgl. Martinot & Kuhlemeier, in voorbereiding).

Ten behoeve van de steekproeftrekking is het Cito-adressenbestand van scholen voor voortgezet onderwijs ingedeeld naar opleidingstype (LBO, MAVO, HAVO en VWO). Scholengemeenschappen konden, al naar gelang het aantal binnen de school vertegenwoordigde opleidingstypen, in twee of meer van de vier deelbestanden voorkomen. Middels a-selecte trekking van opleidingen uit dit steekproefkader is voor elk leerjaar een aparte steekproef van opleidingen samengesteld.

De scholen die de opleidingen uit deze steekproeven verzorgden, zijn schriftelijk benaderd met een verzoek om zich in te schrijven voor het onderzoek. Tijdens deze werving is per opleiding een overzicht verkregen van alle klassen uit het desbetreffende leerjaar. Uit elk overzicht is op toevalsbasis één klas geselecteerd voor deelname aan het onderzoek. Aan de scholen die zich voor het onderzoek hebben ingeschreven, is begin november 1986 het onderzoeksmateriaal voor docenten en leerlingen toegestuurd. In ruim 95 procent van de gevallen is dit materiaal ingevuld en aan het Cito geretourneerd. Er is dus slechts in zeer geringe mate van uitval sprake geweest. Een overzicht van de werving, inschrijving en uiteindelijke deelname per normgroep is weergegeven in Tabel 1.

Tabel 1. Overzicht werving, inschrijving en deelname³

	normgroep leerjaar 1	normgroep leerjaar 2	normgroep leerjaar 3
Aantal aangeschreven opleidingen	225	230	340
Aantal inschrijvingen	102	100	152
Aantal deelnemende klassen	96	96	143

³ Om na te kunnen gaan in welke mate klassen van het HAVO/VWO, MAVO en LBO in het derde leerjaar van elkaar verschillen, is de gewenste omvang van de normgroep voor het derde leerjaar met circa 70 klassen verhoogd.

Onderzoeksinstrumentarium

Het onderzoeksinstrumentarium omvat behalve de BSW een vragenlijst voor de leerling en een vragenlijst voor de docent.

In de *vragenlijst voor de leerlingen* is de leerlingen gevraagd enkele achtergrondgegevens met betrekking tot de schoolvakken wiskunde, Nederlands en Engels te verstrekken. Gevraagd is naar: a) de waardering voor de vakken wiskunde, Nederlands en Engels, telkens uitgedrukt in een cijfer van 1 tot 10, b) de tijd die de leerling gemiddeld besteedt aan huiswerk voor een wiskundeles, een les Nederlands en een les Engels en c) het rapportcijfer dat de leerling voor elk van de drie genoemde vakken op het komende (kerst) rapport denkt te krijgen.

Deze vragen zijn de leerlingen gesteld met het oog op het verkrijgen van meer inzicht in de validiteit van de BSW. Zouden de vier subschalen wat betreft hun correlatie met deze variabelen op substantiële en interpreteerbare wijze van elkaar verschillen, dan vormt dit een ondersteu-

ning voor het maken van een onderscheid tussen de verschillende aspecten van de wiskundebeleving zoals die met de BSW worden gemeten. Voorafgaand aan het onderzoek zijn ten aanzien van het correlatiepatroon de volgende verwachtingen geformuleerd:

- 1 In het cijfer waarmee de leerlingen hun algemene waardering voor het vak wiskunde uitdrukken kunnen alle aspecten van de beleving van het vak wiskunde tot uiting komen. Het cijfer zal echter vooral betrekking hebben op het algemene plezier van de leerlingen in wiskunde. Verwacht wordt dan ook dat dit cijfer met elk van de subschalen van de BSW, maar vooral met de schaal 'plezier', sterk samenhangt.
- 2 De hoeveelheid tijd die aan het huiswerk voor wiskunde wordt besteed kan worden opgevat als een indicator voor de inzet en de interesse van de leerlingen voor het vak. De verwachting is dan ook dat dit gegeven sterker correleert met de subschaal 'inzet/interesse' dan met elk van de overige subschalen. Overigens kan tevens worden verondersteld dat leerlingen voor wie het vak relatief moeilijk is, relatief veel tijd besteden aan het huiswerk. Een negatieve correlatie met de subschaal 'angst/moeilijkheid' zou deze veronderstelling ondersteunen.
- 3 Verwacht wordt dat het rapportcijfer voor het vak wiskunde hoger correleert met de schaal 'angst/moeilijkheid' dan met de overige subschalen. Deze verwachting is gebaseerd op het vermoeden dat laag presterende leerlingen het vak als moeilijk ervaren en in verband hiermee in hogere mate angst voor dit vakgebied zullen beleven.
- 4 Tot slot dienen de gevonden correlaties specifiek te zijn voor het vak wiskunde. Daarom wordt niet alleen verwacht dat de klassescores verband houden met de 'wiskundevariabelen', maar tevens dat dit verband sterker is dan eventuele correlaties met dezelfde gegevens voor de vakken Nederlands en Engels.

Met behulp van de *vragenlijst voor de docenten* zijn gegevens verzameld over de condities waaronder de instrumentafname heeft plaatsgevonden. Voorts is aan de docent gevraagd om de eigen beleving van het lesgeven en de verwachtingen ten aanzien van elk van de klassescores op een vijfpuntsschaal aan te geven.

ONDERZOEKSRESULTATEN

Eerst wordt ingegaan op de interne structuur, betrouwbaarheid en validiteit van de BSW. Vervolgens wordt aandacht besteed aan de normgegevens en de samenhang met leerjaar, opleidingstype en enkele docentvariabelen.

De indeling in subschalen (interne structuur)

Met behulp van factoranalyse is getracht na te gaan in hoeverre de indeling van de BSW-items in de vier subschalen op psychometrische gronden te rechtvaardigen valt.

Eerst is zowel met de leerling als de klas als analyse-eenheid op de itemscores een principale componentenanalyse uitgevoerd (zie tabel 2).

Tabel 2. Het percentage variantie dat door de eerste vijf factoren van een principale componentenanalyse wordt verklaard (A = leerlingniveau; B = klasniveau).

factor	A	B
1	27.3	43.3
2	8.9	11.3
3	6.0	7.2
4	4.1	4.5
5	3.4	3.2

Tabel 2 laat zien dat de eerste factor een groot deel van de totale variantie verklaart. Aan de items van de BSW ligt dus één belangrijke, gemeenschappelijke component ten grondslag. Inspectie van de factormatrix (zie bijlage 1) maakt duidelijk dat het vooral de 8 items van de subschaal 'plezier' zijn die het meest extreem en zuiver op deze eerste factor laden: de 'plezier-items' zijn in hoge mate representatief voor deze algemene factor. Voor de leerlingen lijkt de beleving van het vak wiskunde met andere woorden te kunnen worden samengevat in termen van 'leuk' of 'niet leuk'.

Toch is besloten om niet alleen de items van de subschaal 'plezier' in de uitsprakenlijst van de BSW op te nemen. De overige subschalen maken deel uit van de BSW om, naast het algemene plezier in wiskunde, meer specifieke aspecten van de beleving van het vak wiskunde aan de orde te kunnen stellen, waardoor de gebruiksmogelijkheden van het instrument toenemen (vgl. Kuhlemeier en Martinot, 1987; Martinot, 1986). Deze meer specifieke aspecten hebben overigens alle met het plezier in wiskunde te maken: de factormatrix in bijlage 1 laat zien dat praktisch elk item van de BSW substantieel laadt op de algemene factor. Indien een leerling dus veel plezier heeft in wiskunde zullen niet alleen de plezieritems, maar ook de meeste overige items min of meer positief worden beoordeeld.

Om na te gaan of het onderscheid dat tussen de drie overige, meer specifieke subschalen wordt gemaakt gerechtvaardigd is, is op de items uit deze subschalen een aparte factoranalyse uitgevoerd. De resultaten zijn weergegeven in bijlage 2. Uit de aldaar weergegeven 3-factoroplossing blijkt dat zowel op leerling- als op klasniveau de items van de drie subschalen ieder op een *afzonderlijke* factor (hoog) laden. Iedere schaal neemt met andere woorden een eigen factor voor zijn rekening. Deze bevinding vormt een belangrijke ondersteuning voor de indeling van de items in de schalen 'angst/moeilijkheid', 'inzet/interesse' en 'nut/relevantie'.

Uit deze resultaten kunnen de volgende conclusies worden getrokken. Allereerst blijkt dat naast het algemene aspect 'plezier in wiskunde', ook duidelijk te onderscheiden, meer specifieke, aspecten in de beleving van de leerlingen een rol spelen. Het begrip 'beleving van wiskunde' kan met andere woorden op zowel een algemene als een meer specifieke wijze worden benaderd. Beide benaderingen zijn in de BSW (door middel van de algemene schaal 'plezier', en de meer specifieke schalen 'angst/moeilijkheid', 'inzet/interesse' en 'nut/relevantie') geoperationaliiseerd.

Op de tweede plaats kan worden vastgesteld dat de structuur van de schaal op leerlingniveau dezelfde is als op klasniveau. De 'beleving van het vak wiskunde' kan dan ook als een zowel op leerling- als op klasniveau meetbaar construct worden opgevat.

Nu evidentie gevonden is voor de gehanteerde indeling in subschalen, is het geoorloofd de scores per subschaal te sommeren tot 'schaalscores'. Wanneer we kijken naar de correlaties tussen de schaalscores (zie tabel 3) zien we dat de subschaal 'plezier' sterk samenhangt met elk van de overige subschalen. Daarnaast blijken de schalen 'nut/relevantie' en 'inzet/interesse' duidelijk gecorreleerd.

Tabel 3. De produkt-moment correlaties tussen de schaalscores van de BSW (onder en boven de diagonaal op leerlingniveau (N = 7815) resp. klasniveau (N = 335)).

	1 plezier	2 angst/ moeilijkheid	3 inzet/ interesse	4 nut/ relevantie
1 plezier	–	.60	.76	.65
2 angst/moeilijkheid	.52	–	.32	.36
3 inzet/interesse	.65	.22	–	.66
4 nut/relevantie	.53	.25	.51	–

Betrouwbaarheid

In onderstaande tabel is de interne consistentie (Cronbach's alpha) voor de vier subschalen van de BSW zowel op leerling- als op klasniveau weergegeven.

Tabel 4. De betrouwbaarheid (Cronbach's alpha) van de vier subschalen van de BSW (A = leerlingniveau, B = klasniveau).

	A	B
1 plezier	.88	.95
2 angst/moeilijkheid	.80	.88
3 inzet/interesse	.74	.86
4 nut/relevantie	.80	.89

Gezien het kleine aantal items per subschaal kan de betrouwbaarheid relatief hoog worden genoemd.

Validiteit

Wanneer de variantie binnen klassen relatief klein is ten opzichte van de variantie tussen klassen, biedt een klassescore een zinvolle weergave van de beleving van de 'klas als geheel'. Om na te gaan in hoeverre er sprake is van een dergelijk klasse-effect, is een univariate variantie-analyse uitgevoerd. Hierbij fungeerde de klas als onafhankelijke variabele en de leerlingsscores op de schaal 'plezier', die zoals vermeld betrekking heeft op het meest centrale aspect van de beleving van het vak wiskunde, als afhankelijke variabele. De resultaten ervan zijn per leerjaar weergegeven in onderstaande tabel.

Tabel 5. Resultaten van de univariate variantie-analyse voor verschillen tussen klassen (N = 7804; analyse-eenheid: leerling).

Leerjaar	variantiebron	df	SS	MS	F-ratio	P
1	tussen klassen	95	32748	344.72	7.03	.0000
	binnen klassen	2177	106684	49.01		
	totaal	2272	139432			
2	tussen klassen	95	27160	285.89	5.06	.0000
	binnen klassen	2247	126849	56.45		
	totaal	2342	154009			
3	tussen klassen	142	43268	304.71	5.71	.0000
	binnen klassen	3045	162431	53.34		
	totaal	3187	205700			

Uit tabel 5 valt op te maken dat de klassen significant verschillen in hun beleving van het wiskunde-onderwijs. Voorts blijken de verschillen tussen klassen in verhouding tot die binnen klassen, gezien de F-waarden, relatief groot te zijn⁴. De intraklassecorrelatie Rho (Kish, 1965) bedraagt voor leerjaar 1, 2 en 3 respectievelijk .20, .14 en .17.

Deze gegevens betekenen dat er sprake is van een duidelijk klasse-effect: hoewel leerlingen onderling uiteraard verschillen in de wijze waarop het vak wiskunde wordt beleefd, blijkt er tussen de leerlingen van dezelfde klas in dit opzicht een zekere overeenstemming te bestaan. De klassescores lijken dus waardevolle informatie te kunnen verschaffen over de beleving van de 'klas als geheel'.

⁴ Indien deze analyses per leerjaar voor elk van de opleidingstypen afzonderlijk worden uitgevoerd, worden vergelijkbare gegevens verkregen (vgl. Martinot & Kuhlemeier, in voorbereiding).

Teneinde meer te weten te komen over de validiteit van de BSW als instrument ter meting van beleving van het vak wiskunde, zijn de correlaties berekend tussen enerzijds de klassescores op de vier subschalen en anderzijds de per klas gemiddelde gegevens die met behulp van de vragenlijst voor de leerlingen zijn verzameld. Deze correlaties zijn weergegeven in tabel 6.

Tabel 6. Product-moment correlaties tussen de schaalscores van de BSW en de variabelen uit de vragenlijst voor de leerlingen (N = 335; analyse-eenheid: klas)

	1 plezier	2 angst en moeilijkheid	3 inzet en interesse	4 nut/ relevantie
waardering wiskunde	.86**	.65**	.68**	.65**
waardering Nederlands	-.04	.00	.09	.06
waardering Engels	.09	.01	.25**	.22**
huiswerk wiskunde	.16*	-.15*	.33**	.22*
huiswerk Nederlands	.11	.10	.15*	.06
huiswerk Engels	.18**	.09	.21**	.15*
rapportcijfer wiskunde	.46**	.68**	.30**	.28**
rapportcijfer Nederlands	.04	.09	.06	.07
rapportcijfer Engels	.09	.13	.14	.12

* $p < .01$; ** $p < .001$

Uit deze tabel kan worden opgemaakt dat de, eerder in dit artikel beschreven, voorspellingen ten aanzien van zowel de sterkte als de richting van de veronderstelde relaties worden bevestigd. De klassescores van de BSW blijken, zoals verwacht, ieder betrekking te hebben op een specifiek aspect van de beleving van het vak wiskunde.

Verschillen tussen leerjaren en opleidingstypen

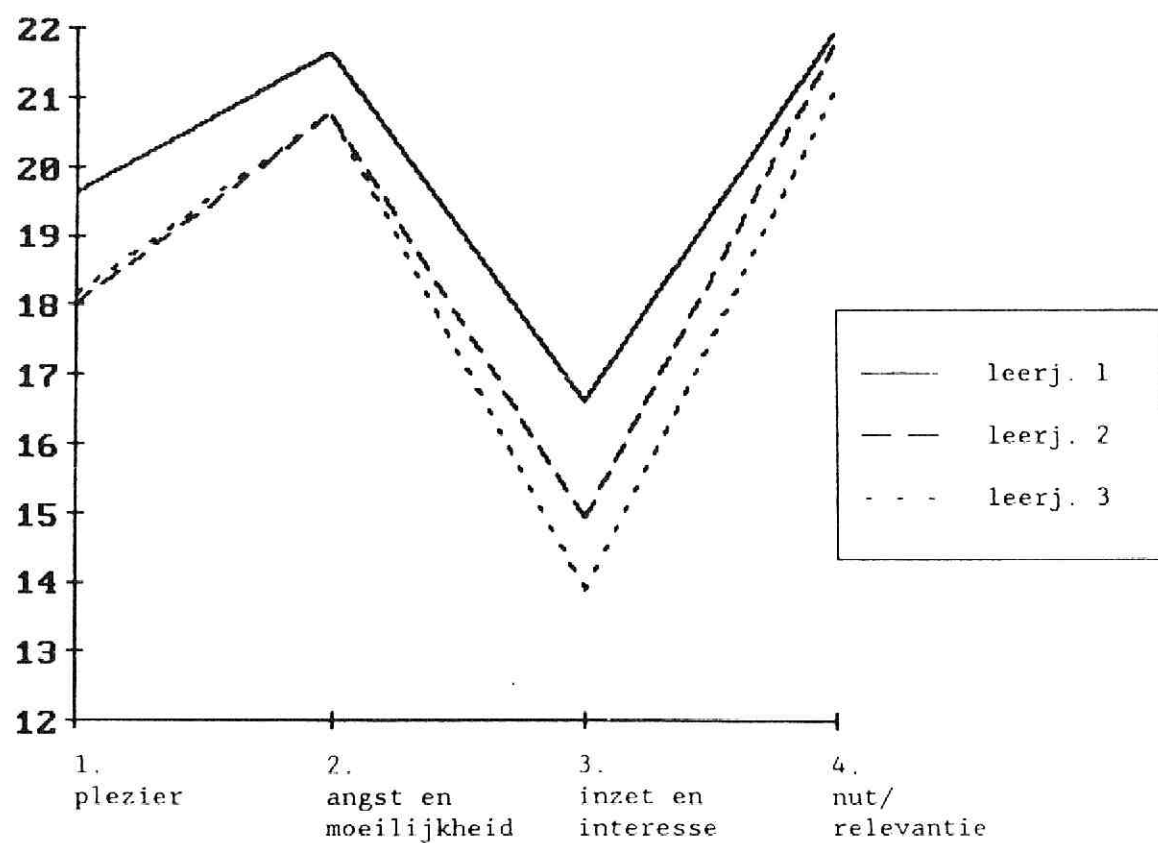
Met behulp van multivariate variantie-analyse is de samenhang onderzocht tussen de klassescores van de BSW (afhankelijke variabelen) en leerjaar en opleidingstype (onafhankelijke variabelen). De variabele opleidingstype omvatte de categorieën LBO, MAVO en HAVO/VWO; combinaties van deze drie opleidingstypen (15.8%) werden van tevoren uit het bestand verwijderd. De figuren 1 en 2 geven een beeld van de gemiddelden per leerjaar en per opleiding.

Eerst is nagegaan of het interactie-effect van leerjaar en opleidingstype significant was. Zoals verwacht, bleek dit niet het geval (Wilks' lambda = .92; $F = 1.39$; $p = .14$).

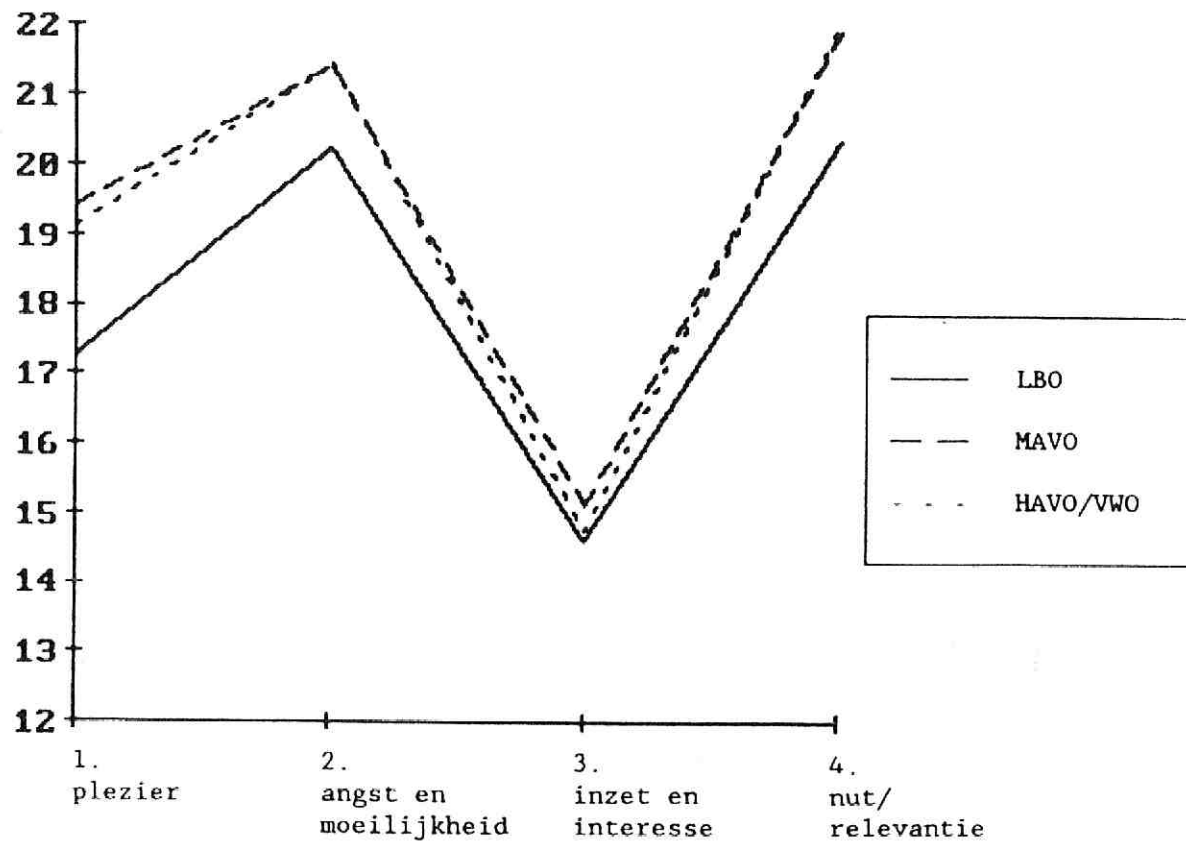
Daarna zijn de hoofdeffecten van leerjaar en opleidingstype eveneens multivariaat geanalyseerd. Het hoofdeffect van leerjaar bleek significant (Wilks' lambda = .73; $F = 11.63$; $p = .000$). Hetzelfde gold het hoofdeffect van opleidingstype (Wilks' lambda = .83; $F = 6.83$; $p = .000$). Gezien de grootte van de F-waarden kunnen we stellen dat het effect van leerjaar groter is dan dat van opleidingstype.

De normgegevens

De belangrijkste beschrijvende gegevens van het normeringsgedeelte van het onderzoek zijn weergegeven in Tabel 7. Hoe kunnen we de hier gerapporteerde landelijke gemiddelden op de vier subschalen interpreteren? Anders geformuleerd: welke conclusies kunnen we trekken met betrekking tot de wijze waarop de gemiddelde klas in Nederland het wiskunde-onderwijs beleeft? In dit verband dient opgemerkt te worden dat de klassescores in beginsel variëren tussen nul (als alle leerlingen bij alle acht items de meest negatieve categorie hebben gekozen) en 32 (als alleen de meest positieve categorie is aangekruist). Een landelijk gemiddelde van 16 betekent dat de leerlingen gemiddeld een neutrale positie hebben ingenomen bij het invullen van de desbetreffende subschaal. In dit licht bezien wijzen de data op een verhoudingsgewijs hoge waardering



Figuur 1: Gemiddelden voor leerjaar 1, 2, en 3.



Figuur 2: Gemiddelden voor het LBO, MAVO en HAVO/VWO

Tabel 7. Gegevens normeringsonderzoek per subschaal en leerjaar (analyse-eenheid: klas)

	leerjaar 1	leerjaar 2	leerjaar 3
plezier			
N	96.	96.	143.
gemiddelde score	19.64	18.02	18.15
standaarddeviatie	3.83	3.46	3.77
standaardmeetfout	.93	1.16	1.19
angst/moeilijkheid			
N	96.	96.	143.
gemiddelde score	21.66	20.77	20.76
standaarddeviatie	2.59	2.19	2.51
standaardmeetfout	1.39	1.16	1.16
inzet/interesse			
N	96.	96.	143.
gemiddelde score	16.62	14.92	13.89
standaarddeviatie	2.24	2.68	2.67
standaardmeetfout	1.43	1.35	1.39
nut/relevantie			
N	96.	96.	143.
gemiddelde score	21.95	21.77	21.09
standaarddeviatie	1.84	2.34	2.57
standaardmeetfout	1.18	.97	1.02

voor het nut en de relevantie van het vak wiskunde. Wat betreft de angst voor en de moeilijkheid van het vak, kan de beleving eveneens als redelijk positief worden gekarakteriseerd. Het plezier dat leerlingen in wiskunde beleven en de inzet voor en interesse in dit vakgebied, ligt daarentegen dicht bij de buurt van de neutrale categorie. Opvallend is de duidelijke afname van de inzet en interesse van leerlingen naarmate het leerjaar hoger wordt. Al met al geven de bevindingen echter geen aanleiding tot het veronderstellen van een extreem negatieve beleving van het wiskunde-onderwijs in de onderbouw van het voortgezet onderwijs.

De samenhang met enkele docentvariabelen

Tenslotte is in het onderzoek gekeken naar eventuele samenhang van de BSW-scores met enkele docentvariabelen. De resultaten zijn weergegeven in tabel 8.

Tabel 8. Product-moment correlaties tussen de schaa scores van de BSW en enkele docentvariabelen (N = 335; analyse-eenheid: klas)

	1 plezier	2 angst/ moeilijkheid	3 inzet/ interesse	4 nut/ relevantie
beleving lesgeven	.30**	.19**	.26**	.21**
scoreverwachting t.a.v.				
- plezier	.38**	.28**	.34**	.27**
- angst/moeilijkheid	.06	.18**	.02	.06
- inzet/interesse	.32**	.24**	.31**	.35**
- nut/relevantie	.18*	.17*	.11	.24**

* $p < .01$; ** $p < .001$

De gevonden correlaties zijn, hoewel in de meeste gevallen significant, niet bijzonder hoog. De samenhang met de beleving van het lesgeven is als volgt te beschrijven: naarmate de docent zijn eigen lesgeven als positiever ervaart, beleeft de klas het wiskunde-onderwijs eveneens als positiever. Voorts blijken docenten de (gemiddelde) beleving van hun klas, zoals gemeten met de BSW, maar matig te kunnen voorspellen. Deze bevinding is overigens in overeenstemming met resultaten van het eerder uitgevoerde implementatieonderzoek (vgl. Martinot, 1986, p. 25). Met name lijkt de docent een onvolledig beeld te hebben van de mate waarin de klas bang is voor wiskunde en het vak als moeilijk ervaart.

CONCLUSIES EN AANBEVELINGEN

Wat betreft de psychometrische kenmerken van de BSW zijn een drietal conclusies te trekken. Ten eerste blijkt de interne structuur van de BSW door de resultaten van het onderzoek te worden ondersteund. Ten tweede is de betrouwbaarheid van het instrument alleszins redelijk te noemen. Ten derde wijzen de bevindingen in de richting van een bevredigende validiteit. Op grond van de in dit onderzoek verzamelde gegevens lijkt het verantwoord de BSW ter beschikking te stellen aan wiskundeleraars in de onderbouw van het voortgezet onderwijs. De gerapporteerde normen bieden de docent de mogelijkheid de wiskundebeleving van zijn klas te vergelijken met die van een steekproef van klassen uit hetzelfde leerjaar. Andere aanbevelingswaardige gebruiksmogelijkheden liggen op het terrein van het wetenschappelijk onderzoek, zoals summatieve leerplanevaluatie en landelijke peiling van het onderwijsniveau. Wat betreft het laatste zou de BSW een empirische bijdrage kunnen leveren aan een discussie in brede kring over de doelstellingen van het wiskunde-onderwijs op affectief gebied.

LITERATUUR

- Angoff, W.H. (1984). Scales, norms and equivalent scores. Princeton, NJ: Educational Testing Service.
- Crombach, M.J., Voeten, M.J.M. & Feenstra, H.J.M. (1986). Ontwikkeling en validering van een computer-attitudeschaal. *Tijdschrift voor Onderwijsresearch*, 11, 301-311.
- Kish, L. (1965). Survey sampling. New York: Wiley.
- Korff, J., Kamphorst, J., Werf, M.P.C. van de, & Clason, C.E. (1986). Meisjes en wiskunde: het HEWET-project (interimrapport). Groningen: RION, Instituut voor Onderwijsresearch.
- Kremers, E.J.J. (1987). Affectieve doelstellingen in het onderwijs en een toepassing voor het vak wiskunde. Cito-memo Nr. 277. Arnhem: Cito.
- Kremers, E.J.J. (1981). De wiskunde-attitudeschaal: een voorbeeld van een instrument voor het evalueren van affectieve doelstellingen. In P. Weeda (red.), *Aspecten van leerplanevaluatie* (p. 208-232). Den Bosch: Malmberg.
- Kremers, E.J.J. (1979). Wat vinden leerlingen van wiskunde? *Euclides*, 54, 301-310.
- Kuhlemeier, J.B. & Martinot, M.J. (1987). Verslag van het enquête-onderzoek met de Wiskundebelevingsschaal. Interne Documentatie Nr. 240. Arnhem: Cito.
- Martinot, M.J. (1986). Verslag van het vooronderzoek met de Wiskunde-belevingsschaal: De interviewgegevens. Interne Documentatie Nr 206. Arnhem: Cito.
- Martinot, M.J. & Kuhlemeier, J.B. (in voorbereiding). De normering van de Belevingsschaal voor wiskunde: Verslag van het onderzoek naar de betrouwbaarheid, validiteit en normen voor de eerste drie leerjaren van het voortgezet onderwijs. Arnhem: Cito.
- Raat, J.H. & Vries, M. de (1986). What do girls and boys think of technology? Report Patt-workshop. Eindhoven: University of Technology.

Manuscript ontvangen 13-7-1987.

Definitieve versie ontvangen 2-2-1987.

Bijlage 1 De faktoroplossing volgens de principale componentenanalyse (2 factoren, ongeroteerd) voor de items van de BSW (A = leerlingniveau, B = klasniveau).

A			B		
item behorende bij subschaal	Faktor		item behorende bij subschaal	faktor	
	1	2		1	2
1	-.77	.	1*	.88	.
1*	.74	.	1*	.88	.
1*	.73	.	1	-.88	.
1*	.69	.	1*	.82	.
1	-.67	.	1*	.80	.
1*	.65	.	3*	.79	-.24
1	-.61	.	1	-.77	.
3*	.57	.	4*	.76	-.25
2	-.56	-.46	1	-.74	.
2*	.56	.52	3*	.73	.
1	-.56	.	4*	.71	-.21
4*	.56	-.25	3	-.70	.31
3*	.54	.	1	-.70	.
3	-.52	.28	4	-.69	.37
4*	.52	-.25	3*	.69	-.35
3	-.51	.	2*	.67	.52
4	-.51	.31	4*	.65	-.21
4*	.51	-.26	4	-.64	.32
2*	.49	.39	2	-.64	-.49
4	-.48	.30	3	-.63	.
2	-.47	-.46	2*	.63	.43
2	-.43	-.35	2	-.62	-.45
4	-.42	.24	2	-.55	-.47
3	-.42	.26	3	-.54	.29
3*	.42	-.29	4	-.54	.25
4	-.40	.29	4*	.54	.
4*	.39	-.25	3*	.49	-.49
3	-.38	.	4	-.49	.
3*	.30	-.40	3	-.44	.22
2*	.30	.57	2*	.34	.68
2*	.	.54	2	-.21	-.61
2	.	-.42	2*	.	.67

* = negatief geformuleerd item

. = $-.20 < \text{faktorlading} < .20$

Bijlage 2 De 3-factoroplossing (orthogonaal, varimax geroteerd) voor de BSW-items uit de subschalen 2, 3 en 4 (A = leerlingniveau, B = klasniveau).

A				B			
item behorende bij subschaal	factor			item behorende bij subschaal	factor		
	a	b	c		a	b	c
2*	.76	.	.	2*	.80	-.25	.23
2	-.71	.21	.	2	-.77	.31	.
2	-.67	.	.	2*	.77	.	.
2*	.66	.	.	2	-.72	.31	.
2*	.62	-.22	.	2*	.72	-.35	.
2*	.58	.	.	2	-.71	.26	.
2	-.55	.	.	2*	.70	.22	.
2	-.48	.	.	2	-.62	-.28	.
3	.	.65	-.22	3	.	.76	.
3*	.	-.61	.	3	.	.73	-.36
3*	.	-.59	.	3*	.	-.72	.21
3	.	.57	.	3*	.	-.71	.40
3	.	.55	.	3*	.	-.65	.44
3	.	.51	-.23	3	.	.60	.
3*	.	-.49	.30	3	.	.54	-.36
3*	.26	-.48	.	3*	.34	-.50	.34
4	.	.	-.71	4*	.	.	.82
4	.	.	-.70	4	.	.	-.82
4*	.	.	.68	4*	.	-.33	.79
4*	.	.	.67	4	.	.39	-.76
4*	.	.	.66	4*	.	-.32	.69
4	.	.33	-.52	4*	.	-.42	.61
4*	.	-.32	.51	4	.	.45	-.61
4	.	.26	-.45	4	.	.28	-.46

a = faktor "angst/moeilijkheid"

b = faktor "inzet/interesse"

c = faktor "nut/relevantie"

* = negatief geformuleerd item

. = $-.20 < \text{faktorlading} < .20$

Inter- en Criterium- Codeer-Overeenstemming van Beroepenclassificaties

J.A. Gras* en G.J. Mellenbergh**

Faculteit der Psychologie, Universiteit van Amsterdam.

ABSTRACT

In an experimental study the reliability of the coding of occupations was investigated. The independent variables are the occupational classification system (ITS/CBS/Simple System), the listing of the occupation (yes/no) in the ITS- and CBS-systems, the extremity (extreme/non-extreme) of the occupational prestige, and the familiarity (familiar/non-familiar) of the occupations with the coders. For each of the eight combinations of the independent variables Listing, Extremity, and Familiarity five occupations were chosen. The occupations were coded by a group of subjects chosen from the DSWO pool of interviewers. The subjects were randomly assigned to the occupational classification systems: six coders used the ITS-classification, six coders the CBS-system, and six coders the Simple System. Moreover, two CBS-experts coded the occupations using the CBS-classification system, and the first author coded the occupations using the ITS-system; these codings were considered to be criterion codings. The dependent variables are the coding time, the interobserver agreement, and the observer-criterion agreement. The mean coding time is much higher for the CBS than the ITS-classification. Although the classification systems (CBS and ITS) are an improvement on the Simple System the agreement is rather low. The agreement is higher for listed occupations than for non-listed ones, and the interobserver agreement is higher for familiar occupations than for non-familiar ones.

1. INLEIDING

In onderwijskundig onderzoek wordt veelvuldig gebruik gemaakt van de variabele "milieu van de leerling", zoals in onderzoek naar de samenhang tussen milieu en studieprestaties, milieu en schoolkeuzeadviezen etc. Bij de constructie van een milieu-index wordt meestal het beroep van de ouders gehanteerd. Deze beroepen worden gecodeerd en omgezet in beroepsprestige-scores. De bruikbaarheid van een dergelijke milieu-index wordt dus mede bepaald door de kwaliteit van de beroepscoderingen.

In het verleden zijn verschillende schalen voor het meten van het beroepsprestige ontwikkeld; voor een overzicht van de Nederlandse onderzoekstraditie op dit gebied zie Swanborn (1978, par. 1). Van meer recente datum zijn de ITS-beroepenklapper (Van Westerlaak, Kropman & Collaris, 1975) en de beroepsprestigeladder van Sixma en Ultee (1983). Op deze laatste schaal werd kritiek geleverd door Pompe en Rutges (1984). Deze auteurs wijzen ook op een aantal methodologische problemen bij het beroepsprestige-onderzoek. Popping (1986) deed nader onderzoek naar een drietal methodologische problemen. Hij onderzocht o.a. de overeenstemming tussen de indeling van beroepen door verschillende personen die met twee andere beroepenclassificaties werkten; deze bleek in het algemeen niet hoog.

In dit artikel wordt eveneens onderzoek gerapporteerd naar de betrouwbaarheid van de codering van beroepen. Aan de betrouwbaarheid van coderingen kan men verschillende aspecten

* Dienst Sociaal Wetenschappelijk, Onderzoek (DSWO), Rijksuniversiteit Leiden.

** Faculteit der Psychologie, Universiteit van Amsterdam, Weesperplein 8, 1018 XA Amsterdam.

Met dank aan de heren J. Oud en Ph. Andela van het Centraal Bureau voor de Statistiek voor hun medeweking.

ten onderscheiden; zie bij voorbeeld Swanborn, (1978 par. 3). Hier wordt aandacht besteed aan de overeenstemming tussen de coderingen van verschillende codeurs (intercodeur-overeenstemming) en de overeenstemming van de coderingen van codeurs met criterium-coderingen (criterium-codeur-overeenstemming).

In onderzoek wordt de codering van het beroep vaak verricht door codeurs die hiertoe niet speciaal zijn opgeleid (interviewers, studenten etc.). Er werd aangesloten bij deze praktijk en de betrouwbaarheid van de codering werd onderzocht bij onervaren codeurs.

Een veronderstelling is dat de betrouwbaarheid van de codering mede beïnvloed kan worden door kenmerken van het beroep en de gebruikte beroepenclassificatie; enkele van deze kenmerken werden daarom in het onderzoek systematisch gevarieerd.

2. OPZET EN UITVOERING VAN HET ONDERZOEK

In een experiment werden de kwaliteiten van twee beroepenclassificaties onderzocht; bovendien werd nagegaan wat de invloed van enkele beroepskenmerken is. In deze paragraaf wordt de opzet en uitvoering van het experiment besproken.

2.1. Beroepenclassificaties

De ITS-beroepenklapper (Van Westerlaak et al. 1975) en de CBS-beroepenclassificatie (CBS, 1985) worden in Nederland vaak gebruikt in sociaal en gedragswetenschappelijk onderzoek. De ITS-klapper bevat een lijst met 764 gecodeerde beroepen. De codeur plaatst een beroep in één van zes beroepsniveau-categorieën, lopend van ongeschoolde arbeid tot hogere beroepen; daarnaast heeft de codeur de mogelijkheid aan te geven dat het beroep niet ingedeeld kan worden in één van deze categorieën. De CBS-classificatie bevat een lijst met ongeveer 18500 beroepsnamen. Elk beroep heeft een code van vier cijfers: het eerste cijfer geeft de beroepssector in acht categorieën (na onderverdeling in tien categorieën), het tweede de beroepsklasse in 87 categorieën, het derde de beroepsgroep in 317 categorieën en het vierde cijfer de beroepscode in 914 categorieën; daarnaast kan de codeur aangeven dat het beroep niet ingedeeld kan worden. Met behulp van de viercijferige code kan het beroep ingedeeld worden in de acht categorieën van de zogenaamde "sociale groepsindeling", lopend van ongeschoolde arbeiders tot hogere employe's (CBS, 1982a). Tevens kan uit de viercijferige code met behulp van een codeersleutel de score op de beroepsprestigeschaal van Sixma en Ultee (1983) worden afgeleid.

Ter vergelijking met de ITS en CBS werd nog een derde methode voor beroepenclassificatie gebruikt, aangeduid als Eenvoudige-Beroepen-Classificatie (EBC). In de EBC wordt de codeur, zonder verdere hulpmiddelen, verzocht het beroep in te delen in een van vijf categorieën lopend van een laag tot een hoog beroepsniveau.

2.2. Beroepen

Er werd een steekproef getrokken van 298 beroepen uit de volgende lijsten: beroepenklapper ITS, beroepenlijst CBS, Elseviers (1983) beroepenalbum en de boeken van Martens (1981) en Wiegersma (1976); dit aantal werd aangevuld met twee nieuwe beroepen (subsidioloog en natuurgeneeskundige). Hieruit werden de beroepen geselecteerd die zowel in de ITS-klapper als de CBS-beroepenlijst voorkomen en de beroepen die in geen van beide lijsten voorkomen; deze worden aangeduid als gelijste, respectievelijk, niet-gelijste beroepen. Door de eerste auteur werden deze beroepen gecodeerd met de CBS-classificatie. De codes werden met behulp van de codeertabel van de beroepenladder van Sixma en Ultee (1983) omgezet in beroepsprestigescores. Vervolgens werden de beroepen ingedeeld in een groep met een extreem beroepsprestige (≤ 31 of ≥ 68) en een groep met een middelmatig beroepsprestige (tussen 31 en 68) op de schaal van Sixma en Ultee. Tenslotte werden door de eerste auteur, met behulp van gegevens uit de Arbeidskrachtentelling (CBS, 1982b), de beroepen verdeeld in relatief bekende en onbekende beroepen. Op deze wijze werden de beroepen verdeeld in 2 (Gelijst/Niet-Gelijk) $\times$ 2 (Extreem/Niet-Extreem Beroepsprestige) $\times$ 2 (Bekend/Onbekend) = 8 categorieën. Uit elk van

deze acht categorieën werden vijf beroepen geselecteerd. Deze 40 beroepen werden gecodeerd door twee deskundigen van het CBS. De codes van de deskundigen werden volgens boven beschreven procedure omgezet in een extreem versus niet-extreem beroepsprestige. Op grond van de coderingen door de deskundigen werd besloten slechts 38 beroepen in de analyse te betrekken; alle beroepen zijn vermeld in Tabel 1.

Tabel 1 Naam met omschrijving en kenmerken van de 40 gecodeerde beroepen

Naam met omschrijving	Kenmerken		
	Prestige	Bekendheid	Lijsting
Automatiseringsdeskundige: adviseren procedures automatisering en aanschaf apparatuur, begeleiding programmeurs; loondienst/salaris (niet-overheid); leidinggevend.	Extreem	Bekend	Niet-Gel
Bedrijfsconsulent: advies en consult over bedrijfsontwikkeling slagersvak; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Onbekend	Niet-Gel
Kleurenadviseur: advies kleurkeuze woninginrichting en andere interieurs; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Bekend	Niet-Gel
Subsidioloog: inventariseren en aanboren potentiële subsidiebronnen; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Onbekend	Niet-Gel
Sales-engineer: marktontwikkeling; loondienst/salaris (niet-overheid); leidinggevend.	Extreem	Onbekend	Niet-Gel
Bedrijfsleermeester; verzorgt bedrijfsopleidingen schilders-personeel; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Onbekend	Niet-Gel
Innovator: leidt vakleerkrachten op voor onderwijs bewegingsleer; overheid; niet-leidinggevend.	Extreem	Onbekend	Niet-Gel
Pasdame: showroomverkoopster; past kleding; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Bekend	Niet-Gel
Standbouwer: tentoonstellings-standbouw; zelfstandig met twee personen in dienst.	Extreem	Bekend	Niet-Gel
Provisor: zaakwaarnemer apotheek; loondienst/salaris (niet-overheid); leidinggevend.	Niet-Extr.	Onbekend	Niet-Gel
Jeugdschaakleider: geeft schaakles aan jongeren, organiseert toernooien; loondienst/salaris (niet-overheid); leidinggevend.	Niet-Extr.	Bekend	Niet-Gel
Hotelemployé: opvang, begeleiding hotelgasten, informatie en organisatie uitstapjes; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Bekend	Niet-Gel
Veiligheidsdeskundige: houdt toezicht op veiligheid en gebruik gevaarlijke stoffen; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem.	Onbekend	Niet-Gel
Hoofdkoper: bepaalt inkoopbeleid; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Bekend	Niet-Gel
Beleidsmedewerker: ondersteuning en voorbereiding ministerieel beleid arbeidsvoorwaarden onderwijzend personeel; overheidsdienst; niet-leidinggevend.	Niet-Extr.	Onbekend	Niet-Gel
Natuurgeneeskundige: consulten volgens natuurgeneeskunde; zelfstandig zonder personeel.	Niet-Extr.	Bekend	Niet-Gel

Tabel 1 (vervolg)

Naam met omschrijving	Kenmerken		
	Prestige	Bekendheid	Lijsting
Consulent bij jeugd-en jongerenwerk: voor hulpverlening bij methodiek- en beleidszaken; overheidsdienst; niet-leidinggevend.	Niet-Extr.	Onbekend	Niet-Gel
* Credietadviseur: adviseert over mogelijke credietverlening loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Bekend	Niet-Gel
Huisbezoekster: bezoekt patienten i.v.m. TBC-bestrijding; overheidsdienst; niet-leidinggevend.	Niet-Extr.	Onbekend	Niet-Gel
Kantooremployé: bijhouden van administratie im- en exportbedrijf in de landbouw; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Bekend	Niet-Gel
Arbeider: in de landbouw; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Bekend	Niet-Gel
Voorwerker: grondwerk wegenbouw; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Onbekend	Gelijst
* Hoofdwerktuigkundige: draaiende houden turbines electriciteitscentrales; overheid; leidinggevend.	Niet-Extr.	Onbekend	Gelijst
Slager: zelfstandige met drie personen in dienst.	Niet-Extr.	Bekend	Gelijst
Assembleerder: assembleert; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Onbekend	Gelijst
Architect; zelfstandige zonder personeel.	Extreem	Bekend	Gelijst
Dakdekker: rietdekker van daken; zelfstandige met twee personen in dienst.	Extreem	Bekend	Gelijst
Natuurkundig ingenieur: laboratoriumkolenverbranding; overheid; niet-leidinggevend.	Extreem	Onbekend	Gelijst
Verkoper schoenenzaak; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Bekend	Gelijst
Vliegtuigbouwer: houtconstructies vliegtuigen; loondienst/salaris (niet-overheid); niet-leidinggevend.	Extreem	Onbekend	Gelijst
Statistisch assistent: opmaken van statistische tabellen en andere beschrijvingen; overheid; niet-leidinggevend.	Niet-Extr.	Onbekend	Gelijst
Dienstleider N.S.: organiseert planning diensten rijdend personeel; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Bekend	Gelijst
Veeboer: veestapel van 100 koeien; zelfstandig met twee personen in dienst.	Niet-Extr.	Bekend	Gelijst
(Consulent-) inspecteur kinderbescherming: toezicht op naleven wet op kinderbescherming; overheid; niet-leidinggevend.	Niet-Extr.	Bekend	Gelijst
Monteur hoogspannings apparatuur: aansluiten electromotoren en zware elektrische apparaten; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Onbekend	Gelijst
Beleggingsadviseur: advies over beleggingen en vermogensbeheer; zelfstandige zonder personeel.	Extreem	Bekend	Gelijst

Tabel 1 (vervolg)

Naam met omschrijving	Kenmerken		
	Prestige	Bekendheid	Lijsting
Effectenadministrateur: administratie en bewaring effecten particuliere bankrelaties; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Onbekend	Gelijst
Kartograaf: kaartredacteur bij kartografisch bedrijf; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Onbekend	Gelijst
Kluisbeheerder: beheer en toegangscontrole grote bankkluis; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Bekend	Gelijst
Scheepsladingmeter: berekening omvang scheepsladingen bij laden en lossen; loondienst/salaris (niet-overheid); niet-leidinggevend.	Niet-Extr.	Onbekend	Gelijst

* Deze beroepen zijn niet gebruikt bij de berekening van kappa's.

De indelingen Extreem/Niet-Extreem Beroepsprestige zoals vastgesteld met behulp van de coderingen door de CBS-deskundigen en Gelijst/Niet-Gelijst beroep werden gebruikt als onafhankelijke variabelen. De onafhankelijke variabele Bekend/Onbekend zoals vastgesteld door de eerste auteur werd later vervangen door de indeling Bekend/Onbekend bij de codeurs; dit wordt in het vervolg van deze paragraaf besproken (zie 2.6).

2.3. Codeurs

De beroepen werden gecodeerd door 18 codeurs. Om de experimentele situatie zo veel mogelijk te laten aansluiten bij de gangbare praktijk van het codeerwerk werd gekozen voor betrouwbare krachten die wel ervaring met sociaalwetenschappelijk onderzoek hebben maar geen bijzondere kennis van beroepen of ervaring met coderen. De codeurs werden geworven uit het interviewerscorps van de Dienst Sociaal-Wetenschappelijk Onderzoek (DSWO) van de Rijksuniversiteit Leiden; zij werden betaald voor hun werkzaamheden.

2.4. Data-verzameling

De codeurs werden aselekt toegewezen aan de beroepenclassificaties: zes werkten met de ITS, zes met de CBS en zes met de EBC. Na een inleiding en instructie codeerde elke codeur alle 40 beroepen. Elk beroep werd aangeboden op een formulier met (1) de naam van het beroep, (2) een korte omschrijving van de aard van de werkzaamheden, (3) het dienstverband, (4) het aantal personen in dienst bij zelfstandigen en (5) het al dan niet leiding geven in geval van loondienst; deze gegevens staan vermeld in Tabel 1. Daarnaast beoordeelde de codeur de mate waarin het beroep hem/haar bekend was op een schaal van 0 tot en met 100.

De beschikbare tijd was 90 minuten. De ITS- en EBC-codeurs waren klaar met hun werkzaamheden binnen de gestelde tijd. Na 110 minuten had geen der CBS-codeurs alle beroepen gecodeerd. Vier van deze codeurs zijn een keer teruggekomen om het werk af te maken. Twee codeurs zijn geen tweede keer verschenen, hun werk is afgemaakt door twee nieuwe codeurs uit het DSWO-interviewersbestand.

2.5. Indices

Er werden drie indices gebruikt om de experimentele effecten te onderzoeken. In de eerste plaats werd voor alle codeurs de codeertijd in minuten bepaald. Daarnaast werden twee typen overeenstemming gebruikt.

De overeenstemming tussen codeurs die met dezelfde beroepenclassificatie hadden gewerkt werd bepaald met de variant van Cohen's (1960) coëfficiënt kappa voor een groep codeurs (Popping, 1983, par 3.3.3; Schouten, 1985; par 2.2; 1986).

Voor de CBS-classificatie is de codering door twee deskundigen beschikbaar; deze is gebruikt als criterium. Voor de ITS-classificatie werd de codering door de eerste auteur als criterium gebruikt; voor de EBC kan geen criterium vastgesteld worden. De overeenstemming met het criterium werd bepaald met de gewogen kappa voor de overeenstemming van een groep codeurs met een criterium (Popping, 1983, par. 3.3.6).

De kappa's werden berekend over het aantal categorieën van de beroepenclassificatie, vijf voor de EBC, zeven (inclusief "niet te coderen") voor de ITS en 914 voor de CBS. In geval van de CBS worden de meeste codes niet gebruikt omdat de 38 beroepen slechts in een beperkt aantal van deze categorieën worden ingedeeld; de niet-gebruikte categorieën leveren geen bijdrage aan kappa.

2.6. Proefopzet

De opzet van het experiment kan beschreven worden als een 3x2x2x2 factoriële proefopzet met de volgende onafhankelijke variabelen: Beroepenclassificatie (ITS/CBS/EBC), Lijsting(Gelijst/Niet-Gelijst), Beroepsprestige (Extreem/Niet-Extreem) zoals vastgesteld met behulp van de CBS-coderingen door deskundigen en Bekendheid (Bekend/Onbekend). Bij de samenstelling van de groep beroepen werd de bekendheid bepaald door de eerste auteur met behulp van de Arbeidskrachtentelling. Na afloop van het experiment kon de variabele bekendheid worden vastgesteld met behulp van gegevens van de codeurs. Voor elk beroep werd de gemiddelde bekendheid van het beroep (schaal 0 tot en met 100) over de 18 codeurs berekend. Deze gemiddelden werden rondom de mediaan (63,5) gedichotomiseerd in relatief bekende en onbekende beroepen. In Tabel 1 is aangegeven tot welk van de acht categorieën elk beroep behoort.

De afhankelijke variabelen zijn codeertijd, de intercodeur-overeenstemming (ico) en de criterium-codeur-overeenstemming (cco). De codeertijden zijn alleen bekend per beroepenclassificatie over alle beroepen heen. De criterium-codeurovereenstemming is alleen bekend voor de ITS en CBS omdat voor de EBC geen criterium opgesteld kan worden. De codeertijden zijn gemeten voor alle 40 beroepen; bij de berekening van de overeenstemmingsmaten zijn slechts 38 beroepen betrokken omdat twee beroepen verwijderd zijn op grond van de oordelen van de CBS-deskundigen.

2.7. Hypothesen en data-analyse

Er werd verwacht dat de codeertijden hoger zijn naarmate de beroepenclassificatie gecompliceerder is, dat wil zeggen hoger bij de CBS dan bij de ITS en hoger bij de ITS dan bij de EBC. De codeertijden werden geanalyseerd met éézijdige t-toetsen voor onafhankelijke waarnemingen.

De CBS en ITS zijn ontworpen om houvast te bieden bij het coderen, terwijl de EBC dat niet doet. Er werd dan ook verondersteld dat de overeenstemming groter is bij de CBS en ITS dan bij de EBC. De CBS en ITS verschillen onderling zowel in de lengte van de beroepenlijst als in het aantal categorieën waarin gecodeerd wordt. De overeenstemming kan dus verschillen tussen deze twee beroepenclassificaties, maar het is niet a priori duidelijk wat het effect van de verschillen in lengte en aantal categorieën op de overeenstemming zal zijn; er werden daarom geen gerichte hypothesen over het verschil in overeenstemming geformuleerd. De hypothesen werden onderzocht met behulp van contrasten tussen kappa's.

Bij gelijste beroepen kan de codeur de code van het beroep opzoeken, terwijl dat bij niet-gelijste beroepen onmogelijk is. Er werd dan ook verondersteld dat de overeenstemming bij gelijste beroepen hoger is dan bij niet-gelijste beroepen. De grotere bekendheid van een beroep kan ertoe leiden dat codeurs meer hetzelfde beeld van een beroep hebben dan bij onbekende beroepen. Er werd daarom verondersteld dat de overeenstemming groter zal zijn bij bekende dan bij onbekende beroepen. Beroepen met extreem prestige hebben meer uitgesproken eigenschappen dan beroepen met middelmatig prestige. Dit kan ertoe leiden dat codeurs meer hetzelfde beeld hebben van beroepen met extreem prestige dan die met middelmatig prestige. Er

werd verondersteld dat de overeenstemming groter is bij beroepen met extreem prestige. Deze hypothesen werden onderzocht met behulp van contrasten tussen kappa's.

De kappa's werden berekend over verschillende groepen beroepen, bij voorbeeld: voor de vergelijking van ITS en CBS werden kappa's berekend over alle 38 beroepen; voor de vergelijking tussen gelijste en niet-gelijste beroepen werden per beroepenclassificatie (ITS of CBS) kappa's berekend over de 20 gelijste beroepen en kappa's over de 18 niet-gelijste beroepen. De variantie van de kappa's werd bepaald met de standaard knipmes ("jackknife") methode; zie Schouten (1985, par. 4.4.; 1986).

De verschillen tussen kappa's werden geanalyseerd met contrasten (Winer, 1962, p. 66).

Een contrast van n kappa's (K_i) is gedefinieerd als $c = \sum_{i=1}^n w_i k_i$, waarbij de som van de gewichten w_i nul is ($\sum_{i=1}^n w_i = 0$). Onder de nulhypothese dat het contrast nul is ($H_0: c = 0$) is het geschatte contrast bij benadering normaal verdeeld met gemiddelde 0 en variantie:

$$\text{var}(\hat{c}) = \sum_{i=1}^n w_i^2 \text{var}(\hat{K}_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n w_i w_j \text{cov}(\hat{K}_i, \hat{K}_j) \quad (1)$$

Dit betekent dat $z = \hat{c} / \sqrt{\text{var}(\hat{c})}$ bij benadering standaard normaal verdeeld is.

De variantie van elke kappa werd geschat met de standaard knipmesmethode; de covarianties zijn echter niet bekend. Deze covariantie-termen zijn echter in het algemeen groter of gelijk nul. Stel dat in één contrast vier kappa's samengenomen worden, bij voorbeeld: K (ITS, Gelijst) berekend over 18 gelijste beroepen gecodeerd met ITS, K (CBS, Gelijst) berekend over 18 gelijste beroepen gecodeerd met CBS, K (ITS, Niet-Gelijst) berekend over 20 niet-gelijste beroepen gecodeerd met ITS, en K (CBS, Niet-Gelijst) berekend over 20 niet-gelijste beroepen gecodeerd met CBS. Tussen deze vier kappa's bestaan zes verschillende covariantie-termen. De termen $\text{cov}\{K(\text{ITS, Gelijst}), K(\text{CBS, Niet-Gelijst})\}$ en $\text{cov}\{K(\text{ITS, Niet-Gelijst}), K(\text{CBS, Gelijst})\}$ zijn nul omdat de kappa's berekend zijn bij verschillende codeurs en verschillende beroepen. De termen $\text{cov}\{K(\text{ITS, Gelijst}), K(\text{CBS, Gelijst})\}$ en $\text{cov}\{K(\text{ITS, Niet-Gelijst}), K(\text{CBS, Niet-Gelijst})\}$, ($K(\text{CBS, Gelijst})$), zijn positief omdat ze weliswaar berekend zijn bij verschillende codeurs maar over dezelfde beroepen. De termen $\text{cov}\{K(\text{ITS, Gelijst}), K(\text{ITS, Niet-Gelijst})\}$ en $\text{cov}\{K(\text{CBS, Gelijst}), K(\text{CBS, Niet-Gelijst})\}$ zijn positief omdat ze weliswaar berekend zijn over verschillende beroepen maar bij dezelfde codeurs. Bij de hier gebruikte contrasten is het aannemelijk dat

$$\text{var}(\hat{c}) \leq \sum_{i=1}^n w_i^2 \text{var}(\hat{K}_i), \quad (2)$$

zodat $z = \hat{c} / \sqrt{\sum_{i=1}^n w_i^2 \text{var}(\hat{K}_i)}$ in het algemeen een conservatieve schatter van de z -waarde is.

3. RESULTATEN

Wat betreft de codeertijd werd verwacht dat deze hoger zou zijn bij de CBS dan bij de ITS en bij de ITS hoger dan bij de EBC. De gemiddelde codeertijden in minuten zijn: 163.3 (sd.: 29.1) voor de CBS, 38.8 (sd.: 25.2) voor de ITS en 29.2 (sd.: 10.2) voor de EBC. De éénzijdige t -toets tussen CBS en ITS is significant op 5% niveau ($t = 10.30$, $df = 10$), evenals die tussen ITS en EBC ($t = 2.04$, $df = 10$).

De tweede verwachting is dat de beroepenclassificaties CBS en ITS tot een hogere overeenstemming leiden dan de EBC. Hiertoe werd de gemiddelde kappa van CBS en ITS, berekend over alle 38 beroepen, vergeleken met die van de EBC. Het volgende contrast werd gehanteerd:

$$1/2 \{K (ITS) + K (CBS)\} - K (EBC).$$

De covariantie-termen in dit contrast zijn positief omdat ze berekend zijn bij dezelfde beroepen. Van de drie verschillende covariantie-termen hebben er twee een negatief teken zodat Formule 2 aannemelijk is. De kappa's staan vermeld in Tabel 2.

Tabel 2 Kappa's en tussen haakjes standaardafwijkingen berekend per beroepenclassificatie over verschillende groepen beroepen

Groep	ICO				CCO		
	ITS	CBS1	CBS	EBC	ITS	CBS1	CBS
38 beroepen	0,620 (0,062)	0,748 (0,048)	0,495 (0,053)	0,274 (0,049)	0,631 (0,069)	0,772 (0,054)	0,568 (0,056)
18 gelijste beroepen	0,786 (0,076)	0,793 (0,071)	0,582 (0,081)	0,320 (0,072)	0,854 (0,060)	0,819 (0,074)	0,655 (0,080)
20 niet-gelijste beroepen	0,451 (0,088)	0,700 (0,072)	0,416 (0,066)	0,240 (0,077)	0,409 (0,102)	0,726 (0,080)	0,491 (0,075)
19 bekende beroepen	0,694 (0,081)	0,802 (0,069)	0,557 (0,083)	0,289 (0,075)	0,641 (0,107)	0,874 (0,049)	0,675 (0,073)
19 onbekende beroepen	0,541 (0,091)	0,693 (0,069)	0,434 (0,066)	0,265 (0,068)	0,607 (0,095)	0,673 (0,090)	0,464 (0,079)
18 beroepen met extreem prestige	0,570 (0,089)	0,723 (0,078)	0,461 (0,076)	0,310 (0,066)	0,570 (0,103)	0,772 (0,077)	0,533 (0,083)
20 beroepen met niet-extreem prestige	0,665 (0,093)	0,764 (0,063)	0,523 (0,074)	0,224 (0,064)	0,695 (0,094)	0,762 (0,082)	0,597 (0,077)

De nulhypothese dat het contrast nul is werd voor de intercodeur-overeenstemming éénzijdig getoetst op 5% niveau. De nulhypothese werd verworpen ($z = 4,45$), zodat wordt geconcludeerd dat het gebruik van een empirisch gefundeerde beroepenclassificatie tot hogere intercodeur-overeenstemming leidt.

De ITS en CBS verschillen zowel in de lengte van de beroepenlijst als in het aantal categorieën waarin gecodeerd wordt. De grotere lengte van de beroepenlijst en het grotere aantal categorieën van de CBS kunnen tot grotere of kleinere nauwkeurigheid leiden bij het coderen. De verwachting dat CBS en ITS verschillen werd onderzocht met het contrast:

$$K (ITS) - K (CBS).$$

De ene covariantie-term in dit contrast is positief omdat dezelfde beroepen gecodeerd zijn maar heeft een negatief teken zodat Formule 2 aannemelijk is. De nulhypothese dat het contrast nul is werd tweezijdig getoetst op 5% niveau, zowel voor de intercodeur- als criterium-codeur-overeenstemming. De nulhypothese werd in geen van beide gevallen verworpen (ico: $z = 1,53$; cco: $z = 0,71$).

In het bovengenoemde contrast werd het gelijktijdige effect van de lengte van de beroepenlijst en het aantal categorieën onderzocht. Het effect van alleen de lengte van de beroepenlijst kan ook onderzocht worden. De CBS-classificatie heeft een viercijferige code met 914 categorieën. Het eerste cijfer van de code geeft een groepering in slechts elf categorieën, hetgeen vergelijkbaar is met de zeven categorieën van de ITS (voor beide beroepenclassificaties het totaal aantal categorieën en "niet te coderen"). Een vergelijking tussen de overeenstemming van de CBS in elf

categorieën (aangeduid als CBS1) en de ITS geeft een indruk van het effect van uitsluitend de lengte van de beroepenlijst. In het bovengenoemde contrast werd de overeenstemming voor de volledige CBS vervangen door die van de CBS1. De nulhypothese dat het contrast nul is werd tweezijdig getoetst op 5% niveau. In geen van beide gevallen werd de nulhypothese verworpen (ico: $z = -1,63$; cco: $z = -1,61$); bij een ongeveer gelijk aantal categorieën is er dus geen duidelijk effect van de lengte van de beroepenlijst.

Wat betreft Lijsting werd verondersteld dat de overeenstemming groter zal zijn bij de gelijste beroepen dan bij de niet-gelijste. De nulhypothese dat het contrast

$$\{K(\text{ITS}, \text{Gelijst}) + K(\text{CBS}, \text{Gelijst})\} - \{K(\text{ITS}, \text{Niet-Gelijst}) + K(\text{CBS}, \text{Niet-Gelijst})\}$$

nul is werd éénzijdig getoetst op 5% niveau. Twee covariantie-termen zijn nul en van de vier positieve termen hebben twee een positief teken en de andere twee een negatief teken; het lijkt dan ook redelijk te veronderstellen dat de z -waarde bij benadering correct geschat wordt. De toets dat het contrast nul is heeft alleen zin als er geen interactie is tussen Lijsting en Beroepen classificatie. Daartoe werd de nulhypothese dat het contrast

$$\{K(\text{ITS}, \text{Gelijst}) + K(\text{CBS}, \text{Niet-Gelijst})\} - \{K(\text{CBS}, \text{Gelijst}) + K(\text{ITS}, \text{Niet-Gelijst})\}$$

nul is tweezijdig op 5% niveau getoetst. Twee covariantie-termen zijn nul en de vier positieve covariantie-termen hebben een negatief teken, zodat Formule 2 van toepassing is. Het interactie-effect is in geen van beide gevallen significant (ico: $z = 1,08$; cco: $z = 1,74$). Het hoofd-effect van Lijsting is in beide gevallen significant (ico: $z = 3,20$; cco: $z = 3,75$), m.a.w. gelijste beroepen leiden tot grotere overeenstemming dan niet-gelijste beroepen.

De effecten van de onafhankelijke variabelen Beroepsprestige en Bekendheid werden op dezelfde wijze onderzocht als dat van Lijsting. Wat betreft Beroepsprestige werd verwacht dat de overeenstemming groter zal zijn bij de codering van beroepen met een extreem prestige dan die met een niet-extreem prestige. De interactie van Beroepsprestige met Beroepen classificatie is in geen van beide gevallen significant (5% tweezijdig; ico: $z = -0,20$; cco: $z = -0,34$). Het hoofdeffect van Beroepsprestige is in geen van beide gevallen significant (5% éénzijdig; ico: $z = -0,94$; cco: $z = -1,05$), m.a.w. er is geen verschil in overeenstemming tussen beroepen met een extreem en niet-extreem prestige.

Wat betreft Bekendheid werd verwacht dat de overeenstemming groter zal zijn bij bekende beroepen dan bij onbekende. De interactie van Bekendheid met Beroepen classificatie is in beide gevallen niet significant (5% tweezijdig; ico: $z = 0,19$; cco: $z = -0,99$). Het hoofdeffect van Bekendheid is voor de intercodeurovereenstemming significant (5% éénzijdig; ico: $z = 1,71$) en is voor de criterium-codeur-overeenstemming niet significant (5% éénzijdig; cco: $z = 1,37$).

Verder werd nagegaan in hoeverre de beroepen classificaties gevoelig zijn voor variaties in beroepskenmerken. Daartoe werden de kappa's berekend per cel van de proefopzet, d. w. z. acht kappa's per beroepen classificatie; het aantal beroepen per cel varieert van 3 tot 8. De gemiddelden en standaardafwijkingen staan in Tabel 3.

De gemiddelden stemmen sterk overeen met de kappa's berekend over alle 38 beroepen per beroepen classificatie (zie Tabel 2). Uit de standaard afwijkingen blijkt dat vooral de ITS-classificatie gevoelig is voor variaties in de beroepskenmerken.

De overeenstemming is in het algemeen aan de lage kant; de kappa's, van alle classificaties berekend over 38 beroepen zijn altijd lager dan 0,80; ook de kappa's berekend over de 18 gelijste beroepen zijn verre van perfect (zie Tabel 2). De kappa's geven weer de overeenstemming tussen codeurs of tussen codeurs en criterium, maar zij geven geen inzicht in de aard en de grootte van de discrepanties. Om hierin enig inzicht te verkrijgen werden de criterium-codeur gegevens, van de ITS en CBS1 nader geanalyseerd. Aangezien er een significant effect is van Lijsting op de overeenstemming werd de analyse afzonderlijk gedaan voor de gelijste en de niet-gelijste beroepen. Voor elke codeur-codering die niet overeenstemt met de criterium-codering werd de richting en de grootte van het verschil bepaald; hierbij werd de categorie "niet te coderen"

Tabel 3 Gemiddelde en standaardafwijking per beroepenclassificatie berekend over de acht kappa's van de proefopzet

Beroepenclassificatie	ICO		CCO	
	Gem	Sd	Gem	Sd
ITS	0,640	0,208	0,653	0,234
CBS1	0,752	0,114	0,776	0,160
CBS	0,490	0,128	0,570	0,155
EBC	0,275	0,158	—	—

buiten beschouwing gelaten. Bijvoorbeeld stel dat een beroep met een criterium - codering van 3 door een codeur gecodeerd wordt als 2; de codeur-codering is dan één stap lager dan de criterium-codering. Als een andere codeur dit beroep een 5 toekent dan is de codeur-codering twee stappen hoger. De gegevens zijn vermeld in Tabel 4.

Tabel 4 Frequenties van aantal stappen dat codeur-codering hoger of lager is dan criterium-codering bij niet-overeenstemmende coderingen, 20 gelijste en 18 niet-gelijste beroepen gecodeerd met CBSI en ITS.

Aantal stappen	CBS1				ITS			
	Gelijst		Niet-Gelijst		Gelijst		Niet-Gelijst	
	Hoger	Lager	Hoger	Lager	Hoger	Lager	Hoger	Lager
1	0	1	6	7	3	3	7	18
2	1	1	1	5	0	3	4	18
3	0	8	0	3	0	1	3	5
4	0	0	0	1	0	0	0	0
5	0	0	0	4	0	0	0	0
6	0	0	0	1	-	-	-	-
7	0	0	0	0	-	-	-	-
8	0	0	1	0	-	-	-	-
9	1	0	1	0	-	-	-	-
Totaal	2	10	9	21	3	7	14	41

Opn.: Het maximale aantal stappen verschil bij de ITS is 5; een streepje duidt aan dat het aantal stappen niet voor kan komen.

De overeenstemming is bij gelijste beroepen groter dan bij niet-gelijste beroepen, uit Tabel 4 blijkt dat de discrepanties bij de gelijste beroepen ook geringer zijn. De tabel toont ook dat de discrepanties aanzienlijk kunnen zijn: op het totaal van 107 niet overeenstemmende coderingen hebben 62 twee of meer stappen verschil met het criterium. Tenslotte blijkt uit de tabel dat de codeurs de neiging hebben lagere coderingen toe te kennen: in totaal zijn 79 coderingen lager dan het criterium en 28 hoger.

4. CONCLUSIES

De vergelijking van de ITS- en CBS-classificaties met de EBC maakt duidelijk dat het gebruik van deze classificaties de overeenstemming doet toenemen.

De overeenstemming neemt toe bij beroepen die op de beroepenlijst staan in vergelijking met beroepen die daar niet op voorkomen. De intercodeurovereenstemming is hoger bij bekende beroepen dan bij onbekende. De extremititeit van het beroepsprestige blijkt niet veel effect te hebben op overeenstemming.

De vergelijking van de ééncijferige CBS-codering (CBS1) met de viercijferige (CBS) doet vermoeden dat een toename in het aantal categorieën de overeenstemming doet afnemen. Deze hypothese kon niet getoetst worden omdat CBS1 en CBS sterk afhankelijk zijn: een codeerfout in het eerste cijfer werkt door in de volledige CBS-codering.

Onervaren codeurs werken zeer snel met de ITS-classificatie: de gemiddelde codeertijd is minder dan 10 minuten hoger voor de ITS dan voor de EBC. De CBS kost echter ongeveer viermaal zoveel tijd als de ITS. De ITS is echter gevoeliger voor beroepskenmerken dan de CBS.

De algemene conclusie is dat het gebruik van een beroepenclassificatie nodig is maar dat codering, althans door onervaren codeurs, veel te wensen overlaat. Een verbetering in dit opzicht kan verwacht worden van het gebruik van automatische procedures voor codering via (micro) computers. In dit verband is ook het voorstel van Saris en De Pijper (1986) van belang om via een computer respondenten gerichte vragen te stellen over hun beroepswerkzaamheden en met behulp van hun antwoorden het beroep automatisch te coderen.

LITERATUUR

- CBS (1982a). *Schoolloopbaan en herkomst van leerlingen bij het voorgezet onderwijs. Deel 2. cohort 77, schoolkeuze*. 's-Gravenhage: Staatsdrukkerij.
- CBS (1982b). *Arbeidskrachtentelling 1979*. 's-Gravenhage: Centraal Bureau voor de Statistiek.
- CBS (1985). *Beroepenclassificatie 1984*. 's-Gravenhage: Centraal Bureau voor de Statistiek.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Elsevier (1983). *Beroepenalbum*. Amsterdam: Elsevier.
- Martens, A. (1981). *Nieuwe gids voor school en beroep*. Haarlem: De Toorts.
- Pompe, J. H., & Rutgers, A. (1984). Het meten van beroepsprestige; Enige kanttekeningen bij een beroepsprestigeschaal voor de jaren tachtig. *Mens en Maatschappij*, 59, 292-300.
- Popping, R. (1983). *Overeenstemmingsmaten voor nominale data*. Ongepubliceerd proefschrift Rijksuniversiteit Groningen.
- Popping, R. (1986). Opnieuw beroepsprestige-onderzoek: Een onderzoeksantekening. *Sociologische Gids*, 33, 29-40.
- Saris, W. E. & Pijper, M. de (1986). Computer assisted interviewing using homecomputers. *European Marketing Research*, 14, 144-150.
- Schouten, H. J. A. (1985). *Statistical measurement of interobserver agreement* Ongepubliceerd proefschrift, Erasmus Universiteit Rotterdam.
- Schouten, H. J. A. (1986). Nominal scale agreement among observers. *Psychometrika*, 51, 453-466.
- Sixma, H., & Ultee, W. (1983). Een beroepsprestigeschaal voor Nederland in de jaren tachtig. *Mens en Maatschappij*, 58, 360-382.
- Swanborn, P. G. (1978). De kunst van het meten, of de operationalisering van beroepsprestige. In J. L. Peschar & W. C. Ultee (Red.), *Sociale stratificatie: Op weg naar empirisch-theoretisch stratificatieonderzoek in Nederland*, 40-64. (Boekaflevering Mens en Maatschappij). Deventer: Van Loghum Slaterus.
- Westerlaak, J. M. van, Kropman, J. A., & Collaris, J. W. M. (1975). *Beroepenklapper* Nijmegen; Instituut voor Toegepaste Sociologie.
- Wiegersma, S. (1976). *De wereld van beroepen*. Haarlem: De Toorts.
- Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Manuscript ontvangen 19-6-1987.

Definitieve versie ontvangen 28-10-1987.

The Influence of a Teacher Training on Student Perception of Classroom Climate*

Pieter C. van der Sijde
University of Twente, The Netherlands

ABSTRACT

Classroom climate can be operationalized in different ways and it appeared from many studies that classroom climate affects the student learning outcomes. In this study the effect of a short teacher training to enhance student learning on the classroom climate is investigated. Two conditions are compared: in the first condition the teachers participated in the training at the beginning of the schoolyear, while the teachers in the second condition did so in January. It appeared from the results that the topics which were central in the training (classroom management and instruction) were not perceived by the students as being different for the teachers in both conditions. However, a difference in the perception of the relationships was found in favor of the teachers in the first condition.

INTRODUCTION

It was Lewin who first formulated the well-known formula, $B = f(P, E)$, which says that Behavior is function of the Person and the Environment (Stern, 1964). An implication of this formula is that if the environment changes, the behavior changes. The classroom environment has a physical (e.g. class-size, equipment etc.), a social (e.g. SES of the students) and a psychological aspect. In this study we will restrict ourself to the classroom psychological environment, which is defined by Walberg (1987) as the climate or atmosphere of the class as a social group. This aspect is referred to as the classroom climate. The classroom climate influences the study behavior of the students; this has been experimentally confirmed by a number of studies (e.g. O' Reilly, 1975; Cort, 1979; Kelly, 1980; Walberg, Singh & Rasher, 1977). The teacher is part of learning environment as perceived by the students and as such his teaching behavior affects the students' perception of the classroom climate.

A large number of instruments have been developed to measure characteristics of learning environments. The most widely used instruments for education at the secondary level are the Learning Environment Inventory (LEI; Fraser, Anderson & Walberg, 1982), the Classroom Environment Scale (CES; Moos & Trickett, 1974) and the Individualized Classroom Environment Questionnaire (ICEQ; Fraser, 1985). Although classroom climate can be operationalized in many different ways (which also becomes visible in the labels of the scales of the three instruments mentioned above), the underlying structure of the instruments is the same. They all start out from the framework formulated by Moos (e.g. Moos, 1979).

From extensive research into the functioning of diverse environments (hospital wards, prisons, military companies etc.) conducted by Moos and his coworkers it appeared that the same three general categories can be used in conceptualizing the dimensions characterizing psychosocial environments. These three dimensions are:

1. *the relationship dimension*: this dimension identifies the nature and intensity of personal relationships within the environment; it assesses the extent to which people are involved in the environment and the extent to which they support and help each other.
2. *The personal development division*: this dimension assesses the basic directions along which personal growth and self-enhancement tend to occur.
3. *the system maintenance and system change dimension*: this dimension involves the extent to

* Deze studie is verricht in het kader van het Leeromgevingsonderzoek, gesubsidiëerd door het Instituut voor Onderzoek van het Onderwijs, SVO (SVO projectnummer 1036).

Instrument	Items per scale	Relationship dimension	Personal development dimension	System maintenance & change dimension
Learning Environment Inventory (LEI)	7	Cohesiveness Friction Favoritism Cliquesness Satisfaction Apathy	Speed Difficulty Competitiveness	Diversity Formality Material Environment Goal Direction Disorganization Democracy
Classroom Environment Scale (CES)	10	Involvement Affiliation Teacher Support	Task Orientation Competition	Order and Organization Rule Clarity Teacher Control Innovation
Individualized Classroom Environment Questionnaire (ICEQ)	10	Personalization Participation	Independence Investigation	Differentiation

Figure 1. Structure of and scales in the three classroom climate instruments.

which the environment is orderly, clear in expectations, maintains control and is responsive to change.

These dimensions form the underlying structure of the instruments for classroom climate which were mentioned above.

There is a growing number of studies in which measures for classroom climate are used to evaluate curricular and instructional innovations (see e.g. Walberg, 1975; Fraser, 1986, for an overview and summary of those studies). In the Netherlands only a few of these studies have been carried out (Kuhlemeier, 1983; Veenman, 1975; Wierstra, 1984). Kuhlemeier (1983) and Wierstra (1984) evaluated the PLON ('Project Leerplan Ontwikkeling Natuurkunde') curriculum using an adapted version of the ICEQ: it appeared that a new physics curriculum caused a more favorable perception of e.g. participation and differentiation, but not of relationship dimensions, such as helpfulness and friendliness. Veenman (1975) investigated as part of his study the influence of a teacher training on the interaction aspects of teaching behavior (Flanders, 1970) on classroom climate as measured by the Minnesota Pupil Attitude Inventory (MPAI). He hypothesized that in the classrooms of the teachers trained on the basis of interaction-analysis a more positive classroom climate could be created. Veenman could not confirm his hypothesis.

In the present study a measure of classroom climate is used to evaluate the effect of a one-day teacher training. This teacher training is based on a teaching script (described in detail Van der Sijde, 1987) in which the organization and execution of the lesson is reflected as the product of a teacher thinking process. The training centers around a new teaching script and recommendations on instruction and classroom management, which are based on the results of the Dutch Classroom Environment Study - correlational phase (Krammer, 1984; Tomic, 1985). The content of the training is organized around three topics:

- the format of the lesson (the script)
- management of the classroom
- instruction (the subjects form a conceptual whole).

The content of the training is written down in a 40-page manual entitled "Management en instructie in wiskundelessen" ("Management and instruction in mathematics lessons") (see Table 1).

Table 1

A summary of instructional and managerial rules

Lesson phase model: phase 1	Review of homework	(5 à 10 min)
phase 2	Presentation	(10 à 15 min)
phase 3	Monitored practice	(± 5 min)
phase 4	Guided practice	(±10 min)
phase 5	Homework/tutoring	(±10 min)

Classroom management

- Use a beginning-of-the lesson signal.
- Make transitions between two lesson phases as smooth as possible.
- Make sure that there is a continuous stream of activities in the classroom.
- Be withit.
- Maintain group focus.
- Let students be responsible for their work.

Instruction

- Discuss all homework assignments.
- Start presentation with a review of relevant concepts.
- Refresh prior knowledge through questioning.
- Use, during presentation, concrete examples.
- Indicate (un)important passages in the textbook.
- Indicate what is important (oral/blackboard).
- Give clear assignments for monitored practice.
- React positive to students.
- Check the assignments during monitored practice.
- Redirect questions and use probes.
- At the end of the lesson give time to students to work on their homework assignments
- Tutor at the end of the lesson students who encountered problems.

The intent of the teacher training is twofold. First to change the teachers teaching script. Second as a consequence to change the actual teaching behavior. The underlying assumption of a research conducted according to the descriptive-correlational-experimental loop (Rosenshine & Furst, 1973) and of which this study is a part, is that only visible teacher behavior has an impact on student learning. It appeared from results reported in Van der Sijde (1987) that the training had no effect on students' attitude, but it did have an effect on students' achievement and on the teachers' behavior. There were a number of important differences in the teacher performance. For a detailed analysis of the differences in teaching behavior we refer to Van der Sijde (1987). If it also has an effect on student perception of classroom climate, it should become visible on those dimensions of an instrument for classroom climate which can be characterized as 'instruction' - and 'management' - related.

METHOD

Subjects

In this study the classes (second grade Havo and Vwo) of 20 mathematics teachers (mean number of years of experience 5.05) using the textbook Sigma participated. The mean number of students per class was approximately 28 students. Twelve teachers participated in a teacher training at the beginning of the schoolyear (condition 1) and eight participated in the training after the Christmas holidays (condition 2).

The Classroom Environment Scale

The Classroom Environment Scale has been selected as the instrument to measure classroom

climate, because the scales in this instrument can, better than in the other two instruments (see Figure 1), be characterized as 'instruction'- and 'management'-related.

The Classroom Environment Scale (CES) is one of a number of social-psychological climate instruments developed by Moos and his colleagues. It is a 90-item true-false questionnaire containing 9 scales divided over three dimensions:

- Relationship dimension: involvement, affiliation and teacher support;
- Personal development dimension: task orientation and competition;
- System dimension: order and organization, rule clarity, teacher control and innovation.

The CES has been translated into Dutch (Van der Sijde, 1986). From a principal component analysis followed by a Varimax rotation carried out on the data sampled in the present study it appeared that six factors, of which each exceeded an eigenvalue of 1.00, accounted for 40.5% of the variance. Almost all factors identified by Trickett and Quinlan (1979) in their analysis ('rule emphasis'; 'order and organization'; 'friendly and easygoing teacher'; 'innovative, student-oriented teaching approach'; 'student competition'; 'student affiliation') could also be identified in the Dutch CES (see Table 2). There are two exceptions. The first concerns the factor 'friendly and easygoing teacher' of Trickett and Quinlan, which is the factor 'friendly teacher' in the Dutch sample. The second exception concerns the factor 'student competition' which could not be identified; in the Dutch sample a factor 'teacher support' could be labelled. In this analysis all 90 items are used and each factor is considered as a Likert-scale. (scale one contains 23 items, scale two 13 items, scale three 16 items, scale four 22 items, scale five 7 items and scale six 9 items). Nevertheless the results of this factor analysis support the original conception of the CES: the system maintenance dimension is presented by 'rule emphasis', and, 'order and organization' the relationship dimension by the 'friendly teacher', 'teacher support' 'student affiliation' and the personal development dimension is represented by 'innovative student-oriented teaching'.

Table 2

The scales of the Dutch CES with items to illustrate them

Scale	Examples
Order & Organization	• This class will often be in an uproar • The teacher will often have to tell students to calm down
Friendly teacher	• The teacher will be more like a friend than an authority • The teacher will often take time out from the lesson plan to talk about other things
Rule emphasis	• There will be a clear set of rules for students to follow • If a student breaks a rule in this class he will sure get into trouble
Teacher support	• Sometimes the teacher will embarrass students for not knowing the answer • This teacher will not thrust students
Student affiliation	• There are groups who won't get along in class • Some students in this class won't like each other
Innovative student-oriented teaching approach	• It will be easy to get a group together for a project • Students will sometimes present something they have worked on to the class

Procedure

There are two conditions in this experiment. The teachers in the first condition participated in the training in September. After the training they were systematically observed in the period September-December. The teachers in the second condition were also systematically observed in the period September-December, but they participated in the training in January. During the one-day training session the manual and the recommendations were discussed with the teachers in the light of their own experiences. In the training given in January the results of the systematic observation were used to give the teachers individual feedback on their teaching behavior. In the period January-April no (systematic) observation took place. The CES was administered in April.

Design and data

The design used in this study is depicted in Figure 2.

To investigate the classroom climate, the mean scores on the scales of the Dutch CES are considered as class-scores: the level of analysis is the classroom and not the individual student. Because the focus of the training is on the subjects classroom management and classroom instruction, which are also to be found in the CES, a score for each of these subjects was calculated. The score for classroom management is a combination of the scales 'order and organization' and 'rule emphasis', while the score for classroom instruction is a combination of the scales 'teacher support' and 'innovative, student-oriented teaching approach'. The other two scales 'student affiliation' and 'friendly teacher' are combined into a score for relationship. To analyse the difference between the conditions the Mann-Whitney U-test was applied.

X_1 O

X_2 O

(X_1 = training followed by observations; X_2 = index observations followed by a training plus individual feedback; O administration of CES).

Figure 2. Design of the study

RESULTS

There is a significant difference in the perception of classroom climate between the students in the classrooms of the teachers in condition one and those in condition two on one scale of the Dutch CES: student affiliation. There are no other differences in the perceptions of the students. There are no differences in the perception of classroom management, nor of classroom instruction. There is a significant difference in the perception of relationships in general (see Table 3).

DISCUSSION

On one scale of the Classroom Environment Scale there is a difference in the perception of classroom climate by the students. From the perceptions of the students (as depicted in Table 3) it appears that they do not perceive the teachers in both conditions to be different in paying more attention to 'order and organization', 'rule emphasis', 'teacher support' and 'innovative student-oriented teaching approach' (which was expected), but they rather perceive differences in the relationships among themselves, and to a somewhat lesser degree with the teacher: the teachers in the first condition are perceived as friendlier, which could mean that they take more (real) interest in the students than the teachers in the second condition. The goal of the teacher

Table 3

The mean rank of the class-scores of the two groups of teachers and the significance of the difference on the six scales of the Dutch CES, and on the dimensions instruction, classroom management, and relationship

	mean rank condition 1	mean rank condition 2	p*
Scale 1: Order & Organization	10.10	9.89	.93
Scale 2: Friendly teacher	8.50	11.67	.22
Scale 3: Rule emphasis	11.50	8.33	.22
Scale 4: Teacher support	10.20	9.78	.87
Scale 5: Student affiliation	6.30	14.11	.01
Scale 6: Innovative student-oriented teaching approach	9.10	11.00	.46
Instruction	9.60	10.44	.74
Management	10.40	9.56	.74
Relationship	7.50	12.78	.04

* p-value calculated with the Mann-Whitney test

training was to make mathematics teaching more effective, to enhance student achievement, and to create a more positive attitude towards mathematics. Or to state it differently: the goal of the training was to change the teacher's teaching script and, through such a change in conception, his actual teaching behavior concerning classroom management and instruction. The students did not perceive the classroom climate to be different with reference to instruction and classroom management; they did perceive differently concerning the relationships. A comparison of the results of this study with the results on student achievement (Van der Sijde, 1987) shows that the most effective training condition is the one with the tailored training, the training in which the teachers received individual feedback on the observed teaching behavior. This study indicates that the classes of the teachers in condition one were perceived to have better relationships.

At this point a question should be raised, namely whether perceptions can be changed and how difficult this is. How persistent can a once-formed perception be? The student perception of classroom climate is based on many observations and when a certain perception of classroom climate has been formed, how penetrative must a training be to change an already-performed perception? Two answers are possible.

First, the perception of classroom climate is formed during the first months of the school year. Once a perception has been formed, it is difficult to change it (pro-active inhibition). The students in the classroom of the teachers in condition one have formed their perception of the classroom climate after the teachers had taken part in the training, whereas the students of the teachers in condition two started to form their perception of the classroom climate when their teachers had not participated in the training at that moment.

Second, it takes quite some time to form a perception of the classroom climate. The teachers in condition one were able to behave according to the teaching script presented to them in the training for almost a whole school year (consistent behavior). The teachers in the second condition, however, behaved differently during the first half and the second half of the school year. The average perception during a whole school year is different for the teachers in both conditions.

It appears from this study that the teachers in condition one were able to build up better relationships with and among their students than the teachers in condition two. It is of importance that a particular classroom procedure (the consistent execution of a certain script) is carried out over a longer period of time to affect classroom climate.

In conclusion, a few remarks are made. A questionnaire on classroom climate such as the Classroom Environment Scale gives information that is supplementary to the more 'classic' product measures. Student perceptions of classroom climate are high inference measures based

on their experiences as participants in the particular setting, while systematic classroom observations by trained observers (low inference measures) are observations based on predetermined categories of behavior. Students perceived better relationships among themselves and with their teacher, no differences were found with respect to their perception of instruction and classroom management.

REFERENCES

- Borg, W. (1975). *Instructor handbook for protocol modules classroom management*. Logan: Utah State University.
- Cort, H.R. (1979). A social study evaluation. In Walberg, H.J. (ed), *Educational environments and effects: evaluation, policy and productivity*. Berkeley: Mc. Crutchan.
- Duffy, G.G., & Roehler, L.R. (1985). *Constraints on teacher change*. East Lansing: IRT.
- Flanders, N.A. (1970). *Analyzing teaching behavior*. Reading: Addison-West.
- Fraser, B.J., Anderson, G.J., & Walberg, H.J. (1982). *Assessment of Learning. Environments: Manual for Learning Environment Inventory and My Class Inventory*. Perth: WAIT.
- Fraser, B.J. (1985). *Individualized Classroom Environment Questionnaire*. Melbourne: Australian Council for Educational Research.
- Fraser, B.J. (1986). *Classroom environment*. London: Crook Helm.
- Kelly, A. (1980). Exploration and authority in science learning environments an international study. *European Journal of Science Education*, 2, 161-174.
- Kounin, J.S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Krammer, H.P.M. (1984). *Leraar en leerboek*. Harlingen: Flevodruk
- Kuhlemeier, H. (1983). Vergelijkend onderzoek naar de perceptie van de leeromgeving in PLON-onderwijs en regulier natuurkunde-onderwijs. *Tijdschrift voor Onderwijsresearch*, 8, 1-16.
- Moos, R.H., & Trickett, E.J. (1974). *Manual Classroom environment scale*. Palo Alto: Consulting Psychologists Press, Inc.
- O'Reilly, R.O. (1975). Classroom climate and achievement in secondary school mathematics classes. *Alberta Journal of Educational Research*, 21, 241-248.
- Rantoul, A.J., & Fraser, B.J. (1979). Conceptualizing of enquiry-based or open classroom learning environments. *Journal of Curriculum Studies*, 11, 233-245.
- Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers, (Ed.), *Second handbook for research on teaching*. Chicago: Rand McNalley.
- Stern, G.G. (1964). *Bf(P, E)*. *Journal of Personality Assessment*, 28, 161-168.
- Tomic, W. (1984). *Docentgedrag en leerprestatie*. Enschede: Universiteit Twente (dissertatie).
- Trickett, E.J., & Quinlan, D.M. (1979). Three domains of classroom environment: factor analysis of the Classroom Environment Scale. *American Journal of Community Psychology*, 7, 279-291.
- Van der Sijde, P.C. (1985). *Management en instructie in wiskundelessen*. Enschede: Universiteit Twente.
- Van der Sijde, P.C. (1986). *Leeromgevingsvragenlijst*. Enschede: Universiteit Twente, IST-MEMO-86-10.
- Van der Sijde, P.C. (1987). *Tranining the teaching script*. Almere: Versluys (dissertatie Universiteit Twente).
- Veenman, S.A.M. (1975). *Training op basis van interactieanalyse*. Tilburg: Zwijsen.
- Walberg, H.J. (1975). Educational process analysis. In M.W. Apple, M.J. Subkoviak & H.S. Lufner, (Eds.), *Educational evaluation analysis and responsibility*. Berkeley: Mc. Crutchan.
- Walberg, H.J. (1987). Psychological environment. In M.J. Dunkin, (Ed.), 1987, *The international encyclopedia of teacher education*. Oxford: Pergamon.
- Wierstra, R. (1984). A study on classroom environment and on cognitive and affective outcomes of the PLON-curriculum. *Studies in Educational Evaluation*, 10, 273-282.

Manuscript ontvangen 30-1-1987

Definitie versie ontvangen 2-2-1988

Notities en Commentaren

AANTEKENINGEN BIJ 'ANATOMIE VAN EEN LEEROMGEVING, EEN ONDERWIJS-ECONOMISCHE ANALYSE VAN UNIVERSITAIR ONDERWIJS', K.J.D.M. VAN DER DRIFT EN P. VOS (DISSERTATIE LEIDEN, 1987).

J.W. Holleman

Afdeling Onderzoek en Ontwikkeling van Onderwijs, Rijksuniversiteit Utrecht, Postbus 80125, 3508 TC Utrecht

Hfst.1. De auteurs onderscheiden twee produktieprocessen in het onderwijs (1.3), elk met eigen produktiemiddelen en een eigen produkt: het instructieproces van de docenten (gericht op het creëren van een leeromgeving) en het leerproces van de student (gericht op het creëren van gedragsverandering). Produktietheoretisch gezien, is er nog een derde proces: de diplomering (het examineren van het leerresultaat in het licht van de eisen die aan het diploma zijn verbonden). Zowel de examencommissie als de examinandus moeten daartoe produktiemiddelen inzetten.

Hfst.2. De auteurs geven een empirische beschrijving van patronen van tijdsbesteding van studenten binnen onderdelen van het universitaire studieprogramma. De geleverde studie-inspanning (geoperationaliseerd in wekelijks bestede studie-uren) wordt uitgezet op een tijd-as van 52 weken. Het is verwarrend dat deze *feitelijk geleverde* inspanning wordt aangeduid als studielast. Met dat begrip verwijst men gewoonlijk naar de *benodigde* tijd of inspanning: hoeveel tijd heeft een student nodig om te slagen? In de gerapporteerde gegevens over de geleverde studie-inspanning is geen onderscheid gemaakt tussen geslaagde en gezakte studenten.

De gerapporteerde patronen betreffen de wekelijkse rekenkundige gemiddelden, berekend over de groep studenten die aan het programma deelnemen. Het analyse-object van de auteurs is echter niet de tijdsbesteding van studentengroepen doch van individuele studenten. De auteurs melden niet of de gerapporteerde patronen inderdaad bij een modale student kunnen worden aangetroffen: waren de onderzochte studentengroepen voldoende homogeen wat de individuele studeerpatronen betreft?

Hfst.3. In 3.6 memoreren de auteurs dat de studielast van een curriculum meestal wordt uitgedrukt in een jaarlijks aantal uren, waarbij zij waarschijnlijk denken aan de wettelijke normatieve studielast van 1700 uur per cursusjaar voor de mediane student. Maar zij behandelen vervolgens een ander onderwerp: is er een leeromgeving denkbaar waarin de studenten geneigd en in staat zijn zoveel inspanning te leveren dat het rekenkundig gemiddelde van de feitelijk per studiejaar geleverde inspanning binnen de studentenjaargroep 1700 uur benadert? De keuze van het rekenkundig gemiddelde ligt in deze probleemstelling niet erg voor de hand, aangezien die maat van centrale tendentie zo gevoelig is voor de scores in de hoogste en laagste decielen. Als men de ijver binnen een studentenjaargroep wil verhogen, lijkt de mediaan een nuttiger maat: is er een leeromgeving denkbaar waarin de studenten geneigd en in staat zijn

zoveel inspanning te leveren dat tenminste 50% van de groep jaarlijks 1700 of meer uren aan de studie besteedt?

Hfst.4. Dan worden gegevens gerapporteerd over de feitelijke studielast van de gemiddelde student per cursusjaar, en over het verband tussen onderwijstijd en zelfstudietijd. Jammergenoeg betreffen deze gegevens uitsluitend diegenen die de propedeuse (of het tweede cursusjaar) in één jaar met succes hebben doorlopen. Het is denkbaar dat de mediane studielast per cursusjaar veel hoger uitvalt als men ook kijkt naar de studenten die méér dan één jaar nodig hebben gehad om het cursusjaar met succes te voltooien.

De auteurs rapporteren niets over de grootte en representativiteit van hun steekproeven ten opzichte van het universum van alle studenten die aan de onderzochte jaarprogramma's deelnamen. Daardoor blijven hun conclusies in de lucht hangen.

Hfst.5. De auteurs trachten vervolgens te verklaren waarom de gemiddelde studielast van de groep studenten die de propedeuse in één jaar met succes hebben doorlopen niet meer dan 1300 uur netto bedraagt. Zij concluderen dat een hogere belasting maatschappelijk niet aanvaardbaar wordt geacht (voor werknemers en studenten). Daarmee gaan zij voorbij aan het feit dat de gemiddelde studielast van een groep geslaagde studenten niet alleen wordt bepaald door de belastbaarheid van de gemiddelde student maar bijvoorbeeld ook door de spreiding van de individuele studielast binnen de studentenpopulatie. Zo zullen biologiestudenten die het vak Biologie niet in hun VWO-eindexamenpakket hadden opgenomen (naast de voor hun toelating verplichte Natuurkunde en Scheikunde en de aanbevolen Wiskunde) een relatief grote studielast hebben in de propedeuse. Als de studieleiding een programma wil aanbieden dat ook voor hen nog enigszins uitvoerbaar is in de cursusduur, zal zij met een relatief lage gemiddelde studielast per cursusjaar genoegen moeten nemen.

Hfst.6. Hoe kan de studieleiding zonder aanwending van extra produktiemiddelen de jaarlijks geleverde studie-inspanning doen stijgen? Bij het beantwoorden van deze probleemstelling (6.2) besteden de auteurs geen aandacht aan het verschijnsel van gedwongen onderbelasting en gedwongen leegloop. Volgens hen wordt de ijver vooral bepaald door de mate van gedragscontrole die in het onderwijs- en examenprogramma is ingebouwd (2.6). Zij gaan voorbij aan Carroll's 'opportunity factor': de studieleiding kan, om te beginnen, de in het rooster ingebouwde belemmeringen tot het leveren van studie-inspanning wegnemen.

De auteurs betwijfelen of het verstandig is om herkansingstentamens tijdens de onderwijsperiodes te programmeren, omdat de voorbereiding daarvan interfereert met de studie-inzet voor de andere vakken. Daar stellen zij de hypothese tegenover dat uitstel van een herkansing de totale studielast van het desbetreffende vak zal verhogen. Bij het post-hoc experiment ter falsificering van deze hypothese (6.2.3) wordt jammergenoeg niet gecontroleerd op de hoeveelheid leerwinst die reeds behaald was voor de eerste tentamenpoging (bv. te operationaliseren als het gemiddelde tentamencijfer van de gezakten).

Hoofdstuk 6 overstijgt in het algemeen nauwelijks het niveau van kabbelende borrelpraat. De auteurs stellen dat de sociale wetenschappen tot nu toe geen gegeneraliseerde aanwijzingen hebben opgeleverd voor de verbetering van de interne doelmatigheid van onderwijzen en leren, maar zij adstrueren dat niet. Dan volgt hun eigen theorie: het resultaat van instructie is gelegen in de geleverde inspanning die per eenheid 'elapsed time' wordt uitgelokt bij de studenten. Uiteindelijk komen zij tot de conclusie dat hun benadering toch wat te mager is: instructie kan er ook toe bijdragen dat studenten meer leerwinst behalen per eenheid geleverde studie-inspanning. Hiermee hebben zij het wiel van Carroll (1963) opnieuw uitgevonden: het gaat niet alleen om 'time spent' maar ook om 'time needed', dus om de arbeidsproductiviteit (studieproductiviteit) van studenten, gerekend per netto besteed uur.

Hfst.7. In 7.3 stellen de auteurs een produktiefunctie voor het leerproces op. De student moet twee schaarse produktiemiddelen mobiliseren om leerwinst te verwerven: bestede studietijd en

begeleiding bij het studieproces. Waarom wordt de hoeveelheid geleverde studie-inspanning per studie-uur (de studie-intensiteit) niet in de produktiefunctie betrokken? Of, fundamenteeler gesproken, waarom besteden de auteurs bij de uitwerking van de produktiefunctie alleen aandacht aan de hoeveelheid en niet aan de kwaliteit van deze produktiemiddelen?

Hfst.9. In figuur 9.2 worden de factoren samengevat die voor studenten van belang zijn bij hun studieplanning. Het theoretische referentiekader dat daaraan ten grondslag ligt, vertoont weinig systematiek. *Voorkennis* en *streefniveau* kunnen beter worden vervangen door de te behalen leerwinst; *leerstijl* en *intelligentie* beter door de studieproductiviteit (leerwinst per eenheid inspanning). Verder wordt de *tijdbehoefte* niet rechtstreeks bepaald door de zojuist genoemde factoren, maar door de benodigde studie-inspanning in combinatie met de studie-intensiteit (geleverde inspanning per studie-uur). En behalve met hun *beschikbare tijdbudget*, moeten studenten rekening houden met de 'opportunity factor': krijgen zij gelegenheid hun beschikbare tijd inderdaad te vullen met nuttige studie-activiteiten?

Hfst.8/10. De auteurs geven in hun onderwijs-economische analyse weinig aandacht aan het produkt dat vele studenten in de eerste plaats tot stand willen brengen: het diploma. Ook de studieduur blijft als aspect van de interne doelmatigheid van de studieloopbaan onderbelicht. Studievertraging is toch een belangrijke kostenfactor voor diegenen die zich ten doel stellen zo snel mogelijk het diploma te verwerven en op basis daarvan een baan te vinden. Tijd is geld. Dat geldt niet alleen voor 'time-on-task', maar ook voor 'elapsed time'.

Ontvangen 16-9-1987.

NASCHRIFT

K.J.D.M. van der Drift en P. Vos
Rijksuniversiteit Leiden

In de hoofdstuksgewijze behandeling van ons proefschrift door J.W. Holleman zijn een aantal thema's te herkennen.

Operationalisatie-kwesties

- Holleman is van mening dat diplomering een op zichzelf staand proces is. Wij hebben de verschillende aspecten van diplomering ondergebracht in de twee hoofdprocessen van onderwijs: instructie en leren.
- Waar hij voorstelt om de modale, respectievelijk de mediane student tot object van onderzoek te nemen, kan worden opgemerkt dat het voor onze analyse uitsluitend van belang is dat één maat voor centrale tendentie in alle vergelijkingen gebruikt wordt; wij hanteren het rekenkundig gemiddelde.
- Voorts stelt hij bij hoofdstuk 9 voor, een aantal door ons gehanteerde begrippen door andere te vervangen. Ons boek laat zien, dat de door ons gehanteerde systematiek tot duidelijke conclusies leidt. Holleman toont niet aan dat zijn suggesties tot andere, betere resultaten zouden leiden. Een dergelijke conclusie kan ook getrokken worden voor zijn suggestie, de intensiteit van leren los van tijdbesteding in de analyse te betrekken. Aangezien hij niet aangeeft hoe dat geoperationaliseerd zou moeten worden, blijft zijn voorstel in de lucht hangen.

Kortzichtigheid

- Holleman vindt het verwarrend dat de feitelijk geleverde inspanning wordt aangeduid als studielast. In het betrokken hoofdstuk 2 wordt studielast gehanteerd in termen van 'wekelijkse studielast per vak'. Dit begrip kan toch moeilijk worden verward met de benodigde tijd om voor een cursus te slagen.
- Ook vindt hij het bezwaarlijk dat wij bij onze heranalyse niet rapporteren over grootte en representativiteit van de steekproeven. Om dergelijke informatie voor de lezer toegankelijk te maken, hebben wij uitvoerig verwezen naar de oorspronkelijke publikaties.
- Op verschillen in voorkennis tussen studenten, zoals biologiestudenten die het vak Biologie niet in hun eindexamenpakket hebben, gaan wij uitvoerig in, zij het niet in hoofdstuk 5 waarin alleen het gemiddelde een rol speelt.
- Holleman mist in paragraaf 6.2 aandacht voor het verschijnsel van gedwongen onderbelasting en gedwongen leegloop, terwijl dat juist het onderwerp van die paragraaf is, negen bladzijden lang.
- Verder is hij van mening dat wij in hoofdstuk 6 niet adstrueren dat de sociale wetenschappen tot nu toe geen gegeneraliseerde aanwijzingen hebben opgeleverd voor de verbetering van de interne doelmatigheid van onderwijzen en leren. Hij ziet daarbij over het hoofd, dat wij in hoofdstuk 1 op grond van een uitvoerig literatuuronderzoek tot een verklaring komen voor het feit dat dergelijk doelmatigheidsonderzoek tot dusverre niet succesvol is gebleken.
- Tenslotte meent Holleman dat wij studievertraging als kostenfactor onderbelichten. Omdat wij de tijd als kostenfactor benaderen vanuit het leerproces, komt het begrip studievertraging als zodanig in ons boek weinig voor. Op de oorzaken van studievertraging gaan wij op vele plaatsen uitvoerig in.

Missers

- In zijn bespreking van hoofdstuk 2 deelt Holleman mee, dat ons analyseobject niet de tijdbesteding van groepen studenten is, maar van individuen. Niet alleen is het onderwerp van dat hoofdstuk de tijdbesteding van groepen studenten per cursus, maar ook wordt in het onderschrift van vrijwel elke figuur de gemiddelde studielast expliciet genoemd.
- Onder de kop hfst. 6 wordt ons een eigen theorie toegedicht: het resultaat van instructie zou zijn gelegen in de geleverde inspanning die wordt uitgelokt. Hiermee negeert Holleman één van de kernelementen van ons boek: Instructieproces en leerproces zijn vergaand onafhankelijk van elkaar; instructie is een aspect van de studieomgeving, maar resulteert niet in inspanning.
- Er blijft uiteindelijk maar één kritiekpunt uit de waslijst van Holleman over: in hoofdstuk 9 zou een verwijzing naar Carroll op zijn plaats zijn geweest.

Ontvangen 29-2-1988

Boekbesprekingen

Ton Mooij, *Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations* (1987). 's-Gravenhage, Het Instituut voor Onderzoek van het Onderwijs SVO en Nijmegen, Instituut voor Toegepaste Sociale Wetenschappen (ISBN 90-6472-112-2) (234 pp).

De multilevel benadering van het onderwijs is sinds korte tijd voorwerp van een groeiende belangstelling. Lange tijd leek de onderzoeksmethodologie daarvan te gecompliceerd te zijn om oplossingen te kunnen bieden voor de praktijk van het onderzoek. De studies die men vanwege deze bezwaren met 'second-best' procedures uitvoerde, vertoonden tegenvallende resultaten. Toch is de multilevel benadering van belang voor het onderzoek van het onderwijs omdat onderwijssystemen steeds meerdere niveaus bevatten: leerlingen behoren tot klassen, die zijn weer georganiseerd binnen scholen, die op hun beurt behoren tot een inspectiedistrict, enzovoort. Daarbij komt, dat het in het onderwijs om vrij gemakkelijk afgrensbare gehelen (individueen, klassen, scholen, districten) gaat, waardoor de niveaus betrekkelijk gemakkelijk tegenover elkaar kunnen worden afgebakend.

In het multilevel onderzoek wordt getracht kenmerken van klassen of scholen in verband te brengen met kenmerken van leerlingen, ofwel algemener geformuleerd, kenmerken van eenheden van een 'hoger sociaal niveau' in verband te brengen met kenmerken van eenheden van een 'lager sociaal niveau'. Vraagstellingen voor empirische studies naar dergelijke samenhangen zijn er in het onderwijsonderzoek legio aanwezig, temeer omdat men veelal tracht via sturing van klas- en schoolvariabelen de uitkomsten van leerlingen te beïnvloeden. Een voorbeeld van een dergelijke vraagstelling is, of de heterogeniteit van de samenstelling van een klas bijdraagt tot de studiestatistiek van de leerlingen. Gezien het belang van deze soort vraagstelling werd (en wordt) er dan ook volop daadwerkelijk onderzoek aan de hand van multilevel vraagstellingen verricht. De wijdverbreide traditie van dit multilevel onderzoek nam echter niet weg, dat op de gebieden van theorievorming, researchdesign en modelconstructie en -toetsing van het multilevel onderzoek er nog tal van onbeantwoorde vragen bestonden (en nog steeds bestaan). Een van deze vragen is hoe verklaringsmodellen met meerdere niveaus eruit zien en hoe de coëfficiënten daarvan dienen te worden geschat: hoe plausibel is de aanname van directe effecten? Daarmee hangt samen de vraag naar de keuze van de juiste analyse-eenheid: dient deze de leerling te zijn, waarbij het gevaar van autocorrelatie tussen gemiddelden en de individuele gegevens op de loer ligt? Of dient de analyse-eenheid de school of de klas te zijn, met het gevaar dat een groot deel van de binnengroepvariatie verdwijnt met als gevolg dat de coëfficiënten onjuist (veelal te hoog) worden geschat (de aggregatiefout)? Een andere vraag is naar de verschuiving in betekenis van de variabelen wanneer er groepsmaten op basis van individuele gegevens worden geconstrueerd (bijvoorbeeld door gemiddelden of percentages te berekenen), of, omgekeerd, wanneer groepsgegevens aan individuen worden gehecht.

Sinds kort is het multilevel onderzoek belangrijk verbeterd door de opvatting dat de kenmerken van het hoger niveau niet zozeer directe effecten uitoefenen op de individuele 'leeruitkomsten' bij leerlingen, zoals men lange tijd heeft aangenomen, maar dat deze kenmerken inwerken op de individuele processen die tot die leeruitkomsten leiden. Deze hoger niveau kenmerken vormen de stimulerende of remmende condities voor de onderwijsleerprocessen die zich bij de leerlingen afspelen. Deze opvatting, hoe vanzelfsprekend die ook lijkt, is zozeer revolutionair en veelbelovend, dat de benadering die daarop is gebaseerd als 'moderne multilevel benadering' wordt aangeduid.

Mooij heeft binnen de moderne multilevel benadering een empirische studie gedaan naar problemen van leerlingen in het voortgezet onderwijs. De studie is getiteld: *'Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational*

situations'. In zijn studie doet Mooij een poging aan de hand van een concreet geval de genoemde vragen naar theorievorming, researchdesign en modelconstructie en -toetsing te behandelen. Daarmee reikt het belang van zijn studie verder dan alleen de titel suggereert. Niet alleen kan deze worden beschouwd als trend-settend voor het 'moderne multilevel onderzoek', maar ook de inhoudelijk interessante conclusies die zijn onderzoek heeft opgeleverd, verdienen nadere uitwerking.

Het doel van Mooij's studie is 'het exploreren van theoretische aspecten van een interactionele multi-niveau benadering van gedrag, prestatie, bekwaamheid en gerichtheid van leerlingen, waarbij drie niveaus worden onderscheiden: een leerlingniveau, een groepsniveau en een schoolniveau; en het onderzoeken van relevante processen en effecten in de onderwijspraktijk' (blz.160). Uit deze omschrijving blijkt, dat Mooij een dubbel doel heeft beoogd: theorievorming en onderzoek.

Het eerste doel, de theorievorming, wordt als volgt nagestreefd. Na een vrij uitvoerige uiteenzetting van de multilevel benadering van het onderwijs volgt een behandeling van het theoretisch uitgangspunt dat er een wisselwerking bestaat tussen (kenmerken van) de leerling en (kenmerken van) zijn of haar onderwijsomgeving; deze wisselwerking voltrekt zich door de tijd heen. De begripsvorming weerspiegelt deze wisselwerking. De begrippen van bijvoorbeeld studiegedrag, studieprestatie, studiebekwaamheid en studiegerichtheid van leerlingen, die vanouds uitsluitend aan de geïsoleerde leerlingen werden toegeschreven, hebben daarmee niet alleen maar betrekking op de leerling maar ook op zijn of haar omgeving. De omgeving van een leerling wordt gezien als een multilevel systeem, waarvan de onderwijsleersituatie (de klas) en de school deel uitmaken. Mooij noemt deze opvatting de 'interactionele multi-level benadering'. Deze benadering wordt vervolgens uitgewerkt voor het gedrag, de prestatie, de bekwaamheid en de gerichtheid van leerlingen, waarbij drie niveaus worden onderscheiden: een leerlingniveau, een groepsniveau en een schoolniveau. Wil de geschetste wijze van begripsvorming adequaat en vruchtbaar zijn, zo is het basisidee van Mooij's studie, dan zal op basis van empirisch onderzoek moeten kunnen worden vastgesteld of de met die begrippen aangeduide verschijnselen van de verschillende onderwijsniveaus door de tijd op elkaar inwerken.

Het design dat bij een non-experimenteel onderzoek naar een dergelijke inwerking hoort, zal, om de wisselwerking in beeld te kunnen krijgen, longitudinaal van aard dienen te zijn. Ook zal het een voldoende aantal leerlingen van een voldoende aantal onderwijsleersituaties van verschillende scholen dienen te bevatten. De eenheden van onderzoek dienen zodanig te zijn gekozen dat er voldoende variatie is op de onafhankelijke en de afhankelijke variabelen (het 'extreme groups design'). Dit raakt het tweede doel van de studie, een empirisch onderzoek naar de relevante verschijnselen en processen die samenhangen met het gedrag, de prestatie, de bekwaamheid en de gerichtheid van leerlingen met daarbij de drie genoemde niveaus: dat van de leerling, de groep en de school.

In Mooij's studie gebeurde de realisering van een dergelijk design bij het lager beroepsonderwijs (LBO). Er werden daar tien scholen geselecteerd. De klassen werden opgevat als onderwijsleersituaties, bestaande uit specifieke leerlinggroepen en bepaalde schoolvakken, die door verschillende docenten worden gegeven. Het onderzoek was vooral gericht op de schoolvakken Nederlands, wiskunde/rekenen en handvaardigheid bij 39 leerlinggroepen in het eerste leerjaar en, een jaar later, in het tweede leerjaar. 860 leerlingen waren bij het onderzoek betrokken. De gegevens werden overwegend verzameld door middel van uitgebreide systematische observaties. In het schooljaar 1982-1983 en 1983-1984 werden telkens per klas en per leervak twee observatiemomenten georganiseerd. Bij de selectie van de klassen was rekening gehouden met een maximale variatie in cognitieve prestatie van de leerlingen.

De belangrijkste variabelen op leerlingniveau zijn: taakgedrag, schoolvakprestatie, schoolvakbekwaamheid en schoolvakoriëntatie. De variabelen van de onderwijsleersituatie zijn de volgende: taakgedrag tijdens de les, organisationele leskenmerken (zoals organisatie van de leerlingen, soort taak en herhaling), didactische procedures, managementgedrag van de docent, en de gemiddelde en de standaarddeviaties op een aantal leerlingkenmerken. Schoolkenmerken zijn, onder andere, de mate van samenwerking en het percentage uitval van leerlingen.

Bij de analyse is vooral gebruik gemaakt van het twee-stappen model voor multilevel analyse. In de eerste stap werd voor alle leerlingen per leervak en per leerjaar nagegaan in hoeverre de schoolvakprestatie, de schoolvakbekwaamheid, en schoolvakoriëntatie op het tweede meetmoment afhankelijk is van de diverse metingen op het eerste meetmoment en enkele exogene variabelen. Dit gebeurde met behulp van het LISREL-programma en met gewone multiple regressierekening. Uit de resultaten werd geconcludeerd, dat op leerlingniveau het veronderstelde interactieproces inderdaad enigszins aanwezig is: de prestaties aan het begin van het schooljaar beïnvloeden de schoolcijfers tijdens het schooljaar, terwijl er ook, omgekeerd, directe effecten van schoolvakcijfers op schoolvakbekwaamheid en schoolvakprestatie werden gevonden.

In de tweede analysestap kwam de nadruk te liggen op de stabiliteit van de drie genoemde leerlingvariabelen per leerjaar, per klas en per vak. De mate van deze stabiliteit werd afgelezen aan de regressiecoëfficiënten tussen de metingen op de twee observatiemomenten. Het was gebleken, dat deze regressiecoëfficiënten van onderwijsleersituatie naar onderwijsleersituatie aanzienlijke verschillen vertoonden. Nagegaan wordt in hoeverre die stabiliteit afhankelijk is van de samenstelling van de leerlingen van de klas en de manipulatiekarakteristieken van het lesgedrag van de docenten. Daartoe werden de binnengroepregressiecoëfficiënten als afhankelijke variabelen in de tussengroepregressies ingevoerd. De resultaten toonden relatief hoge percentages verklaarde variantie (R^2 variërend van .26 tot .68, met een vijfde deel van de coëfficiënten hoger dan .50). Hiermee is aangetoond dat de classesamenstelling en de manipulatiekarakteristieken van de docent in belangrijke mate van invloed zijn op het proces dat leidt tot de 'uitkomsten' op de leerlingvariabelen op het tweede observatiemoment. Met andere woorden: de onderwijsleersituatie is van invloed op de onderwijsuitkomsten van de leerlingen.

In het laatste deel van de studie is een nadere analyse op schoolniveau aan deze bevinding toegevoegd. Daarbij waren de tien scholen betrokken. Een der voorlopige conclusies is dat er een verband bestaat tussen gebrek aan samenwerking in een school en het percentage leerlingen dat de school voortijdig verlaat, daarmee illustrerend dat ook schoolkenmerken van invloed zijn op de onderwijsuitkomsten.

De studie van Mooij kan worden gezien als gericht op de vakspecifieke onderwijsleerprocessen per leerlinggroep en per leerjaar. Door de opzet is een vergelijking van de invloeden mogelijk tussen deze drie type-indelingen van leersituaties.

De studie illustreert daarmee de vruchtbaarheid van de moderne multilevel benadering voor het onderwijsonderzoek. Door de gekozen uitgangspunten was het mogelijk om, zonder gevaar voor de artefacten van autocorrelatie en de aggregatiefout, het verband tussen een aantal kenmerken van de onderwijsleersituatie en de 'leeruitkomsten' van de leerling vast te stellen. Daartoe bleek het noodzakelijk om vooraf het veronderstelde proces bij de leerling te specificeren. De studie van Mooij geeft hiermee aan hoe een adequate opzet en een theorievorming van onderzoek naar leeromgevingen er uit zou kunnen zien. Bovendien blijkt uit de onderzoeksopzet van Mooij's studie dat onderzoek naar 'attitude treatment interaction' (ATI), over de perspectieven waarvan sommigen de laatste jaren somber gestemd waren, wel degelijk mogelijkheden biedt voor empirisch onderzoek.

De theoretische sterkte van Mooij's studie wijst tegelijk op een aantal zwakke plekken. Aan de inhoudelijke interpretaties van de geaggregeerde maten en aan de inhoudelijke beschrijving van de gevonden coëfficiënten is weinig aandacht besteed. Hoe inzichtelijk de analyse in twee stappen ook is, aan de uitwerking daarvan kleven enkele schoonheidsfouten, die wellicht zijn veroorzaakt doordat de geschikte computerprogramma's voor toetsing van het model en de schatting van de coëfficiënten (zoals VARCL en HLM) ten tijde van het onderzoek nog niet beschikbaar waren.

Deze opmerkingen nemen niet weg dat deze studie getuigt van nauwkeurigheid van presentatie en van een overzicht van relevante literatuur. De studie van Mooij kan worden aangemerkt als onontbeerlijk voor het onderwijsonderzoek met een multilevel vraagstelling van de komende jaren, ook in het buitenland.

P. van den Eeden V.U. Amsterdam

H. Bonset: '*Onderwijs in heterogene groepen*'. Proefschrift Rijksuniversiteit Utrecht, 1987. Tevens verschenen te Purmerend bij uitgeverij Muusses.

In het inleidende *hoofdstuk 1* blijkt dat de dissertatie een onderdeel is van een omvangrijker onderzoeksproject over onderwijs in heterogene groepen. 'De doelstelling van dit project is bij te dragen aan praktijkdidactische theorie omtrent wat in de onderwijswerkelijkheid plaatsvindt bij onderwijs in heterogene groepen, in het bijzonder brugklassen, in het vigerende voortgezet onderwijs.' (p.24). De te beantwoorden algemene onderzoeksvragen luiden:

1. Wat geeft het onderwijsleergebeuren te zien
 - a. in een reeks lessen;
 - b. door één, aan zekere eisen van bekwaamheid beantwoordende, leraar/lerares;
 - c. in een heterogene brugklas;
 - d. van een school die tot het Nederlandse vigerende onderwijs behoort;
 - e. met inbegrip van de opvattingen van de schoolstaf en de ouders?
2. Wat draagt dit bij aan de praktijkdidactische theorie omtrent wat in werkelijkheid op het gebied van het hier bedoelde onderwijs plaatsvindt?' (p.25).

Hoofdstuk 2 gaat over de onderzoeksopzet. Vraag 1 wordt beantwoord met behulp van een case-study met betrekking tot het schoolvak Nederlands in een breed-heterogene brugklas (eerste klas voortgezet onderwijs lbo-mavo-havo-vwo). De keuze voor één case-study is gebaseerd op de wens om de onderwijswerkelijkheid in zoveel mogelijk facetten te beschrijven en deze niet bij voorbaat te reduceren. Het antwoord op vraag 2 wordt aan de orde gesteld door de informatie uit de case-study aan te vullen dan wel te confronteren met informatie uit een aansluitend literatuuronderzoek over heterogeniteit en interne differentiatie in het schoolvak Nederlands.

In de case-study wordt getracht antwoord te geven op vragen die geïnteresseerde leraren zich stellen over het lesgeven in een breed-heterogene klas. In dit verband is de eerste algemene onderzoeksvraag (zie 1a-1e) uitgewerkt in 16 specifiekere onderzoeksvragen (p.36). Deze vragen worden beantwoord met behulp van informatie verkregen via verschillende methoden bij verschillende relevante personen of groepen. Er worden een reguliere scholengemeenschap en hierbinnen een case-leraar geselecteerd en gecontracteerd; er worden interviews afgenomen bij de leraar, leerlingen, brugklas- en sectiefunctionarissen, en ouders; er wordt een enquête afgenomen bij de leerlingen; 22 lessen Nederlands van de leraar worden geobserveerd; bij de leerlingen worden learner reports afgenomen; er vindt bestudering plaats van relevante documenten (toetsen, cijferlijsten, onderwijsleermateriaal, jaarplan van de sectie, gegevens van de leerlingenadministratie, informatie ten behoeve van de ouders); en er worden rapportvergaderingen van leraren over beoordeling en selectie van leerlingen bijgewoond. Bij de observatie van lessen zijn drie observatoren betrokken: één voor observatie van de leraar en twee voor observatie van de leerlingen. Bij de verslaglegging van deze observaties gaan deskundige *tegenlezers* na of de lesbeschrijvingen van de observatoren overbrengen wat deze willen overbrengen. Zij geven bovendien suggesties voor het observeren in volgende lessen (uitgaande van de doelstellingen van de leraar, die vooraf in een interview verkregen waren). De case-leraar geeft op- en aanmerkingen bij de verslaglegging van zowel de case-study als het literatuuronderzoek. Bonset acht dit van belang voor de verhoging van de authenticiteit, validiteit en betrouwbaarheid van de gegevens en de hierbij aansluitende interpretaties en conclusies, en een uiting van respect ten opzichte van de leraar (p.49-50).

In *hoofdstuk 3* staat de validering van in vorig onderzoek (van Bonset) gespecificeerde onderzoeksvragen centraal. Bij de validering worden drie informatie-bronnen gebruikt: zes voorgestructureerde en zes open interviews bij leerkrachten Nederlands uit het voortgezet onderwijs; schriftelijke reacties van twee verenigingen van leraren Nederlands; en informatie uit

ander onderzoek. De resultaten (zie p.112-113) zijn relatief kleine bijstellingen van de oorspronkelijke onderzoeksvragen en zijn identiek aan de reeds op p.36 geformuleerde 16 onderzoeksvragen.

In *hoofdstuk 4* worden de 16 onderzoeksvragen beantwoord met behulp van steeds een selectie uit de variëteit aan methoden. Bij elk antwoord wordt ook het commentaar van de case-leraar op het antwoord van de onderzoekers vermeld. In het hoofdstuk wordt veel gedetailleerde informatie gepresenteerd, hetgeen mede tot uiting komt in het aantal pagina's (van p.114 tot en met p.300). Uit de interviews met de case-leraar blijkt dat zijn hoofddoelstellingen zijn: leerlingen scholen in communicatief opzicht/sociaal vormen; leerlingen mondig maken; leerlingen leren taal te hanteren als expressie-middel. De hoofddoelstelling van de leraar met betrekking tot het werken met heterogene groepen luidt: 'Rekening houden met het totaal van verschillen dat tussen de leerlingen voelbaar is, en dit niet alleen op het gebied van kennen en kunnen, maar ook op dat van bereid zijn' (p.126). De informatie uit de observaties maakt duidelijk dat de leraar deze hoofddoelstellingen lijkt te realiseren (op de pagina's 310-322 geeft Bonset in 94 punten een samenvatting van de bevindingen uit de case-study). Mijns inziens zijn vooral de volgende resultaten uit hoofdstuk 4 van *inhoudelijk belang* voor de problematiek van de breed-heterogene brugklas:

- er wordt wel in subgroepen samengewerkt door de leerlingen, maar de groepssamenstelling vindt plaats op basis van wisselende criteria en vaak is de samenwerking tussen leerlingen een 'lichte samenwerking' (in subgroepen werken aan een individuele taak);
- de leraar differentieert in vrijwel al zijn lessen naar tempo, en afhankelijk van een leerstofonderdeel of van een concrete taak differentieert hij tevens naar belangstelling of leervaardigheden van leerlingen;
- differentiatie naar presentatie en verwerking van de leerstof komt nauwelijks voor;
- het onderdeel grammatica is opgezet volgens een basisstof-herhalingsstof-verrijkingsstof-model (BHV-model) en alleen bij dit onderdeel zijn minimumdoelen geformuleerd;
- de leraar slaagt er niet in de zwakkere leerlingen in evenredige mate te betrekken bij klassikale leerprocessen, maar biedt hen wel stimulerende individuele hulp;
- de leraar laat goede leerlingen de relatief grootste hoeveelheid bijdragen aan het gezamenlijk onderwijsleerproces voor hun rekening nemen, terwijl hij hen regelmatig afremt in zowel klassikale leerprocessen als bij de individuele hulp;
- de leraar heeft oog voor sociale processen in de klas en speelt hierop in of houdt er rekening mee (een probleem is dat lbo- en mavo-leerlingen eerder 'autoritair' leraargedrag vragen, en havo- en vwo-leerlingen meer zelfstandig kunnen en willen werken);
- in de loop van het eerste leerjaar neemt de selectiedruk op leraar en leerlingen toe: prestaties van leerlingen gaan beoordeeld worden volgens twee niveau's (A en B);
- de gebruikte toetsen blijken slechts een geringe inhoudsvaliditeit te bezitten, waardoor de leerlingen soms weinig of niet vooraf oefenen met betrekking tot in een toets gevraagde vaardigheden (dit is vooral nadelig voor de zwakke leerlingen);
- in rapportenvergaderingen spelen niet alleen de toetsresultaten maar ook verwachtingen en inschattingen met betrekking tot een leerling een belangrijke rol bij selectie-beslissingen;
- ongeveer de helft van de leerlingen verlaat de breed-heterogene brugklas met een ander schoolkeuze-advies dan waarmee hij of zij de klas binnenkwam (er is een trend tot verhoging van het advies, met name voor de goede leerlingen);
- de ouders van de leerlingen weten weinig van de werkwijze van de school met betrekking tot de breed-heterogene brugklas en lijken ook niet zo geïnteresseerd.

In het *slothoofdstuk 5* worden de bevindingen uit de case-beschrijving geconfronteerd met de literatuur over heterogeniteit en interne differentiatie, in het bijzonder wat betreft het vak Nederlands in het Nederlandse voortgezet onderwijs. Het aantal Nederlandse publikaties met informatie gebaseerd op andere bronnen dan vragenlijst-onderzoek is klein.

Uit de literatuurstudie blijkt onder andere dat een breed-heterogene brugklas van lbo-mavo-

havo-vwo nog relatief zeldzaam is in het Nederlandse reguliere onderwijs (dus met voorbijgaan van de 'middenschoolen'). De case-school is daarom te beschouwen als een relatieve voortrekker in de richting van middenschool of voortgezet basisonderwijs. Daarnaast levert de literatuur aanwijzingen dat de meer gesloten vormen van differentiatie (naar niveau, tempo) de meer open vormen (naar belangstelling) overheersen. Binnen de gesloten vormen is het BHV-model dominant. De case-school past in het eerste heterogene leerjaar interne differentiatie volgens het BHV-model toe. In het vak Nederlands gebeurt dit met name bij de onderdelen spelling en ontleden; bij de overige onderdelen wordt een meer open differentiatie gehanteerd. Bonset gaat vervolgens in op: drie interpretaties van gelijke kansen (gelijke kansen voor gelijkbegaafden (meritocratie); gelijke kansen door vrijheid van ontplooiing (liberaal-humanistische opvatting); gelijke kansen door gelijke onderwijsresultaten (socialistische opvatting); thematisch-cursorisch onderwijs; de recent opgekomen stellingname dat moedertaalonderwijs communicatie-onderwijs moet zijn; en onderzoek naar leerstijlen van leerlingen in relatie tot interne differentiatie. Tenslotte worden de eisen aan de leraar in verband met het werken met breed-heterogene groepen aan de orde gesteld. In tegenstelling tot de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) is de auteur van mening dat interne differentiatie in een breed-heterogene brugklas wel realiseerbaar is voor de gemiddeld-bekwame leraar Nederlands in het reguliere voortgezet onderwijs (p.465).

De *positieve punten* in het proefschrift zijn de volgende. Ten eerste: Bonset geeft een uitvoerige, zorgvuldige beschrijving van een deel van de complexe onderwijswerkelijkheid in en met betrekking tot één heterogene brugklas lbo-mavo-havo-vwo, en vergelijkt de uit diverse bronnen afkomstige feiten en inzichten zo verantwoord mogelijk met elkaar. Dit lijkt ook relevant voor 'de praktijkdidactische theorie', die genoemd wordt in de tweede algemene onderzoeksvraag. Ten tweede: de samenwerking met de case-leraar vergroot de validiteit van de onderzoeksresultaten wat betreft de interpretatie en conclusies, die gebaseerd worden op de observaties. Ten derde: door de concentratie op de onderwijspraktijk geeft de auteur informatie over de kern waarop onderwijsonderzoek gebaseerd moet zijn, en vermijdt hij een overdaad aan (subjectieve) opvattingen over of waarderingen van bepaalde betrokkenen of belanghebbenden. Ten vierde: Bonset maakt in zijn onderzoek als het ware indirect duidelijk welke aspecten en knelpunten van belang zijn bij een verder inhoudelijk doordenken en vormgeven van interne differentiatie in het schoolvak Nederlands. Voor de beantwoording van de vraag naar: 'Hoe verder?' geeft het proefschrift essentiële aanwijzingen, nuances en overwegingen rondom vernieuwing van het vak Nederlands en de noodzaak van op onderzoek gebaseerde leerplanontwikkeling. Zeer belangrijk hierbij is, naar mijn mening, de constatering van de case-leraar dat er (nog) geen adequaat gedifferentieerd curriculum is (p.265) en dat beoordeling van leerlingen volgens individuele normen zou moeten geschieden (p.285; vgl. ook p.387 en 389). Een samenhangend probleem betreft de dubbele functie van activiteiten van subgroepen leerlingen tijdens lessen Nederlands: enerzijds beogen deze een vergroting van individuele leerprestaties, anderzijds dienen deze ter realisatie van sociale doelstellingen (p.436-438). Deze dubbele functie vraagt om explicitering van twee soorten doelstellingen en beoogde effecten in twee (ten dele samenhangende) leerplannen.

Ik wil de volgende *kanttekeningen* plaatsen bij het proefschrift. De omvang van het boekwerk (525 pagina's) zou kleiner gekund hebben door een strakkere opzet, minder herhalingen in de tekst en door hoofdstuk 3 te reduceren tot een paragraaf binnen hoofdstuk 2. De relevantie van hoofdstuk 3 wordt door de auteur zelf ook gereduceerd, doordat de 16 onderzoeksvragen in hoofdstuk 2 (p.36) dezelfde zijn als de vragen die resulteren uit hoofdstuk 3 (p.112-113). Ten tweede: bij de beantwoording van de onderzoeksvragen in hoofdstuk 4 zou het overzicht bevorderd worden door een schema, waarin per vraag aangegeven is welke informatie via welke onderzoeksmethode(n) verzameld is. Ten derde: over de observatie-procedure en de mogelijke effecten van de observaties op de leraar bestaat nog enige onduidelijkheid. Op p.54-55 wordt beschreven dat observatoren niet mochten praten met de leraar over de lessen wegens eventuele

beïnvloeding; op p.56 staat te lezen dat de leraar na ongeveer 15 geobserveerde lessen de observatie als zeer drukkend begon te ervaren vanwege de continue observaties en de na-interviews die zich veelal in de pauzes tussen zijn lessen afspeelden. Blijkbaar werd de leraar steeds na elke geobserveerde les geïnterviewd over de les. Deze interviews zullen de leraar en zijn volgende lessen toch wel beïnvloed hebben!? Ten vierde: in de mate waarin de drie observatoren een verschillende taakstelling hadden (zie p.42), draagt het werken met drie observatoren *niet* bij aan de betrouwbaarheid van de observatiegegevens (vgl. p.44). Ten vijfde: het inschakelen van tegenlezers in de observatieprocedure (p.43) is van belang voor het vergelijken van leraardoelstellingen en observatieresultaten, maar tegenlezers dragen *niet* bij aan de validiteit van observatiegegevens zelf (vgl. p.44). Ten zesde: het onderscheid tussen de termen leerresultaat, leereffect en leerervaring (p.180) komt te weinig naar voren in de formulering van de hypothesen en de omschrijving van de onderzoeksresultaten op p.182-197, in het bijzonder wanneer door de onderzoeksopzet wel samenhangen maar geen oorzakelijke relaties beargumenteerd kunnen worden. Ten zevende: op p.60 is het niet duidelijk welk criterium gehanteerd is bij de selectie van de zes ouderparen (ten behoeve van een interview). Ten achtste: Bonset is geneigd het breed-heterogene werken positief te waarderen, onder andere wegens de opwaartse trend in advisering voor het type voortgezet onderwijs bij een vergelijking van de adviezen gegeven in de basisschool met die, welke gegeven worden aan het einde van het eerste breed-heterogene leerjaar in de brugklas (zie p.256, 400-401). Ik denk echter dat deze opwaartse trend net zo goed verklaard kan worden uit binnenschoolse selectie-mechanismen (binnen klas of school in het basisonderwijs of het voortgezet onderwijs), of uit verschillen tussen basisscholen in procedures in verband met deze advisering. Ten negende: Bonset is vóór onderwijs in breed-heterogene groepen, omdat hij dit onderwijs ziet als een mogelijkheid om gelijke kansen door gelijke onderwijsresultaten te realiseren. Vanwege de onderwijskundige en onderwijspolitieke relevantie van dit punt, wil ik hierbij drie argumenten geven waarom ik deze voorkeur prematuur acht:

1. uit het onderhavige onderzoek van Bonset blijkt dat er zich in de breed-heterogene groep verschillende *problemen* voordoen: er is sprake van benadeling van relatief zwakke maar ook van benadeling van relatief goede leerlingen; de lbo-leerlingen hebben het gevoel naar beneden toe uitgeselecteerd te zijn (p.279); er bestaan nog aanzienlijke problemen met het (ontwikkelen van het) juiste curriculum;

2. de case-leraar vermijdt doelbewust kennis te nemen van de over de leerlingen afkomstige informatie uit de basisschool. Dit lijkt mij veroorzaakt te zijn doordat hij, zelfs met die informatie, toch nog niet weet wat een leerling in de basisschool gehad heeft en in welke mate een leerling bepaalde leerstof beheerst. Bovendien verschillen basisscholen in het onderwijsaanbod aan leerlingen. Om deze redenen zou het nodig zijn een landelijk kerncurriculum te ontwikkelen (voor elk 'selectievak' in het basisonderwijs en voortgezet (basis)onderwijs), waarmee de mate van beheersing en de leervorderingen van elke leerling in een vak direct afleesbaar zijn. Met behulp van een landelijk kerncurriculum is onder andere een goede aansluiting mogelijk tussen basis- en voortgezet onderwijs. De informatie uit het basisonderwijs kan dan gebruikt worden om de leerprocessen en leereffecten bij elke leerling zo *snel* en zo *adequaat* mogelijk te continueren in het voortgezet onderwijs; dit is *het meest van belang voor de relatief zwakke leerlingen*. In de huidige situatie wekt het werken met heterogene groepen in het voortgezet onderwijs soms de indruk dat men het basisonderwijs wil herhalen, en dit kan niet de bedoeling zijn;

3. onderwijs dient – volgens de wet – te leiden tot ontwikkeling van de cognitieve, affectieve, creatieve, en esthetische vaardigheden van elke leerling. Onderwijsonderzoek kan hierbij ondersteunend werken door na te gaan welke onderwijs- of curriculumcondities in interactie met welke leerlingkenmerken (bijv. reeds aanwezige kennis (vaardigheden) of de structurering hiervan, bepaalde leerstijlen of leerstrategieën) via welke processen leiden tot welke leerervaringen en welke leereffecten. Voor elke leerling, dus ook voor de relatief zwakke en relatief goede leerlingen, zouden relatief optimale onderwijsleersituaties gecreëerd moeten worden. Met behulp van onderwijsonderzoek dient achterhaald te worden of, en zo ja in welke mate, brede

heterogeniteit hierbij een rol speelt of kan spelen. Op dit moment is mijns inziens over onderwijsprocessen en onderwijseffecten in Nederland nog *niet voldoende* informatie aanwezig om een *onderwijskundig of onderwijspsychologisch verantwoorde keuze* te maken inzake de wenselijkheid van breed-heterogene klassen.

Ton Mooij

Katholieke Universiteit Nijmegen
Instituut voor Toegepaste Sociale wetenschappen
Graafseweg 274
6532 ZV Nijmegen

Fred Wester, '*Strategieën voor kwalitatief onderzoek*'. Muiderberg: Coutinho, 1987

Je moet nog in 'oude stijl' zijn afgestudeerd om je de lange en moeizame discussies te kunnen herinneren over het kwalitatieve onderzoek. Maar wie zo'n tien jaar geleden een vaste bezoeker was van de universiteit, kent ongetwijfeld dit debat dat vooral op aktiebijeenkomsten en in het grijze stencilcircuit werd uitgevochten. De felheid waarmee dit debat werd gevoerd, kan achteraf worden verklaard vanuit het opkomend belang van externe subsidies voor de uitvoering van onderzoek. Maar wellicht was het juist die heftigheid en vaak ook ongenueanceerdheid die de kwalitatieve onderzoekers uiteindelijk in de marge deden belanden.

De reacties van de betrokkenen liepen uiteen. Sommige besloten het door hen gewenste onderzoek gewoon te gaan uitvoeren, geruggesteund door het feit dat kwalitatief onderzoek ook vrij goedkoop kan worden uitgevoerd. De herontdekking van de ethnografische traditie en de voorbeelden van ondermeer het C.C.C.S. in Birmingham en de antropologische stadstudies uit Utrecht deden hier hun werk. Anderen wijdden zich aan een filosofische herbezinning op zaken als het causaliteitsbeginsel, de empirische cyclus en de rationaliteit van wetenschappelijke fora. Daartussenin bewoog zich een derde groep, die zich meer pragmatisch richtte op de explicitering van de kwalitatieve methode. Recentelijk verschenen van deze kant een aantal dissertaties en daarop gebaseerde handboekjes. Voor onderwijsonderzoekers is vooral dat van Wester interessant, omdat het uit hun eigen specialisatie voortkomt en daar ook de voorbeelden aan ontleent.

Het grootste probleem van ieder die zich aan een dergelijke onderneming waagt, is de definitie van het adjectief 'kwalitatief'. Hier wreekt zich het verleden. De voorstanders van het kwalitatieve onderzoek pinden zich immers in hun verzet tegen de tradities van het survey-onderzoek en het experimentele onderzoek, bijna onwrikbaar vast op een principiële verwerping van elke becijfering en calculatie van de sociale werkelijkheid. Wester ruimt gelukkig snel het misverstand uit de weg, dat 'kwalitatief' zou staan voor de uitsluiting van mathematische principes. Hij geeft niet alleen aan dat het hier gaat om een verkeerd geformuleerde tegenstelling, maar ruimt ook plaats in voor de behandeling van meer en minder bekende statistische technieken. Wester kiest voor een meer inhoudelijke afbakening en plaatst deze op het scheidsvlak van een meer objectiverende en een meer interpretatieve benadering. De objectiverende benadering gaat ervanuit, dat aan het menselijk handelen theoretisch te formuleren, objectieve structuren ten grondslag liggen; de interpretatieve benadering stelt dat veel menselijk gedrag kan worden begrepen vanuit de actieve zin- of betekenisgeving door de betrokkenen zelf. Voldoende argumentatie levert een dergelijk onderscheid niet op. Want al beperkt het kwalitatieve onderzoek zich tot onderzoek naar zingevingen, dan beoogt zij toch daar de 'dieperliggende' structuren van bloot te leggen. En zodra een dergelijke structuur theoretisch duidelijker wordt, wordt zij het mogelijke onderwerp van een objectiverende benadering. Het is deze paradox van het kwalitatieve onderzoek, die haar steeds weer in de rol dringt van vooronderzoek of de leverancier van geschikte illustraties. Er is dus een extra

argument nodig om het kwalitatieve onderzoek als een zelfstandige vorm van onderzoek te legitimeren. Het standaard argument daarvoor is, dat betekenisgevingen plaatsvinden in een eigen domein, dat van het 'dagelijkse leven', het 'sociale', of in meer fenomenologische termen 'de leefwereld'. Een domein, waarin veranderingen een bijna permanent karakter dragen. Wester maakt dit argument, zij het niet al te expliciet, tot het zijne door een keuze te maken voor het symbolisch interactionisme. In een compleet hoofdstuk wordt deze theoretische basis weergegeven.

Hoewel deze keuze legitiem is, leidt zij toch tot inperkingen. Zaken als 'oral history', biografisch onderzoek en ook meer actuele stromingen als taal- en 'discours'-analyse vallen min of meer uit de boot. Het is ook echter duidelijk niet de bedoeling van Wester om de lezer een compleet overzicht te verschaffen van alle vormen van kwalitatief onderzoek. Zijn grootste zorg blijkt de kwaliteit van het kwalitatieve onderzoek en daarmee het weerleggen van de twee belangrijkste kritiekpunten die tegen het kwalitatieve onderzoek zijn ingebracht: die van oncontroleerbaarheid en gebrek aan generaliserend vermogen. Wat het laatste betreft, gaat Wester uitgebreid in op de ideeën die ten grondslag liggen aan het trekken van steekproeven, of 'monsters' zoals hij ze in dit geval noemt. Nog uitgebreider en nog meer zorg besteedt hij aan het probleem van de controleerbaarheid. Hier grijpt hij terug op de benadering van de 'grounded theory', zoals die is uitgewerkt door Glaser en Strauss. De 'grounded theory' - of GT-benadering kan worden omschreven als een poging om de 'logic-in-use', de pragmatische manier waarop verklarende theoretische termen in de loop van een empirisch onderzoek worden ontwikkeld en aangepast, bewust te systematiseren.

De GT-aanpak begint niet met uitgewerkte begrippen, maar met meer vage 'sensitizing concepts' en tracht deze door 'empirische verzadiging' verder uit te werken tot een verklaringsmodel. Wester geeft een nauwkeurige indeling in fasen en subfasen en behandelt per fase de mogelijkheden die de onderzoeker ten dienste staan en de risico's die moeten worden vermeden. Dit alles steeds op grond van methodologische argumenten. In twee afzonderlijke hoofdstukken behandelt hij uitgebreid wat er zoal komt kijken bij de waarnemings- en verzamelingsfase en de ordenings- en analysefase.

Wester gaat tevens in op de rapportage, een netelig punt bij het kwalitatieve onderzoek. Veel geslaagde onderzoeksrapporten uit deze hoek hebben immers vaak iets particularistisch – ze rusten op een geheel eigen begrippenkader – en een literaire opbouw. Op die manier ontwijken ze de moeizame opdracht hun lezers mee te voeren langs de eigen onderzoeksroute. Voor 'objectiverend onderzoek' bestaat dit probleem niet. Onderzoeksverslagen uit die hoek kennen immers hun eigen 'logic of presentation'. In twee pagina's tracht Wester iets dergelijks ook voor het kwalitatieve onderzoek aan te geven. Dit is ongetwijfeld het meest zwakke gedeelte van zijn boek, dat blijft steken in een aantal vage aanwijzingen. Een dergelijk cruciaal probleem had meer aandacht en ruimte verdiend.

De hamvraag blijft inmiddels of er in het onderwijsonderzoek behoefte bestaat aan kwalitatief of misschien beter gezegd leefwereld-onderzoek. Dat de school een interessant domein is voor leefwereldonderzoekers staat buiten kijf, maar moeten onderwijsonderzoekers daarom ook leefwereldonderzoek doen? In het onderwijsonderzoek is het kwalitatieve onderzoek tot nu toe marginaal gebleven. Het biedt zelden de neutrale cijfers waarmee de onderwijspolitieke discussie zo graag stoeit: het initiatief tot dit type onderzoek berustte vaak bij de meer kritisch ingestelde onderzoekers en tot slot is het onderwijs een leerinstituut, waar het dagelijks leven geacht wordt zich buiten af te spelen. Er lijkt echter sprake van een groeiend inzicht, dat leefwereldkenmerken een belangrijke invloed uitoefenen op leerprocessen. De vaardigheden waarover individuen beschikken om zich in leefwereldsituaties staande te houden, zouden niet alleen samenhangen met morele, maar ook met communicatieve competenties en op grond daarvan weer met cognitieve prestaties. Onderzoek dat op een dergelijke vraagstelling is gebaseerd zou veel kunnen hebben aan Westers inleiding.

Het boek van Wester is als geheel beschouwd zeker een verdienstelijk werkstuk. Het bevat een aanzet tot een theoretisch-methodologische plaatsing van het leefwereldonderzoek, het biedt een verdere systematisering van de 'grounded theory'-aanpak en het staat – zoals het hoort in

een goed handboek – boordevol praktische tips en aandachtspunten voor onderzoekers. Het ondergraft de twee belangrijkste kritiekpunten die tegen het kwalitatief onderzoek zijn ingebracht; die van oncontroleerbaarheid en gebrek aan generaliserend vermogen. Het is in een geduldig vriendelijke, discursieve toon geschreven, en samen met de instructieve vragenbatterij die het boek afsluit, maakt het dat tot een ideale inleiding.

Niettemin bevat het boek, zoals aangegeven, enkele lacunes. Sommige gepresenteerde principes, zoals bijvoorbeeld 'triangulatie', blijven onduidelijk en zijn mogelijk ook geheel overbodig. Aan cruciale zaken in het kwalitatief onderzoek, zoals bijvoorbeeld typologieconstructie en het cruciale probleem van de presentatie, had theoretisch en methodologisch best nog wat meer ruimte besteed mogen worden. En teleurstellend zijn ook de voorbeelden uit het eigen onderzoek, die vaak weinigzeggend zijn weergegeven. Ook door zijn omvang blijft het boek eerder een inleiding dan een handboek. Als onderzoek in deze stijl verder van de grond komt, zal er te zijner tijd toch een nieuwe en uitgebreidere tekst moeten worden geschreven. Wester heeft zich met dit boekje alvast gekwalificeerd als de potentiële auteur van een dergelijk werk.

Ger Tillekens
R.U. Groningen

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 OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 12, nr. 2

April 1988

Inhoudsopgave

Convocatie ledenvergadering VOR	110
Beknopt verslag verenigingsjaar 1987	111
VOR-bijdrage aan inhoudelijke bijeenkomsten	112
VOR begroting 1988	114
Toelichting op de VOR-begroting 1988	115
De gastsprekers tijdens de ORD-Leuven	116
Kwaliteitsbepaling wetenschappelijke produktie in de Onderwijskunde <i>J. Ax i.s.m. B. Creemers</i>	118
Congressen, Symposia, Cursussen, Trainingen	122
Mededelingen	126
Proefschriften	127
Ontvangen publikaties	128

Convocatie ledenvergadering VOR

- Tijdens de ORD 1988 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, 30 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. 26 mei 1987 (zie VOR bulletin november 1987)
3. Jaarrekening 1987 en verslag van de kascontrolecommissie. (De jaarrekening is twee weken vóór de ledenvergadering op te vragen bij de penningmeester, drs. E. Roede, (tel. 020-264547). Ter vergadering zal de kascontrolecommissie verslag doen van haar bevindingen)
4. Begroting 1988 (zie dit VOR-bulletin)
5. Beknopt verslag verenigingsjaar 1987 (zie dit VOR-bulletin)
6. Samenstelling kascontrolecommissie 1988
7. W.v.t.t.k.
8. Sluiting

Beknopt verslag verenigingsjaar 1987

1. Bestuurssamenstelling

Het bestuur was in 1987 als volgt samengesteld:

- | | |
|---------------------|----------------------------------|
| a. Voorzitter: | Prof. Dr. J. Claessen |
| b. Vice-voorzitter: | Mevr. Dr. B. van Hout Wolters |
| c. Secretaris: | Mevr. Drs. M. P. C. van der Werf |
| d. Penningmeester: | Drs. E. Roede |
| e. Leden: | Drs. K. Beintema |
| | Dr. W. Gerritsen van der Hoop |
| | Prof. Dr. J. Scheerens |
| | Prof. Dr. P. Span |

2. Internationale contacten

- Ook in 1987 zijn op verzoek van de VOR door de AERA divisies Nederlandse reviewers ingeschakeld bij de beoordeling van papers. Met de voorzitter van de AERA, de heer Shavelson, zijn afspraken gemaakt over intensivering van de contacten tussen Nederlandse en Amerikaanse onderwijsonderzoekers.
- Een groot aantal VOR-leden bezocht in 1987 de AERA-bijeenkomst te Washington DC.
- Op de EARLI conferentie in Tübingen was de aanwezigheid van VOR-leden opvallend.
- Voor de overeenkomst tussen de VOR en Pergamon Press dat VOR-leden voor £ 10,- een abonnement kunnen nemen op het International Journal of Educational Research (IJER) bleek grote belangstelling. Het vereiste aantal van 75 abonnees voor 1988 was op 31 december 1987 al vrijwel bereikt. In 1987 verschenen zes nummers. Het Nederlandse nummer werd samengesteld door Tillema en Veenman.

3. Onderwijs Research Dagen

De Onderwijs Research Dagen vonden plaats in Groningen. Prof. Dr. B. Creemers was voorzitter van de organisatiecommissie. Het aantal deelnemers bedroeg 575.

De bundels over de thema's Curriculum (Harskamp en Nijhof), Effectiviteit Onderwijsorganisaties (Scheerens en Stoel), Evaluatiemethodologie (Hofstee) en Scholing en Nascholing (Tillema en Peters) verschenen nog in 1987.

4. Tijdschrift voor Onderwijs Research/VOR bulletin

Prof. Dr. B. Creemers was ook in 1987 redactiesecretaris van het Tijdschrift voor Onderwijs Research. Redactiesecretaris van het VOR-bulletin was Dr. P. van der Sijde. De opname van het bulletin in het tijdschrift is geëffectueerd. Het geplande aantal van zes nummers is in 1987 ook inderdaad verschenen. Evaluatie van de integratie van het tijdschrift en het bulletin zal in 1988 plaatsvinden.

5. VOR-aanmoedigingsprijs

De jury, onder leiding van Prof. Dr. W. van der Linden, kende de prijs van f 1000,- toe aan Mevr. Drs. P. Willemsen.

De titel van het paper luidde: Overleving in de bovenstroom van het Voortgezet Onderwijs: veranderde onderwijskansen van jongens en meisjes.

6. Gedragsregels voor onderwijsonderzoekers

Op initiatief van het RION is een werkgroep ingeteld, bestaande uit mevr. Drs. S. Blaauw-boer namens de onderzoeksinstituten, Mevr. Dr. B. Van Hout Wolters en Drs. K. Beintema namens de VOR en drs. Ph. Groeneweg als waarnemer van de SVO, voor het opstellen van gedragsregels voor onderwijsonderzoekers.

7. Ledenwerving

Ten behoeve van de werving van nieuwe leden is een folder samengesteld en actief verspreid. Deze ledenwerving is succesvol gebleken. De werving leverde 80 nieuwe leden op. Dit bracht het aantal VOR-leden in 1987 op 469.

8. Financiën

Zie kasverslag 1987 van de penningmeester van de VOR.

Greetje v.d. Werf
Secretaris

VOR-bijdrage aan inhoudelijke bijeenkomsten

Met ingang van 1988 heeft het bestuur van de VOR voor VOR-leden de mogelijkheid gecreëerd om voor inhoudelijke bijeenkomsten een beroep te doen op financiële ondersteuning door de VOR. Voor 1988 is hiervoor een bedrag van f 1.500.= op de begroting gereserveerd. Aanvragen voor een bijdrage kunnen via de secretaris bij het bestuur worden ingediend. Hoewel het niet om een erg groot budget gaat, verwacht het VOR-bestuur hiermee inhoudelijke bijeenkomsten op het terrein van de onderwijsresearch te stimuleren. Vooral wordt daarbij gedacht aan landelijke bijeenkomsten, maar ook voor internationale samenwerking kan een beroep op de VOR worden gedaan voor een (beperkte) bijdrage.



**NAAR EEN
STRATEGISCHE PROGRAMMERING VAN
ONDERWIJSONDERZOEK**

Onder redactie van
K. Beintema
P.C. van den Dool
E. Roede
P.C. van der Sijde

VOR Vereniging voor Onderwijsresearch

Aan deze bundel werken mee:

Van Amersfoort
Beintema
Blok
Emmelot
Hetebrij
Van Hoesel
De Jong
Koornstra
Kotte
Van der Linden
Lodewijks
Monsma
Van Oijen
Roede
Schepel
Van der Sijde
Zijlstra

Deze bundel bevat de bewerkte artikelen over programmering die in eerdere nummers van het VOR-Bulletin zijn verschenen, alsmede de bijdragen aan het symposium over dit onderwerp die op de ORD'87 werden gepresenteerd.

De bundel is een uitgave van de VOR (ISBN 90-249-1604-6) en kan besteld worden door overmaking van fl 20,- op gironummer 3415813 t.n.v. penningmeester VOR, o.v.v. Bundel Onderzoeksprogrammering.

VOR Begroting 1988

Inkomsten

100 Contributie

110 leden	ingeschr.	betalend	
120 studenten	450	430 x 105.=	f 45.150.=
	25	20 x 25.=	f 500.=
			<u>f 45.650.=</u>

300 Intern. Journ. Ed. Res.

310 abonneementen	75 x f 40.=		f 3.000.=
320 garantiesubsidie SVO	max.		<u>f 5.500.=</u>

500 SOR bijdrage VOR prijs
700 Inkomsten TOR (Royalties)
800 SVO subsidie TOR
900 Diversen

	f 8.500.=
	f 1.000.=
	f 1.800.=
	f 5.000.=
	f 600.=
	<u>f 62.550.=</u>

Uitgaven

400 Activiteiten	
410 ledenwerving	f 500.=
420 ORD-programmaboek	f 1.000.=
440 colloquia	f 1.500.=
	<u>f 3.000.=</u>

500 Tijdschriften

510 TOR/VOR bulletin	
511 Collect. abonnem.	f 37.600.=
leden 470 x 80.=	f 1.325.=
stud. 25 x 53.=	<u>f 38.925.=</u>

512 Redactiekosten TOR

513 Redactiekosten	f 4.000.=
VOR-bulletin	f 1.500.=
	<u>f 44.425.=</u>

520 I.J.E.R.

150 abonneementen	
a f 40.=	f 6.000.=
linguistic editing	<u>f 2.500.=</u>

600 Bestuur

700 VOR prijs	f 8.500.=
800 Ledenadministratie	f 2.000.=
900 Diversen	f 1.000.=
	f 3.000.=
	f 625.=
	<u>f 62.550.=</u>

Toelichting op de VOR-begroting 1988

In grote lijnen is de begroting als voor 1987. Wijzigingen betreffen de inkomsten uit de abonnementen voor het IJER, de ledenwerving, de uitgaven voor het toezenden van het ORD-programmaboek aan alle leden, redactiekosten VOR-bulletin en de ledenadministratie.

Toelichting per post:

Inkomsten

- 100 Wellicht mede door de folder is het aantal leden groter geworden. Voor 1988 wordt gerekend op $\pm$ 475 leden, waarvan 25 student. Gezien de blijkbaar niet geheel te vermijden betalingsproblemen, wordt gerekend op 430 betalende gewone en 20 betalende studentleden. De contributie is hetzelfde gebleven als voor 1987, nl. f 105.= voor gewone leden en f 25.= voor student-leden.
- 300 Met uitgeverij Pergamon en de SVO is een overeenkomst gesloten, waarbij de VOR 150 collectieve abonnementen op het International Journal of Educational Research (IJER) neemt. Om aanloopverliezen te kunnen opvangen heeft de SVO een garantiesubsidie beschikbaar gesteld. Voor 1988 bedraagt deze maximaal £ 750,=. Daarnaast subsidieert de SVO maximaal f 2.500.= voor linguistic editing ten behoeve van het IJER.
- 500 De VOR-prijs wordt gefinancierd door de Stichting Werkgroep Onderwijsresearch (SOR).
- 700 Van Swets & Zeitlinger ontvangt de VOR 10% van de TOR abonnementsinkomsten van niet-VOR-leden. Voor 1988 wordt dit geschat op f 1.800.=.
- 800 Ten behoeven van het TOR ontvangt de VOR jaarlijks F 5.000.= van de SVO.

Uitgaven

- 410 Na de intensievere ledenwerving in 1987 door middel van een op ruime schaal verspreide brochure zijn de wervingsactiviteiten voor 1988 in financieel opzicht lager ingeschat, nl. f 500.= in plaats van f 3.500.=. In het bestuur is besloten dat er naar gestreefd zal worden om het programmaboek voor de ORD met ingang van 1988 vóór de ORD aan alle VOR-leden toe te sturen - ongeacht of zij zich voor de ORD hebben opgegeven of niet. De hieraan verbonden kosten worden geschat op maximaal F 1.000.=.
- 440 Colloquia zijn weer opgevoerd voor f 1.500.=. Hieruit kunnen tevens andere kosten t.b.v. inhoudelijke activiteiten worden betaald.
- 500 De uitgaven voor de tijdschriften (TOR/VOR-bulletin en IJER) zijn naar verwachting hetzelfde als voor 1987, met dien verstande dat:
- er meer leden, dus meer abonnementen op TOR/VOR-bulletin moeten worden betaald;
 - voor redactiekosten voor het VOR-bulletin maximaal f 1.500.= beschikbaar zijn.
- 600 De uitgaven ten behoeve van bestuur zijn hetzelfde begroot als voor 1987. Een belangrijk deel hiervan is bestemd voor reiskosten van enkele bestuursleden.
- 700 De jaarlijks tijdens de ORD toe te kennen VOR-prijs bedraagt maximaal f 1.000.=.
- 800 De uitgaven voor de ledenadministratie zijn ten opzichte van 1987 verhoogd ten behoeve van investeringen in verdere automatisering van leden- en financiële administratie.

E. Roede,
(penningmeester).

De gastsprekers tijdens de ORD-Leuven

Dr. Gavriel SALOMON

Op uitnodiging van de organisatiecommissie zal Dr. Gavriel Salomon tijdens de ORD 1988 te Leuven op maandag 30 mei een lezing houden getiteld: 'Technology for the Mind: Partner and nourisher'. Van de aanwezigheid van Dr. Salomon zal tevens gebruik gemaakt worden om hem actief te betrekken bij een symposium over de problematiek van het verwerven van algemene (meta)cognitieve vaardigheden.

Uit zijn curriculum vitae blijkt dat Dr. Salomon (geboren in 1938) reeds een rijkge vulde staat van dienst heeft, zowel in het onderwijskundig onderzoek als in de opleiding van onderwijskundigen en leraren. Ik zal me er hier toe beperken beknopt enkele hoofdzaken te vermelden.

Dr. Salomon kreeg zijn basisopleiding aan de Hebrew University of Jeruzalem, Israel, waar hij in 1964 een B.A. haalde in 'Education and Geography' en in 1966 een M.A. in 'Education and Psychology'. Hij vervolgde zijn vorming aan Stanford University waar hij in 1968 zijn Ph. D. behaalde met een werk op het domein van 'Educational Psychology and Communication'.

Sinds 1984 is Dr. Salomon 'full professor' aan Tel Aviv University, waar hij hoofd is van het 'Graduate Program on Communication and Technology in Education'. Dit academiejaar heeft hij echter zijn activiteiten in Israel onderbroken en is hij 'visiting professor' aan het College of Education van de University of Arizona in Tucson, waar hij Dr. David Berliner vervangt tijdens diens sabbatsverlof. Van 1968 tot 1984 bekleedde Dr. Salomon academische posities aan diverse universiteiten in Israel maar ook in de V.S. (Harvard University, University of Southern California, Stanford University, University of Michigan, Indiana University).

Dr. Salomon werd in juli 1986 verkozen als 'President-elect of the Educational, Instructional and School Psychology Division of the International Association of Applied Psychology.' In 1983 verwierf hij de status van Fellow van de 'Division on Educational Psychology' van de American Psychological Association. Hij was één van de gastsprekers tijdens de Second European Conference for Research on Learning and Instruction (EARLI) gehouden te Tübingen in september 1987 en werd uitgenodigd om een 'keynote address' te houden tijdens het '24th International Congress of Psychology' dat te Sydney, Australië, gehouden wordt eind augustus 1988.

Deze ORD-gastspreeker heeft uiteraard ook vele wetenschappelijke publikaties op zijn naam, waaronder 3 boeken, 14 hoofdstukken in boeken en meer dan 40 tijdschriftenartikelen. Voor één van zijn boeken, *Interaction of media, cognition and learning* ontving hij in 1981 de 'First James W. Brown Publication Award' van de Association for Educational Communication and Technology (U.S.A.). De titel van dit boek is meteen indicatief voor de meest centrale thematiek uit Dr. Salomon's onderzoekswerk. Hij is trouwens co-auteur (met Dr. R. E. Clark) van het hoofdstuk over 'Media in teaching' in het in 1986 verschenen *Handbook of research on teaching, Third Edition* (onder redactie van M.C. Wittrock). Zijn voornaamste interesse kan als volgt samengevat worden: Welke relatie is er tussen culturele fenomenen, inzonderheid media en computers enerzijds en de verstandelijke ontwikkeling anderzijds, en hoe kan deze relatie door opvoeding en onderwijs beïnvloed worden? Verwant met het voorafgaande is Dr. Salomon's belangstelling voor het verband tussen 'cognition and volition', zoals dit tot uiting komt in de hoeveelheid mentale inspanning ('mindful versus mindless') die mensen investeren in het verwerken van informatie. Zo is één van zijn interessante bevindingen dat kinderen TV-kijken, in tegenstelling tot lezen, opvatten als een gemakkelijke activiteit die weinig inspanning vereist. Consistent daarmee constateerde hij dat kinderen weinig mentale energie investeren ook in het verwerken van een TV-programma dat nochtans cognitief veeleisend is; als gevolg daarvan leren ze er ook weinig van.

Het verband tussen informatietechnologie en de cognitieve ontwikkeling is vooral de laatste jaren een meer belangrijke plaats gaan innemen in Dr. Salomon's speurwerk. Daarbij is hij recent een zeer belangrijk participant in de discussie over de actuele vraag of computerervaring, inzonderheid leren programmeren, kan bijdragen tot het verwerven van veralgemeenbare denk- en leervaardigheden. Centraal in zijn theorie over transfer staat trouwens de notie 'mindfulness' opgevat als 'the volitional', metacognitively guided employment of non-automatic, usually effort demanding processes'; transfer van hogere cognitieve vaardigheden vergt volgens Salomon's visie 'mindful abstraction'.

In zijn ORD-lezing beoogt hij een visie te schetsen op de interactie tussen technologie en cognitie, gebaseerd zowel op eigen theoretisch en empirisch werk als op de beschikbare literatuur. Wie zich door lectuur enigszins op Dr. Salomon's ORD-lezing wil voorbereiden raden wij de lectuur aan van de volgende twee artikelen:

- Salomon, G. & Gardner, H. (1986). The computer as educator: Lessons from television research. *Educational Researcher*, 15, no. 5, 13-19.
- Salomon, G. & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. In E. De Corte (Ed.), *Acquisition and transfer of knowledge and cognitive skills*. (*International Journal of Educational Research*, 11, 623-637.)

Erik De Corte

Dr. Michael FULLAN

Michael Fullan is hoogleraar in de sociologie en dekaan van de Faculteit Onderwijskunde van de Universiteit van Toronto. Hij heeft vele functies vervuld: onderzoeker, consultant, trainer en beleidsadviseur op een breed scala van projecten waar onderwijskundige veranderingen werden doorgevoerd (schoolsystemen, lerarenorganisaties) zowel nationaal als internationaal. Hij is auteur van 'The Meaning of Educational Change' (OISE Press) en co-auteur van drie recente beleidsstudies voor de overheid van Ontario: 'Teacher Education in Ontario: Current Practice and Options for the Future' (samen met Connelly, Heller en Watson), 'The Supervisor Officer in Ontario: Current Practice and Recommendations for the Future' (samen met Paul, Williams, Allison, Walker en Watson), en 'Strategies for implementing Microcomputers in Schools' (samen met Miles and Anderson).

Kwaliteitsbepaling wetenschappelijke productie in de Onderwijskunde

J. Ax i.s.m. B. Creemers (namens Landelijk Overleg Onderwijskunde)

1. INLEIDING

Over de beoordeling van wetenschappelijke productie binnen een wetenschapsgebied, in casu de onderwijskunde, bestaan uiteenlopende opvattingen. Dit betreft zowel beoordelingsprocedures als beoordelingsmaatstaven. Naarmate beoordelingen in sterkere mate gehanteerd worden als beleidsinstrument van de overheid en er in sterkere mate consequenties aan de beoordeling verbonden worden in de sfeer van het alloceren van middelen, is de noodzaak sterker aanwezig overeenstemming te bereiken. Overeenstemming tussen de wetenschapsbeoefenaren, de beoordelende instantie en de overheid die conclusies verbindt aan de beoordelingen. Het is in dit verband van groot belang dat betrokken partijen de legitimiteit van elkaars standpunt niet bestrijden. Eerst dan is het mogelijk tot een algemeen geaccepteerde beoordelingssystematiek te komen. Het Landelijk Overlegorgaan Onderwijskunde (L.O.O.) wil via deze notitie een bijdrage leveren aan het bereiken van de gewenste overeenstemming. De bijdrage is in niet onbelangrijke mate ingegeven door de bezorgdheid van het L.O.O. over recentelijk gehouden beoordelingsronden.

Het L.O.O. hecht vooral waarde aan duidelijkheid omtrent de functie van beoordelingsprocedures en aan het respecteren van specifieke kenmerken van het eigen wetenschapsgebied. Deze twee aspecten worden behandeld in respectievelijk par. 2 en par. 3. In de par. 4 en 5 worden voorstellen gedaan voor een kwantitatief en een kwalitatief beoordelingssysteem.

2. DE FUNCTIE VAN BEOORDELEN WETENSCHAPPELIJKE PRODUCTIE

Het zou nuttig zijn in dit verband onderscheid aan te brengen tussen:

- a. een externe beoordeling aan de hand van algemene criteria, zodat een onderlinge vergelijking van verschillende onderzoeksgroepen mogelijk wordt gemaakt,
- b. een externe beoordeling om vast te kunnen stellen of onderzoeksgroepen voldoen aan bepaalde produktienormen,
- c. een gecombineerde intern-externe beoordeling van onderzoeksgroepen met het oogmerk de kwaliteit door te lichten en het doen van aanbevelingen ter verbetering van de kwaliteit.

Beoordelingen van het type a en b zullen vaak in combinatie voorkomen. Gezien de functie van dit type beoordeling is het van belang dat de gehanteerde systematiek

1. inzichtelijk en controleerbaar is (objectief toepasbaar)
2. praktisch goed uitvoerbaar is
3. kan steunen op een brede acceptatie.

In geval het gaat om type a- en b-beoordelingen dan is het wenselijk dat vooraf duidelijkheid bestaat over de normen die gehanteerd worden. Tot op heden ontbreekt het hieraan. In de praktijk blijkt men (de betreffende groep onderzoekers) in hoge mate in het ongewisse te verkeren over de uitkomst van externe beoordelingen, hetgeen onwenselijk is. Het zou immers zo moeten zijn dat onderzoekseenheden van te voren weten wat van hen verwacht mag worden, zodat het interne onderzoeksmanagement daarop kan worden afgestemd.

De combinatie van de drie eisen, zoals boven genoemd, maakt het nagenoeg onmogelijk dat er een wetenschappelijk-inhoudelijke beoordeling plaats vindt. Het systeem van kwantitatieve weging van produktie lijkt in dit verband (type a en b) het meest voor de hand liggend.

Het meest belangrijke waar het bij beoordelingen om zou moeten gaan is het verhogen van de onderzoekskwaliteit. Het wetenschappelijk forum vervult in deze een belangrijke functie. Toch kan het forum maar een beperkte rol spelen. Immers, forumdiscussies leiden niet tot een systematische beoordeling van onderzoekseenheden, doen geen (zelden) uitspraken over onderzoeksprogramma's en geven geen (zelden) aanbevelingen tot gewenste en haalbare bijstellingen e.d. Type c-beoordelingsprocedures kunnen echter niet volstaan met kwantitatieve maten en kunnen evenmin tot stand gebracht worden zonder nauw contact met de te beoordelen groep onderzoekers.

3. DE AARD VAN HET WETENSCHAPSGBIED: ONDERWIJSKUNDE

Onderwijskunde als wetenschap is gericht op het ontwikkelen van theorieën. Daarbij is zij nauw betrokken op de onderwijspraktijk en stelt zich tot doel een bijdrage te leveren aan de verbetering van de onderwijskwaliteit. Deze praktijk gerichtheid heeft in verband met de beoordelingsproblematiek de volgende consequenties:

- a. Grote betrokkenheid op het Nederlandse onderwijs leidt ertoe dat in veel gevallen internationale/buitenlandse tijdschriften niet het aangewezen communicatiemiddel zijn. Bovendien zal de georiënteerdheid op het Nederlandse onderwijs de toegang tot internationale fora belemmeren.
- b. Het ontwikkelen van praktisch toepasbare produkten moet als een noodzakelijke vorm van dienstbetoon aan de praktijk worden gezien en moet evenzeer noodzakelijk geacht worden voor voortgaande ontwikkelingen in wetenschappelijke zin.
- c. De aanwezigheid in niet strikt wetenschappelijke fora (vaktijdschriften, symposia e.d.) is noodzakelijk en moet als zodanig ook als een wetenschappelijke output worden beoordeeld.
- d. Een relatief groot aandeel 2e en 3e geldstroomonderzoek binnen het totale onderzoeksbudget. Veel wetenschappelijke output vindt plaats in de vorm van onderzoeksrapporten die meestal (vaak niet) hun weg naar het publiek vinden via uitgeverijen e.d. Gezien de huidige ontwikkelingen waarin er naar gestreefd wordt het onderzoek binnen verschillende geldstromen meer op elkaar af te stemmen (vergelijk ook het SVO-beleid in deze) is het van belang aan dit punt aandacht te besteden.

Als gevolg van deze vier factoren is het gewenst beoordelingscriteria te hanteren waarbij wetenschappelijke produkten in relatie tot a, b, c en d in principe alle gehonoreerd worden. Het is gewenst een zodanig systeem te ontwikkelen dat er in relatie tot a, b, c en d een evenwichtige honorering van wetenschappelijke produktie plaats vindt. Een al te grote mate van waardering voor publicaties in internationale wetenschappelijke tijdschriften doet te kort aan hetgeen onderzoeksgroepen in feite opleveren. Men kan de kwaliteit en kwantiteit van de onderwijskunde niet bepalen door na te gaan wat de redactietafel van respectabele internationaal georiënteerde uitgeverijen passeert.

4. VOORSTEL TOT EEN KWANTITATIEF BEOORDELINGSSYSTEEM

In het onderstaande worden een aantal voorstellen gedaan om te komen tot een systeem van kwantitatieve beoordeling van wetenschappelijke productie binnen de onderwijskunde. Het gaat hierbij om beoordelingen van het type a en b zoals in paragraaf 1 omschreven.

Voorgesteld wordt de volgende categorieën te hanteren

1. op het wetenschappelijk forum gerichte produkten
2. veldgerichte produkten
 - a. publicaties
 - b. praktisch toepasbare produkten
3. rapportage opdrachtonderzoek

Produkten binnen deze drie categorieën moeten in redelijke mate gehonoreerd worden, zodat aan groepen onderzoekers een zekere vrijheid kan worden gelaten. Wel zou het te overwegen zijn een zekere spreiding over de verschillende categorieën als een kwaliteitskriterium aan te merken. Dit lijkt een betere methodiek dan het hanteren van een sterk gedifferentieerd systeem van wegingsfactoren per categorie. Dat leidt onvermijdelijk tot verdringsverschijnselen. Producten binnen categorie 3 moeten gehonoreerd worden afhankelijk van de omstandigheid in hoeverre de subsidieaanvraag en het eindproduct beoordeeld zijn door deskundigen. Binnen het systeem van de SVO, ZWO en de WRR is daarin voorzien. Problematisch in dit verband zijn andere subsidieverstrekkers/opdrachtgevers. Voorts zou rekening gehouden moeten worden met de omvang van de studie. Het gaat daarbij alleen om het eindrapport.

Categorie 1: op wetenschappelijk forum gerichte producten

soort produkt	aantal punten
1. dissertatie	8
2. boek	8
3. bijdrage aan een boek in de vorm van een hoofdstuk of afzonderlijk onderdeel naar omvang	2-3
4. redigeren boek	2
5. tijdschriftartikel	2/4*
6. congresbijdrage (opname in bundel)	2/4*
7. begeleiding promovendi	1
8. deelname aan begeleidingscommissies etc.	1

* Voorgesteld wordt een onderscheid te maken naar wetenschappelijke status van het tijdschrift/congres. Er zou een lijst opgesteld moeten worden van a- en b- tijdschriften/congressen. In principe moet dit geschieden los van de vraag of het gaat om Nederlandse of buitenlandse tijdschriften/congressen.

Categorie 2a: praktijk/veldgerichte producten

soort produkt	aantal punten
1. boek + b.v. opleiding/onderwijs	8
2. tijdschriftartikel	1
3. bijdrage congres/symposia	1

Categorie 2b: Praktisch toepasbare producten (De bedoeling is hiervoor een door SVO ontwikkelde indeling te hanteren. Is reeds aangevraagd, maar nog niet beschikbaar)

Categorie 3: opdrachtonderzoek

opdrachtgever	aantal punten
1. SVO/ZWO	6
2. WRR	6
3. Ministeries	6
4. OECD	6

Over het vaststellen van een minimum norm van wat verwacht mag worden dient de hoeveelheid onderzoeksformatie als uitgangspunt genomen te worden. Een gemiddelde van 2 punten per 0,5 onderzoeksformatie per jaar.

In gevallen waar sprake is van meer dan één auteur binnen eenzelfde onderzoeksgroep is er voor de toekenning van punten geen probleem. Betreft het auteurs uit verschillende onderzoeksgroepen (te beoordelen eenheden) dan zijn er verschillende mogelijkheden.

1. Iedere auteur krijgt het totale aantal punten

2. De auteurs verdelen een aantal punten waarbij meegewogen wordt wie de hoofdauteur is. In geval er sprake is van twee auteurs verdient het aanbeveling de onder 1. genoemde mogelijkheid te kiezen.

5. KWALITATIEVE BEOORDELING

Kwantitatieve beoordelingen hebben slechts een beperkte waarde. Er is immers geen systeem van punten toekenning denkbaar waarbij aan de wetenschappelijke kwaliteit van de output van onderzoeksgroepen volledig recht wordt gedaan. Een kwantitatieve beoordeling geeft weinig aanknopingspunten voor een inhoudelijke evaluatie. Gevolg daarvan is dat kwantitatieve beoordelingen maar in heel beperkte mate kunnen bijdragen aan de inhoudelijke verbetering van het werk van onderzoeksgroepen. Daarom wordt een gecombineerde interne en externe beoordelende instantie voorgesteld. Het zal daarbij moeten gaan om een groep gerenommeerde vakdeskundigen, eventueel van internationale samenstelling. Het is niet wenselijk een dergelijke beoordelingsprocedure te binden aan nauwkeurig omschreven criteria. Evenmin is het de bedoeling dat een kwalitatieve beoordeling aanleiding geeft tot (re)allocatie beslissingen op landelijk niveau. De rapportage is primair bedoeld voor de beoordeelde instelling.

Congressen, Symposia, Cursussen, Trainingen.

CONGRES 'SCHOOL EN PREVENTIE'

De *Nederlandse Vereniging van Leerlingbegeleiders (NVL)* viert op 26 mei 1988 haar tienjarig bestaan met een jubileumcongres, gewijd aan het thema 'School en preventie: de plaats van de leerlingbegeleiding'. Het congres heeft plaats in 'De Vereeniging' te Nijmegen.

Sinds haar oprichting in 1978, pleit de NVL voor een begeleiding die in de structuur van scholen voor voortgezet onderwijs is opgenomen en die zich richt op alle leerlingen. Deze begeleiding, een taak van alle leerkrachten waaronder een aantal gespecialiseerde leerlingbegeleiders, helpt de jongeren beter en efficiënter te studeren, goede keuzen te maken, de eigen problemen op te lossen en, zo mogelijk, te voorkomen. Een van de kerntaken van de leerlingbegeleiding is preventie. Ook in het huidige WVC-beleid gaat preventie een steeds grotere rol spelen, daarom heeft de NVL het thema 'School en preventie' als centraal uitgangspunt van het congres gekozen.

Programma

In het ochtendgedeelte zullen diverse inleiders de ontwikkelingen vanuit verschillende invalshoeken beschrijven, onder andere:

- drs. W. J. Deetman, Minister van Onderwijs en Wetenschappen;
- dr. Clemens M. H. Hosman, Universitair Hoofddocent Preventieve Psychologie, K.U. Nijmegen;
- drs. Douwe de Vries, rector RSG 'de Meergronden', Almere;
- dr. Nathan Deen, medewerker van de Vakgroep Onderwijskunde R.U. Utrecht (voormalig hoofd van de sectie 'Leerlingbegeleiding');

's Middags komt de praktijk aan bod, in de vorm van: presentaties van scholen en projecten, begeleidingsmodellen, een informatiemarkt en forumdiscussie.

Voor meer informatie:

Annelies Spierenburg, congrescoördinator

Zevenblad 15

5345 KX Oss

tel. 04120-41547.

BEDRIJFSOPLEIDINGEN IN EUROPA

Corporate Training in European Perspective (CoTEP)

Op 30 september 1988 vindt aan de Universiteit Twente een internationale conferentie plaats over bedrijfsopleidingen in Europa. Directe aanleiding van de conferentie is het verschijnen van het onderzoeksrapport van de Faculteit der Toegepaste Onderwijskunde over Bedrijfsopleidingen in Nederland. Dit rapport is geschreven in opdracht van het Instituut voor Onderzoek van het Onderwijs, SVO. Het onderzoek duurde van 1986 tot 1988 en er hebben meerdere onderzoekers aan gewerkt.

Tijdens de uitvoering van het onderzoek bleek dat ook elders in Europa onderzoeken gaande waren over vergelijkbare onderwerpen. Op communautair niveau bestaat tot nu toe echter geen overzicht over de opleidingsinspanningen die bedrijven en instellingen zich getroosten ten behoeve van het menselijk talent in de organisatie.

De conferentie heeft ten doel een eerste uitwisseling mogelijk te maken van de resultaten van diverse landelijke onderzoeken in verschillende landen van de Europese Gemeenschap. Voorlopig wordt niet naar volledigheid gestreefd, maar het gaat om het multipliceren van de in verschillende EG-staten opgebouwde expertise inzake onderzoek naar bedrijfsopleidingen.

Wellicht is het mogelijk op deze manier een Europees forum te creëren waarin voor de Gemeenschap relevante inzichten omgezet kunnen worden in beleidsaanbevelingen voor het

stimuleren van de verdere opleiding van werkenden.

De volgende thema's zullen aan de orde komen middels presentaties van verschillende experts op het gebied van bedrijfsopleidingenonderzoek.

1. Statistische gegevens

Hoe omvangrijk is het terrein uitgedrukt in deelnemers, de kosten, het percentage van de bruto loonsom, het percentage van de totale werktijd?

2. Het aanbod

Welke opleidingen worden gegeven? Hangt dit samen met het bedrijfsbeleid? Welke relatie is er tussen inspanning en performance van de organisatie?

3. Trends

Welke ontwikkelingen zijn er te onderkennen op dit terrein? Neemt de inspanning toe? Neemt de proportie publiek geld toe? Neemt de samenwerking tussen particulier initiatief en overheid toe? Wordt de inspanning gezien tegen de achtergrond van de sociaal-economische ontwikkeling?

4. Knelpunten

Welke knelpunten zijn er te onderkennen aan beleid, beheer, organisatie, ontwikkeling, uitvoering en evaluatie van bedrijfsopleidingen?

De conferentie zal uitmonden in een congresbundel waarin de gepresenteerde papers zijn uitgewerkt tot hoofdstukken van het boek.

De doelgroep van de conferentie bestaat enerzijds uit personen die beleids- en onderzoeksmatig werk verrichten met betrekking tot functiegerichte educatie. Anderzijds is de conferentie interessant voor opleidingsmanagers van grote ondernemingen en instellingen en van particuliere opleidingsinstellingen.

De volgende sprekers zullen een presentatie verzorgen.

Frankrijk	- Dr. Philippe Méhaut Universiteit van Nancy.
Groot-Brittannië	- dr. Timothy Down Manpower Services Commission.
Bondsrepubliek Duitsland	- dr. Edgar Sauter Bundesinstitut für Berufsbildung (Berlijn) - dr. Burkart Sellin CDEFOP (Berlijn)
Nederland	- dr. C. Boeren Vereniging van Nederlandse Ondernemingen - drs. Martin Mulder Universiteit Twente, Toegepaste Onderwijskunde - drs. Netty Bentvelsen en drs. J. Akkerman Onderzoekcentrum Toegepaste Onderwijskunde

Naar aanleiding van deze conferentie zullen de proceedings worden gepubliceerd, die op 30 september voor de deelnemers beschikbaar zijn.

Nadere informatie te verkrijgen bij mevr. A. Manoch, OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede (tel: 053-893584).

VIJFDE NATIONAAL CONGRES DYSLEXIE**14 en 15 JUNI 1988**

Dinsdag 14 en woensdag 15 juni 1988 vindt in de Koninklijke Nederlandse Jaarbeurs te Utrecht het vijfde Congres Dyslexie plaats. Dit vijfde congres, georganiseerd door de Stichting Dyslexie, beoogt de wetenschappelijke kennis en de praktijk-kennis over dyslexie bij elkaar te brengen, om zo tot uitwisseling van kennis en ervaring te komen.

Het congres is bedoeld voor zowel wetenschappers die zich met onderzoek over dyslexie bezighouden, als dyslectici zelf, ouders van dyslectische kinderen, leerkrachten in het gewoon- en speciaal onderwijs, leraren in het voortgezet onderwijs, docenten aan pedagogische academies, hulpverleners van schooladviesdiensten enz.

De opzet van het congres zal breed zijn met als titel:

Dyslexie in het basis- en het voortgezet onderwijs.

Congres

Tijdens het congres wordt uiteengezet welke de principes van de behandeling (orthodidactiek) behoren te zijn en welke opvattingen de toets van de wetenschappelijke kritiek kunnen doorstaan. Tijdens een praktijkgerichte informatiemarkt en postersessies zal de deelnemer geïnformeerd worden over wat er beschikbaar is aan directe en praktische hulp ten behoeve van dyslectici.

Inschrijving

De inschrijving voor dit vijfde Congres Dyslexie start eind januari 1988 middels inschrijffolders. De inschrijving sluit op 15 maart 1988.

VOOR EEN WERELD VAN MORGEN EEN ONDERWIJS VANDAAG

Het Stedelijk Pedagogisch Centrum Antwerpen bestaat 25 jaar. Bij deze gelegenheid organiseert zij een onderwijsbeurs op 19, 20 en 21 mei 1988 in de stadsfeestzaal aan de Meir te Antwerpen rond het genoemde thema.

De **onderwijsbeurs** is gericht op de toekomstgerichte aspecten van het basis- en secundair onderwijs en wordt gehouden met medewerking van:

educatieve uitgeverijen;

firma's van educatief materieel en speelgoed;

firma's van didactisch materieel;

computerfirma's;

bankinstellingen met educatieve diensten;

het stedelijk onderwijs Antwerpen.

Er worden **vier symposia** ingericht.

De toekomst in onderwijsland.

Onderwijs/industrie naar een symbiose?

Voor een wereld van morgen een onderwijs vandaag.

De onderwijspolitieke actualiteit.

De sprekers komen uit verschillende universiteiten, grote bedrijven, werkgevers- en werknemersorganisatie, politieke partijen...

Er worden ook **tien workshops** georganiseerd rond pedagogische, onderwijskundige of didactische thema's.

De toegang is gratis.

Voor meer inlichtingen gelieve zich te wenden tot het Stedelijk Pedagogisch Centrum, Meerlenhoflaan 30,

2710 Antwerpen-Hoboken, tel. 03/830 01 11.

SOCIETY FOR MULTIVARIATE BEHAVIOR IN THE BEHAVIORAL SCIENCES
(SMABS)

Groningen, The Netherlands, December 18-21, 1988

The Society for Multivariate Analysis in the Behavioral Sciences will hold its biennial conference on the development and application of methods for multivariate data collection and analysis in the behavioral sciences in Groningen, The Netherlands, December 18-21, 1988. The Organization Committee invites you to present a paper. Topics on which papers are especially welcome include: further developments in, or applications of, structural equation models, loglinear models, latent trait models, graph theory, correspondence analysis, and data collection procedures. Abstracts, not exceeding 100 words, should be typewritten in English, and include the author's name, address, title, affiliation, telephone number, and, if available, electronic mail address. They should be sent to:

Organizing Committee SMABS, Dr. M.G.H. Jansen and Dr. W.H. van Schuur Department of
Statistics and Measurement Theory

Faculty of Social Sciences, University of Groningen

Oude Boteringestraat 23, 9712 GC Groningen, The Netherlands

tel. +31-50-636189/636184. BITNET address: WSCHUUR @ HGRRRUG5, EARN

Request for additional information can also be sent to the above address.

Mededelingen

Geldpremies voor overzichtsartikelen alfa- en gammawetenschappen

Auteurs van overzichtsartikelen op het terrein van de geesteswetenschappen en de gedrags- en maatschappijwetenschappen kunnen ook in 1988 in aanmerking komen voor een geldpremie.

Voor de bekroningen is er dit jaar voor elk der beide vakgebieden een bedrag van f 15.000 beschikbaar. De prijzen bestaan uit bedragen van f 2.000 à f 3.000. De toekenning gebeurt door de Commissie Geesteswetenschappen en de Sociaal-Wetenschappelijke Raad van de Koninklijke Nederlandse Akademie van Wetenschappen.

Onder geesteswetenschappen worden verstaan: taal- en letterkunde, historische wetenschappen, wijsbegeerte en godgeleerdheid.

Het begrip gedrags- en maatschappijwetenschappen wordt ruim opgevat: ook rechtswetenschap en economie vallen eronder.

Om voor bekroning in aanmerking te komen moet het artikel in 1987 of 1988 zijn verschenen in een Nederlands wetenschappelijk tijdschrift.

De volgende maatstaven worden aangelegd:

1. Het artikel moet inhoudelijk een aanmerkelijk segment van een erkende discipline bestrijken.
2. De hoofdlijnen van de ontwikkelingen in een bepaald tijdvak moeten erin worden beschreven, zonder dat er relevante literatuur of jurisprudentie is overgeslagen.
3. De auteur dient eigen uitgangspunten op methodologisch gebied en andere theoretische veronderstellingen van principiële aard ten minste impliciet duidelijk te maken.
4. Er dient te worden ingegaan op het theoretisch gehalte van de ontwikkelingen in (het segment van) de discipline.

Auteurs die voor bekroning in aanmerking wensen te komen, moeten hun artikel **vóór**

1 augustus 1988 inzenden aan:

Commissie Geesteswetenschappen,
resp. Sociaal-Wetenschappelijke Raad,
Herengracht 410-412,
1017 BX Amsterdam.

Proefschriften

W. van de Grift: **Implementatie van vernieuwingen: de rol van de schoolleider**

(Handelseditie: De rol van de schoolleider bij onderwijsvernieuwingen, VUGA Uitgeverij, 's Gravenhage, 1987).

Promotores: prof. dr. P. van den Broek

prof. dr. M. van de Vall

Referent: prof. dr. W. J. van der Linden

Plaats: Rijksuniversiteit Leiden, 9 september 1987

Korte samenvatting

Het boek bestaat uit drie delen. In deel I wordt ingegaan op het innovatieprobleem: Waarom zijn de resultaten van vernieuwingspogingen in het onderwijs zo vaak teleurstellend? De voorhanden zijnde onderzoeksliteratuur leert, dat met name de volgende zaken belangrijke verklaringen vormen voor dit probleem:

- de vernieuwingsdoelen zijn onduidelijk geformuleerd;
- het probleem waarvoor het vernieuwingsprogramma een oplossing beoogt te zijn, sluit niet, of niet goed, aan bij de feitelijke probleemsituatie;
- het evaluatie-onderzoek is niet goed uitgevoerd;
- het vernieuwingsprogramma is niet geschikt om de doelen te bereiken waarvoor het ontworpen is;
- het vernieuwingsprogramma is niet in praktijk gebracht.

In de laatstgenoemde verklaring wordt het innovatieprobleem verlegd naar het implementatieprobleem: Hoe komt het dat zoveel vernieuwingsprogramma's niet of slechts gedeeltelijk in praktijk gebracht worden? De oplossing voor het implementatieprobleem wordt blijkens de recente onderzoeksliteratuur gezocht in vier groepen van mogelijke oorzaken: macrosociopolitieke factoren; de kenmerken van het vernieuwingsprogramma; de kenmerken van de innoverende eenheid; en de kenmerken van innovatie-strategieën.

In deel II van het boek wordt de constructie beschreven van twee onderzoeksinstrumentaria op het terrein van innovatiestrategieën: de schoolleiderinterventielijst en de schoolleiderinterventieschaal.

De schoolleiderinterventielijst bestaat in zijn definitieve vorm uit zeven Likertschalen, die tezamen 63 items tellen. Met de zeven schalen van de schoolleiderinterventielijst wordt beoogd om vast te stellen in welke mate schoolleiders interventies ondernemen op de volgende terreinen:

- Het initiëren van de vernieuwing.
- Het onderhouden van externe contacten over de vernieuwing.
- Teambegeleiding aan onderwijsgevendendie met de vernieuwing werken.
- Individuele begeleiding aan onderwijsgevendendie met de vernieuwing werken.
- Klassikale begeleiding aan onderwijsgevendendie met de vernieuwing werken.
- Het evalueren en bewaken van het vernieuwingsproces.
- Het bijhouden van de leerlingvorderingen met de vernieuwing.

Deze zeven Likertschalen bleken zowel homogeen als stabiel te zijn bij schoolleiders uit basis- en voortgezet onderwijs. Ook werd er bij deze groepen schoolleiders evidentie gevonden voor de constructvaliditeit van dit instrumentarium.

De schoolleidersinterventieschaal bestaat uit 16 dichotome items, en is bedoeld om de begeleidingsvaardigheid van schoolleiders bij onderwijsvernieuwingen te meten. De schoolleiderinterventieschaal bleek niet alleen te voldoen aan de eisen die vanuit de klassieke testtheorie aan een schaal gesteld kunnen worden, maar ook aan de eisen die vanuit het Raschmodel aan een schaal gesteld kunnen worden.

In deel III van het boek wordt ingegaan op de rol van de schoolleider bij het invoeren van vernieuwingen in het onderwijs. Het gaat daarbij met name om de vraag: Wat doen schoolleiders in het basis- en voortgezet onderwijs om vernieuwingspogingen op hun school te begeleiden en te ondersteunen?

ONTVANGEN PUBLIKATIES

- Adviescommissie Onderzoek Minderheden (ACOM) (1986). Onderzoek minderheden in opdracht van de Rijksoverheid. 's-Gravenhage: Ministerie van Binnenlandse Zaken.
- Boekaerts, M. (1987). Psychologie van de leerling en het leerproces. Nijmegen: Dekker & Van de Vegt.
- Bollen, R. & Hopkins, D. (1987). School based review: towards a praxis. Leuven, Amersfoort: ACCO (International School Improvement Project, no. 3).
- Braster, J. F. A. (1987). De inspecteur en de school. Een onderzoek naar de relatie tussen de Rijksinspectie en schoolleiders van scholen voor voortgezet onderwijs. Rotterdam: Vakgroep Onderwijssociologie en Onderwijsbeleid.
- Gerris, J. R. M. & Van Acker, J. (Red.) (1987). Gezin: onderzoek en hulpverlening. Lisse: Swets & Zeitlinger (Gezinsonderzoek nr. 1).
- Graaf, P. M. de (1987). De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen. Nijmegen: ITS (Dissertatie).
- Jong-Gierveld, J. de & Zouwen, J. van der (Red.) (1987). De vragenlijst in het sociaal onderzoek. Een confrontatie van onderzoekspraktijk en -methodiek. Deventer: Van Loghum Slaterus.
- Lagerweij, N. A. J. & Vos, J. F. (Red.) (1987). Onderwijskunde, een inleiding. Groningen: Wolters-Noordhoff.
- Putten, C. M. van (1987). Leerlingen van het individueel beroepsonderwijs nader beschouwd. Leiden (Dissertatie).
- Roeland, Th. & Veenman, J. (1987). Minderheden in Nederland. Achtergrondstudie 1987 bij het rapportagesysteem Toegankelijkheid en Evenredigheid. Rotterdam: ISEO.
- Symposiumbundel ORD 1987. (1987). De ontwikkeling van de leesvaardigheid in longitudinaal perspectief. Nijmegen: Instituut voor Onderwijskunde.
- Verijdt, H. & Diederens, J. (1987). Determinanten van beroepskeuze. 's-Gravenhage: Organisatie voor Strategisch Arbeidsmarktonderzoek.
- Wester, F. (1987). Strategieën voor kwalitatief onderzoek. Muiderberg: Coutinho.
- Wijers, G. A. (1987). Een ontwikkelingspsychologisch concept van beroepskeuze en arbeidsmarktprognoses. 's-Gravenhage: Organisatie voor Strategisch Arbeidsmarktonderzoek.
- Raad voor de Volwassenen-educatie (1987). Advies over het voortgezet algemeen volwassenenonderwijs. Den Haag: RVE.
- Koetsier, C. P. & Ten Brinke, J. S. (1987). Experiment individuele eindstage-1 op de afdeling Didactiek van de Biologie-analyse van de interviews met student-leerkrachten. Utrecht: Afdeling Didactiek van de Biologie, RUU.
- Koetsier, C. P., Van Driel, C. & Janzen, R. W. (1987). De begrippen exploreren en reflecteren nader bekeken... Utrecht: Afdeling Didactiek van de Biologie, RUU.
- SWIDOC (1987). Periodieke parade, oktober 1987.

Opsporen van Misconcepties bij Middelbare Scholieren

Henk Schmidt, Gerard Spaaij en Willem de Grave

Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Rijksuniversiteit Limburg

ABSTRACT

In two experiments, a method was tested for discovering misconceptions in students' knowledge and ideas about science subject-matter as taught in secondary school. The method consists of presenting small groups of students with a description of a set of natural phenomena. They are requested to produce as many explanations as possible for these phenomena and discuss these explanations critically. The discussions were audiotaped and verbatim protocols were screened for misconceptions. The method appeared to be successful. In Experiment 1 on osmosis, twelve qualitatively distinct misconceptions were uncovered. Experiment 2, on the behavior of a plane taking off, unraveled three misconceptions concerning aerodynamics. In both experiments, performances of novices and experts were compared. Only in the case where the subject-matter was explicitly part of the curriculum of the experts (osmosis), clear differences between novices and experts could be observed. Because differences in performance between novices and experts on the aerodynamics problem were largely absent, it is suggested that subjects have a poor ability to elaborate on relevant knowledge available in memory if they do not recognize a problem as a special case of something else. In addition to characteristics reviewed from the literature, three other attributes of misconceptions emerged from the experiments. First, misconceptions seem to be often not just available in memory for use when appropriate, but are actively constructed in response to a novel situation. Second, misconceptions appear to be highly metaphorical in nature. And third, while constructing a misconception, students tend to be mainly sensitive to the perceptual aspects of a situation.

Unlearning of preconceptions might prove to be the most determinative single factor in the acquisition and retention of subject-matter knowledge.

Ausubel, Novak en Hanesian (1978, p. 372)

De kennis die een leerling in eerdere leersituaties heeft opgedaan, lijkt in belangrijke mate te bepalen wat die leerling in daaropvolgend onderwijs nog leren kan en daarmee op zijn of haar leerprestaties. Voorkennis blijkt in veel onderzoek verantwoordelijk voor ongeveer 30 tot 60% van de variantie in leerprestaties. Zo rapporteert Lodewijks (1981) dat voorkennis in een door hem uitgevoerd onderwijsexperiment 36% van de variantie op een natoets verklaarde. Weeda (1982) vindt in één van de condities van zijn onderzoek een correlatie van 0,82 tussen de resultaten op een voortoets en die op een natoets (gelijk aan 67,8% verklaarde variantie). In eigen onderzoek waarin voorkennis bij leerlingen door middel van groepsdiscussie geactiveerd werd, verklaarde deze manipulatie 24% van de variantie in de data (Schmidt, 1984).

Uit laboratoriumonderzoek naar dit fenomeen, kan worden afgeleid dat voorkennis zo'n sterke invloed op het leren uitoefent omdat het een kader verschaft waarbinnen nieuwe informatie betekenis gegeven kan worden (Bransford & Johnson, 1972, Pichert & Anderson,

De auteurs spreken hun dank uit aan Jos Moust, Maurice en Rita de Volder, Bert Kerkhofs, Irma Kokx, Els Boshuizen, Jan Beliën, Herman Nuy en de grote groep student-assistenten betrokken bij het onderzoeksproject 'Cognitieve en motivationele effecten van probleemgestuurd onderwijs', voor hun hulp en advies bij het voorbereiden en uitvoeren van de gerapporteerde experimenten.

Adres: Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

1977). Bestaande cognitieve structuren bieden aanknopingspunten voor de nieuwverworven informatie; de mate waarin nieuwe kennis in al aanwezige geïntegreerd kan worden, bepaalt hoeveel er geleerd wordt.

Voorkennis heeft echter niet in alle gevallen een positieve invloed op het leren, het openingscitaat van Ausubel et al. (1978) verwijst daar al naar. In sommige gevallen, en met betrekking tot sommige onderwerpen, houden leerlingen er ideeën op na die hen in feite hinderen bij het begrijpen van nieuwe informatie. Veel geciteerd is bijvoorbeeld dat nogal wat mensen de impliciete opvatting huldigen dat in een bewegend voorwerp een kracht aanwezig is die dat object als het ware 'voortbeweegt', terwijl dat volgens de mechanica-theorie van Newton onjuist is (Clement, 1982; Driver & Easley, 1978).

Over deze onjuiste impliciete kennis, 'naïve theorieën' of *misconceptions*, handelt deze bijdrage. Eerst zal een overzicht worden gegeven van opvattingen over het ontstaan van misconcepties, hun kenmerken en de oorzaken voor hun hardnekkigheid. Daarna zal een relatief nieuwe methode gepresenteerd worden waarmee misconcepties van leerlingen in een klassesituatie kunnen worden opgespoord: het probleemgestuurde groepsgesprek. Deze methode zal vervolgens gecontrasteerd worden met twee bestaande procedures. Tenslotte zullen resultaten van twee experimenten naar de bruikbaarheid van die aanpak worden besproken.

ONTSTAAN, KENMERKEN EN INSTANDHOUDING VAN MISCONCEPTIES

Piaget (1954) is wellicht de eerste geweest, die erop gewezen heeft dat mensen al op zeer jonge leeftijd verklaringen ontwikkelen voor verschijnselen om zich heen, verklaringen die strijdig kunnen zijn met wetenschappelijk geaccepteerde verklaringen. Zo wijten kinderen onweer vaak aan het op elkaar botsen van wolken, waarschijnlijk omdat hun eigen ervaringen met donderend lawaai zich beperken tot het tegen elkaar botsen van voorwerpen. DiSessa (1983) schrijft het ontstaan van deze verklaringsmodellen toe aan het streven van de mens de wereld rondom hem te ordenen op een wijze die hem in staat stelt de verschijnselen die zich aan hem voordoen een zinvolle betekenis te geven. Deze ordeningsbehoefte heeft een 'survival value', omdat mensen door betekenisgeving in staat gesteld worden op een adequate wijze te reageren op veranderingen die zich in hun wereld voordoen. Volgens Hunt (1971) bezitten mensen een intrinsieke behoefte om incongruentie tussen kennis die ze hebben van hun omgeving en uit die omgeving binnenkomende informatie op te heffen. Menselijke cognitie is er dus op gericht theorieën te construeren ter verklaring van de verschijnselen die worden waargenomen.

Afgaande op de aard van het verrichte onderzoek naar misconcepties zou haast geconcludeerd moeten worden dat die zich voornamelijk beperken tot onjuiste opvattingen over natuurverschijnselen¹. Het merendeel van het onderzoek vindt plaats in het domein van de natuurwetenschappen en richt zich vooral op ideeën die leerlingen erop na houden met betrekking tot verschijnselen als beweging, energie en zwaartekracht (Clement, 1979; Driver, 1973; Driver & Easley, 1978; Green, McCloskey & Caramazza, 1980; Gunstone & White, 1981; Lebouttet-Barrel, 1976; Vionnet, 1979). Een enkele auteur houdt zich met biologische onderwerpen bezig, zoals natuurlijk selectie (Brumby, 1984). Op zichzelf is deze belangstelling wel begrijpelijk, omdat juist op deze terreinen wetenschappelijke opvattingen zich sterk verwijderen van intuïtieve, zeg maar naïve, ideeën.

Misconcepties worden in de literatuur meestal betrekkelijk vaag en algemeen besproken; een heldere definitie ontbreekt veelal. Auteurs beperken zich vaak tot het geven van voorbeelden. Niettemin zijn er uit die voorbeelden toch wel enkele steeds terugkerende kenmerken van misconcepties af te leiden.

¹) In werkelijkheid houden mensen er natuurlijk misvattingen op na op zowat alle terreinen des levens. Voorbeelden van misconcepties omtrent de oorzaken van eigen en andermans gedrag levert het sociaal-psychologische attributieonderzoek.

- misconcepties ontwikkelen zich in een zeer vroeg stadium van iemands ontwikkeling, nog voordat formele instructie heeft plaatsgevonden,
- misconcepties ontwikkelen zich op basis van praktische ervaringen die de betrokkene met de betreffende verschijnselen opdoet,
- misconcepties wijken af van wetenschappelijk aanvaarde verklaringsmodellen,
- misconcepties hebben een betrekkelijk consistent karakter in de verschillende bevolkingsgroepen, leeftijdscategorieën en nationaliteiten; ze kunnen min of meer beschouwd worden als universalia,
- misconcepties zijn weinig gedifferentieerd van karakter, en
- misconcepties zijn stabiel en daardoor moeilijk te wijzigen. Champagne, Gunstone en Klopfer (1983) noemen als belangrijkste reden daarvoor dat dezelfde misconcepties in heel verschillende kennisgebieden een rol spelen.

Misconcepties kunnen dus beschouwd worden als concrete uitingvormen van dieperliggende – wellicht meer abstracte – cognitieve structuren, die niet domeinspecifiek zijn, maar in belangrijke mate ons perspectief zouden structureren op de werkelijkheid als geheel. Deze fundamentele opvattingen omtrent de werkelijkheid ontstaan min of meer 'spontaan', en in een vroeg stadium, in de dagelijkse omgang met de dingen, en zijn in strijd met wetenschappelijke conceptualisering van diezelfde werkelijkheid. De schijnbare resistentie van misconcepties tegen informatieoverdracht zou volgens deze redenering voortvloeien uit het feit dat de bestrijding van concrete misconcepties hun dieperliggende basis onaangetast laat. Een andere veel genoemde reden voor de stabiliteit van sommige misconcepties is het feit dat ze goed voldoen in het dagelijks leven van degene die ze bezit (Viennot, 1979). Waarom zou je je druk maken over het feit dat een veer en een steen dezelfde valsnelheid hebben, als in de werkelijkheid de veer (als gevolg van de luchtweerstand) de aarde later bereikt dan de steen. Een laatste reden die wordt aangevoerd voor de persistentie van misconcepties is dat de betrokkene zich niet bewust is van de ongerijmdheid van zijn ideeën met hoe de werkelijkheid in elkaar zit (Cohen, 1984).

METHODEN VOOR HET OPSPOREN VAN MISCONCEPTIES

Het zal uit het voorgaande duidelijk zijn dat een docent er in zijn onderwijs verstandig aan doet rekening te houden met de naïeve theorieën die zijn leerlingen erop na houden, wil hij cognitieve verandering tot stand brengen. Daarbij kan het van belang zijn dat hij weet *welke* misconcepties er vigeren over het onderwerp dat hij wil gaan behandelen. In de literatuur worden twee methoden beschreven om misconcepties op te sporen.

De eerste is die van het diepteinterview van één enkele of een paar studenten. De geïnterviewde krijgt daarbij meestal één of meer opgaven voorgelegd waarover hij gevraagd wordt *hardop* na te denken. De interviewer beperkt zich tot het stellen van verhelderende vragen (Gentner & Stevens, 1983; Nussbaum & Novak, 1976; Osborne & Gilbert, 1980).

Deze methode levert meestal wel gedetailleerd inzicht op omtrent de aard van de betreffende misconceptie, maar lijkt nauwelijks geschikt om een mogelijke *diversiteit* aan opvattingen binnen een groep op te sporen omdat veelal slechts een beperkt aantal proefpersonen ondervraagd wordt; noch lijkt ze erg bruikbaar voor toepassing in de praktijk van het onderwijs, omdat ze erg arbeidsintensief is. De tweede methode maakt gebruik van meerkeuze items. Proefpersonen worden daarbij geconfronteerd met opgaven, ieder voorzien van een aantal antwoordmogelijkheden. Misconcepties die in de onderzochte groep leven, worden daarbij als afleiders gepresenteerd (Doran, 1972; Strike & Posner, 1982). Alhoewel deze aanpak het in principe mogelijk maakt op een eenvoudige wijze grotere groepen leerlingen te onderzoeken, heeft ze ook een voor de hand liggend nadeel: de onderzoeker zal nooit andere misconcepties op het spoor komen dan die welke hij van te voren zelf al heeft bedacht.

In deze bijdrage wordt een methode van probleemanalyse in kleine groepen gepresenteerd en geëvalueerd, die docenten (en onderzoekers) kan helpen om gedetailleerd de misvattingen die er in een groep leerlingen leven, op te sporen. De veronderstelling daarbij is, dat via een probleemgestuurde discussie op een meer efficiënte wijze informatie kan worden verzameld over binnen een leeftijdsgroep levende onjuiste opvattingen dan met de bestaande methoden mogelijk is: groepsgewijze afname omzeilt de beperkingen die kleven aan individuele interviews, terwijl vrije discussie over een probleem misconcepties aan het licht kan brengen waaraan onderzoeker of docent niet al van te voren hebben gedacht. Daarmee lijken in principe de voordelen van beide bestaande methoden te worden gecombineerd, zonder de nadelen mede over te erven.

In twee experimenten is geprobeerd een indruk te krijgen van de bruikbaarheid van deze nieuwe aanpak. In die experimenten werd aan groepjes middelbare scholieren een verzameling natuurverschijnselen voorgelegd met als opdracht via discussie zoveel mogelijk verklaringen voor die verschijnselen te bedenken. Deze discussies werden op geluidsband opgenomen, en met behulp van protocolanalyse werden concepties opgespoord die deze leerlingen hebben van processen, principes of mechanismen die aan de verschijnselen ten grondslag liggen. Zowel leerlingen zonder, als leerlingen met specifieke voorkennis werden onderzocht. Deze groepen worden hier voor het gemak novieten en gevorderden genoemd; al zal straks blijken dat een dergelijk onderscheid soms wat al te eenvoudig is. Doel daarvan was, na te gaan in hoeverre de voorgestelde methode sensitief is voor de invloed van onderwijs. Immers, van gevorderden in een bepaald kennisdomein mag verwacht worden dat zij minder, en minder eenvoudige misconcepties in dat domein ten toon spreiden. De gevoeligheid van de gevolgde methode voor dergelijke verschillen kan beschouwd worden als een maat voor zijn validiteit.

Een tweede variabele die wellicht het vóórkomen van misconcepties in een bepaalde groep beïnvloedt, is de mate waarin de gevorderden gedurende het onderwijs expliciet in aanraking gekomen zijn met een bepaald onderwerp. Binnen de beperkte tijd die een onderwijsprogramma biedt, kan immers slechts aan bepaalde onderwerpen aandacht gegeven worden. Men veronderstelt dan dat het begrijpen van verwante onderwerpen door middel van transfer wordt gerealiseerd. Voor Experiment 1 werd daarom een biologisch onderwerp gekozen dat expliciet deel uitmaakt van het VWO-curriculum; voor Experiment 2 een aerodynamisch onderwerp, dat daarvan geen deel uitmaakt, maar waarvoor relevante kennis uit andere vakgebieden kan worden afgeleid.

EXPERIMENT 1

Misconcepties met betrekking tot osmose en diffuse

Methode

Proefpersonen. Proefpersonen waren 22 derdeklassers en 21 vierdeklassers van een atheneum in Maastricht. Gemiddelde leeftijd van de derdeklasse scholieren was 14,9 jaar; die van de vierdeklassers 16,1 jaar. De scholieren uit de derde klas hadden geen specifieke kennis van osmose en diffusie, alhoewel ze de beschikking hadden over algemene biologische kennis, waaronder globale kennis van de structuur van cellen. Ze worden hier verder novieten genoemd. De vierdeklassers hadden het onderwerp enige weken daarvoor gehad; zij waren de gevorderden in dit onderzoek. De scholieren ontvingen een vergoeding voor hun medewerking.

Materiaal. Het materiaal bestond uit een beschrijving van het volgende probleem: 'Een rode bloedcel (een rood bloedlichaampje) wordt onder een microscoop in zuiver water gebracht. De bloedcel zwelt in korte tijd op en springt uiteindelijk kapot. Een andere bloedcel wordt in een oplossing van zout in water gebracht. Hij schrompelt ineen. Hoe zijn deze verschijnselen te verklaren?'

Procedure. Novieten en gevorderden werden binnen hun eigen conditie aselekt in zes kleine groepen van ongeveer gelijke grootte ondergebracht. Aan elk van deze groepen werd een proefleider toegewezen. Deze legde kort uit wat er van de proefpersonen verlangd werd, aan de hand van een voorbeeld. Dit voorbeeld bestond uit een beschrijving van het gedrag van een plant die zuurstof afgeeft bij daglicht, maar niet in het donker. In een uit 400 woorden bestaande toelichting werd dit fenomeen geanalyseerd vanuit verschillende perspectieven en werd een aantal mogelijke verklaringen besproken. De proefleider betrok de scholieren actief bij zijn uitleg om er zeker van te zijn dat zij begrepen wat er straks van hen verlangd zou worden bij de analyse van het bloedcelprobleem. Hij benadrukte dat van hen verwacht werd verklaringen te bedenken voor het probleem en die verklaringen kritisch te bediscussiëren. Ze werden aangemoedigd alles wat er bij hen opkwam naar voren te brengen.

Deze introductie nam 5 tot 10 minuten in beslag. Vervolgens kregen de scholieren 15 minuten om verklaringen te bedenken voor het bloedcelprobleem. De proefleider fungeerde daarbij als gespreksleider, die de verschillende ideeën van tijd tot tijd samenvatte. Hij of zij had de opdracht te vermijden er een socratisch leergesprek van te maken. Er werd verwacht dat hij of zij op non-directieve wijze de productie van zoveel mogelijk verschillende ideeën aanmoedigde. De discussie werd op band opgenomen. De opname vormde de basis voor het onderzoek naar misconcepties, maar werd ook gebruikt om te kunnen controleren of de proefleider geen informatie over het probleem verschaft. De geluidsbanden werden na afloop uitgetypt. De resulterende protocollen vormden de basis voor de analyses. Geen van de groepen had meer dan 10 minuten nodig voor de analyse van het probleem (Minimum: 8 minuten).

Resultaten en discussie

Doel van het onderzoek was inzicht te krijgen in de aard van de verklaringen die novieten en gevorderden geven voor het bloedcelprobleem. Daartoe werden de protocollen van de discussies onderzocht op (partiële) verklaringen voor het gedrag van de bloedcel. Twee representatieve excerpten laten zien op welke wijze verklaringsmodellen in de protocollen verschijnen. Het eerste fragment is afkomstig uit een novietengroep, het tweede uit een groep met gevorderden².

Proefleider (P): 'Tja, wil iemand daar wat over zeggen?' (stilte gedurende enige seconden) 't Is moeilijk, he.'
Proefpersoon 1 (Pp1): 'Ja, het enige wat ik weet is dat er twee bloedlichaampjes zijn, een rode en een witte, he.'

²) Om de uitlatingen van de leerlingen in het betreffende experiment op hun waarde te kunnen schatten, is het wellicht nuttig een korte uitleg te verschaffen over het osmotisch proces dat aan het zwellen, barsten en krimpen van de bloedcel ten grondslag ligt. Als twee oplossingen van zout in water zijn gescheiden door een halfdoorlaatbare wand (of semipermeabele membraan), zullen de watermoleculen, als gevolg van de brownse beweging, vrij door de membraan heen bewegen, terwijl de zoutmoleculen er niet doorheen kunnen. Als de concentraties aan beide zijden van de membraan van elkaar verschillen, zal meer water van de kant van de kleinere concentratie vloeien naar de kant van de grotere concentratie dan omgekeerd, totdat beide concentraties gelijk zijn. Op moleculair niveau kan dat als volgt begrepen worden door een semipermeabele membraan voor te stellen waartegen, op elk moment en aan beide zijden, 100 moleculen botsen. Als aan één zijde 80 water- en 20 zoutmoleculen botsen, en aan de andere kant 50 water- en 50 zoutmoleculen, dan zal een netto hoeveelheid van 30 watermoleculen naar één kant vloeien totdat beide concentraties gelijk zijn. Daarom zal, wanneer een bloedcel met een concentratie equivalent 0,9 percent zout in zuiver water geplaatst wordt, osmose optreden in de richting van de cel. De cel zwelt daardoor op. En omdat in zuiver water geen equilibrium tussen concentraties in en buiten de cel bereikt kan worden (en omdat een bloedcel – in tegenstelling tot een plantencel – geen wand bezit die een tegendruk – turgor – uitoefent) barst hij uiteindelijk. Het krimpen wordt veroorzaakt door osmose van water uit de cel.

P: 'Ja.....'

Pp1: 'Dat was het enige.'

P: 'En hier gaat het blijkbaar over een rood bloedlichaampje; één van die twee. Kan iemand uitleggen wat ermee gebeurt?'

Pp2: 'In zout, daar droogt 'ie misschien uit.'

P: 'Ja...'

Pp2: 'Dat het daarom krimpt.'

P: 'Ja, je zegt...'

Pp2: 'Zout onttrekt... eh... onttrekt vocht, want dat is ook als je bijvoorbeeld een wijnvlek in een tafelkleed of zo iets hebt, dan gooi je er gelijk zout op, en dat neemt dat dan op, dus gaat het uit het tafellaken. Dus ik denk dat dan (...) bij het bloedlichaampje, als het in het zout is dat het zout ook het bloed onttrekt aan de bloedcel, en dat die daarom inkrimpt.'

[Excerpt uit PROTOCOL Novieten 3 (N3), illustratie van misconceptie 8, zie tabel 1]

Het volgende, constrasterende, fragment is afkomstig van een groep gevorderden.

Proefleider: 'Nou Roland, begin maar.'

Pp1: 'Als de bloedcel in zuiver water wordt gebracht, dan zit in de bloedcel een grotere concentratie zout dan in het water en diffundeert het water in de cel. Om de concentratie een beetje gelijk te krijgen; de concentratie in de bloedcel en in het water, daardoor zwelt hij op. De bloedcel heeft ook geen celwand; daarom springt 'ie kapot als hij teveel opzwelt. En als een rode bloedcel in een oplossing van zout in water wordt gebracht, dan zit in het water meer zout dan in de bloedcel, dan is daar de concentratie groter, dan diffundeert water uit de bloedcel en schrompelt hij in elkaar.'

[PROTOCOL Gevorderden 1 (G1)]

In tabel 1 zijn de voornaamste verklaringen samengevat die door novieten gegeven werden.

Nadere analyse van de geproduceerde verklaringen leidt tot de volgende tentatieve conclusies:

De meeste misconcepties worden slechts één of twee keer genoemd. Er is dus sprake van een aanzienlijke variatie in opvattingen omtrent de oorzaken van het gedrag van de bloedcel binnen en tussen de groepen.

De novieten hebben de neiging het bloedcelprobleem op te delen in kleinere eenheden en vervolgens verklaringen te bedenken die alleen maar van toepassing zijn op die kleinere eenheden. Ze concentreren zich bijvoorbeeld exclusief op het opzwellen van de cel en bedenken dan dat dat wellicht verklaard kan worden door aan te nemen dat de cel minuscule sponsjes bevat. Of de cel is maar alléén in een waterige omgeving en moet daarom teveel vocht opnemen waardoor hij noodzakelijkerwijze barst. Blijkbaar heeft het probleem voor de novieten een niveau van complexiteit waar ze niet goed mee overweg kunnen. Ze reduceren daarom het probleem tot een omvang die wel hanteerbaar lijkt. De gevorderden daarentegen passen een oplossingsstrategie toe die eruit bestaat een theorie te construeren die de verschillende fenomenen met behulp van één of twee principes verklaart, zie het hierboven weergegeven fragment. Deze bevinding komt overeen met resultaten van onderzoek naar het oplossen van medische diagnostische problemen, waarin bleek dat middelmatig ervaren medische studenten een beschrijving van klachten en symptomen van een patiënt opdeelden in afzonderlijke klachten en voor elk van de klachten 'minitheorietjes' ontwierpen, terwijl experts zochten naar concepten die de klachten in onderlinge samenhang konden verklaren (Boshuizen, Schmidt & Coughlin, 1987). Overigens is er onder de novieten een bewustzijn van de beperkingen van hun verklaringen. Ze stuiten in hun conversaties regelmatig op onverklaarbare zaken zoals:

Tabel 1

Naïeve concepties van de processen die aan het bloedcelprobleem ten grondslag liggen

Opzwellen

1. De cel is gevuld met kleine sponsjes die het water opzuigen
2. De cel neemt water op door onbekend mechanisme omdat de wand poreus is. Echter, de wand bevat kleppen die verhinderen dat het water er weer uitkomt
3. Rode bloedcellen vervoeren zuurstof. Cel onttrekt zuurstof aan het water en zwelt op
4. De cel bevat in vocht opgelost zout. De oplossing oefent druk uit op de wand groter dan de tegendruk die door zuiver water wordt uitgeoefend
5. De inname van water zet een onbekende chemische reactie in de cel op gang

Barsten

6. Bloedcellen nemen gewoonlijk slechts kleine hoeveelheden vocht op, omdat in het menselijk lichaam veel cellen aanwezig zijn. In dit voorbeeld is er maar één cel, die teveel water moet opnemen
7. Levende objecten hebben maar een beperkte levensduur

Inkrimpen

8. Water of ander vocht worden aan de cel onttrokken door de hygroscopische (wateraantrekkende) eigenschappen van het zout
9. Zout water oefent een grotere druk uit op de celwand dan de inhoud van de cel
10. Zout tast de wand aan door de moleculen van de wand te beschadigen. De cel gaat lekken
11. Het zout gaat de cel in en vreet de cel van binnen aan

Opzwellen en inkrimpen in combinatie

12. De cel bevat zout dat water aan de omgeving onttrekt vanwege zijn hygroscopische eigenschappen. Als het water in de omgeving echter een hogere concentratie zout bevat, dan zal vocht aan de cel onttrokken worden

Pp1: '...Maar gewoon, als je eet en zo, dat... daar zit ook zout in. Dat komt ook allemaal in het bloed en... dat 't dan in het lichaam niet.. niet doet...' [noot auteurs: bedoeld wordt: dat bloedcellen in het lichaam niet ineenschrompelen ondanks de aanwezigheid van zout.]
[PROTOCOL N3].

Ten derde suggereren de gegevens dat de concepties die door de novieten geproduceerd worden, grotendeels op algemene wereldkennis en prewetenschappelijke ideeën gebaseerd zijn, getuige het intensieve gebruik van metaforen, als ballonnen die worden opgeblazen tot ze barsten of lichamen die in zee drijven, en de afwezigheid van wetenschappelijke begrippen in hun conversatie. Novieten zoeken vooral naar bekende analogieën, gebaseerd op oppervlakenkenmerken van het probleem, en passen die vervolgens toe op de nieuwe situatie. Bijvoorbeeld: het zwellen van de cel en de aanwezigheid van water herinnert ze aan het gedrag van een spons, en dus postuleren ze kleine sponsjes in de cel (conceptie 1). Of ze proberen bekende, maar in deze situatie irrelevante, kenmerken van de elementen van het probleem aan elkaar te relateren, zoals in conceptie 8 waarin vermeende hygroscopische eigenschap van zout verantwoordelijk wordt gesteld voor het krimpen; alsof het zout niet *in oplossing* aanwezig is! Deze misconceptie is overigens zeer diep verankerd, want hij werd ook twee maal in de protocollen van de gevorderden gevonden:

Proefleider: 'Nou, kan jij er wat van zeggen, of Carien of zo?'

Pp1: 'Ik kan er van alles van zeggen.'

Pp2: 'Ja, ik denk, die rode bloedcel ja, of ja, hier bij dat zout ja, dan, zout heeft gewoon een sterkere osmose of zo, *en zout trekt water aan* [cursivering van de auteurs], dus dan gaat dat water uit die cel en dan schrompelt die ineen.'

[PROTOCOL G3]

Daarnaast zijn er ook voorbeelden terug te vinden van wat quasi-logisch denken genoemd kan worden. Conceptie 6 is daar een voorbeeld van: normaal zijn er veel bloedcellen in een omgeving aanwezig die allemaal maar een beetje van het werk hoeven te doen, maar omdat in het voorbeeld er maar één bloedcel aanwezig is wordt die natuurlijk overbelast! Curieus is een teleologische verklaring die door iemand geproduceerd werd:

Pp1: 'Wanneer nou iets alleen maar een bepaalde tijd leeft, ...net zoals een vis, die schiet kuit en dan gaat 'ie dood. Misschien is dit hetzelfde.' [Conceptie 7, PROTOCOL N1]

In contrast daarmee postuleren gevorderden een niet direct observeerbaar fysisch proces dat aan het probleem ten grondslag zou liggen, waarmee ze trachten de fenomenen te verklaren. Zij herkennen het probleem als een speciaal geval van een meer algemeen principe en passen dit principe toe om aan te tonen dat dat inderdaad het geval is. Tenslotte; hoewel de verklaringen van novieten veel elementen bevatten die nuttig zijn bij het begrijpen van het probleem, bijvoorbeeld de aanwezigheid van 'kleppen' in de celwand, zijn ze zonder uitzondering incorrect. Zelfs de enige verklaring waarin rekening gehouden werd met verschillen in druk aan beide zijden van de membraan die door de waterverplaatsing veroorzaakt worden (conceptie 12), is niettemin gebaseerd op de misconceptie dat de waterverplaatsing door de wateraantrekkende eigenschap van het zout veroorzaakt wordt.

EXPERIMENT 2

Misconcepties met betrekking tot aerodynamica

Doel van Experiment 2 was na te gaan of ook in een kwalitatief verschillend domein als de aerodynamica, de in Experiment 1 gebruikte methode bruikbaar is om misconcepties op te sporen. Ook hier wordt nagegaan welke misconcepties gegenereerd worden en of er verschillen optreden tussen leerlingen met een verschillend opleidingsniveau. Alhoewel het betreffende onderwerp niet expliciet behandeld was, zou verwacht kunnen worden dat gevorderden op basis van hun grotere kennis van de mechanica (die in hun studiejaar aan de orde komt) beter dan novieten kunnen afleiden welke processen voor de voorgelegde fenomenen verantwoordelijk zijn.

Methode

Proefpersonen. Proefpersonen waren 45 scholieren van een Maastrichts lyceum: 24 derdeklassers en 21 vierdeklassers, met dezelfde karakteristieken als de scholieren in Experiment 1.

Materiaal. Het materiaal bestond uit een probleembeschrijving met de volgende tekst: 'Men kan op Schiphol waarnemen hoe vliegtuigen, die verschillende tonnen wegen, opstijgen, blijkbaar zonder moeilijkheden. Hoe kan dit verschijnsel verklaard worden?'

Procedure. De procedure was identiek aan de in Experiment 1 gevolgde. De scholieren werden aselekt aan zes kleinere groepen toegedeeld (drie met novieten en drie met gevorderden) die een korte instructie ontvingen over wat er van hen bij de analyse van het experimentele probleem verwacht werd. Vervolgens werd gedurende ongeveer 10 minuten aan het probleem gewerkt. De discussie werd weer op band opgenomen en geanalyseerd.

Resultaten en discussie

De protocollen van de derde- en vierdeklassers zijn bij de discussie van het vliegtuigprobleem veel moeilijker van elkaar te onderscheiden, vermoedelijk omdat het onderwerp niet expliciet in het onderwijs aan de orde geweest is. Toch is dat wel wat verrassend, omdat verwacht werd dat vierdeklassers in ieder geval meer achtergrondkennis van de mechanica hebben, die hen in staat zou moeten stellen een oplossing te construeren. Als transfer van kennis optreedt, dan zouden de resultaten daarvan in de protocollen terug te vinden moeten zijn. Dit nu is nauwelijks het geval.

Veelal werd de discussie begonnen met het noemen van elementen die wellicht belangrijk zijn voor de verklaring, zonder echter die elementen met elkaar in verband te brengen: snelheid, lengte van de startbaan, draagvermogen van de vleugels, luchtdruk, luchtweerstand, vorm van de vleugels, stand van de vleugels. Het kostte de proefleiders enige moeite de groepen zo ver te krijgen dat ze een verklaring zochten in procestermen. Slechts in twee van de zes groepen, in één novietengroep en één gevorderdengroep, werd een min of meer juiste oplossing gepresenteerd. Naar het (achteraf gevraagde) oordeel van de proefleiders ging het in beide gevallen om groepen waarin leerlingen zaten die een hobby hadden, betrekking hebbend op vliegen en vliegtuigen. Hun kennis was dus opgebouwd los van de schoolse context. Hieronder volgt een fragment uit de discussie in de betreffende novietengroep.³:

Pp1: 'Ze moeten eerst als ze snelheid hebben met die vleugels, ...dan krijg je bovenop een soort, ja vacuum of een lagere luchtdruk. En dan wordt die luchtdruk [onder de vleugel, auteurs] hoger, dus dan wordt het vliegtuig omhoog geduwd.'

[PROTOCOL N2]

De betreffende proefpersoon maakt geen melding van de condities waaronder die onderdruk ontstaat, en daarover ontspint zich even verderop een woordenwisseling:

Pp2: 'De vleugel is onder toch ...holler, ja boller. Boven is 'ie plat. Als dan de lucht komt, dan wordt 'ie eronder, dan wordt 'ie aangetrokken en boven wordt 'ie weggetrokken. Ja, dan krijg je boven een vacuum en wordt zo omhoog getrokken.' [PROTOCOL N2]

Blijkbaar veronderstelt de spreker dat er zich onder de vleugel door de bolle vorm een ophoping van lucht voordoet, terwijl de lucht boven over het platte vlak vrijelijk kan wegstromen wat dan het vacuum veroorzaakt. Na enige verwarde kreten zegt vervolgens de oorspronkelijke leerling.

Pp1: 'Nee, het is net andersom.'

Proefleider (P): 'Net andersom?'

Pp1: 'Ja, want als 'ie van boven plat is en onder bol, dan krijg je onder een vacuum.'

P: 'Wat gebeurt er dan?'

Pp1: 'Dan gaat 'ie naar onderen.'

Anderen: 'Ha, ha, ha...'

[PROTOCOL N2]

³) Een verklaring waarom een vliegtuig kan opstijgen komt vereenvoudigd hierop neer: De vleugels zijn van boven bolvormig en van onderen afgeplat. Zodra een toestand van snelheid is bereikt, komt een luchtstroom op gang. Die luchtstroom moet over de vleugel een langere weg afleggen dan onder de vleugel; er treedt dus boven de vleugel een 'verdunning' van lucht moleculen op, vergeleken met de situatie onder de vleugel. Deze relatieve onderdruk achter op de vleugel is verantwoordelijk voor het opstijgen (en in de lucht blijven).

In één van de gevorderdengroepen wordt wel onmiddellijk de relatie gelegd tussen de vorm van de vleugel en het opstijgen van het vliegtuig:

Pp1: 'De vleugels zijn zo gebouwd dat ze van boven bol zijn en van onderen of recht, of een beetje hol. Dan, de snelheid boven de vleugels zal veel hoger zijn en daardoor ontstaat er een lagedrukgebied boven de vleugel, en onder ontstaat ... een hogedruk gebied omdat de lucht daar stilstaat of langzamer beweegt dan d'r boven. En door het verschil wordt het ding een beetje opgetild.'
[PROTOCOL G6]

Tabel 2 bevat in samengevatte vorm de naïeve concepties van zowel novieten als gevorderden over waarom een vliegtuig, hoewel zwaar, toch kans ziet op te stijgen.

Tabel 2

Naïeve concepties van de processen die aan het vliegtuigprobleem ten grondslag liggen.

1. de afduwmetafoor

De onder de vleugel doorgaande lucht wordt door naar beneden gerichte kleppen naar de grond geleid, en duwt daarmee het vliegtuig af.

2. de raketmetafoor

Door ontwikkeling van enorme kracht door de straalmotoren krijgt het toestel zoveel snelheid dat het aan de zwaartekracht ontsnapt.

3. de drijfmetafoor

De vleugels zijn van voren iets naar voren gericht (als een soort boeg), dan wel de vleugels zijn bol van onderen, waardoor de onder de vleugel komende lucht deze opduwt (zoals een speedboot op het water).

Ook bij dit probleem ziet men dat door scholieren zonder specifieke voorkennis niet zozeer een onderliggend model geconstrueerd wordt dat de verschijnselen kan verklaren, maar dat naar een oplossing gezocht wordt door een bekend verschijnsel te nemen dat oppervlakkige – perceptuele – kenmerken met het te verklaren verschijnsel gemeen heeft (een boot die vaart op het water; een raket of kanonskogel die wordt afgeschoten), om vervolgens dat bekende verschijnsel te transponeren naar de te verklaren situatie.

ALGEMENE DISCUSSIE

In dit artikel werden twee experimenten gerapporteerd naar de bruikbaarheid van een methode van probleemanalyse in een kleine groep voor het opsporen van misconcepties die leerlingen er in het middelbaar onderwijs op na houden met betrekking tot natuurwetenschappelijke fenomenen. In de inleiding werd vastgesteld dat voorkennis in belangrijke mate bepaalt wat iemand kan leren, omdat nieuwe informatie in bestaande geïntegreerd moet worden. Als die voorkennis echter bestaat uit misconcepties, dan kan het van belang zijn die misconcepties in een bepaald domein eerst gedetailleerd in kaart te brengen, opdat de docent in zijn onderwijs zal kunnen aansluiten bij die misvattingen, en methoden kan ontwerpen waarmee ze bestreden kunnen worden. Eveneens werd geconstateerd dat de bestaande methodologie op dit terrein, die zich concentreert op het diepgaand uitvragen van één of enkele leerlingen (Nussbaum & Novak, 1976) of het voorleggen van opgaven in meerkeuzevorm (Strike & Posner, 1982) wellicht onvoldoende zicht biedt op de mogelijke diversiteit van misconcepties binnen een bepaalde groep, en in de praktijk van het klassikale onderwijs moeilijk te hanteren lijkt. De hier

voorgestelde aanpak bestaat daarin dat een groep leerlingen een beschrijving van een aantal natuurfenomenen voorgelegd krijgt met de opdracht zoveel mogelijk verklaringen te bedenken voor die verschijnselen.

Afgaande op de hoeveelheid geproduceerde misconcepties, leek de methode succesvol te zijn. Voor het vliegtuigprobleem werden drie, en voor het bloedcelprobleem niet minder dan twaalf misconcepties opgespoord. Daaruit blijkt dat, in tegenstelling tot wat soms in de literatuur gesuggereerd wordt, de variëteit aan onjuiste opvattingen soms groot kan zijn. Onduidelijk is waarom het bloedcelprobleem zoveel meer misconcepties genereert.

Behalve de mogelijkheid van het kwantificeren van misconcepties in een groep, lijkt de methode vruchtbaar omdat ermee veel kwalitatieve informatie – vergelijkbaar met het diepgaand uitvragen van één enkele leerling – wordt verkregen. Zo blijken leerlingen bijvoorbeeld niet enkel wat ze ooit geleerd hebben of door ervaring weten te reproduceren, maar op een actieve wijze causale verklaringen te construeren die passen op het betreffende probleem. Novieten tonen daarbij de neiging het probleem op te breken in delen en verklaringen voor te stellen voor elk van de verschijnselen afzonderlijk, terwijl de gevorderden meestal één of twee algemene principes postuleren en dan trachten aan te tonen dat de beschreven verschijnselen concrete manifestaties van die principes zijn. Daarnaast suggereren de analyses van de verkregen protocollen dat misconcepties een uitgesproken metafoor karakter hebben. De scholieren zochten veelal naar bekende voorbeelden met dezelfde perceptuele kenmerken als de te verklaren verschijnselen en transposeerden vervolgens die bekende voorbeelden. Soms werd ook wel getracht de elementen uit het probleem op een directe wijze (dus zonder tussenkomst van een te postuleren onderliggend proces) aan elkaar te koppelen.

Nogal verrassend is dat de meeste aan het licht gebrachte misconcepties niet enkel oppervlakkig of imprecies zijn, maar echt onjuist. De vraag doet zich dan ook onmiddellijk voor hoe pertinent onjuiste opvattingen over hoe de wereld in elkaar steekt, toch een positieve bijdrage kunnen leveren aan het leren. Dat dat wel degelijk gebeurt, blijkt bijvoorbeeld uit onderzoek van Schmidt en Patel (1987): novieten die aan het bloedcelprobleem gewerkt hadden, begrepen toch meer van een daarna gepresenteerde tekst over osmose en diffusie, dan novieten die aan een neutraal probleem gewerkt hadden. Met andere woorden: hoe kunnen bestaande, onjuiste, maar diep verankerde opvattingen over hoe de wereld in elkaar steekt, überhaupt veranderen? De beantwoording van deze vraag vergt onderzoek waarbij de in deze bijdrage voorgestelde methode wellicht nuttige antwoorden zou kunnen verschaffen.

REFERENTIES

- Ausubel, D.P., Novak, I.D., & Hanesian, H. (1978). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart and Winston.
- Boshuizen, H.P.A., Schmidt, H.G., & Coughlin, L. D. (1987). *On-line representation of a clinical case and the development of expertise*. Paper presented at the meeting of the American Educational Research Association, Washington, DC.
- Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.
- Brumby, M.N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education*, 68, 493-503.
- Champagne, A.B., Gunstone, R.F., & Klopfer, L. E. (1983). Naive knowledge and science learning. *Research in Science and Technological Education*, 2, 173-187.
- Champagne, A.B., Klopfer, L.E., & Gunstone, R.F. (1982). Cognitive research and the design of science instruction. *Educational Psychologist*, 17, 31-53.
- Clement, J. (1979). Mapping a student's causal conceptions from a problem-solving protocol. In H. Lochhead & J. Clement (Eds.). *Cognitive Process Instruction*. Philadelphia, NJ: Franklin Institute Press.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 1, 66-71.
- Cohen, M. R. (1979). Where does the old man go? *The Science Teacher*, 8, 22-23.

- DiSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner, & A. L. Stevens (Eds.), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum.
- Doran, R. L. (1972). Misconceptions of selected science concepts held by elementary school students. *Journal of Research in Science Teaching*, 9, 127-137.
- Driver, R. (1973). *The Representation of conceptual frameworks in young adolescent science students*. Unpublished Ph.D. thesis. Chicago, IL: University of Illinois.
- Driver, R., & Easley, J.A. (1978). Pupils and paradigms: a review of literature related to concept development in adolescent science students. *Studies in Science Education*, 5, 61-84.
- Driver, R., & Easley, J.A. (1969). Autonomous dynamical thinking of young adolescent physics students. In J. A. Easley (Ed.), *The Use of Mathematics in Science Teaching*. Urbana, IL: University of Illinois.
- Gentner, D., & Stevens, A.L. (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.
- Green, B., McCloskey, M., & Caramazza, A. (1980). Curvilinear motion in the absence of external forces: naive beliefs about the motion of objects. *Science*, 210, 1139-1141.
- Gunstone, R. F., & White, R. T. (1981). Understanding of gravity. *Science Education*, 65, 291-300.
- Hunt, J. McV. (1971). Intrinsic motivation: information and circumstances. In H. M. Schroder & P. Suedfeld (Eds.), *Personality theory and information processing*. New York, NY: Ronald.
- Lebouttet-Barrell, L. (1976). Concepts of mechanics in young people: *Physics Education*, 11, 462-466.
- Lodewijks, J.G.L.C. (1981). *Leerstofsequenties*. Academisch proefschrift. Tilburg: Katholieke Hogeschool.
- Nussbaum, J. & Novak, J.D. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education*, 60, 535-550.
- Osborne, R.J., & Gilbert, J.K. (1980). A technique for exploring students' views of the world. *Physics Education*, 15, 376-379.
- Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.
- Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology*, 69, 305-315.
- Schmidt, H. G. (1984). Activatie van voorkennis en tekstverwerking. *Nederlands Tijdschrift voor de Psychologie*, 39, 335-347.
- Schmidt, H.G., & Patel, V.L. (1987). *Effects of knowledge activation through small-group discussion on the processing of science text*. Paper presented at the meeting of the American Educational Research Association, Washington, DC.
- Strike, K., & Posner, G. (1982). *Epistemological assumptions of College students: an initial report*. Paper presented at the Annual Convention of the Northeastern Education Research Association, Ellenville, N.Y.
- Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, 1, 205-221.
- Weeda, W. C. (1982). *Beheersingsleren: een model getoetst in de tijd*. Academisch proefschrift. Tilburg: Katholieke Hogeschool.

Manuscript ontvangen 6-4-1987. Definitieve versie ontvangen 25-1-1988.

Effecten van Klasseconsultatie op Curriculumimplementatie en Leerresultaten

J. Snippe

*RION, Instituut voor Onderwijsonderzoek,
Rijksuniversiteit Groningen.*

ABSTRACT

*In a (quasi-) experimental study** we examined the effectiveness of facilitator-strategies in improving the implementation of a curriculum as well as the learning results of pupils. Thirty teachers from the fifth grade of the elementary school implemented a new math curriculum. Three groups were formed, two experimental and one control group:

1. *Consultation group*: teachers attended meetings where they received information about the curriculum and discussed problems with fellow teachers. They also received classroom consultation: a classroom consultant observed their lessons and provided them with feedback on their behaviour.
2. *Information group*: teachers attended meetings without classroom consultation.
3. *Control group*: teachers implemented the curriculum without any support.

The results of the experiment show that the consultation group attained the highest implementation as measured by the content covered and the amount of time spent on the curriculum. The learning results of the pupils were also highest in the consultation group. The teachers from the experimental groups exhibited significantly more curriculum related behaviour than did the teachers of the control group. Classroom consultation appears to be an effective strategy to influence curriculum implementation as well as learning outcomes of pupils.

1. INLEIDING

De opbrengsten van onderwijsvernieuwingsprojecten zijn betrekkelijk teleurstellend te noemen. Er zijn diverse factoren aan te wijzen die hiertoe bijdragen (Van de Grift, 1987). Als een van de belangrijkste oorzaken wordt de gebrekkige implementatie van een onderwijsvernieuwing genoemd. Onder implementatie verstaan we het feitelijk gebruik van een vernieuwing in de praktijk (Fullan en Pomfret, 1977). Voor het welslagen van een onderwijsvernieuwingsproject is een goed verloop van het implementatieproces van cruciaal belang.

In dit artikel wordt verslag gedaan van een onderzoek waarin is nagegaan hoe de implementatie door middel van een gerichte begeleiding aan leerkrachten kan worden verbeterd. Hiertoe zijn in een veldexperiment twee begeleidingsinterventies met elkaar vergeleken op hun effectiviteit ten aanzien van de implementatie van een curriculum 'cijferend vermenigvuldigen' en ten aanzien van de leerprestaties.

2. THEORETISCHE ACHTERGRONDEN

Uit onderzoek naar implementatie van onderwijsvernieuwingen blijkt dat leerkrachten over het algemeen de vernieuwing niet of slechts gedeeltelijk uitvoeren zoals bedoeld.

Er zijn verschillende factoren aan te wijzen die bijdragen tot het ontstaan van een dergelijk implementatieprobleem (Van de Grift, 1987; Nijhof en Harskamp 1987). In de eerste plaats zijn er factoren in de context waarbinnen een onderwijsvernieuwing wordt geïntroduceerd. Een tweede belangrijke bron voor het ontstaan van implementatieproblemen wordt gevormd door

* Dit onderzoek was mogelijk dankzij een subsidie van het Instituut voor Onderzoek van het Onderwijs (SVO) te Den Haag, projectnummer 201.4226.

Adres: Westerhaven 15, Postbus 1286, 9701 BG GRONINGEN

de aard van de vernieuwing. De vaagheid die in vele onderwijsvernieuwingsprogramma's is te bespeuren omtrent de doelstelling en de wijze waarop implementatie dient te geschieden is niet bevorderlijk voor de implementatie (Creemers, 1987).

Als derde factor van belang voor implementatie noemen we de gebruiker van de vernieuwing, de leerkracht. In diverse onderzoeken is gewezen op het feit dat leerkrachten niet of in onvoldoende mate over de kennis en vaardigheden beschikken om een onderwijsvernieuwingsprogramma c.q. methode te gebruiken volgens de bedoelingen van de auteurs (Appelhof, 1979; Nijhof, 1978). Begeleiding van leerkrachten bij de invoering van methoden zou een belangrijke rol kunnen spelen in het voorkomen van implementatieproblemen.

De vierde en laatste belangrijke factor in het implementatieproces is de strategie waarmee een onderwijsvernieuwingsprogramma in de praktijk wordt geïntroduceerd. Hierbij denken we zowel aan innovatiestrategieën die door de diverse begeleidingsinstanties worden gehanteerd als aan concrete begeleidingsinterventies die door schoolleiders en/of schoolbegeleiders worden gebruikt (Van de Grift, 1987).

In het experiment waarvan we hier verslag doen staat de beïnvloeding van het gedrag van leerkrachten centraal als middel om implementatie te bevorderen. Alvorens over te gaan tot de beschrijving van dit experiment, gaan we eerst nader in op de inhoud en vormgeving van een doelmatige curriculumbegeleiding.

2.1. Inhoud van curriculumbegeleiding

Met een doelmatige curriculumbegeleiding wordt een zodanige beïnvloeding van het gedrag van de leerkracht beoogd dat deze in staat wordt gesteld tot het implementeren van het curriculum volgens de bedoelingen van de auteurs. Om vast te stellen welke leerkrachtgedragingen essentieel zijn voor de implementatie van het curriculum gaan we uit van de veronderstelling dat curriculuminnovatie uiteindelijk tot doel heeft te komen tot verbetering van de leerprestaties van leerlingen. We gaan daarom na welke componenten van het curriculum bij toepassing in de klassesituatie zullen leiden tot een verbetering van de leerprestaties van leerlingen.

Uit onderzoek is inmiddels naar voren gekomen, dat de hoeveelheid leerstof die wordt behandeld, de hoeveelheid tijd die hieraan wordt besteed, alsmede de mate waarin de leerlingen taakgericht bezig zijn tijdens de les, belangrijke factoren zijn voor het verklaren van verschillen in leerprestaties (Veenman, 1986; Harskamp & Suhre, 1986).

We leggen vervolgens de relatie met leerkrachtgedrag door na te gaan welke leerkrachtgedragingen leiden tot een grotere hoeveelheid behandelde leerstof, lestijd en tot een hoge taakgerichtheid van leerlingen. Rosenshine en Stevens (1986) noemen in dit verband het regelmatig controleren van leerlingen en hun leerprestaties, het duidelijk stellen van doelen voor leerlingen, duidelijkheid in het geven van opdrachten en het stellen van vragen aan leerlingen waarop directe feedback mogelijk is. Dergelijke leerkrachtgedragingen zijn gelegen in de voorwaardelijke sfeer en zijn voor de implementatie van elk curriculum noodzakelijk.

Daarnaast zijn er componenten aan te wijzen die specifiek voor het juiste gebruik van één bepaald curriculum zijn. Zij geven bij toepassing in de klas vorm aan het unieke en onderscheidende karakter van dit ene curriculum. In de moderne reken-wiskunde methoden bijvoorbeeld is het voeren van een leergesprek met de leerlingen naar aanleiding van een probleemsituatie een vorm van curriculumspecifiek leerkrachtgedrag dat essentieel is voor een juiste uitvoering van dergelijke methoden. Van deze curriculumspecifieke gedragingen is niet vast komen te staan dat ze bij toepassing in de klas leiden tot betere leerprestaties van leerlingen (zie Harskamp & Suhre, 1986).

De inhoud van een effectieve curriculumbegeleiding zal derhalve gericht moeten zijn op zowel algemeen onderwijskundige als curriculumspecifieke gedragingen.

2.2. Vormgeving van curriculumbegeleiding

In de jaren '70 heeft een aantal onderzoeken plaatsgevonden naar in-service training van leerkrachten. Door middel van het oefenen van geïsoleerde onderwijsvaardigheden bracht men gedragsveranderingen bij leerkrachten teweeg. Een voorbeeld hiervan is de minicursus (zie

bijvoorbeeld Veenman en Heeringa, 1977; de Vries, 1980).

Om de transfer van de geoefende vaardigheden naar de klaspraktijk te bevorderen, wordt de laatste tijd meer nadruk gelegd op een vorm van in-service training waarbij het geven van feedback aan leerkrachten over de toepassing van vaardigheden in de klas centraal staat (Tillema, 1986).

In een analyse van een aantal onderzoeken naar begeleiding van leerkrachten geven Joyce en Showers (1980) aan dat begeleiding van leerkrachten in de eigen classesituatie een onmisbare aanvulling op cursusbijeenkomsten is, indien het doel van de begeleiding het invoeren van een nieuw curriculum betreft. Het gaat dan om een intensievere vorm van begeleiding, klasseconsultatie genaamd, die tot doel heeft om aan de hand van gestructureerd observeren en het geven van feedback over het onderwijsgedrag van de leerkracht, te komen tot een verandering van het lesgeefgedrag in de richting die het curriculum vereist. Daar de leerkracht in de eigen classesituatie wordt geobserveerd, is het mogelijk om gerichte feedback te verschaffen over de mate waarin deze het bedoelde gedrag vertoont en over aanpassingen ten behoeve van de context waarin de leerkracht werkzaam is. Hoewel vooral in Amerika klasseconsultatie als begeleidingsmiddel sterk wordt gepropageerd, ontbreekt de empirische basis omtrent de effectiviteit ervan vrijwel geheel. Showers (1982) doet verslag van een experiment waarin klasseconsultatie als aanvullend begeleidingsmiddel is onderzocht op de training van onderwijsvaardigheden. Hiertoe bezochten twee groepen leerkrachten cursusbijeenkomsten waar men informatie ontving over nieuwe onderwijsvaardigheden. Na afloop ontving de ene groep leerkrachten klasseconsultatie; de andere groep fungeerde als controlegroep. Uit klasseobservaties bleek dat de klasseconsultatiegroep significant meer van de gewenste gedragingen vertoonde dan de controlegroep. Ten aanzien van de leerlingprestaties in beide groepen was er slechts sprake van een gering verschil. De onderzoeker concludeert dat klasseconsultatie een noodzakelijke aanvulling op de bijeenkomsten vormt, wil er sprake zijn van overdraagbaarheid van de geleerde vaardigheden naar de classesituatie. Ten aanzien van de opzet van een effectieve curriculumbegeleiding, kunnen we naar aanleiding van het bovenstaande twee begeleidingsvormen onderscheiden:

1. cursusbijeenkomsten met klasseconsultatie en
2. cursusbijeenkomsten zonder klasseconsultatie.

Wij voerden een experiment uit met deze twee begeleidingsvormen als experimentele condities en daarnaast een controlegroep zonder cursusbijeenkomsten.

De onderzoeksvragen voor dit experiment luiden als volgt:

1. Leidt de begeleiding van leerkrachten tot een grotere mate van implementatie van het curriculum dan geen begeleiding?
2. Leidt de begeleiding in de vorm van cursusbijeenkomsten, aangevuld met klasseconsultatie tot een grotere mate van implementatie van het curriculum dan cursusbijeenkomsten zonder klasseconsultatie?
3. Leidt begeleiding van leerkrachten tot hogere leerprestaties bij leerlingen dan geen begeleiding?
4. Leidt begeleiding in de vorm van cursusbijeenkomsten, aangevuld met klasseconsultatie tot hogere leerprestaties bij leerlingen dan cursusbijeenkomsten zonder klasseconsultatie?

3. ONDERZOEKSOPZET

Het onderzoek kent een (quasi-) experimenteel design. Leerkrachten van groep 5 van de basisschool voeren een nieuw curriculum 'cijferend vermenigvuldigen' in de klas in. Zij zijn verdeeld over drie begeleidingscondities, twee experimentele condities en een controleconditie. In de eerste experimentele conditie volgen leerkrachten cursusbijeenkomsten die worden aangevuld met klasseconsultatie (consultatiegroep). In de tweede experimentele conditie volgen leerkrachten eveneens cursusbijeenkomsten; er wordt geen klasseconsultatie toegevoegd (cususgroep). De leerkrachten uit de controlegroep ontvangen geen begeleiding bij de invoering van het curriculum. De onafhankelijke variabele wordt gevormd door de begeleidingsconditie en de afhankelijke

variabelen zijn respectievelijk de mate van implementatie van het curriculum en de leerprestaties van leerlingen.

3.1. Het curriculum

In het experiment wordt gebruik gemaakt van een curriculum 'cijferend vermenigvuldigen', dat binnen het RION is ontwikkeld. In de 20 lessen die dit curriculum telt, wordt op inzichtelijke wijze het cijferend vermenigvuldigen behandeld. Dit gebeurt volgens het principe van progressieve schematisering: vanuit concreet visueel weergegeven schema's van handeling ontwikkelt zich langzamerhand via het aanbrengen van verkortingen in de schema's het eindalgoritme. De rekenopgaven worden aan de leerlingen aangeboden in verhaalvorm. Tevens wordt de leerlingen geleerd om systematisch gebruik te maken van schattingen als controle op de uitgevoerde rekenbewerking.

Het werken met deze leergang is op een aantal punten verschillend van het werken met een traditionele rekenmethode. In de leergang heeft de instructie een interactief karakter: de leerstof wordt in de vorm van een leergesprek met de leerlingen aan de orde gesteld. In traditionele methoden heeft de leerkracht tijdens de instructie vaak een sterk docerende rol. Na de instructie werken de leerlingen met de leergang aan een aantal oefeningen die zeer gevarieerd zijn samengesteld, in tegenstelling tot traditionele methoden waar de zelfstandige verwerkingsopgaven per les gelijksoortig van aard zijn.

3.2. Inhoud van de begeleidingscondities

De leerkrachten uit de beide experimentele condities (de consultatiegroep en de cursusgroep) volgen vijf cursusbijeenkomsten in het tijdvak van drie maanden waarin het curriculum wordt uitgevoerd.

Tijdens deze bijeenkomsten worden de leerkrachten geïnformeerd over de achtergronden en inhoud van het curriculum en krijgen ze de gelegenheid om ervaringen uit te wisselen met collega-leerkrachten en knelpunten te bespreken bij de uitvoering van de leergang. In deze bijeenkomsten worden tevens de lessen van de leergang voorbesproken, waarbij aandacht wordt besteed aan de curriculumspecifieke gedragingen, namelijk het geven van interactieve instructie en het organiseren van zelfstandig werken van de leerlingen aan gevarieerde leertaken en aan de gedragingen in de voorwaardelijke sfeer, namelijk het optimaal besteden van de beschikbare lestijd, duidelijkheid in het geven van opdrachten, controleren van de leerlingen. Er worden voorbeelden gegeven van deze gedragingen die vervolgens worden besproken.

In de consultatiegroep worden de cursusbijeenkomsten aangevuld met een drietal klassebezoeken van de klasseconsultant. Deze observeert een les van de leerkracht aan de hand van een checklist die items bevat met betrekking tot het algemeen onderwijskundig handelen van de leerkracht tijdens de les en met betrekking tot de curriculumspecifieke gedragingen van de leerkracht.

Na afloop van de les vindt er een nabespreking met de leerkracht plaats, aan de hand van de punten uit de checklist. In overleg met de leerkracht worden aanbevelingen gedaan ter verbetering van eventuele knelpunten die zijn gesignaleerd. Het feedback-gesprek wordt afgesloten met een aantal concrete afspraken over de aanbevelingen die zijn gedaan.

Zowel de cursusbijeenkomsten als de klasseconsultaties zijn verzorgd door één onderzoeksmedewerker, die verbonden is aan het project. De leerkrachten uit de controlegroep hebben – evenals de overige leerkrachten – het curriculum inclusief de handleiding ontvangen. In een begeleidende brief is de technische gang van zaken rond het onderzoek toegelicht. Eens in de drie weken zijn ze telefonisch benaderd door de onderzoekers om de voortgang te bewaken. De controlegroep heeft geen inhoudelijke begeleiding ontvangen, noch was er sprake van systematische overdracht van kennis en vaardigheden.

3.3. Samenstelling van de steekproef

De leerkrachten zijn door de onderzoekers telefonisch benaderd met het verzoek om medewerking aan het onderzoek. In verband met het bezoeken van de cursusbijeenkomsten, heeft de werving

per regio plaatsgevonden.

De in totaal 30 leerkrachten zijn allen werkzaam in groep 5 van de basisschool en afkomstig uit het noorden van het land.

De leerkrachten die deelname aan het onderzoek hebben geweigerd, deden dit in de eerste plaats vanwege een gebrek aan tijd en in de tweede plaats vanwege de inhoud van het curriculum. (Men heeft geen interesse in het vak rekenen of men gebruikt al een moderne reken/wiskundemethode.) Het is derhalve niet ondenkbaar, dat er enige vorm van selectie heeft plaatsgevonden bij de samenstelling van de steekproef. Mogelijkerwijs zijn de leerkrachten uit de steekproef iets meer gemotiveerd om hun rekenonderwijs te vernieuwen dan de leerkrachten die hebben geweigerd.

De begeleidingscondities zijn willekeurig aan de regio's toegewezen.

4. INSTRUMENTEN

4.1 Implementatie-instrumenten

De eerste twee vragen voor dit experiment hebben betrekking op de invloed van (de soort) begeleiding op de mate van implementatie van het curriculum.

Om de afhankelijke variabele, de implementatie, te operationaliseren is in de eerste plaats een korte vragenlijst voor leerkrachten ontwikkeld, waarmee wordt nagegaan of de leergang daadwerkelijk in gebruik is genomen. Er wordt aan de leerkrachten gevraagd hoeveel lessen ze van de leergang hebben behandeld en hoeveel tijd ze aan de leergang hebben besteed. Deze vragenlijst is na afloop van het experiment door alle leerkrachten ingevuld.

In de tweede plaats is een observatiesysteem ontwikkeld om aan de hand hiervan klasse-observaties te verrichten. Het observatiesysteem bestaat uit een checklist met acht items (dichotoom) die zijn ondergebracht in de twee subgroepen. Voor de ene subgroep van curriculumspecifieke gedragingen wordt de somscore van 5 items berekend. De tweede subgroep van de checklist – de algemeen onderwijskundige gedragingen – behoort tot de gedragingen in de voorwaardelijke sfeer; hiertoe wordt de somscore van 3 items berekend. Tevens wordt het controlerend gedrag van de leerkracht in kaart gebracht door per tijdseenheid van 5 minuten te noteren of de leerkracht controlerend gedrag vertoont; per les wordt hier een somscore van berekend. Ook dit controlerend gedrag hoort tot de gedragingen in de voorwaardelijke sfeer.

Van de leerlingen wordt tijdens de les de niet-taakgerichtheid gescoord als indicator voor de actieve leertijd. Dit gebeurt door elke vijfde minuut te noteren hoeveel leerlingen er zichtbaar niet met de leertaak bezig zijn. Per les wordt dan het percentage niet-taakgerichtheid van de hele klas berekend.

De vier observatoren die zijn ingeschakeld voor de dataverzameling, hebben een training ontvangen in het hanteren van het observatiesysteem. (De onderzoeksmedewerker die de cursusbijeenkomsten en de klasseconsultaties verzorgde is niet betrokken geweest bij de observaties.) Teneinde de interbeoordelaarsbetrouwbaarheid te bepalen is per subgroepje van de gedragingen de Cohen's Kappa berekend. De resultaten van de vier observatoren zijn als volgt:

- curriculumspecifieke gedragingen: Kappa = .91
- algemeen onderwijskundige gedragingen: Kappa = .83
- controlerend gedrag: Kappa = .74

Er blijkt sprake van een redelijke mate van overeenstemming tussen de observatoren. Bij alle leerkrachten heeft in de laatste weken van het experiment een observatie plaatsgevonden.

4.2. Prestatietoetsen

De laatste twee vragen voor dit experiment hebben betrekking op de invloed van (de soort) begeleiding op de leerprestaties van leerlingen. Voor beantwoording van deze vragen is er bij de leerlingen een aantal prestatietoetsen afgenomen.

Om het beginniveau van de leerlingen te kunnen inschatten, is voorafgaand aan het experiment een toets algemene rekenvaardigheid van het CITO afgenomen (M3B, midden leerjaar 3, multiple choice) en een toets Redactie-opgaven die op het RION is ontwikkeld. Deze toets bevat tien

betrekkelijk eenvoudige vermenigvuldigings- en deelsommen (open items). De interne consistentie van deze toets is vastgesteld door de betrouwbaarheidscoëfficiënt alpha te berekenen; deze bedraagt . 71. De toets is redelijk consistent.

Na afloop van het experiment is opnieuw een CITO-toets algemene rekenvaardigheid afgenomen, (E3A, eind leerjaar 3, 40 multiple choice items) en een RION-toets. Deze toets bevat 15 redactie-opgaven (open items) die voor het merendeel betrekking hebben op het cijferend vermenigvuldigen. Bij de helft van de items wordt van de leerlingen gevraagd om een schatting van het antwoord op de opgave te geven. De belangrijkste doelstellingen van het curriculum zijn in deze toets vertegenwoordigd. De interne consistentie van deze toets is hoog (alpha = . 87). De afname van de toetsen heeft zowel bij de voor- als bij de nameting plaatsgevonden in twee gescheiden dagdelen.

5. RESULTATEN

De gegevens betreffende de implementatie van het curriculum, zijn geanalyseerd om de vraag naar een eventueel differentieel effect van de begeleidingsconditie op de mate van implementatie van het curriculum te beantwoorden.

In de eerste plaats is gekeken naar de gegevens van de vragenlijst over de uitvoering van het curriculum. In onderstaande tabel worden de antwoorden op de vraag naar het aantal lessen dat de leerkracht van het curriculum heeft behandeld en de hoeveelheid tijd die hieraan is besteed in gemiddelden weergegeven, met tussen haakjes de standaarddeviaties vermeld.

TABEL 1 *Hoeveelheid behandelde leerstof van het curriculum en hoeveelheid tijd, besteed aan het curriculum. Uitgedrukt in gemiddelden en standaarddeviaties.*

	gem. aantal behandelde lessen (maximaal 20)	gem. bestede lestijd (in minuten)
consultatiegroep (n=13)	17.7 (3.3)	1263.1 (311.4)
cursusgroep (n=10)	12.8 (3.3)	893.0 (207.0)
controlegroep (n=7)	11.4 (6.9)	777.9 (386.2)

Er is een statistisch significant verschil tussen de condities met betrekking tot het aantal behandelde lessen (F-waarde (2,27) = 6.2; p=.0). Ook het verschil ten aanzien van de hoeveelheid tijd die aan de leergang is besteed, is significant (F-waarde (2,27) = 7.4; p=.01). Het blijkt dat de consultatiegroep significant meer lessen vn het curriculum heeft behandeld en meer tijd eraan heeft besteed dan de overige condities. De controlegroep scoort het laagst op de behandelde lesstof en de bestede lestijd.

Vervolgens zijn de gegevens geanalyseerd die uit de observatie zijn verkregen. Het gaat hier om

TABEL 2: *Observatiescores per conditie, uitgedrukt in gemiddelden en standaarddeviaties.*

	curriculumspeci- fieke gedragingen (range 0-5)	algemeen onderw. kundige gedragin- gen (range 0-3)	controllerend gedrag (range 1-12)	niet-taakgerichtheid van leerlingen (range 1-100%)
consultatie groep (n=12)	3.4 (0.7)	1.0 (0.9)	4.2 (3.1)	17.6 (13.3)
cursusgroep (n=9)	3.6 (1.0)	1.6 (0.9)	4.7 (2.6)	17.1 (10.1)
controle- groep (n=4)	1.8 (1.3)	1.3 (1.3)	2.3 (2.1)	13.4 (8.3)

de curriculumspecifieke gedragingen, de algemeen onderwijskundige gedragingen en het controlerend gedrag van de leerkracht alsmede de niet-taakgerichtheid van de leerlingen.

In tabel 2 worden de observatiegegevens per conditie voor de onderscheiden categorieën vermeld, uitgedrukt in gemiddelde scores en standaarddeviaties. In totaal zijn er observatiegegevens van 25 leerkrachten beschikbaar.

Er is sprake van een significant verschil tussen de condities op de categorie curriculumspecifieke gedragingen. (F -waarde $(2,22) = 6.22$; $p = .0$). Dit verschil is gelegen tussen de consultatie- en cursusgroep enerzijds en de controlegroep anderzijds. De controlegroep scoort het laagst op de curriculumspecifieke gedragscategorie.

Samenvattend kunnen we ten aanzien van de eerste twee onderzoeksvragen concluderen dat begeleiding leidt tot een grotere mate van implementatie in de zin van hoeveelheid behandelde leerstof en lestijd. Er is tevens sprake van een extra effect van klasseconsultatie op de implementatie.

Begeleiding van leerkrachten leidt er eveneens toe dat zij meer curriculumspecifieke gedragingen vertonen en derhalve het curriculum meer uitvoeren in overeenstemming met de bedoelingen. Op dit punt is er geen extra effect van klasseconsultatie gebleken. Ten aanzien van de overige leerkrachtgedragingen zijn er geen significante verschillen tussen de begeleidingscondities geconstateerd, noch ten aanzien van de taakgerichtheid van leerlingen.

We gaan vervolgens na wat de invloed is van de begeleidingsconditie op de leerprestaties van leerlingen. Hiertoe beschouwen we eerst de relatie tussen de mate van implementatie en de leerresultaten van leerlingen. De mate van implementatie is in dit experiment geoperationaliseerd als de hoeveelheid behandelde leerstof en de hoeveelheid tijd die aan de leerstof is besteed en de mate waarin de leerkracht curriculumspecifieke en algemeen onderwijskundig gedrag vertoont.

Om zicht te krijgen op de samenhang tussen de implementatiefactoren enerzijds en de leerprestaties van de leerlingen anderzijds, zijn correlaties berekend. De resultaten worden in onderstaande tabel weergegeven. Hierbij is voor beide toetsen gecorrigeerd voor de resultaten van de leerlingen op de voormeting van de CITO-toets.

TABEL 3: *Partiële correlaties tussen implementatiefactoren en leerprestaties (gecorrigeerd voor CITO-voortoets)*

	natoets CITO	natoets Redactie-opgaven
aantal behandelde lessen uit de leergang	.19 ($p=.196$)	.65 ($p=.000$)
hoeveelheid tijd besteed aan de leergang	.23 ($p=.145$)	.55 ($p=.003$)
curriculumspecifieke gedragingen	.03 ($p=.447$)	-.08 ($p=.367$)
algemeen onderwijskundige gedragingen	-.14 ($p=.267$)	.30 ($p=.079$)
controlerend gedrag	.48 ($p=.010$)	.30 ($p=.084$)
niet-taakgerichtheid van leerlingen	-.33 ($p=.064$)	-.22 ($p=.154$)

Uit de tabel blijkt dat het controlerend gedrag van de leerkracht significant samenhangt met de (gecorrigeerde) CITO-natoetsscores. Met de (gecorrigeerde) scores op de natoets Redactie-opgaven hangen de hoeveelheid behandelde lessen uit de leergang en de hoeveelheid tijd die hieraan is besteed, significant samen.

Alvorens we deze implementatiefactoren in verband brengen met de leerprestaties van de leerlingen, gaan we eerst na of er verschillen zijn tussen de leerlingen op de toetsscores. We geven een overzicht van de gemiddelde klassescores op de voor- en natoetsen van de drie condities. In dit overzicht zijn ter vergelijking tevens de scores opgenomen van de leerlingen uit de landelijke steekproef die door het CITO als normgroep wordt gehanteerd.

TABEL 4: *Gemiddelde klassescores op de voor- en natoetsen van de drie condities, met tussen haakjes de standaarddeviaties.*

	voortoets CITO (40 items)	natoets CITO (40 items)	voortoets Redactie- opgaven (10 items)	natoets Redactie- opgaven (15 items)
consultatie groep (n=13)	28.8 (4.0)	26.5 (4.2)	4.6 (1.4)	8.0 (2.4)
cursusgroep (n=10)	29.1 (2.5)	23.4 (3.8)	4.6 (1.5)	6.4 (2.8)
controle- groep (n=6)	28.3 (3.9)	23.1 (4.5)	4.1 (0.8)	5.4 (3.2)
normgroep CITO voor- toets (687 leerlingen)	31.5 (7.1)			
normgroep CITO na- toets (852 leerlingen)		24.9 (7.8)		

Het blijkt dat onze steekproef een lagere startpositie had dan de landelijke steekproef van het CITO; dit verschil is ingehaald op de natoets. De moeilijkheidsgraad van de CITO-natoets ligt hoger dan van de CITO-voortoets.

Er zijn op de voormeting geen verschillen tussen de drie condities, noch op de voortoets Redactie-opgaven (F -waarde $(2,26) = .33$; $p = .723$), noch op de voortoets CITO (F -waarde $(2,26) = .11$; $p = .898$). Er is ook geen sprake van een significant interactie-effect tussen begeleidingsconditie en voormeting, noch voor de toets Redactie-opgaven, noch voor de CITO-toets. Vervolgens is getoetst of er verschillen in gemiddelde klassescores op de natoets bestaan die zijn toe te schrijven aan de onafhankelijke variabele 'begeleidingsconditie'. Met behulp van covariantie-analyse zijn eerst de klassescores op de natoets Redactie-opgaven als afhankelijke variabele genomen, waarbij de klassescores op de voortoets CITO als covariaat zijn ingevoerd. De resultaten van deze analyse geven een significant effect van begeleidingsconditie te zien op de natoets Redactie-opgaven (F -waarde $(2,22) = 3.47$; $p = .047$).

Deze analyse is eveneens uitgevoerd voor de klassescores op de natoets CITO en hier blijkt ook dat er een significant verschil bestaat dat terug te voeren is op de begeleidingsconditie (F -waarde $(2,22) = 8.10$; $p = .002$).

Samenvattend kunnen we ten aanzien van de derde en vierde onderzoeksvraag concluderen dat begeleiding van leerkrachten in de vorm van cursusbijeenkomsten aangevuld met klasseconsultatie leidt tot hogere leerprestaties van leerlingen, dan begeleiding zonder klasseconsultatie en geen

begeleiding. Het verschil in leerprestaties tussen de cursusgroep en de controlegroep is niet significant.

5.1. Verbanden tussen begeleidingsconditie, implementatie en leerresultaten.

Nu gebleken is dat er verschillen bestaan in leerprestaties voor de verschillende condities, is het zinvol om na te gaan via welke implementatiefactoren deze verschillen kunnen worden verklaard. Uit eerdere analyses was gebleken dat de drie condities van elkaar verschillen op twee belangrijke voorspellende variabelen ten aanzien van leerprestaties, namelijk de 'hoeveelheid tijd aan de leergang besteed' en 'het aantal behandelde lessen'. Het is derhalve zinvol om deze variabelen als controle variabelen in een analyse in te voeren.

Er is een covariantie-analyse uitgevoerd met de klassescores op de natoets Redactie-opgaven als afhankelijke variabele en de klassescores op de CITO-voortoets als covariaat, waarbij door middel van step-down toetsing de invloed van de controlevariabelen - de hoeveelheid tijd aan de leergang besteed en het aantal behandelde lessen - op de leerprestaties is uitgeschakeld.

De resultaten van deze analyse geven aan dat er geen significante verschillen meer zijn tussen de condities op de nameting van de RION-toets (F -waarde $(2,20) = .12$; $p = .889$). Het oorspronkelijk gevonden effect van de begeleidingsconditie op deze curriculumspecifieke natoets kan worden toegeschreven aan de invloed van de variabelen 'hoeveelheid tijd aan de leergang besteed' en 'hoeveelheid behandelde leerstof'.

Bovenstaande covariantie-analyse met step-down toetsing is eveneens uitgevoerd met de klassescores op de natoets CITO als afhankelijke variabele.

De resultaten van deze analyse geven een significant effect van begeleidingsconditie op de natoets CITO te zien. (F -waarde $(2,20) = 6.43$; $p = .006$). Hieruit blijkt dat na uitpartialisering van de variantie door de ingevoerde controlevariabelen (hoeveelheid tijd en behandelde leerstof) er nog een significant deel van de variantie wordt verklaard door de factor begeleidingsconditie. Dit effect kan niet worden verklaard met de variabelen die in het onderzoek zijn opgenomen. Daar de leerkrachten uit de verschillende condities niet verschillen op die gedragingen die voorspellende waarde bezitten ten aanzien van de leerprestaties van leerlingen, is het effect van conditie van daaruit niet te verklaren.

Het is echter mogelijk dat de leerkrachten (in de tijd waarin het experiment is uitgevoerd) meer leerstof uit de eigen methode hebben behandeld of op intensievere wijze met de eigen methode werkten. Dit is een factor die in het onderzoek niet is onderzocht. We hebben hier kennelijk te maken met een soort transfer-effect van het werken met het curriculum 'cijferend vermenigvuldigen' op de algemene rekenvaardigheid van de leerlingen. Dit transfer-effect is mogelijkwerwijs tot stand gekomen via leerkrachtgedragingen die we niet onderzochten. Het is eveneens mogelijk dat leerkrachten tijdens de klasseconsultaties zijn geweest op een aantal knelpunten in het lesgeefgedrag, welke bij opheffing tot verbetering van de resultaten van de leerlingen leiden.

6. CONCLUSIE EN AANBEVELINGEN

De resultaten van het experiment geven aan dat de leerkrachten die begeleiding hebben ontvangen in de vorm van cursusbijeenkomsten, aangevuld met klasseconsultatie, tot een hogere mate van implementatie van het curriculum komen in de zin van de hoeveelheid behandelde leerstof en de hoeveelheid bestede lestijd. In die zin draagt deze begeleidingsinterventie bij tot een hogere mate van implementatie en tevens tot hogere leerprestaties van de leerlingen.

Voorts blijkt dat leerkrachten die begeleiding hebben ontvangen tot een hogere mate van implementatie komen in die zin dat zij meer curriculumspecifieke gedragingen vertonen dan de controlegroep. Dit wijst erop dat door het gezamenlijk voorbereiden en bespreken van lessen bij de leerkrachten een duidelijker beeld is ontstaan van het gewenste gedrag volgens het curriculum, dan in de controlegroep. Daar er geen verschil in leerkrachtgedragingen bestaat tussen de consultatie- en de cursusgroep, kunnen we vooralsnog de stelling, dat klasseconsultatie via gerichte feed-back in de klas tot een grotere gedragsverandering bij leerkrachten leidt, niet ondersteunen.

Het effect van klasseconsultatie is, dat de leerkrachten meer lessen uitvoerden en meer tijd aan

aan de leergang besteedden. Het is mogelijk, dat de leerkrachten door de klasseconsultaties extra gemotiveerd raakten tot het uitvoeren van de lessen. Een andere verklaring is dat de leerkrachten de klasseconsultaties als vorm van controle op de uitvoering hebben beschouwd.

In een volgend experiment zou hierover meer informatie kunnen worden verschaft door twee condities op te nemen waarin klasseconsultatie wordt toegepast.

REFERENTIES

- Appelhof, P. (1979). *Begeleide onderwijsvernieuwing*. Tilburg: Zwijzen.
- Creemers, B.P.M. (1987). Formatieve curriculumevaluatie, een heroverweging van de CURVO-strategie. *Pedagogische Studiën*, 64, p. 59-66.
- Fullan, M. en A. Pomfret (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, Vol. 47, no. 1, p. 335-397.
- Grift, W. Van de (1987). *Implementatie van vernieuwingen: de rol van de schoolleider*. Dissertatie. Leiden: Rijksuniversiteit.
- Harskamp, E. en C. Suhre (1986). *Vergelijking van rekenmethoden in het basisonderwijs*. Eindrapport, SVO-project 9.4035. Groningen: RION.
- Joyce, B. en B. Showers (1980). Improving in-service training: the messages of research. *Educational Leadership*, februari, p. 379-385.
- Nijhof, W.J. (1978). *Interne differentiatie als een innovatie*. SVO-reeks, no. 9.
- Nijhof, W.J. en E. Harskamp (1987). Van curriculumontwerp naar curriculumeffect. In: Harskamp, E. en W.J. Nijhof (red.) *Curriculum: ontwerp, implementatie en evaluatie*. Bijdragen aan de Onderwijsresearch, 13. Lisse: Swets en Zeitlinger.
- Rosenshine, B. en R. Stevens (1986). Teaching functions. In: Wittrock, M.C. (ed) *Handbook of research on teaching*. Third edition. New York.
- Showers, B. (1982). *Transfer of training: the contribution of coaching*. Beverly: Oregon University, Eugene. Center for Educational Policy and Management, December.
- Snippe, J. (1986). *Effecten van klasseconsultatie op curriculum-implementatie en leerresultaten*. Subsidie-aanvraag OTG-curriculum. Groningen: RION.
- Snippe, J. en E. Harskamp (1987). *Effecten van in-service training op curriculum-implementatie en leerresultaten*. Eindrapport SVO-project 201.4226. Groningen: RION.
- Tillema, H.H. (1986). *Effectiviteit van nascholingstrainingen voor leerkrachten*. Onderzoeksaanvraag. Groninge: RION.
- Veenman, S. en J. Heeringa (1977). Ontwikkeling en evaluatie van de minikursus: Het organiseren van zelfstandig leren. *Pedagogische Studiën*, 54, p. 176-189.
- Veenman, S. en P. Lem (1986). Groepen en differentiëren in combinatieklassen. In: A. Reints en P. Span (red.) *Differentiatie in het onderwijs*. Bijdragen aan de Onderwijsresearch, no. 11. Lisse: Swets & Zeitlinger.
- Vries, A.K. de (1980). Inhouden en achtergronden van de in-service trainingscursussen in het GEON-project. *Pedagogische Studiën*, 59, p. 449-459

Manuscript ontvangen 29-4-1987

Definitieve versie ontvangen 16-3-1988

De relatie tussen leerlingkenmerken en schoolverloop bij IBO-leerlingen

D.A. van Damme, J.M.T. Eulderink-Mulder en H.E. de Jonge
Rijksuniversiteit Leiden, Vakgroep Onderwijspsychologie

ABSTRACT

In 1980 the IBO-schoolcareer-research-project (IBO=Individueel Beroeps Onderwijs) collected data on characteristics of students of individual junior vocational education. Five years later, in 1985, the schoolcareers of 1189 students of that sample were assessed and related to those characteristics. The most important finding is that social-emotional characteristics, in particular the factor dedication, are more strongly correlated with the student's schoolsuccesses than cognitive or background characteristics. It is therefore concluded that teachers in individual junior vocational education should give special efforts to motivating students. It is feared that negative effects on the motivation of IBO-students may result if plans to integrate individual junior vocational education within the schools for secondary education are carried through. This on account of the fact that, with the large number of students in these integrated schools, supervision will decrease. A plea is held to keep up the separate approach of IBO-students within the secondary education, an approach which is characterized by individuality.

1. INLEIDING EN VRAAGSTELLING

Het IBO-schoolcarrière-onderzoek is uitgevoerd in het kader van het Leidse IBO-project (Van den Broek en Sietaram, 1983) en is een vervolg op het leerling-bevolkingsonderzoek (Van Putten en Hendriks, 1982; Van Putten, 1987).

In dit onderzoek is van ruim 2000 leerlingen aan het begin van hun IBO-schoolloopbaan een aantal cognitieve, sociaal-emotionele, en achtergrondkenmerken vastgesteld. Van Putten (1985) vindt een verband tussen met name sociaal-emotionele kenmerken, zoals inzet, aandacht opeisend en storend gedrag en de doorstroom na 1 jaar IBO-onderwijs. Van deze sociaal-emotionele kenmerken blijkt het leerkrachtoordeel de meest voorspellende waarde te hebben.

Neuwahl (1979) vindt een verband tussen de indruk die docenten van leerlingen hebben en de prestaties van deze leerlingen bij de AVO-vakken.

In het verlengde van het onderzoek van Van Putten wordt in dit onderzoek de vraag centraal gesteld welk type leerlingkenmerk de meest voorspellende waarde heeft voor de schoolcarrière van IBO-leerlingen: de schoolcarrière is 5 jaar na aanvang vastgesteld.

Hierbij hebben wij de volgende verwachtingen:

1. De sociaal-emotionele leerlingkenmerken van de IBO-leerlingen zullen een sterker verband te zien geven met de schoolcarrière dan de cognitieve kenmerken.
2. Van deze sociaal-emotionele kenmerken zullen de leerkrachtoordelen de meest voorspellende waarde hebben.
3. Wat betreft de achtergrondkenmerken wordt verwacht dat deze een minder sterk verband met de schoolcarrière hebben dan de twee overige typen leerlingkenmerken.

2. ONDERZOEKSOPZET

2.1 Onderzoekspopulatie

In het leerlingbevolkingsonderzoek (Van Putten) werd een landelijk-representatieve, aselechte steekproef van IBO-opleidingen (1e stap: gestratificeerd naar provincie en schooltype) en van IBO-leerlingen (2e stap: bij enkele gemengde schooltypen gestratificeerd naar sexe), getrokken.

In de onderzoekspopulatie van dit onderzoek zijn die leerlingen uit het leerlingbevolkingsonderzoek (Van Putten, 1987) opgenomen van wie alle gegevens met betrekking tot de door ons geselecteerde beginkenmerken bekend zijn. Daarvan deed 87% van de scholen en 69% van de leerlingen aan ons onderzoek mee (zie tabel 2.1).

Tabel 2.1

Aantal scholen en leerlingen

gerealiseerde	steekproef B.Oz.		selectie C.Oz		onderzoeksgroep C.Oz	
	school	II	school	II	school	II
A.O.	15	227	13	166	12	141
H.N.O.	35	630	32	459	27	349
T.O.	38	819	35	613	32	551
S.G.	13	415	7	210	4	148
totaal	101	2091	87	1448	75	1189

(Lln = leerling, A.O. agrarisch onderwijs, H.N.O. = huishoud- en nijverheidsonderwijs, T.O. = technisch onderwijs, S.G. = scholengemeenschap; B.Oz = bevolkingsonderzoek, C.Oz = carrièreonderzoek).

2.2 Onderzoeksvariabelen en operationalisering

In het IBO-schoolcarrièreonderzoek wordt het begrip schoolcarrière gedefinieerd als de weg die de leerling volgt vanaf het moment dat hij tot deze IBO-opleiding is toegelaten tot het moment dat hij deze opleiding verlaat.

Een schoolcarrière kan op verscheidene manieren verlopen. Een veel voorkomende indeling van schoolcarrières is die van geslaagden enerzijds en falers anderzijds (Van der Ree, 1970). In de CBS-statistieken worden verschillende begripsomschrijvingen gegeven, waarvan de hiernavolgende voor het IBO-schoolcarrièreonderzoek van belang zijn:

- Doorstroom: leerlingen die in twee opeenvolgende jaren bij hetzelfde schooltype staan ingeschreven; leerlingen die naar een hoger leerjaar gaan en zittenblijvers.
- Uitstroom: leerlingen die in het nieuwe schooljaar niet meer bij hetzelfde schooltype als in het voorafgaande jaar staan ingeschreven; leerlingen die naar een hoger schooltype gaan, leerlingen die naar een lager schooltype gaan en leerlingen die van school af gaan.

Door de verschillende, uit literatuurstudie voortkomende gegevens, (Hermans, 1981; CBS, 1982; Van der Ree, 1970) met elkaar te combineren is in dit onderzoek tot de volgende indeling van schoolcarrières gekomen:

1. Falers: leerlingen die de IBO-opleiding zonder diploma verlaten om naar een lager of onbekend onderwijstype te gaan, die hun schoolloopbaan voortijdig beëindigen of die op deze IBO-opleiding een of meer keren zijn gedoubleerd.
2. Gelijktromers: leerlingen die de IBO-opleiding verlaten om naar een gelijkwaardige, andere IBO-opleiding te gaan.
3. Geslaagden: leerlingen die hun IBO-diploma binnen de gestelde tijd (4 jaar) behalen.
4. Opstroomers en profielverbeteraars: respectievelijk leerlingen die naar een hoger onderwijstype gaan voordat ze het IBO-diploma hebben behaald en leerlingen die een extra 4e jaar ten behoeve van profielverbetering, vakkenpakketverbreding of extra stage volgen.

De vier hierbovengenoemde schoolcarriëretypen worden in het IBO-schoolcarrièreonderzoek gerelateerd aan de volgende leerlingkenmerken (factorscores), welke zijn vastgesteld op het moment dat de betreffende leerlingen aan hun IBO-opleiding begonnen (Van Putten en Hendriks, 1982; Van Putten, 1987).

1. Cognitieve leerlingkenmerken
 - a. Schoolvorderingen rekenen (hb = hoofdbewerkingen, mm = maten en meten, tp = toepassingen)
 - b. Redeneervermogen (red.verm.)
 - c. Verbaal vermogen (verb.verm.)

- d. Werktempo
- e. Technisch lezen (t-lezen).
- 2. Sociaal-emotionele kenmerken
 - Op grond van leerling-uitspraken (II.) zijn de volgende leerlingkenmerken bepaald:
 - a. Motivatie voor school en leren (motiv.)
 - b. Relatie met medeleerlingen (medell.)
 - c. Faalangst.
 - Op grond van leerkracht-uitspraken (Ik.) zijn de volgende leerlingkenmerken bepaald:
 - d. Inzet en belangstelling (inzet.)
 - e. Storend gedrag (stoor.)
 - f. Afhankelijk gedrag (afhank.).
- 3. Achtergrondkenmerken
 - a. Vooropleiding van de leerling (nonderw.)
 - b. Opvatting van de ouders over de vervolgopleiding van de leerling (vervolg.)
 - c. Sociaal-economische status van het gezin: opleidings- en beroepsniveau van beide ouders (SES)
 - d. Afwezigheid van vader of moeder in het gezin (man./vrouw.)
 - e. Sociaal-economisch gezinstype: het al dan niet werkloos, in loondienst of zelfstandig zijn van een of beide ouders (gezin.)
 - f. Urbanisatiegraad van de woonplaats van de leerling (urbangn.)
 - g. Huidig schooltype; welke richting binnen het IBO (nschoolt.).

De cognitieve leerlingkenmerken zijn gemeten door middel van schriftelijke tests. De sociaal-emotionele- en de achtergrondkenmerken zijn door middel van schriftelijke vragenlijsten bepaald. Voor de verantwoording en de precieze definiëring van deze kenmerken wordt verwezen naar het onderzoeksrapport van Van Putten en Hendriks (1982) en de dissertatie van Van Putten (1987).

2.2.1. Het onderzoeksinstrument

Voor de verzameling van gegevens over de schoolcarrière is gekozen voor een schriftelijke vragenlijst met geprecodeerde, gesloten vragen. Deze lijst diende door de schoolleiding/administratie te worden ingevuld. De benodigde informatie kon grotendeels met behulp van het schoolarchief worden gevonden. Via deze vragenlijst is per leerling informatie verkregen over de volgende punten:

- jaar waarin het diploma is behaald
- eventuele aantal doublures
- jaar van door- of uitstroom
- al dan niet profielverbeteren (extra 4e jaar).

2.3 Analysetechnieken

De aldus verzamelde gegevens hebben een gemengd meetniveau:

- de schoolcarrière-gegevens van de IBO-leerlingen zijn van ordinaal meetniveau
- de leerlingkenmerken (factorscores) hebben een intervalniveau
- de achtergrondvariabelen hebben een nominaal meetniveau (met uitzondering van de SES met een ordinaal niveau).

Inzicht in de mogelijke verbanden tussen schoolcarrière en leerlingkenmerken wordt verkregen op de volgende manieren:

- gemiddelde scores en standaarddeviaties
- correlaties tussen schoolcarrière en afzonderlijke leerlingkenmerken
- discriminantanalyse (met het programma OVERALS) van de schoolcarrière op de verzameling leerlingkenmerken.

3. RESULTATEN

3.1 Carriëretypen

De variabele schoolcarrière kent vier categorieën. In tabel 3.1 zijn de aantallen leerlingen per categorie van schoolcarrière weergegeven.

Duidelijk is dat de geslaagden de grootste groep leerlingen vormen.

Tabel 3.1

Aantal en percentage leerlingen per type schoolcarrière

	aantal	percentage
1. falers	351	29,5
2. gelijkstromers	96	8,1
3. geslaagden	512	43,1
4. opstromers	230	19,3
totaal	1189	100

3.2 Factorscores per schoolcarrière

In tabel 3.2 worden de gemiddelden en de standaardafwijkingen van de factorscores gegeven, per schoolcarrière en voor de totale groep. Een hoge score betekent een goede prestatie, een positieve uitspraak en een gunstig oordeel.

Voor standaardscores op de factoren, zie Van Putten, 1987, p. 66, 67, 87, en 107.

Tabel 3.2 Gemiddelde (M) en standaardafwijking (Sd) van de factorscores per type schoolcarrière

schoolcarrière	M	Sd	M	Sd	M	Sd	M	Sd
	rek	hb	rek	mm	rek	tp		
1.	5.97	3.38	6.38	3.18	6.79	3.58		
2.	6.91	3.65	8.10	3.59	8.54	3.68		
3.	7.43	3.45	8.05	3.37	8.34	3.57		
4.	7.53	3.52	8.05	3.61	8.29	3.52		
	red. verm.		werktempo		verb.verm.*		t lezen*	
1.	7.03	3.64	8.39	3.60	6.24	2.71	8.30	3.73
2.	8.76	4.12	8.53	3.61	7.41	2.80	8.66	3.58
3.	8.36	3.68	8.93	3.38	7.18	2.84	8.48	3.47
4.	8.37	3.77	8.77	3.42	7.57	2.98	8.60	3.27
	motivati		medell		faalangst			
1.	7.67	3.98	7.40	3.87	8.41	3.63		
2.	8.05	4.16	7.56	4.10	7.56	3.90		
3.	8.36	3.97	7.97	4.19	8.27	3.89		
4.	8.81	3.86	7.60	4.14	8.44	3.91		
	inzet		stoort		afhank			
1.	6.41	3.82	6.79	4.13	6.57	4.05		
2.	6.28	3.85	5.80	4.14	6.43	3.86		
3.	8.39	3.83	8.16	4.00	8.12	4.01		
4.	9.22	3.46	8.77	3.58	8.72	3.69		

Schoolcarrière: 1. falers

N = 351

2. gelijkstromers

N = 96

3. geslaagden

N = 512

4. opstromers

N = 230

* N is anders (i.v.m. missing data): vv 1 = 344, 2 = 96, 3 = 510, 4 = 227; t lezen 1 = 348, 2 = 96, 3 = 509, 4 = 228.

Uit de tabel is op te maken dat gelijkstromers relatief meer last van faalangst hebben. Het leerkrachtoordeel valt voor de gelijkstromers negatiever uit dan voor de falers, terwijl deze gelijkstromers hoog scoren op de meeste cognitieve factoren. Wat het leerkrachtoordeel betreft valt ook het grote verschil tussen enerzijds de groepen 1 en 2 en anderzijds de groepen 3 en 4 op. Daarnaast valt op dat de falers op alle cognitieve factoren een slechte positie innemen.

3.3 Relatie tussen schoolcarrière en afzonderlijke factorscores

In tabel 3.3 zijn de correlaties weergegeven tussen schoolcarrière enerzijds, en de afzonderlijke factorscores/leerlingenkenmerken anderzijds.

Tabel 3.3

Correlaties schoolcarrière/leerlingenkenmerken

Cognitief		Sociaal-Emotioneel	
rekenen hb	.17	ll. motivat.	.10
rekenen mm	.19	medell.	.03
rekenen tp	.16	faalangst	.01
red. verm.	.13	lk. inzet	.28
verb. verm.	.16	stoor.	.20
werktempo	.05	afhank.	.21
t-lezen	.03		

(ll. = leerlinguitspraak, lk. = leerkrachtoordeel)

Uit de tabel blijkt dat de cognitieve factoren niet hoog correleren en dat daarvan rekenen (meten en maten) met .19 het hoogst correleert met de IBO-schoolcarrière.

Van de sociaal-emotionele kenmerken zijn het de leerkrachtoordelen die het sterkste verband te zien geven met schoolcarrière. De sociaal-emotionele (leerkracht) factoren geven een duidelijk sterker verband met schoolcarrière weer dan de cognitieve factoren.

Het was niet mogelijk correlaties te berekenen tussen schoolcarrière en de achtergrondvariabelen in verband met het nominale meetniveau van de achtergrondvariabelen.

In tabel 3.4 zijn de correlaties per schooltype weergegeven.

Tabel 3.4

Correlaties schoolcarrière/factorscores per schooltype

	A.O.	H.N.O.	T.O.	S.G.
Cognitief				
rekenen hb	.28	.15	.15	.30
rekenen mm	.29	.12	.24	.30
rekenen tp	.19	.21	.13	.31
red.verm.	.06	.11	.18	.22
verb.verm.	.21	.17	.10	.30
werktempo	.03	-.01	.06	.11
t-lezen	-.06	.05	.00	.02
Sociaal-emotioneel				
ll. motivat.	.05	.12	.09	.11
medell.	.05	.04	.01	.07
faalangst.	.14	.07	.04	.08
lk. inzet	.28	.27	.28	.22
stoor.	.20	.15	.19	.24
afhank.	.21	.14	.23	.27

(A.O. = agrarisch onderwijs, H.N.O. = huishoud- en nijverheidsonderwijs, T.O. = technisch onderwijs, S.G. = scholengemeenschap, ll. = leerlinguitspraak, lk. = leerkrachtoordeel).

Wat betreft de cognitieve factoren zijn het bij alle schooltypen de rekenscores en het verbaal vermogen die het sterkste verband te zien geven. Bij het technische onderwijs is het belang van verbaal vermogen iets minder groot.

Wat de sociaal-emotionele factoren betreft zijn het de leerkrachtoordelen die sterker correleren met schoolcarrière dan de leerlinguitspraken.

Bij de vergelijking van de cognitieve- met de sociaal-emotionele factoren valt op dat de leerkrachtoordelen hoger correleren dan de rekenscores. Uitzondering hierop vormen de scholengemeenschappen; hier is het relatieve belang van de cognitieve factoren (niet alleen rekenscores) groter; bij het agrarisch onderwijs zijn de scores op de cognitieve en sociaal-emotionele kenmerken gelijk.

3.4 Verband schoolcarrière en de leerlingkenmerken als groep.

In par. 3.3. is een overzicht weergegeven van het verband tussen schoolcarrière en de leerlingkenmerken, afzonderlijk.

Het is echter ook interessant het verband na te gaan tussen schoolcarrière enerzijds en alle leerlingkenmerken tezamen anderzijds.

Om dit verband na te gaan is het programma OVERALS gebruikt (Gifi, 1981; Verdegaal, 1985). OVERALS is een techniek voor de analyse van de relaties tussen twee of meer sets variabelen van een gemengd meetniveau. Per dimensie worden de variabelen gekwantificeerd.

OVERALS maakt een gewogen som van de variabelen in de sets op zo'n manier dat deze gewogen sommen maximaal correleren. Deze correlatie wordt uitgedrukt in de canonische correlatiecoëfficiënt.

Er worden in dit onderzoek 2 sets variabelen gerelateerd, te weten schoolcarrière (nominaal) enerzijds, en de factorscores (interval) en achtergrondvariabelen (nominaal) anderzijds. De samenhang wordt uitgedrukt in een canonische correlatie-coëfficiënt (per dimensie).

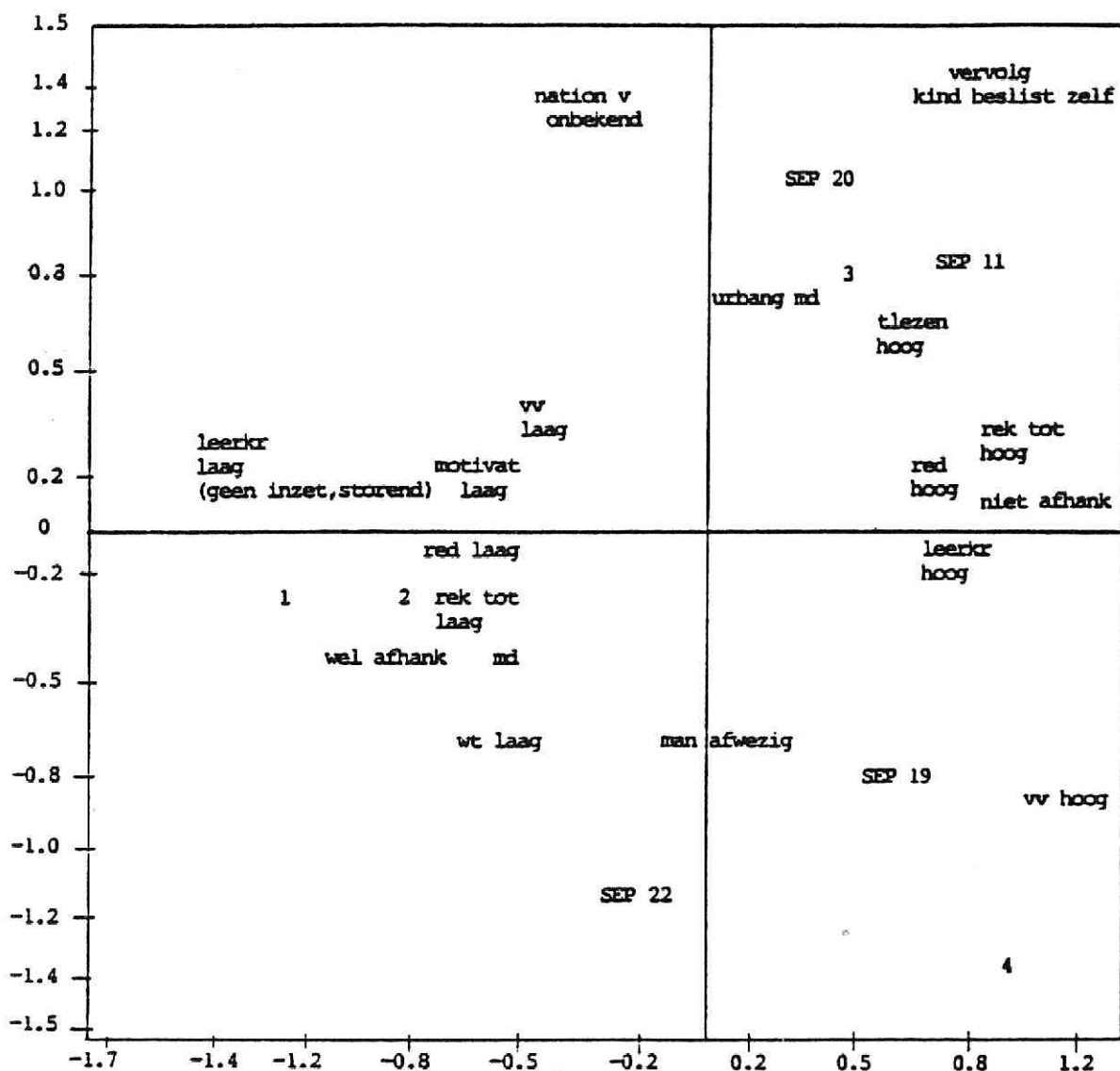
Tabel 3.5
Stress, canonische correlatie -OVERALS

stress	.64			
	dimensie	1	2	3
can. corr		.44	.38	.29

Uit tabel 3.5 is op te maken dat de canonische correlaties hoger zijn dan de hoogste bivariate correlatie uit tabel 3.3; deze was .28. Dit wil zeggen dat de factorscores samen meer zeggen (voorspellen) over schoolcarrière dan enige afzonderlijke variabele. Ook kan de aanwezigheid van achtergrondvariabelen, die bij deze techniek wel gebruikt kunnen worden, meespelen. Er is dan meer informatie op basis waarvan een voorspelling gedaan kan worden.

Ook het feit dat OVERALS meer vrijheid heeft in het quantificeren van de variabelen kan een correlatie-verhogende invloed hebben.

Uit figuur 3.1 zijn de kenmerken van de schoolcarriëretypen af te leiden, waarop deze het sterkst te onderscheiden zijn.



Figuur 3.1

Gemiddelde scores van de afzonderlijke variabelen

De groepen 1 en 2 hebben een laag leerkrachtoordeel (inzet, stoor, afhank), een laag redeneervermogen, lage rekenscores en een lage motivatie.

De groepen 3 en 4 scoren tegengesteld op deze factoren, dat wil zeggen hoog leerkrachtoordeel, hoog redeneervermogen, hoge rekenscores en een hoge motivatie. Een hoog verbaal vermogen is meer specifiek gebonden aan groep 4, - en een hoog technisch leesniveau aan groep 3.

Uit de figuur is op te maken dat ook SES van belang is. Hier zij echter opgemerkt dat deze variabele bij deze techniek nominaal geschaald is, waardoor dat moeilijk interpreteerbaar is.

4. SAMENVATTING EN DISCUSSIE

4.1 Samenvatting

De belangrijkste conclusie van het IBO-schoolcarrièreonderzoek is dat leerlingen die met een, volgens het oordeel van de leerkracht, hogere inzet de IBO-opleiding aanvangen, een grotere kans hebben deze opleiding zonder problemen af te maken.

De eerste hypothese wordt bevestigd: de sociaal-emotionele leerlingkenmerken geven een sterker verband te zien met schoolsucces dan de cognitieve kenmerken.

Ook de tweede hypothese kunnen we bevestigen: de kenmerken bepaald op grond van leerkrachtoordelen geven een sterker verband te zien met de schoolcarrière dan de kenmerken die op grond van leerlinguitspraken zijn bepaald.

Opvallend hierbij is de geringe mate van overeenstemming tussen leerkrachtoordelen over sociaal-emotionele kenmerken en de uitspraken van leerlingen op dit gebied. Zo is er een discrepantie ten aanzien van het meest opvallende sociaal-emotionele kenmerk motivatie en inzet. Het leerkrachtoordeel hierover wijkt af van de eigen inschatting van motivatie voor school en leren.

Anders dan verwacht blijken afzonderlijke correlaties van enkele cognitieve factoren af te wijken van de overige cognitieve factoren: de rekenscores geven relatief hogere correlaties te zien dan alle overige cognitieve factoren.

Ook de derde hypothese kunnen wij bevestigen: de achtergrondkenmerken geven inderdaad een geringer verband te zien met schoolsucces dan de 2 andere typen leerlingkenmerken. Van de achtergrondkenmerken blijkt de Sociaal Economische Status nog het meeste verband te vertonen.

Samenvattend blijken het leerkrachtoordeel over de inzet en, in mindere mate, de vastgestelde rekenvaardigheid, de IBO-leerlingen goed te kunnen onderscheiden in geslaagden en falers. Samen geven zij het sterkste verband te zien met schoolsucces.

4.2 Discussie

Alhoewel inzet ook voor andere, meer algemeen vormende schooltypen een belangrijk leerlingkenmerk zal zijn, geldt daar dat met name de schoolvorderingen op de basisschool een goede voorspelling kunnen geven van schoolsucces. De cognitieve leerlingkenmerken zullen daar wellicht een grotere rol spelen bij schoolsucces.

Uit de resultaten blijkt dat de rekenvaardigheid van een leerling op het IBO een belangrijke factor is: rekenen legt meer gewicht in de schaal dan verbale kwaliteiten. Een verklaring voor dit verschil kan zijn dat aan de rekenvaardigheid van de IBO-leerling een groter belang wordt toegekend in tegenstelling tot de eisen die aan het technisch leesniveau en de verbale intelligentie worden gesteld. Het is ook mogelijk dat scholen zich aanpassen aan het lage taalniveau van de gemiddelde IBO-leerling.

Het grote belang van de inzet in ons onderzoek kan verklaard worden uit de doelstelling van het IBO-onderwijs. Uit eerder onderzoek (SID, 1984) blijkt dat schoolleidingen van mening zijn dat het IBO primair beroepsvoorbereidend dient te zijn: men stelt meer prijs op vakbekwaamheid dan op kennis van schoolse zaken. In samenhang met de besproken resultaten van het carrièreonderzoek kan gesteld worden dat inzet, als aspect van de werkhouding in het algemeen, in verband met de toekomstige beroepsuitoefening eveneens belangrijker wordt geacht dan de meer cognitieve vaardigheden. Dit zou een mogelijke verklaring kunnen zijn voor het feit dat gelijkstromers ondanks goede cognitieve capaciteiten toch negatief beoordeeld worden.

De resultaten gelden ook voor de drie IBO-vormen afzonderlijk, maar niet voor de scholengemeenschappen: de cognitieve leerlingkenmerken zijn op dit schooltype belangrijker dan de sociaal-emotionele voor het schoolsucces van de IBO-leerling. Mogelijk heeft dit te maken met het feit dat op deze scholengemeenschappen meer doorstroommogelijkheden aanwezig zijn: men zal de leerling wellicht dan ook eerder beoordelen op kenmerken die met het cognitief functioneren verband houden. Uit onderzoek dat uitgevoerd is naar de redenen van voortijdig schoolverlaten blijkt overigens dat er vanuit de scholengemeenschappen meer leerlingen uit het IBO doorstromen naar een hoger onderwijstype (Van Damme, De Jonge en Mulder, 1985). Daarnaast zou ook de wat zakelijke sfeer op de doorgaans grotere en massalere scholengemeenschappen een verklaring kunnen bieden. De relatie tussen leerkracht en leerling wordt hierdoor wat afstandelijker dan op een kleinere school: wellicht dat er voor een individuele benadering, met aandacht voor de werkhouding van de leerling, minder ruimte is. Bij bovenstaande verklaringen is voorzichtigheid geboden: er is een relatief klein aantal scholengemeenschappen onderzocht. IBO-leerlingen hebben over het algemeen weinig te verwachten van hun cognitieve vaardigheden. Het zijn zwakke leerlingen, die tekorten kunnen compenseren met hun inzet. Een specifieke aanpak gericht op stimulering van de sociaal-emotionele kenmerken lijkt dan ook voor IBO-leerlingen meer wenselijk dan een benadering waarbij de, geringe, cognitieve kenmerken centraal staan.

Enige kanttekeningen ten aanzien van de onderzoeksresultaten lijken op zijn plaats. De geconstateerde discrepantie tussen de leerkrachtoordelen en de leerlinguitspraken ten aanzien van de motivatie voor school en leren wordt ook signaleerd door Van Putten (1987) als het gaat om IBO-leerlingen bij hun intree in het VO.

Mogelijke verklaringen kunnen gevonden worden in beide databronnen. Ten aanzien van het leerkrachtoordeel kan gesteld worden dat leerkrachten wellicht op grond van een kleine groep zeer ongemotiveerde leerlingen een wat stereotiepe wijze van beoordelen hanteren. Een te stereotiepe beoordeling kan een gedifferentieerde beoordeling van leerlingen verminderen, terwijl juist bij een individuele benadering dit van belang is. Ook in leerlinguitspraken kunnen verklaringen gezocht worden. Allereerst is het mogelijk dat de uitspraken van leerlingen gedeeltelijk sociaal wenselijke antwoorden zijn. Daarnaast bepalen de beleving en attitude ten opzichte van school en leren natuurlijk maar voor een deel het gedrag van de leerling. Gedrag staat ook onder invloed van de sfeer in de klas en de medeleerlingen.

Een tweede kanttekening betreft de invloed die het leerkrachtoordeel kan hebben op de prestaties van leerlingen en het gedrag. Een positief of negatief oordeel van de leerkracht over een leerling kan als een 'self-fulfilling prophecy' gaan werken: de leerling gaat zich gedragen overeenkomstig de verwachtingen die men van hem op school heeft.

In de plannen met betrekking tot de basisvorming in het Voortgezet Onderwijs blijft van de specifieke IBO-aanpak weinig over. Hoewel er in de plannen wel sprake is van een afwijkende leerroute voor zwakke leerlingen blijft het feit bestaan dat ook deze leerlingen de, overwegend cognitieve, eindtermen moeten zien te bereiken. Het onderwijsprogramma doet een sterk beroep op de verbale kwaliteiten van de leerling, wat, naast de omvang van 14 vakken, een extra hindernis vormt voor IBO-leerlingen.

Met dit onderzoek is geprobeerd aan te geven met welke gunstige, respectievelijk ongunstige, startcondities van de IBO-leerling rekening moet worden gehouden, wil deze, zonder verder gedemotiveerd te raken, toch een voortgezette opleiding kunnen volgen. Wij concluderen dat de ontwikkelingen in het Voortgezet Onderwijs, die wijzen in de richting van (door fusie) grotere scholen met een overwegend cognitief onderwijspakket, de faalervaringen van IBO-leerlingen zullen doen toenemen als er geen rekening wordt gehouden met die kwaliteiten waar juist deze leerlingen sterk in kunnen zijn: hun inzet en motivatie.

Bij het vaststellen van een aparte leerroute voor o.a. IBO-leerlingen lijkt het ons raadzaam niet alleen rekening te houden met het tempo waarin de leerling de stof verwerkt: in de plannen wordt nu alleen gesproken over een verlenging van de tijd waarin de leerling de eindtermen moet zien te bereiken. Ons inziens is naast deze maatregel ook een aanpassing qua aanpak gewenst. Een individuele benadering waarbij leerlingen aangesproken worden op hun werkhouding verkleint ons inziens de kans dat deze IBO-leerlingen opnieuw tussen wal en schip raken.

LITERATUURLIJST

- Baljet, M. & U. Sybesma (1984). Door- en uitstroom van leerlingen 1 jaar na aanvang vna het IBO. RU Leiden: Intern rapport, vakgroep Onderwijskunde/Licor.
- Burg, E. van der (1983). CANALS user's guide. Leiden: Rijksuniversiteit, Vakgroep Datatheorie.
- Centraal Bureau voor de statistiek (1982). Statistiek van het beroepsonderwijs 1980/'81, agrarisch onderwijs. 's-Gravenhage, Staatsuitgeverij.
- Centraal Bureau voor de statistiek (1982). Statistiek van het beroepsonderwijs 1980/'81, technisch onderwijs. 's-Gravenhage, Staatsuitgeverij.
- Centraal Bureau voor de statistiek (1982). Statistiek van het beroepsonderwijs 1980/'81, huishoud- en nijverheidsonderwijs. 's-Gravenhage, Staatsuitgeverij.
- Damme, D.A. van, H.E. de Jonge & J.M.T. Mulder (1986). IBO-schoolcarrière-onderzoek 1985. RU Leiden: Vakgroep Onderwijskunde.
- Geer, J.P. van de (1971). Introduction to multivariate analysis for the social sciences. San Francisco: Freeman.
- Gifi, A. (1981). Non-Linear to multivariate analysis. RU Leiden: Vakgroep Datatheorie.
- Hermans, J.J. (1981). Niet Voortgezet Onderwijs: voortijdig schoolverlaten in het Algemeen Voortgezet Onderwijs; omvang, aard en voorspelbaarheid. Lisse, Swets & Zeitlinger b.v.

- Neuwahl, N.M.E. (1979). Het IBO gevolgd, een onderzoek naar het individueel beroepsonderwijs en zijn leerlingen. Lisse, Swets & Zeitlinger b.v.
- Putten, C.M. van (1987). Leerlingen van het Individueel Beroepsonderwijs nader beschouwd. Academisch Proefschrift, RU Leiden.
- Putten, C.M. van (1985). Kenmerken van leerlingen die na 1 jaar het Individueel Beroepsonderwijs verlaten. In: K.B. Koster & N. Kwantes (red.), Schoolbegeleiding, leerproblemen en schooluitval (pp. 133-143) Bijdragen tot de Onderwijsresearchdagen 1984. Lisse, Swets & Zeitlinger.
- Putten, C.M. van, J.A.M. Hendriks (1982). Leerlingen in het VBO/MLK, IBO en LBO. RU Leiden: Vakgroep Onderwijskunde/Licor.
- Ree, E. van der (1970). Verslag van het onderzoek naar de schoolcarrière van leerlingen van het lager land- en tuinbouwonderwijs. Leiden.
- Samenwerkende ITO-direkties in Nederland (SID, 1984). Kinderen die niet goed kunnen leren hebben niet altijd gouden handjes. Delft.
- Verdegaal, R. (1985). Meer-sets analyse voor kwalitatieve gegevens. Datatheorie. Leiden.

Manuscript ontvangen 19-8-1987

Definitieve versie ontvangen 17-3-1988

Stress-situaties op school: Een onderzoek naar alledaagse schoolsituaties die door leerlingen als belastend worden waargenomen.

J. Hendriksen en M. Boekaerts

Onderwijspsychologie/I.S.O., Katholieke Universiteit Nijmegen.

ABSTRACT

Stressful situations in school.

This study concerns pupils' perceptions of school-related sources of stress. A card-sorting task was constructed to measure these perceptions in primary school pupils. The task consisted of 41 different daily school situations, derived from interviews with pupils and open-ended questionnaires. For each situation pupils had to indicate *how bad* they would feel in the respective situations. The card-sorting task was administered to a sample of 468 pupils. Principal Component Analysis revealed four reliable dimensions of stressful school situations: (1) identification with the stress of other pupils (2) social isolation, (3) confrontation with errors and failure, and (4) task overload. The results are in accordance with other research findings suggesting that schoolstress is multidimensional in nature. Further investigation in the study of schoolstress and coping behavior is proposed.

INLEIDING

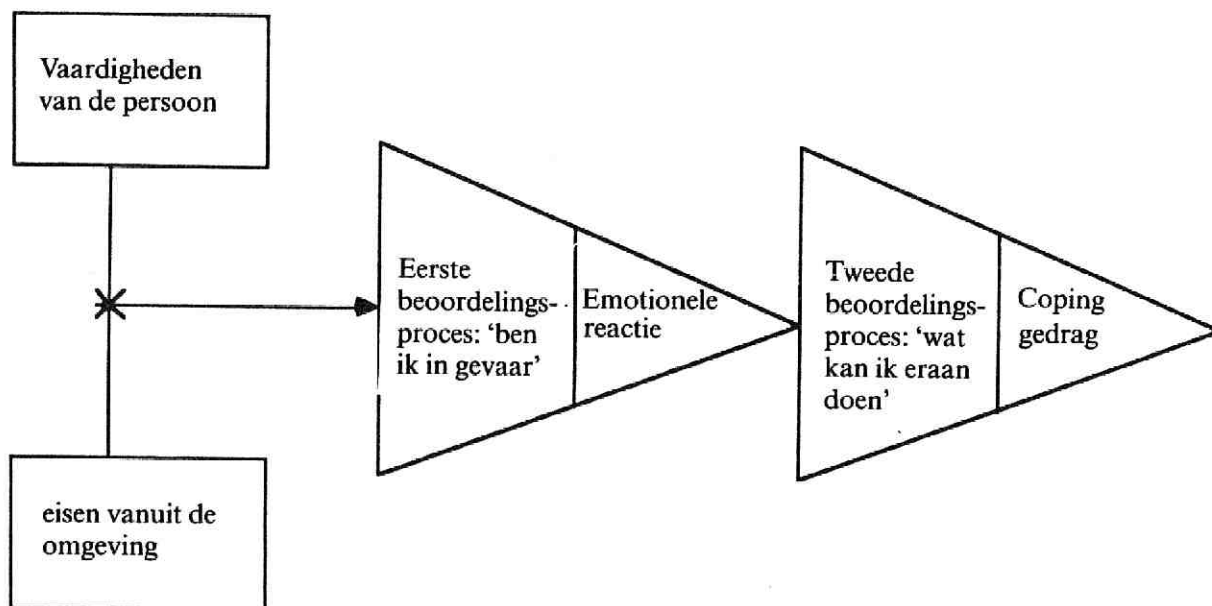
Op de vraag: 'wat is stress?' kan geen eenvoudig antwoord worden gegeven, aangezien er een veelheid van verschillende opvattingen en definities bestaat sinds dit verschijnsel in de jaren zestig in een groeiende wetenschappelijke belangstelling is komen te staan. Voor een goed begrip van stress is het van belang om in navolging van andere auteurs (o.a. Cox, 1978; Lazarus en Folkman, 1984) een drietal benaderingen te onderscheiden: nl. 1. een omgevingsbenadering, 2. een reactiebenadering, en 3. een interactiebenadering. In de eerste benadering wordt de rol van de omgeving centraal gesteld, in de tweede de rol van de persoon, terwijl de derde benadering de wisselwerking tussen persoon en omgeving benadrukt.

Binnen de omgevingsbenadering wordt stress gezien als de mate waarin aspecten van de *objectieve omgeving* een beroep doen op de aanpassingscapaciteiten van een persoon. Met name het onderzoek van Holmes en Rahe (1967) is van groot belang geweest voor deze benadering. Zij gingen na hoeveel sociale aanpassing ieder van 43 ingrijpende levensgebeurtenissen vereist op een schaal van 0 tot 100 (variërend van een kleine wetsovertreding tot het overlijden van de partner). Op basis daarvan ontwikkelden zij een lijst van belangrijke gebeurtenissen: de 'Social Readjustment Rating Scale'. Het bleek dat naarmate mensen meer gebeurtenissen hadden meegemaakt die om sociale aanpassing vroegen, er ook meer klachten waren m.b.t. gezondheid. Vervolgonderzoek heeft echter aangetoond dat dit verband tussen dramatische levensgebeurtenissen enerzijds en lichamelijke en psychische klachten anderzijds minder hoog is dan aanvankelijk werd aangenomen (o.a. Lazarus en Folkman, 1984).

De tweede stressbenadering, die vooral binnen de medisch/biologische wetenschap wordt voorgestaan, vat stress op als *de reactie van het lichaam* op bepaalde belastende omstandigheden. Hans Selye, de grondlegger van het stress onderzoek, ontdekte reeds in 1936 dat het lichaam bij extreme prikkels zoals sterke hitte of koude, een algemeen reactiepatroon vertoont, dat steeds op

dezelfde wijze verloopt (Selye, 1978). Na een aanvankelijke alarmreactie, waarbij tal van lichamelijke veranderingen optreden, wordt het lichaam vervolgens gemobiliseerd om weerstand te bieden en ertegen te vechten. Blijft de extreme situatie voortduren, dan volgt een uitputtingsfase, die kan leiden tot langdurige ontregelingen en soms zelfs tot de dood. Vanuit deze medisch/biologische opvatting bezien, vormt stress een waardevolle aanpassing van het organisme aan de omgeving, in de zin van snel reageren op een bedreigende situatie. Pas wanneer zo'n situatie lang blijft voortduren kan de gezondheid in gevaar komen.

De derde en momenteel meest gangbare opvatting over stress is ontstaan halverwege de jaren zestig. De kritiek op de twee andere benaderingen leidde ertoe dat Richard Lazarus (Lazarus, 1966; Lazarus en Folkman, 1984) het belang van cognitive beoordelingsprocessen sterk ging benadrukken. Hij bracht naar voren dat een situatie pas stressvol is wanneer een persoon deze situatie als stressvol waarneemt en interpreteert. In deze benadering wordt stress dus opgevat als *een wisselwerking tussen de persoon en de omgeving* en is er sprake van stress wanneer de eisen vanuit de omgeving de vaardigheden van de persoon overstijgen. Binnen deze wisselwerking dienen twee beoordelingsprocessen ("appraisal processen") onderscheiden te worden: op de eerste plaats de beoordeling van de situatie zelf, waarbij de vraag centraal staat of de situatie al dan niet een bedreiging vormt voor het welbevinden (m.a.w. 'ben ik in gevaar'). Als deze vraag bevestigend beantwoord wordt, volgt een tweede reeks beoordelingsprocessen, die te maken hebben met de ernst van de situatie en wat het individu eraan denkt te kunnen doen (zie ook Boekaerts, 1987). In Figuur 1 is deze benadering schematisch weergegeven en zien we dat de eerste soort beoordelingsprocessen aanleiding geven tot emotionele reacties en de tweede soort beoordelingsprocessen aanleiding geven tot bepaald gedrag. Dit gedrag waarmee men stress tracht te reguleren of reduceren wordt copinggedrag of stresshantering genoemd.



Figuur 1: Stress opgevat als een wisselwerking tussen de persoon en zijn waargenomen omgeving.

Een belangrijke ontwikkeling binnen het recente stress-onderzoek is dat stress steeds meer in het alledaagse leven wordt bestudeerd, ook binnen het onderwijs. Het aantal praktijkrelevante publikaties is de afgelopen vijf jaar sterk toegenomen. Het belang van het verschijnsel schoolstress wordt steeds duidelijker: zo rapporteren Bäuerle en Kury (1980) een groot aantal lichamelijke klachten en gevoelens van ongenoegen bij leerlingen die zij toeschrijven aan schoolstress. Benders (1987) spreekt in dit kader dan ook over 'de befaamde ijsberg waarvan het aantal toppen en de omvang nog niet aan te wijzen zijn' (p. 27).

De eerste studies m.b.t. stress binnen het onderwijs gingen vooral uit van het gezichtspunt van de volwassene. Er werd gebruik gemaakt van lijsten die een bewerking waren van vragenlijsten bedoeld voor volwassenen (o.a. de eerder genoemde 'Social Readjustment Rating Scale' van Holmes en Rahe), en waarin ouders of leerkrachten gevraagd werden een oordeel te geven over hoe de leerling iets zou ervaren (o.a. de 'Children's Life Events Inventory' van Chandler, 1981). Onderzoek heeft evenwel aangetoond dat de gezichtspunten van ouder/leerkracht en leerling nogal eens van elkaar afwijken (Yamamoto, 1979; Yamamoto & Felsenthal, 1982). Wat de leerkracht als belastend ziet voor de leerling is iets anders dan wat de leerling zelf als belastend ervaart. Recent wordt dan ook steeds meer aandacht besteed aan de belevingswereld en het gezichtspunt van de leerling.

Naast de verschuiving van gezichtspunt, is de aandacht de laatste jaren ook verschoven van dramatische gebeurtenissen naar alledaagse beslommeringen ('daily hassels'). Dramatische levensgebeurtenissen verstoren het evenwicht op vrij dramatische wijze en hebben relatief lang geleden plaats gevonden (bijv. het overlijden van een gezinslid, of het scheiden van de ouders), terwijl dagelijkse beslommeringen minder intens van aard zijn en alledaagse irritaties en frustaties representeren. Deze verschuiving is met name voortgekomen uit de onderzoeksbevinding dat alledaagse beslommeringen een even groot en zelfs een groter beroep kunnen doen op de aanpassingscapaciteiten van het individu dan de dramatische levensgebeurtenissen. Deze bevinding werd zowel bij volwassenen (Monroe, 1983) als bij kinderen (Wertlieb, Weigel & Feldstein, 1987) gevonden.

Schoolstress

Uitgaande van de belevingswereld van de leerling en van alledaagse schoolsituaties, is recent een aantal studies verricht naar het verschijnsel schoolstress. Er werd met name onderzocht welke verschillende soorten schoolsituaties door leerlingen als belastend worden ervaren. Metcalfe, Dobson, Cook en Michaud (1982) gingen via interviews met dertienjarige leerlingen na welke bronnen van aan school gerelateerde stress er voor deze groep bestaan. Op basis hiervan werd de 'Student Stress Inventory' geconstrueerd. De situaties in deze vragenlijst hebben vooral betrekking op organisatorische en meer algemene aspecten van het schoolleven (veel lawaai op school, oninteressant huiswerk, kosten van het schooluniform, straf voor te laat komen e.d.). Middels factoranalyse kon een tiental subschalen onderscheiden worden, die staan weergegeven in Tabel 1.

Een tweede vragenlijst, de 'Children's Own Perceptions and Experiences of Stressors' (COPES) van Colton (1985) werd geconstrueerd op basis van interviews en een literatuurstudie. Behalve alledaagse beslommeringen (zoals uitgelachen worden; je niet kunnen concentreren; moeilijkheden met een vak hebben) werden ook dramatische levensgebeurtenissen, die niet gerelateerd zijn aan de schoolsituatie in deze vragenlijst opgenomen (zoals het scheiden van ouders, problemen met broers en/of zussen). Ook voor deze vragenlijst staan de subschalen zoals die gevonden werden na factoranalyse weergegeven in Tabel 1.

Een derde en laatste vragenlijst, waarnaar we binnen het kader van dit artikel willen verwijzen, is de 'Academic Achievement Inventory' van Grannis (1985), die eveneens op basis van interviews en observaties met leerlingen werd geconstrueerd. De lijst bestaat uit in totaal 34 situaties, waarvan er 18 betrekking hebben op belastende schoolsituaties en 16 op schoolsituaties waarin leerlingen steun ervaren van anderen ('social support'). Grannis kiest in zijn vragenlijst bewust voor uitsluitend school- en klassituaties. Hij verwoordt het als volgt: '... after only a few months in the school which has been selected as the site for the research, the investigator concluded that an approach to the students' homes for information on major life events and daily hassels could not be justified until the school had made some progress with reducing schoolstress. The investigator then turned to creating an instrument that would measure the recurrent stressors and supports experienced by the students in the life of classrooms and the school itself. The project aimed to feed back findings from the instrument to the school in such a way as to contribute to the reduction of students' stress in the school (p. 9)'. Binnen het kader van dit artikel zijn alleen de 18 belastende schoolsituaties van belang. Een aantal van de situaties staat in Tabel 1 weergegeven.

Tabel 1:

Overzicht van de verschillende belastende schoolsituaties afkomstig van een drietal recente vragenlijsten met betrekking tot schoolstress.

Metcalfe et al. (1982)	Colton (1985)	Grannis (1985)
Student Stress Inventory	Children's Own Perceptions and Experiences of Stressors (COPES)	Academic Achievement Inventory
40 items (10 subschalen)	60 items (7 subschalen)	34 items
13- tot 14-jarige leerlingen	8- tot 11-jarige leerlingen	11- tot 13-jarige leerlingen
- zelfbeeld en relatie met anderen - formele schoolorganisatie - informele schoolorganisatie - toekomstperspectief - overgang naar adolescentie - overgang naar andere school - evaluatie en toetsing - studie faciliteiten - financiële overwegingen - straf voor te laat komen	- gevoelens van isolatie - dramatische gebeurtenissen - ontwrichting in familie - cognitieve overbelasting - financiële zorgen - stief-ouders - serieuze schoolproblemen	- als ik me niet veilig voel - als iemand mij slaat/schopt - als ze me uitlachen - als iemand me pest - als ik het werk niet afkrijg - als kinderen in de klas onrustig zijn - als ik geen antwoord krijg op een vraag die ik stel

Alle drie de hier genoemde instrumenten kenmerken zich door het feit dat zij per situatie een intensiteits-oordeel van de leerling vragen. De lijst van Metcalfe et al. bevroegt hoeveel stress zo'n situatie veroorzaakt (vier-puntsschaal), de lijsten van Grannis en Colton bevragen de mate waarin leerlingen van streek raken in zo'n situatie (vijf-puntsschaal). Op deze wijze trachten zij dus de wisselwerking tussen omgeving en persoon te meten in termen van de subjectieve perceptie van zo'n situatie.

Schoolstress blijkt dus multidimensioneel van aard: het omvat niet alleen het falen voor een proefwerk of toets, maar een breed scala van taakgerichte- en sociale situaties (zie Tabel 1). Wij zouden schoolstress dan ook willen omschrijven als de spanning die ontstaat bij leerlingen, wanneer zij geconfronteerd worden met een variëteit van alledaagse schoolsituaties, die als een potentiële bedreiging voor het welbevinden worden waargenomen. Wanneer we de leerling willen bestuderen binnen zijn werkcontext met de bedoeling het onderwijsleerproces te optimaliseren is het van belang kennis te verwerven m.b.t. schoolsituaties die als problematisch worden ervaren en mogelijk stress oproepen.

In Nederland werd nog weinig empirisch onderzoek naar stress bij leerlingen gerapporteerd (zie ook Beusekom, 1985). Het deelonderzoek waarover hier gerapporteerd wordt betreft een onderzoek naar de beoordelingsprocessen van leerlingen van de brugklas en basisschool t.a.v. een variëteit van alledaagse schoolsituaties. Doelstelling van dit deelonderzoek is om na te gaan hoe leerlingen op school problematische situaties waarderen. Dit deelonderzoek betreft de *gepercipieerde situatie* kant van schoolstress en beperkt zich dus tot het eerste beoordelingsproces zoals dat in figuur 1 wordt weergegeven. In het hoofdonderzoek, waarover hier niet wordt gerapporteerd, wordt ook de *response* kant bestudeerd. Daarin worden ook de karakteristieke wijzen waarop leerlingen van deze leeftijdsgroep omgaan met stress (coping) geïnventariseerd. De S-R vragenlijst die hieruit resulteerde (Boekaerts, Hendriksen en Maes, in voorbereiding) is een meetinstrument om stress en coping bij leerlingen te meten. Het onderzoek waarover in dit artikel verslag wordt gedaan is op te vatten als een voorstudie voor de uiteindelijke instrument-constructie. Gezinssituaties blijven buiten beschouwing. In deze studie staat de beleavingswereld van de leerling m.b.t. schoolsituaties centraal. Daarbij zal tevens worden ingegaan op de mogelijke verschillen tussen jongens en meisjes. Dit omdat uit onderzoek hieromtrent gesuggereerd wordt dat meisjes op school meer stress zouden ervaren dan jongens, in de zin van een hogere score op de

respectievelijke vragenlijsten (Metcalf et al., 1982; Fahs, 1986, die gebruik maakte van de door Grannis geconstrueerde 'Academic Achievement Inventory'). Ook zal kort worden ingegaan op verschillen tussen leerlingen van de basisschool en de brugklas. Tot slot zal het verband met een bestaande vragenlijst besproken worden, die een tweetal voor het onderwijsleerproces relevant veronderstelde variabelen meet (prestatie-motivatie en negatieve faalangst). Met name op basis van het onderzoek van Metcalf et al. (1982) mag verwacht worden dat ervaren schoolstress positief gerelateerd is aan prestatie-motivatie. Metcalf et al. vond dit verband overigens alleen bij jongens. Elders (Hendriksen, 1988) zal uitgebreider worden ingegaan op de relatie tussen schoolstress en andere hieraan gerelateerde constructen zoals angstdispositie e.d.

METHODE

Subjecten: Aan het onderzoek werd deelgenomen door 16 willekeurig gekozen scholen (18 verschillende klassen) uit het Zuid-Oosten van Nederland. Er waren 361 leerlingen uit de twee hoogste groepen van het Basisonderwijs (97 leerlingen uit groep zeven en 264 leerlingen uit groep acht), en 107 leerlingen uit de brugklas (MAVO; HAVO/VWO). De leerlingen in de steekproef varieerden in leeftijd van 9 tot 14 jaar.

Materiaal en procedure: Om te komen tot een steekproef van alledaagse problematische situaties, werd in een eerder onderzoek met 90 leerlingen uit groep zeven en acht van de Basisschool een individueel interview gehouden en werd aan 336 leerlingen, eveneens uit groep zeven en acht, een open-vragenlijst voorgelegd. De leerlingen werden gevraagd schoolsituaties te beschrijven waarover zij zich zorgen maakten. Uit de totale pool van situaties werd uiteindelijk een steekproef van 41 relevante, eenduidig formuleerbare situaties afgeleid. Elders (Hendriksen, 1988) wordt deze procedure uitgebreid gerapporteerd.

Deze 41 situaties, die kort beschreven staan in Tabel 2, werden in klassikaal verband ter beoordeling aan de leerlingen voorgelegd. De situatie-omschrijvingen werden met opzet in de 'ik-vorm' geformuleerd teneinde een zo groot mogelijke betrokkenheid van de leerlingen bij de geschetste situaties tot stand te brengen. In de situatie-omschrijvingen zoals aangeboden aan de leerlingen van de brugklas werd het taalgebruik enigszins aangepast. Zo werden de termen 'meester' en 'juf' vervangen door 'docent(e)' en werd de term 'andere kinderen' vervangen door 'andere leerlingen'. De 41 alledaagse schoolsituaties stonden op een apart kaartje en werden afgenomen in de vorm van een sorteertaak. Hierbij moesten de situaties door leerlingen onderverdeeld worden in vier vaste categorieën naar hoe *erg* ze die situatie voor zichzelf waardeerden (1 = helemaal niet erg, tot 4 = heel erg). De leerlingen waren vrij om te bepalen hoeveel situaties ze in een bepaalde categorie onderbrachten. De erg-score wordt dus opgevat als een cognitieve inschatting van het problematische karakter van de betreffende situatie. Daarnaast werd bij de leerlingen uit groep acht van de basisschool en leerlingen van de brugklas de Prestatie-Motivatie Test voor Kinderen (PMT-K) van Hermans (1983) afgenomen. Hiermee worden o.a. twee persoonlijkheidsdisposities gemeten nl. prestatie-motivatie (een relatief stabiele persoonlijkheidsdispositie, die in specifieke situaties leidt tot presteren in de zin van excelleren) en negatieve faalangst (twijfel aan eigen kunnen en angst om te falen, die vooral in relatief ongestructureerde taaksituaties leiden tot onderpresteren).

RESULTATEN

De gemiddelde erg-scores per situatie varieerden van 1.25 (als de meester of juf iets uitlegt) tot 3.23 (als mijn vriendje of vriendinnetje niets meer van mij wil weten), en blijken een redelijke spreiding te hebben. Het gemiddelde, de standaarddeviatie, alsmede de rangorde staan voor iedere situatie weergegeven in Tabel 2.

Tabel 2:

Gemiddelde, standaarddeviatie (s.d) en rangorde van de erg-scores voor ieder van de 41 situaties (N = 468).

	korte situatie-omschrijving	gemiddelde	s.d.	rangorde
01	Als ik een slecht punt haal voor een proefwerk	2.96	0.78	10
02	Als anderen niet mee kunnen komen bij een vak	2.28	0.85	24
03	Als ik van de meester/juf iets overnieuw moet doen	2.27	0.87	25
04	Als iemand uit de klas gepest wordt	2.55	0.89	21
05	Als de meester/juf slechte zin heeft	2.48	1.02	22
06	Als ik iets kwijt ben wat van school is	3.07	0.88	8
07	Als ik niet goed mee kan komen bij een vak	3.20	0.79	2
08	Als andere kinderen ruzie hebben met elkaar	1.99	0.84	28
09	Als ik werk mee naar huis moet nemen	1.94	0.95	30
10	Als ik te laat op school kom	2.12	0.81	27
11	Als anderen mij uitlachen omdat ik iets niet weet	2.76	1.01	17
12	Als ik ruzie heb met mijn vriendje	2.94	0.91	11
13	Als de meester/juf aan het schelden is op mij	3.15	0.95	4
14	Als ik de stof voor een proefwerk niet geleerd heb	3.14	0.81	5
15	Als anderen mij ergens de schuld van geven	3.10	0.79	7
16	Als de meester/juf kwaad is op mij	2.66	0.87	18
17	Als andere kinderen op school mij uitschelden	2.89	0.91	13
18	Als we in de les moeilijke sommen moeten maken	1.78	0.82	35
19	Als iemand mij slaat of schopt	2.64	0.87	20
20	Als we tijdens de les veel moeten opschrijven	1.98	0.96	29
21	Als mijn vriendje niets meer van mij wil weten	3.23	0.92	1
22	Als ik niet op kan schieten met de meester/juf	2.94	0.88	12
23	Als anderen hun huiswerk niet afhebben	1.53	0.69	39
24	Als de meester/juf mij een beurt geeft	1.34	0.69	40
25	Als ik het huiswerk niet af heb	2.78	0.82	16
26	Als de meester/juf kwaad is op andere kinderen	1.88	0.79	31
27	Als ik van anderen niet mee mag doen bij spelen	2.79	0.87	15
28	Als de meester/juf ziek is	1.71	0.94	37
29	Als ik gevallen ben	1.77	0.76	36
30	Als anderen hun sommen/taken niet kennen	1.88	0.80	32
31	Als de meester/juf onrechtvaardig is tegen mij	3.18	0.83	3
32	Als andere kinderen mij pesten	2.86	0.82	14
33	Als ik van de meester/juf netter moet schrijven	1.79	0.82	34
34	Als ik op school iets kapot maak	3.12	0.84	6
35	Als de meester/juf aan het schelden is tegen anderen	2.18	0.91	26
36	Als anderen uit mijn groepje iets niet snappen	1.84	0.74	33
37	Als ik vergeten ben wat ik geleerd heb	3.01	0.83	9
38	Als andere kinderen het antwoord niet weten	1.70	0.70	38
39	Als de meester/juf mij straf geeft	2.66	0.89	19
40	Als ik iets niet snap	2.36	0.89	23
41	Als de meester/juf iets aan het uitleggen is	1.25	0.60	41

Om na te gaan welke betekenisdimensies ten grondslag lagen aan de 41 alledaagse schoolsituaties werden enige Principale componenten analyses met varimax rotatie uitgevoerd. Er werden elf componenten gevonden met een eigenwaarde groter dan 1.00. Van de verschillende geïnspecteerde oplossingen bleek de vier componentoplossing (percentage verklaarde variantie = 35.9%) inhoudelijk het best te interpreteren (zie Tabel 3). Aangezien de sorteertaak niet door alle leerlingen volledig was uitgevoerd, en de procedure van 'listwise deletion' werd toegepast, betreffen de bevindingen van deze analyse 370 leerlingen.

De items 5, 7, 29 en 31 hadden geen lading van betekenis op één der componenten en zullen in verdere analyses buiten beschouwing worden gelaten. De eerste component (eigenwaarde = 6.89; verklaarde variantie = 16.8%) heeft betrekking op gebeurtenissen waarin andere leerlingen een

moeilijke situatie ervaren en de leerling zelf a.h.w. een toeschouwersrol vervult. De leerling observeert met andere woorden stress bij andere klasgenootjes (andere kinderen kennen hun sommen/taken niet, of andere kinderen hebben ruzie met elkaar), en identificeert zich met hun situatie. De component kreeg daarom het label '*Identificatie met stress bij Anderen*'. De tweede component (eigenwaarde = 3.27; verklaarde variantie = 8.0%) heeft betrekking op de relatie van de leerling met belangrijke anderen, zoals de meester, een vriendje, of een klasgenootje. Centraal bij deze situaties staat steeds de sociale afzondering; het niet mee mogen spelen, uitgescholden worden e.d. Deze component zou dan ook als '*Sociale Isolatie*' bestempeld kunnen worden. De derde component (eigenwaarde = 2.49; verklaarde variantie = 6.1%) heeft betrekking op gebeurtenissen waarbij de leerling zelf oorzaak is van de ervaren emoties. De situaties hebben zowel betrekking op taakgerichte zaken ('als ik vergeten ben wat ik geleerd heb') als op niet-taakgerichte zaken ('als ik op school iets kapot maak'). Deze component kan best als '*Confrontatie met falen en fouten*' benoemd worden. Situaties tenslotte, waarin door leerkrachten taakgerichte

Tabel 3:

Factor ladingen van de verschillende situaties bij de viercomponentoplossing na varimax rotatie. (Alleen ladingen hoger dan .30 staan weergegeven; N = 370).

Korte situatie-beschrijving		1	2	3	4
30	Als anderen hun sommen/taken niet kennen	.77			
23	Als anderen hun huiswerk niet afhebben	.72			
38	Als anderen het antwoord op een vraag niet weten	.69			
26	Als de meester/juf kwaad is op andere kinderen	.68			
08	Als andere kinderen ruzie hebben met elkaar	.66	.36		
35	Als de meester/juf aan het schelden is tegen anderen	.61			
36	Als anderen uit mijn groepje iets niet snappen	.61			
02	Als anderen niet mee kunnen komen bij een vak	.57			
28	Als de meester/juf ziek is	.41			
32	Als andere kinderen mij pesten		.73		
17	Als andere kinderen op school mij uitschelden		.69		
27	Als ik van de anderen niet mee mag doen bij spelen		.68		
12	Als ik ruzie heb met mijn vriendje		.60		
21	Als mijn vriendje niets meer van mij wil weten		.59		
15	Als anderen mij ergens de schuld van geven		.58		
19	Als iemand mij slaat of schopt		.46		
13	Als de meester/juf aan het schelden is op mij		.46	.33	
11	Als anderen mij uitlachen omdat ik iets niet weet		.44		
04	Als iemand uit de klas gepest wordt	.39	.40		
14	Als ik de stof voor een proefwerk niet geleerd heb			.59	
37	Als ik vergeten ben wat ik geleerd heb			.59	
25	Als ik het huiswerk niet af heb			.57	
10	Als ik te laat op school kom			.55	
34	Als ik op school iets kapot maak			.52	
06	Als ik iets kwijt ben wat van school is			.46	
16	Als de meester/juf kwaad is op mij		.37	.42	
01	Als ik een slecht punt haal voor een proefwerk			.38	
22	Als ik niet op kan schieten met de meester/juf			.36	
40	Als ik iets niet snap			.36	.33
20	Als we tijdens de les veel moeten opschrijven				.67
18	Als we in de les moeilijke sommen moeten maken				.60
03	Als ik van de meester/juf iets overnieuw moet doen				.57
09	Als ik werk mee naar huis moet nemen				.55
33	Als ik van de meester/juf netter moet schrijven				.53
24	Als de meester/juf mij een beurt geeft				.53
39	Als de meester/juf mij straf geeft			.38	.46
41	Als de meester/juf iets aan het uitleggen is				.40

activiteiten worden opgelegd die een extra werkbelasting voor de leerling inhouden, zoals veel moeten opschrijven, moeilijke sommen moeten maken, werk overnieuw moeten doen e.d., vormen een vierde component (eigenwaarde = 2.08; verklaarde variantie = 5.1%). We noemden deze component 'Taakoverbelasting'. De componenten blijken onderling redelijk onafhankelijk te zijn. Bij oblimin scheve rotatie varieerden de intercorrelaties tussen de componenten van .01 (factor 2 en 3) tot .28 (factor 1 en 4).

Op basis van de hierboven beschreven component analyse kunnen een viertal subschalen van belastende schoolsituaties worden onderscheiden. Deze bestaan uit de in Tabel 3 gepresenteerde situaties, op de situaties 4, 16 en 40 na, die onvoldoende discrimineren tussen de verschillende factoren en derhalve niet worden opgenomen in één der subschalen. Voor ieder van deze subschalen staan in Tabel 4 de gemiddelde score, de standaarddeviatie en de alpha coëfficiënt (Cronbach, 1951) weergegeven. Zoals uit die gegevens kan worden afgeleid, hebben de subschalen een acceptabele spreiding en homogeniteit, hetgeen hun gebruik als subschalen rechtvaardigt. Bovendien blijkt uit deze Tabel dat situaties m.b.t. sociale isolatie en confrontatie met falen als 'erger' worden waargenomen dan situaties waarin de leerling zich identificeert met anderen of overbelasting ervaart.

Tabel 4:

Gemiddelde, standaarddeviatie en alpha coëfficiënten voor de vier subschalen van belastende schoolsituaties. (N=370).

Domein	aantal items	minimum score	maximum score	gemiddelde	s.d.	alpha
Identificatie met anderen	9	9	36	16.9	4.7	.83
Sociale Isolatie	9	9	36	26.4	4.9	.79
Confrontatie met falen/fouten	8	8	32	23.2	3.8	.70
Taakoverbelasting	8	8	32	15.0	3.7	.70

Om na te gaan of er verbanden bestaan tussen de scores op de subschalen van belastende schoolsituaties en de persoonlijkheidsdisposities 'prestatie-motivatie' en 'negatieve faalangst' zoals gemeten door de PMT-K, werden Multiple Regressie analyses uitgevoerd (de gevonden bivariate correlatiecoëfficiënten worden in Tabel 5 gegeven).

Tabel 5:

Bivariate correlaties tussen de scores op de subschalen van belastende schoolsituaties en prestatie-motivatie en negatieve faalangst (N=277).

Domein	Prestatie-motivatie	Negatieve faalangst
Identificatie met anderen	.20***	.03
Sociale Isolatie	.13*	.12*
Confrontatie met falen/fouten	.36***	.18***
Taakoverbelasting	-.35***	.10*

(* $p < .05$; ** $p < .01$; *** $p < .001$)

In de regressie analyses, waarvan de resultaten beschreven staan in Tabel 6, werden eveneens geslacht en leergroep (basisschool/brugklas) als variabele opgenomen. Er worden uitsluitend hoofdeffecten gerapporteerd, aangezien we geen interactie-effecten (geslacht x prestatie-motivatie; geslacht x negatieve faalangst; geslacht x leergroep) vonden. De score op de subschaal 'identificatie met stress bij anderen' laat zowel een statistisch significant leergroep effect, als geslacht effect zien. Meisjes hebben een significant hogere erg-score dan jongens en leerlingen van de basisschool scoren hoger in vergelijking tot leerlingen van de brugklas. Voor de overige drie subschalen werd uitsluitend een geslacht effect gevonden: meisjes scoren significant hoger dan jongens op situaties

met betrekking tot sociale isolatie en confrontatie met falen. Voor taakoverbelasting daarentegen hebben jongens een hogere erg-score dan meisjes. De gemiddelde scores voor de verschillende groepen staan weergegeven in Tabel 7. Tussen leerlingen uit groep zeven en groep acht van de basisschool werden geen statistisch significante verschillen gevonden.

Tabel 6:

Multipale Regressie analyses van de subschalen van belastende schoolsituaties op geslacht, leergroep, prestatiemotivatie en negatieve faalangst (N=277).

	Identificatie met stress bij anderen		Sociale Isolatie		Confrontatie met falen en fouten		Taakoverbelasting	
	beta	F	beta	F	beta	F	beta	F
Geslacht	-.19	10.39**	-.18	9.44**	-.18	10.15**	.14	5.70*
Leergroep	-.27	21.72***	.10	2.59	.04	.59	.09	2.30
prestatie-motiv.	.14	5.85*	.12	4.00*	.36	43.00***	-.31	27.60***
Neg. Faalangst	.09	2.46	.10	2.75	.21	13.87***	.05	.71
R ²		.14		.08		.22		.15

(* $p < .05$; ** $p < .01$; *** $p < .001$)

Er werden zoals verwacht enige statistisch significante verbanden gevonden met prestatie-motivatie en negatieve faalangst. De score voor confrontatie met falen blijkt zowel positief samen te hangen met prestatie-motivatie als met negatieve faalangst (zie tabel 5 en 6). Dit betekent dat hoge prestatie-motivatie en hoge negatieve faalangst samen hangen met het als erg waarnemen van situaties waarin fouten worden gemaakt of falen wordt verwacht. Taakoverbelasting hangt daarentegen negatief samen met prestatie-motivatie. Dit zou erop kunnen wijzen dat leerlingen die laag prestatie gemotiveerd zijn het 'erg' vinden dat ze veel moeten opschrijven, moeilijke sommen moeten maken, iets overnieuw moeten doen en werk mee naar huis moeten nemen, terwijl leerlingen met een hoge prestatie-motivatie dergelijke situaties als minder 'erg' waarnemen. Voor

Tabel 7:

Overzicht van de gemiddelde scores per subschaal voor jongens en meisjes en voor de verschillende leeftijdsgroepen afzonderlijk (om onderlinge vergelijking mogelijk te maken werden de somscores per subschaal gedeeld door het aantal items/situaties dat in een subschaal is opgenomen).

	Jongens	N	Meisjes	N
Identificatie met stress bij anderen:				
- groep zeven	1.93	37	2.04	43
- groep acht	1.79	107	2.04	111
- brugklas	1.54	37	1.78	49
Sociale isolatie:				
- groep zeven	2.79	37	2.98	43
- groep acht	2.82	107	2.97	111
- brugklas	2.83	37	3.21	49
Confrontatie met falen en fouten:				
- groep zeven	2.86	37	3.03	43
- groep acht	2.74	107	2.97	111
- brugklas	2.82	37	3.05	49
Taakoverbelasting:				
- groep zeven	1.95	37	1.85	43
- groep acht	1.89	107	1.75	111
- brugklas	2.14	37	1.88	49

de domeinen Sociale Isolatie en Identificatie met anderen worden minder duidelijke verbanden gevonden hetgeen in de lijn der verwachtingen ligt gezien de aard van de bevraagde situaties in deze beide domeinen. We wijzen er overigens op dat het percentage verklaarde variantie niet hoog uitvalt. Dit betekent in ieder geval dat er andere variabelen zijn die erg-score van de leerlingen mede beïnvloeden.

DISCUSSIE

De resultaten van deze studie geven enige beschrijvende informatie over de cognitieve beoordelingsprocessen van leerlingen ten aanzien van alledaagse schoolsituaties die een potentiële bedreiging kunnen vormen voor het welbevinden. In navolging van Grannis (1985) hebben we ons in deze studie beperkt tot uitsluitend school- en klassituaties. De 41 schoolsituaties die we de leerlingen aanboden blijken uit een viertal onderscheiden, en van elkaar onafhankelijke domeinen te bestaan. Leerlingen ervaren schoolstress tijdens hun eigen functioneren in taaksituaties (taakoverbelasting en confrontatie met falen), maar ook in de sociale omgang met leraar en medeleerlingen (sociale isolatie en identificatie met stress bij anderen). Dat het zinvol is deze domeinen van elkaar te onderscheiden wordt bevestigd door de tegengestelde richting van het verband tussen prestatiemotivatie en confrontatie met falen enerzijds ($\beta = -.36$) en prestatiemotivatie met taakoverbelasting anderzijds ($\beta = .31$). Ook het feit dat de verschillen tussen jongens en meisjes en tussen de verschillende leergroepen domeinspecifiek blijken te zijn, wijst erop dat het zinvol is de vier soorten situaties die stress veroorzaken op school te onderscheiden in verder onderzoek. Gezien de beperkte omvang van de onderzoeksgroep dienen de gepresenteerde resultaten (Tabel 7) met enige voorzichtigheid te worden geïnterpreteerd. Voor meisjes geldt dat zij taakoverbelasting minder erg vinden dan jongens, terwijl zij op de andere drie domeinen een hogere erg-score hebben dan jongens. Voor wat betreft de verschillen tussen de leergroepen kan geconstateerd worden dat het wel en wee van klasgenootjes (identificatie met stress bij anderen) leerlingen van de brugklas minder zorgen baart dan leerlingen van de basisschool. Andere significante verschillen tussen de leergroepen werden niet gevonden. In navolging van Metcalfe et al. (1982) wordt een statistisch significant verband gevonden tussen de erg-score van leerlingen (de mate waarin schoolsituaties als belastend worden waargenomen) en prestatiemotivatie. Dat dit verband geslacht specifiek zou zijn (Metcalfe et al. vond het verband uitsluitend bij jongens) wordt in deze studie niet teruggevonden, blijkens de statistisch niet significant zijnde interactie-termen.

De algemene bevinding uit de onderzoeksliteratuur dat schoolstress multidimensioneel van aard is werd in deze studie bevestigd. De gevonden domeinen zijn evenwel van een andere aard. Er bestaat geen overeenstemming met de door Metcalfe et al. (1982) beschreven subschalen. Van de door Colton (1985) gerapporteerde subschalen worden er twee in onze studie teruggevonden: sociale isolatie en overbelasting. Echter bij de COPES vragenlijst spelen thuissituaties en dramatische levensgebeurtenissen een voorname rol. Wanneer we tot slot, de meer alledaagse schoolsituaties uit de studie van Grannis (1985) vergelijken met de gevonden resultaten blijkt een aantal situaties duidelijk overeen te komen (uitgelachen worden, gepest worden, geslagen of geschopt worden, het werk niet afkrijgen).

Misschien nog belangrijker dan de waardering van de situatie zelf, zijn de pogingen van de leerling om aan stressvolle situaties het hoofd te bieden (zie o.a. Lazarus en Folkman, 1984). De wijze waarop de leerling omgaat met stress en tot stress-reductie komt (copinggedrag) bepaalt immers de gevolgen van dergelijke situaties. Met andere woorden wanneer een leerling er niet in slaagt om op adequate wijze het hoofd te bieden aan een 'als bedreigend waargenomen situatie', kan dit resulteren in het verstoren van het welbevinden. Hierdoor wordt de aandacht van de leerling (tijdelijk) verlegd van leerstofbeheersing naar herstel van het welbevinden (cf. Boekaerts, 1987). Wanneer de druk zo hoog wordt, dat de leerling zelf het gevoel heeft dat hij het niet (meer) aankan, dreigt er gevaar. Het welbevinden komt dan chronisch in het gedrang en we zien symptomen van onbehagen, weerstand tegen het schoolsysteem en een grote verscheidenheid aan lichamelijk klachten zoals inslaap moeilijkheden, hartkloppingen, vermoeidheid, hoofdpijn e.d.

(Bäuerle & Kury, 1980; Benders, 1987). Hier zij overigens opgemerkt dat stress niet per definitie negatief hoeft te zijn. Een leerling kan namelijk een problematische situatie positief waarderen en zich bijvoorbeeld uitgedaagd voelen om prestaties te leveren (zie Compennolle, 1987).

Dat leerlingen geholpen moeten worden om snel tot stress-reductie te komen vraagt geen betoog. Goed onderwijs houdt hiermee rekening. In dit artikel hebben wij getracht een eerste aanzet te geven tot kennis omtrent één aspect van het verschijnsel schoolstress n.l. de situaties die door leerlingen als potentiële bedreiging voor het welbevinden worden gepercipieerd. Verder onderzoek naar stress en het daaraan gerelateerde copinggedrag zal nodig zijn om op adequate wijze in te grijpen op het 'stress-proces' en leerlingen te begeleiden bij efficiënte stress-reductie. Met name zal daarbij aandacht dienen te worden besteed aan stress en copinggedrag in relatie tot het functioneren op school (o.a. studie-prestaties, schoolverzuim, attitude t.a.v. school e.d.).

LITERATUUR

- Beusekom, H.A.M. van (1985). Met stress onder leerlingen valt het nogal mee. *Tijdschrift Nederlands Genootschap van Leraren*, 17, 1120-1121.
- Bäuerle, S. & Kury, H. (1980). Stress in der Schule: eine experimentelle Untersuchung an 13- bis 16-jährigen Schülern. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 29, 70-76.
- Benders, H. (1987). Ik hoef niet naar die school. *School, februari nummer*, 26-27.
- Boekaerts, M. (1987). *Psychologie van de leerling en het leerproces*. Nijmegen: Dekker en van de Vegt.
- Chandler, L.A. (1981). The Source of Stress Inventory. *Psychology in the Schools*, 18, 164-168.
- Colton, J.A. (1985). Childhood stress: Perceptions of Children and Professionals. *Journal of Psychopathology and Behavioral assessment*, 7, 155-173.
- Compennolle, T.H.L. (1987). *Stress op de middelbare school*. Leuven/Amersfoort: Acco.
- Cronbach, L.T. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cox, T. (1978). *Stress*. London: Macmillan.
- Fahs, M.E. (1986). Coping in school: correlations among perceptions of stress, coping styles, personal attributes and academic achievement in inner-city junior high school students. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, California.
- Grannis, J.C. (1985). Students' stress, social support and achievement in an urban middle school. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois.
- Hendriksen, J.G.M. (1988). Dissertatie in voorbereiding. Katholieke Universiteit Nijmegen.
- Hermans, H.J.M. (1983). *PMT-K 83: Prestatie Motivatie Test voor Kinderen*. Lisse: Swets en Zeitlinger.
- Holmes, T.H. & Rahe, R.H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, 11, 213-218.
- Lazarus, R.S. (1966). *Psychological Stress and the Coping Process*. New York: Mc Graw Hill.
- Lazarus, R.S. & Folkman, S. (1984). *Stress, Appraisal and Coping*. New York: Springer Publishing Company.
- Metcalf, R.J.A., Dobson, C.B., Cook, A. & Michaud, A. (1982). The Construction, Reliability and Validity of a Stress Inventory for Children. *Educational Psychology*, 2, 59-71.
- Monroe, S.M. (1983). Major and Minor Life Events as Predictors of Psychological Distress: Further Issues and Findings. *Journal of Behavioral Medicine*, 6, 189-205.
- Selye, H. (1987). *Stress*. Utrecht: Uitgeverij het Spectrum.
- Wertlieb, D., Weigel C., & Feldstein M. (1987). Stress, Social Support and Behavior Symptoms in Middle Childhood. *Journal of Clinical Child Psychology*, 16, 204-212.
- Yamamoto, K. (1979). Brief reports: Children's Ratings of the Stressfulness of Experiences. *Developmental Psychology*, 15(5), 581-582.
- Yamamoto, K. & Felsenthal, H.M. (1982). Stressful Experiences of Children: Professional judgments. *Psychological Reports*, 50, 1087-1093.

Manuscript ontvangen 19-1-1988

Definitieve versie ontvangen 23-3-1988

Boekbesprekingen

P.C. van de Sijde, *Training the teaching script*. Proefschrift Universiteit Twente. Almere: Versluys, 1987, 109 pag. ISBN 90-249-1598-8.

Dit proefschrift vormt het sluitstuk van het Nederlands gedeelte van Het leeromgevingsonderzoek (Classroom Environment Study) dat in 1979 gestart is onder auspiciën van de International Association for the Evaluation of Educational Achievement (IEA). Met dit onderzoek is een project van zeven jaar afgesloten. De twee andere proefschriften uit dit drieluik staan op naam van Krammer (1984) en Tomic (1985). Het doel van dit onderzoeksproject is het vaststellen van lesgerag van wiskundeleraren dat een samenhang vertoont met de leerprestaties en houdingen van leerlingen, om deze samenhangen vervolgens nader te onderzoeken in een correlatieel onderzoek. Dát lesgedrag dat positief samenhangt met de leerprestaties en houdingen van de leerlingen zou dan de basis vormen voor een gerichte training voor leraren welke tenslotte in een experimenteel onderzoek getoetst zou moeten worden. Hiermee wordt voldaan aan de door Rosenshine & Furst (1973) aanbevolen drietrapsacyclus: descriptie, correlatieel onderzoek en experimenteel onderzoek. Krammer heeft gerapporteerd over de samenhang tussen lesgedrag en leerboek, Tomic over de samenhang tussen lesgedrag en leerprestaties. Dit proefschrift beschrijft de training van leraren en het experimentele onderzoek.

Na een korte karakterisering van het leeromgevingsonderzoek in hoofdstuk 1, wordt in hoofdstuk 2 ingegaan op twee benaderingen in het onderzoek naar het onderwijzen. De eerste benadering sluit aan bij het expert-paradigma uit de cognitieve psychologie. Onderwijzen wordt hier opgevat als een complexe cognitieve vaardigheid waar sommige leraren meer bedreven in zijn (de experts) dan andere. Drie kennisdomeinen worden bij onderwijs- 'expertisie' van belang geacht: lesstructuur-kennis, vakinhoudelijke kennis en kennis van het managen van een klas. Lesstructuur-kennis heeft betrekking op de structuur, de activiteiten die in een les worden uitgevoerd. Deze structuur wordt opgevat als een script met een vijftal scenes volgens welke de leraar handelt: uitleg van nieuwe leerstof, verwerkingsopdracht, klassikale opdracht, het geven van individuele hulp en huiswerk nakijken. Met betrekking tot het script zijn er twee opvattingen te onderscheiden: het script als temporele sequentie (wat volgt na wat) en het script als een ordening naar belangrijkheid van de scenes (centraliteit). Bij de temporele opvatting gedraagt de leraar zich volgens het onderwijs-script als hij alle scenes in een bepaalde volgorde doorloopt, bij de centraliteits-opvatting als hij de belangrijkste scenes uitvoert. Door Van der Sijde worden beide opvattingen gehuldigd. Vakinhoudelijke kennis heeft hier betrekking op het vak wiskunde, maar speelt in dit proefschrift geen rol omdat de training gericht is op algemeen didactische vaardigheden. Voor klasmanagement worden de opvattingen van Kounin (1970) centraal gesteld. Hierbij ligt de nadruk op de preventie van ordeproblemen in de klas.

De tweede benadering in het onderzoek van het onderwijzen betreft 'research on teacher/teaching effectiveness' volgens het klassieke proces-product paradigma. In dit paradigma worden de samenhangen onderzocht (correlatieel en causaal) tussen lesgeefgedragingen van leraren en de uitkomsten van de leerlingen. Op interessante en verdienstelijke wijze worden de beide benaderingen geïntegreerd. De drie kennisdomeinen worden ingevuld met de uitkomsten van de proces-product benadering en cognitief geïnterpreteerd. Zo wordt de scene 'uitleg' ingevuld met de uitkomsten van het onderzoek van Good, Grouws & Ebmeier (1983) en Rosenshine & Stevens (1986). Bij klasmanagement wordt gebruik gemaakt van effectief gebleken managementvaardigheden van Kounin (1970) als: 'withitness, overlapping, signal continuity, group alerting, student accountability'. De kennis van de lesstructuur en klasmanagement vormen de basis voor een trainingsprogramma voor wiskundeleraren in de tweede klas Havo en Vwo. In de training wordt aandacht gegeven aan: het onderwijs-script (er wordt een lesfase-model gepresenteerd), specifieke instructieregels voor iedere scene van het script en algemene regels voor management in de klas.

Hoofdstuk 3 geeft de opzet en de resultaten van de experimentele beproeving van de training. Dit gedeelte wordt de implementatie-studie genoemd. Aan het onderzoek hebben negen leraren deelgenomen, die allemaal gebruik maken van eenzelfde wiskundeboek. Het onderzoek is opgezet als een 'pretest-posttest design'. Voorafgaande aan de training zijn van iedere leraar door getrainde observatoren met behulp van een shootcomputer zes lessen geobserveerd. Na deze observaties hebben de leraren een training gevolgd van één dag. Hierna zijn de leraren opnieuw gedurende zes lessen geobserveerd. Op grond van de verkregen gegevens wordt geconcludeerd dat het lesfase-model na de training wordt geïmplementeerd en dat de gemiddelde effectgrootte van de instructie- en managementregels gezamenlijk 0.25 bedraagt, met andere woorden: de training is in staat het lesgedrag te veranderen. Toch kunnen bij deze conclusie een aantal kanttekeningen geplaatst worden. Gezien de interne en externe validiteit van het gekozen design kunnen rivaliserende verklaringen voor het effect niet geheel worden uitgesloten. Een controlegroep ontbreekt in de implementatiestudie. De effectgrootte van 0.25 is berekend op grond van een somscore van alle instructie- en managementregels. De standaarddeviaties van de scores van de management-variabelen zijn veel groter dan hun gemiddelden, wat wijst op veel variatie tussen de leraren. Bij een significantieniveau van 5% is geen enkel verschil significant. Bij de 13 instructievaardigheden treden alleen bij het geven van voorbeelden en het stellen van denkvragen significante verschillen op. De resultaten zijn moeilijk te duiden omdat overzichtstabellen hier ontbreken. Na de training worden wel significant meer lesscenes toegepast, maar dit resultaat is ook moeilijk te interpreteren omdat per scene in navolging van Good, Grouws & Ebmeier (1983) een aanbevolen normtijd wordt gehanteerd. Als voor het geven van een les aanbevolen wordt om te beginnen met het nakijken van huiswerk met als tijdsaanduiding 5-10 minuten en de leraren besteden hieraan vóór de cursus 22 minuten per les en na de cursus 17 minuten, dan wordt aan deze scene wel minder tijd besteed, maar de uitkomst ligt niet in de buurt van de aangegeven norm. Hierbij kan men zich natuurlijk de vraag stellen of deze norm voor elke leraar, voor elke les en voor elke klas moet gelden.

Hoofdstuk 4 beschrijft een tweede experimenteel onderzoek, waarin is nagegaan of als gevolg van de training de leerprestaties en de attitudes van de leerlingen in één leerjaar zijn veranderd. Aan dit onderzoek hebben 33 leraren deelgenomen, verdeeld over twee experimentele en twee controlecondities. De leraren in de eerste experimentele conditie hebben eerst een ééndaagse training gevolgd (met eenzelfde inhoud als in het eerste onderzoek) en zijn vervolgens gedurende 8-10 lessen geobserveerd. De leraren in de tweede experimentele conditie zijn eerst geobserveerd en hebben daarna de eendaagse training gevolgd, waarbij gebruik is gemaakt van deze observatiegegevens. De twee experimentele groepen hebben van september 1985 tot april 1986 aan het onderzoek deelgenomen, de controlegroepen van december tot april. Om veranderingen in leerprestaties en attitudes vast te stellen zijn in de maanden september, december en april door het CITO geconstrueerde wiskunde-toetsen en een attitude-test (m.b.t. plezier aan wiskunde en zelfbeeld) afgenomen. Verschillen in leerprestaties en attitudes bij de twee experimentele condities zijn voor de periode september-december niet gevonden. Verschillen in managementgedragingen zijn niet aangetroffen, bij de instructiegedragingen zijn alleen verschillen aangetroffen voor: het geven van een overzicht van de vorige les, verwijzing naar het boek, geven van positieve feedback. De december-en apriltoetsen zijn eveneens afgenomen bij leraren in de controlecondities. In januari hebben de leraren in de tweede experimentele conditie de training gevolgd, terwijl de leraren in de eerste controleconditie de handleiding per post hebben gekregen. De leerprestaties en de attitudes in de periode december-april blijken tussen de vier condities niet significant te verschillen. Wel wordt geconstateerd dat er een significant verschil bestaat tussen de twee experimentele condities voor wat betreft de toename in leerprestaties over de gehele onderzoeksperiode in het voordeel van de tweede conditie. Dit verschil wordt toegeschreven aan de training van leraren in de tweede experimentele conditie op basis van individuele feedback aan de hand van de observatiegegevens. Of dit inderdaad zo is, laat zich moeilijk beoordelen. Uit de toetsscores blijkt dat deze op de drie meetmomenten in de loop van het schooljaar afnemen. Dit kan liggen aan de toetsen.

Informatie over de aard van de toetsen ontbreekt, m.n. over de vraag in hoeverre de toetsen equivalent zijn wordt niets gezegd. De toename in leerprestaties, waarover gesproken wordt, is moeilijk controleerbaar. De implementatie van de onderwijsvaardigheden is geringer dan in de eerste studie.

In hoofdstuk 5 wordt de invloed van de training op het klasseklimaat beschreven. Hiertoe zijn bij de leerlingen in klassen van de leraren in de experimentele condities vragenlijsten afgenomen, een Nederlandse bewerking van de Classroom Environment Scale van Moos & Trickett. Van de zes subschalen blijkt één schaal te verschillen ten gunste van de leerlingen bij de leraren in de tweede experimentele conditie. Dit verschil in perceptie wordt toegeschreven aan de training. Gezien de zwakte van het gekozen design (one-shot case study – alleen nameting) zijn rivaliserende verklaringen mogelijk. Er is niet gecontroleerd op verschillen in percepties aan het begin van het schooljaar. Deze studie, genoemd een 'pilot study', is het zwakste deel van het proefschrift.

Na lezing van het gehele proefschrift blijft men toch zitten met de vraag of de training nu effect heeft gehad en in welke omvang. De onderzoeker zelf meent deze vraag positief te moeten beantwoorden en baseert zich daarbij op totaalscores. De lezer kan zich hierbij aan enige twijfel niet onttrekken. Dit komt omdat het proefschrift erg beknopt is gehouden. Elk hoofdstuk is zo gecomponeerd dat het als een afzonderlijk artikel voor publicatie kan worden aangeboden. Dit leidt soms tot een cryptische weergave van de resultaten en soms tot redundante informatie. In het gehele proefschrift wordt nergens besproken hoe de ééndaagse training is opgezet en uitgevoerd en wat de leraren er van gevonden hebben. Dit inzicht is toch nodig om het effect van de training te kunnen beoordelen. Ook is niet beargumenteerd waarom gekozen is voor een ééndaagse training, terwijl overzichtsstudies op het terrein van stafontwikkeling of nascholing laten zien dat dergelijke korte bijeenkomsten weinig effect hebben. Verder blijft men zitten met de vraag of de leraren nu getraind zijn in de betreffende vaardigheden of een hoorcollege ondergaan hebben met zelfstudie. Er wordt alleen opgemerkt dat de aanbevelingen in de handleiding bediscussieerd zijn en dat de leraren gevraagd is de vaardigheden te implementeren. Van een dergelijke beschrijving gaat nauwelijks een leereffect voor collega-onderzoekers of lerarenopleiders uit.

Afgezien van deesignaleerde leemtes, enkele slordigheden in de literatuurlijst en onduidelijkheden, getuigt de studie van grote degelijkheid en nauwgezetheid. De observaties zijn zorgvuldig verricht en elke leraar is gedurende een groot aantal lessen geobserveerd, waardoor per leraar betrouwbare uitspraken zijn verkregen over het lesgeefgedrag. De inhoud van het trainingsprogramma, gebaseerd op empirische uitkomsten, verdient serieuze verspreiding en toepassing. Aldus leveren het leeromgevingsproject in het algemeen en deze studie in het bijzonder een belangrijke bijdrage aan de verdere professionalisering van de lerarenopleidingen.

REFERENTIES

- Good, T.L., Grouws, D.A. & Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.
 Kounin, J.S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
 Krammer, H.P.M. (1984). *Leerboek en Leraar*. Harlingen: Flevodruk.
 Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.), *Handbook of research on teaching* (2nd ed., pp. 122-183). Chicago: Rand McNally.
 Rosenshine, N. & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), *Handbook of research on teaching*. (3rd. ed., pp. 376-391). New York: MacMillan.
 Tomic, W. (1986). *Docentgedrag en leerresultaten*. Enschede: Febodruk.

S. Veenman
 Instituut voor Onderwijskunde
 K.U. 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, Faculteit der Toegepaste Onderwijskunde, Vakgroep OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 12, nr. 3

Juni 1988

Inhoudsopgave

VOR-Bulletin, Jrg 12, nr. 3

Proefschriften	176
Congressen	177
Ontvangen publikaties	179

Proefschriften

J. J. H. van den Akker: Ontwerp en implementatie van natuuronderwijs (Lisse, Swets & Zeitlinger, 1988).

Promotoren: Prof.dr. W.J. Nijhof en Prof.dr. Tj. Plomp.

Promotie: Universiteit Twente, 15 januari 1988.

Korte Samenvatting

De algemene thematiek van dit proefschrift betreft de relatie tussen het ontwerp en de implementatie van curricula. Meer in het bijzonder gaat de aandacht uit naar de invloed van (het ontwerp en de vormgeving van) curriculummateriaal op de wijze van lesvoorbereiding en lesuitvoering door leerkrachten in de eerste fase van het implementatieproces. Een en ander is onderzocht binnen de praktische context van het (in ontwikkeling zijnde vak) natuuronderwijs op de basisschool.

Als lastige implementatieproblemen voor leerkrachten kwamen naar voren: ingewikkelde en tijdrovende lesvoorbereiding; gebrek aan benodigde vakinhoudelijke achtergrondkennis; moeilijke wijzigingen in de didactische rol van de leerkracht; onvoldoende realisatie van leereffecten bij leerlingen.

Op grond van een aanvullende, meer theoretische analyse van het leerkrachtperspectief bij curriculumimplementatie, heeft een verscherping plaatsgevonden van de mogelijke functies die voorbeeld-lesmaterialen voor niet-vertrouwde gebruikers dienen te hebben. Een conclusie luidde dat het materiaal een ruime hoeveelheid concrete handelingsaanwijzingen (procedurele specificaties) voor de leerkracht dient te bevatten, toegespitst op de centrale implementatieproblemen. Op grond daarvan zijn ontwerpspecificaties geformuleerd die vervolgens zijn verwerkt in bestaand curriculummateriaal.

In een veldexperiment (met 40 leerkrachten) is daarna het gebruik van dit bewerkte (experimentele) materiaal vergeleken met het oorspronkelijke (controle)materiaal. Uit de resultaten bleek een aanzienlijk effectievere implementatie van het experimentele materiaal, zowel vanuit de optiek van de ontwikkelaars (grotere overeenkomst tussen oorspronkelijk bedoelingen en lesuitvoering) als vanuit het leerkrachtperspectief (meer helderheid over de voorgestelde leerplanvernieuwing en de eigen rol daarin).

Het proefschrift wordt afgesloten met een aantal aanbevelingen aan curriculumontwerpers en -onderzoekers.

Congressen

5TH INTERNATIONAL CONGRESS MUSIC EDUCATION FOR THE HANDICAPPED.

The 5th International Congress on Music Therapy and Music Education for the Handicapped will be held at the Leeuwenhorst Congress Center in Noordwijkerhout, The Netherlands from August 23 till August 27, 1989. The congress will be organized by the Stichting Muziektherapie under the auspices of the International Association of Music for the Handicapped (I.A.M.H.) and the Nederlandse Vereniging voor Creatieve Therapie (N.V.K.T.).

Although music therapy is a widely accepted form of treatment in healthcare institutions, reflections on the different way of using music is now of the first importance. Therefore the main theme of this congress will be the practical and conceptual differences between:

- ☐ music therapy as a form of psychotherapy
- ☐ music therapy in special education and healthcare
- ☐ music education for the handicapped.

Throughout the congress, papers, video's, posters and workshops will be presented, discussing developments in practice and in research.

Music therapists, music educators, psychologists, physicians and remedial educationalists are invited to participate actively in the congress. Active participation is possible by presenting a paper or a video, by leading a workshop or by submitting a proposal for a poster session. You are cordially invited to submit a proposal for a presentation at the congress. The form of the presentation may range from the showing of a video-tape to a theoretical discussion of the theme.

Inlichtingen: Stichting Muziektherapie, c/o Mevr. Loek Stijlen, Henri Polaklaan, 15-1, 1018 CP Amsterdam.

COLOGNE COMPUTER CONFERENCE

The Center for Historical Social Research (ZHSF) of the Central Archive for Empirical Social Research (ZA), Cologne and the Max-Planck-Institut für Geschichte, Göttingen jointly announce:

The International Conference on Data Bases in the Humanities and Social Sciences of 1988 (ICDBHSS/88), the 3rd Annual Conference of the Association for History and Computing (AHC) and the 10th International Conference of the International Federation of Data Organizers for the Social Sciences (IFDO), to be held as a joint conference in

Cologne, September 7 - 10, 1988.

The International Conference on Data Bases in the Humanities and Social Sciences, so far a bi-annual conference series in the United States, has usually brought together 300 - 400 participants from around the world, listening to approx. 100 papers. It invites papers on all aspects of data base technology, as applied to the Humanities and the Social Sciences, specific developments of hardware and software, exemplary project reports, networking and other.

The Annual Meetings of the Association for History and Computing, the professional forum for all aspects of the usage of information technologies in the historical disciplines, has recently been attended by an audience of 300 - 350 participants, predominantly European, which joined in the discussion of ca. 80 presentations. It invites papers on all aspects of computer applications in the Historical disciplines: methodological contributions, technical ones and, again, reports on projects of exemplary interest.

The International Federation of Data Organizations for the Social Sciences (IFDO) - the world wide association of major social science data archives and services - organizes topically centered conferences to promote methodological and technical innovations for empirical social research. It invites papers on all aspects of social science data services and on the development of data protection legislation or data access regulations in particular.

The joint conference will be held at the facilities of the University of Cologne, West Germany, which is the seat of institutes devoted to the application of computers in the Social Sciences as well as the Historical disciplines. The main topics will be: new data bases, software development, data analysis, computer aided instruction, institutional infra-structure.

Proposals for papers should be directed to: Manfred Thaller, Cologne Computer Conference, Max-Planck-Institut für Geschichte, Postfach 2833, D-3400 Göttingen, West Germany; Organizing Secretary: Dagmar Klaff, Zentrum für Historische Sozialforschung, Bachemer Str. 40, D-5000 Köln 41.

OPERATIONALIZATION AND RESEARCH STRATEGY

Op donderdag 8 en vrijdag 9 september 1988 wordt in Amsterdam een symposium georganiseerd over "Operationalization and Research Strategy". Onderwerp van het symposium is de relatie tussen de operationalisering van begrippen en de algemene methodologie van sociaalwetenschappelijk onderzoek, inclusief de veranderende functie van verbale theorieën en mathematische modellen daarin. De nadruk zal daarbij liggen op de algemene concepties van sociaal wetenschappelijk onderzoek die ten grondslag liggen aan de ter discussie gestelde problemen en oplossingen.

Op het symposium zullen onderzoekers die zich (internationaal) met deze problematiek hebben bezig gehouden papers presenteren, waarbij een ruime mogelijkheid tot discussie wordt geboden. De sprekers (voorzover nu bekend) zijn, uit het buitenland: D. Alwin, H. M. Blalock, Jr, J. de Leeuw, (USA); S. Levy (Israel); R. Pawson (UK); Opp, P. Schmidt (BRD); en uit Nederland: J. Dessens, W. Jansen, P. G. Swanborn (RUU); E. E. Roskam (KUN); J. de Jong-Gierveld (VU); J. J. Hox, G. J. Mellenbergh (UvA).

Het symposium wordt georganiseerd door de onderzoeksgroep "Begripsvorming en Researchdesign" van de werkgemeenschap SoMO (Sociaal Methodologisch Onderzoek) van de ZWO stichting SSCW, in samenwerking met het Research Committee on Logic and Methodology van de International Sociological Association. Een informatiebrochure + aanmeldingsformulier kan worden opgevraagd bij: Dr. J. J. Hox, Faculteit PAOW, Prinsengracht 227, 1015 DT Amsterdam, (020-5253315).

STUDIEDAG VOLWASSENENEDUCATIE EN INFORMATIETECHNOLOGIE

De studiedag wordt georganiseerd door INSP Cluster II.7, volwasseneneducatie in samenwerking met het PCBB, het SLO en het SVE.

Datum: 24 juni 1988

Plaats: Nederlands Congresgebouw Den Haag

Tijd: 10.00 uur - 16.00 uur

Doel: Het informeren over de (toekomstige) mogelijkheden van het gebruik van middelen uit de informatietechnologie ten behoeve van onderwijs aan volwassenen. Naast ongeveer 30 nauwkeurig geselecteerde lezingen, presentaties, demonstraties in de deelsectoren basiseducatie, dag-/avondonderwijs, beroepsopleiding/scholing en een internationale workshop 'Vrouwen en nieuwe technologieën' is er gelegenheid een ruim opgezette tentoonstelling te bezoeken, waar, door educatieve uitgeverijen en (INSP)-projecten in demonstraties en voorbeelden de actuele en toekomstige mogelijkheden worden getoond. De dag is bedoeld voor educatieve werkers, docenten, praktijkbegeleiders, coördinatoren, directeuren, ondersteuners, bestuurders, inspecteurs, beleidsambtenaren, politici, kortom iedereen die bij de uitvoering en het beleid van onderwijs en scholing aan volwassenen betrokken is. Kosten: f 65,- per persoon inclusief lunch, koffie, thee e.d.

Secretariaat: Stichting SVE, Myriam Prins, Postbus 351, 3800 AJ Amersfoort.

Ontvangen publicaties

- Akker, J. J. H. van den (1988). Ontwerp en implementatie van natuuronderwijs (Proefschrift Universiteit Twente). Lisse: Swets & Zeitlinger (f 45,00).
- Beishuizen, J. J., Leeuw, L. de & Swart, J. H. de (Red.) (1988). Cognitieve psychologie in theorie en praktijk. Liber Amicorum Prof.dr. S.D. Fokkema. Lisse: Swets & Zeitlinger (f 35,00).
- Beishuizen, M. & Jong, D. de (Red.) (1987). Rekenonderwijs en computer: ontwikkelingsonderzoek en rekenonderwijs met behulp van de computer. Lisse: Swets & Zeitlinger (f 28,50).
- Drift, K. D. J. M. van der & Vos, P. (1987). Anatomie van een leeromgeving: een onderwijseconomische analyse van universitair onderwijs (Proefschrift RU Leiden). Lisse: Swets & Zeitlinger.
- Frith, S. (1987). Jeugdsociologie. Groningen: Wolters-Noordhoff.
- Haars, V. J. E., Verhoeven, A. F. M. & Vreese-Böck, M. F. de (Red.) (1987). Leren met Logo. Nijmegen: Werkgroep Leren met Logo.
- Herfs, P. G. P. (1987). Opvang van allochtone leerlingen in het Utrechts voorgezet onderwijs. Utrecht: Vakgroep Onderwijskunde.
- Janssen, P. J. & Neve, H. de (1988). Studeren en doceren aan het hoger onderwijs. Vakmanschap als leeropdracht. Leuven/Amersfoort: Acco (f 57,50).
- Kanselaar, G. e.a. (1987). Effectiviteitsstudie kennisoverdracht. Het effect van illustraties bij het lezen van teksten uit 'Bouwen Nu'. Utrecht: Vakgroep Onderwijskunde.
- Kreft, G. G. (1987). Models and methods for the measurement of schooleffects (Proefschrift UvAmsterdam). Utrecht: Elinkwijk.
- Min, R. (1987). Computersimulatie als leermiddel: een inleiding in methoden en technieken. Schoonhoven: Academic Service.
- Os, W. van (1987). Evaluatie in het hoger onderwijs. Groningen: Wolters-Noordhoff.
- Parreren, C. F. van (1988). Ontwikkeld onderwijs. Leuven/Amersfoort: Acco (f 28,00).
- Peschar, J. L. (1987). Zo vader - zo zoon, zo moeder - zo dochter? Vergelijkende analyses naar de processen van statusverwerving en onderwijsmobiliteit in Nederland, Hongarije en Polen. Lisse: Swets & Zeitlinger (f 42,00).
- Stokking, K.M., Dekker, E.M. & Leenders, F. J. (1987). Zorgverbreding of selectieverscherping? Utrecht: ISOR.
- Terwel, J., Herfs, P., Dekker, R. & Akkermans, W. (1987). Effecten van interne differentiatie. Utrecht: Vakgroep Onderwijskunde.
- CITO (1988). Toetstechnische begrippenlijst. Arnhem: CITO
- Koetsier, C. P. & Ten Brinke, J. S. (1988). Experiment individuele eindstage -1 op de afdeling Didaktiek van de Biologie. Analyse van de interviews met schoolpraktikumdocenten. Utrecht: Afdeling Didaktiek v.d. Biologie.
- Koetsier, C. P., Van Driel, C. & Janzen, R. W. (1987). De begrippen exploreren en reflektieren nader bekeken ... Utrecht: Afd. Didaktiek van de Biologie.
- Koetsier, C. P. & Janzen, R. W. (1988). Beoordeling van de student in de cursus 'orientatie op het beroep van leraar'. Utrecht: Afd. Didaktiek v.d. Biologie.
- Koetsier, C. P. & Janzen, R. W. (1988). Orientatie op het beroep van leraar. Utrecht: Afd. Didaktiek in de Biologie.
- Stortelder, M., De Graauw, C., & Hutjes-Boelaars, L. (1987). Bruikbaarheid van de editietest als leertaak. Nijmegen: ITS.
- Verijdt, H. J. P. M. & Diederens, J. H. I. W. (1987). Determinanten van beroepskeuze, een verkennend onderzoek. Nijmegen: ITS.
- Buurke, W. J. (1987). Schoolwerkplanontwikkeling taalvaardigheid in de basisschool. Nijmegen: ITS.

- Buis, Th., Mensink, J. & Doesborgh, J. (1987). Feiten en meningen over het schoolonderzoek moderne vreemde talen. Nijmegen: ITS
- Van Kessel, N. (1987). Verlangd voortgezet onderwijs in Maastricht en omstreken. Nijmegen: ITS.
- Diederen, J. (1987). Beroepsloopbaan na KMBO of BBO. Nijmegen: ITS
- Pelkmans, A. & Vrieze, G. (1987). Meer over coördinatie en management in het voortgezet onderwijs. Nijmegen: ITS.
- O&I nieuws december 1987 nr. 0.
- CITO (1987). Meerjareplan/Werkprogramma 1988-1992
- Vernieuwing, Tijdschrift voor Onderwijs en Opvoeding, jrg. 47, nr. 1 januari 1988.
- PAOS cursussen 1988
- SLO Catalogus, aanvulling Voortgezet Onderwijs, opleiding volwasseneneducatie.
- SLO Catalogus aanvulling Basis- en Speciaal onderwijs
- SCO Jaarverslag 1985 en 1986
- Goossens' Consultants (1988) Opleidingsbehoefte in gemeenten. Den Haag: Goossens
- Goossens' Consultants (1988) Opleidingsbehoefte in instellingen in de Gezondheidszorg. Den Haag: Goossens.

Het leren programmeren van computergestuurde gereedschapswerktuigen *

Een afbeeldingstheorie en een informatieverwerkingsmodel

Georg Rakers**, Jules Pieters en Imme de Bruijn
Universiteit Twente

ABSTRACT

To our knowledge no experimental work has addressed programming performance within automated manufacturing system environments. The aim of this paper is to offer a theory of how human subjects learn to program computerized manufacturing systems for discrete products. Several research issues are discussed.

INLEIDING.

Onderzoek dat zich bezig houdt met de wijze waarop mensen omgaan, en leren omgaan, met de computer staat de laatste jaren in de cognitieve wetenschappen sterk in de belangstelling. De reden van deze belangstelling is het inzicht dat wetenschappelijke kennis op het gebied van de 'mens-machine-interactie' (Card, Moran & Newell, 1983; Norman, 1984) niet alleen van belang is voor het *ergonomisch* verantwoord ontwerpen van computers en computersystemen. De vergaarde kennis en inzichten zijn in toenemende mate van belang voor het goed *instrueren* van de aankomende computergebruiker (du Boulay, O'Shea & Monk, 1981; Mayer, 1975, 1976a, 1976c, 1979, 1981; Sime, Arblaster & Green, 1977a, 1977b; En in het verlengde van dit instrueren: Leren met behulp van een computer in het onderwijs (Dirkzwager, Fokkema, van der Veer & Beishuizen, 1984).

Wie echter de relevante literatuur met betrekking tot de interactie tussen mens en werktuig bestudeert zal opmerken dat het onderzoek nogal eenzijdig benadrukt blijkt te zijn. Zo blijkt binnen de grote verscheidenheid van professionele apparatuur en consumenten elektronica de computer veruit de populairste te zijn. Een uitzondering hierop vormt het onderzoek naar het hebben van een mentaal model van de werking van een zakrekenmachientje (Bayman & Mayer, 1984; Norman, 1983; Young, 1981, 1983). Het onderzoek heeft zich daarnaast met name gericht op die computersystemen (zoals mainframe, mini- en personal computer) zoals deze in de praktijk worden gebruikt in onderzoekscentra, banken, kantoren, etc. Mens-machine-interactie onderzoek naar de wijze waarop mensen omgaan, en leren omgaan, met de computer in de industrie is er, zover wij hebben kunnen nagaan, tot op heden nog weinig uitgevoerd. Een ander bezwaar van deze eenzijdige benadering is, dat er vaak taken worden gebruikt die ontleend zijn aan het psychologisch laboratorium, zoals bijvoorbeeld het *onthouden* van computerprogramma's. Een taak die nogal ver afstaat van wat programmeurs in hun dagelijkse beroepspraktijk doen (zie in dit verband ook de bezwaren van Sheppard, Curtis, Milliman & Love, 1979).

* De schrijvers danken G. Beukhof, A. Dirkzwager, L. Koster, H. Krammer, J. van Merrienboer, M. de Leeuw, C. Reijmerink, L. de Vin en E. Warries voor hun kritiek en waardevolle opmerkingen. B. van Hout-Wolters, C. Reijmerink en A. Thijsen willen we hier hartelijk danken voor het kritisch hebben willen doorlezen van enkele voorgaande versies van dit artikel.

Het schrijven van dit artikel werd mogelijk gemaakt door financiële steun door de Stichting voor Onderzoek van het Onderwijs onder contractnr.: 553905, en Prof. dr. S. Dijkstra, die voorzag in de oorspronkelijke ideeën.

** Correspondentie-adres: Nederlandse Philips Bedrijven B.V., Corporate Industrial Design, Applied Ergonomics Group, Gebouw SX, Postbus 218, 5600 MD Eindhoven

Aan het bestuderen van het gebruik van de computer in de industrie is tot nu toe in de computer- en cognitieve wetenschappen veel te weinig aandacht besteed (zie evenwel: Bullinger, Fähnrich & Raether, 1985; Bullinger, Fähnrich & Sprenger, 1984; Das, 1987; Sharit, 1985). En met name op het gebied van het programmeren van computergestuurde gereedschapswerktuigen, zoals frees-, draai-, en boorbanken, blijkt er wat dit betreft, nog weinig terzake bekend te zijn (Rakers, 1987; Pritschow & Viefhaus, 1987; Weule, Löffler & Selinger, 1987). Veelal geschiedt slechts de bestudering van het oplossingsproces van technische problemen en worden deze problemen geïsoleerd van programmeerproblemen (Van der Sanden, 1986).

Het belang van het uitvoeren van onderzoek op het zojuist genoemde gebied wordt echter gevormd door het feit, dat, enerzijds, in het reguliere technisch onderwijs het gebruik van computergestuurde gereedschapswerktuigen in de zeer nabije toekomst een belangrijke rol gaat spelen (Reijmerink & Pieters, 1987; noten 3 en 4). Zo wordt er in het kader van het Informatica-Stimuleringsplan naar gestreefd, dat met ingang van het leerjaar 1988-1989 alle MTS-en in Nederland met een afdeling Werktuigbouwkunde kunnen beschikken over professionele productie-apparatuur. Anderzijds vereisen technologische innovaties nieuwe kwalificaties op het gebied van de moderne produktietechnieken (Mulder & Nijhof, 1987; Nijhof & Mulder, 1986; Reijmerink & Mulder, 1987; Thijsen & Mulder, 1987a, 1987b). Technologische innovaties zullen een aantal aspecten binnen bedrijven doen veranderen. Nijhof & Mulder (1986) noemen onder meer: (1) een veranderde opzet van het productieproces en de productie organisatie; (2) veranderingen in functies en taken van het personeel; (3) afstemming van nieuwe apparatuur op menselijke mogelijkheden en beperkingen. Deze veranderingen verlangen een adequate en flexibele aanpassing van het Middelbaar Technisch Beroepsonderwijs. Leerplannen dienen te worden aangepast aan de nieuwe leerdoelen die als gevolg van de invoering van deze nieuwe technologieën zullen ontstaan. Niet alleen in Nederland, maar ook in het buitenland is deze problematiek onderkend, zie bijvoorbeeld: (*Duitsland*) Laur-Ernst (1982), Weule e.a. (1987); (*Engeland*) Burnes (1984), Kemp & Clegg (1987), Oliver (1984), Senker (1984); (*Zweden*) Elsassner & Lindvall (1984), Reuterswärd (1984); (*Europa, algemeen*) noot 2.

In dit artikel willen we ingaan op de wijze waarop CNC- (*Computer Numerical Controlled*) machines worden geprogrammeerd. Het doel van dit artikel is: (a) een theoretisch kader te formuleren voor de wijze waarop CNC-machines worden geprogrammeerd, en (b) een informatieverwerkingsmodel aan te reiken voor het toetsen van de theorie.

Achtereenvolgens zullen worden besproken: Het theoretisch kader, het informatieverwerkingsmodel, en relevante onderzoeks vraagstellingen op het gebied van de cognitieve psychologie en de instructietechnologie.

HET THEORETISCH KADER.

Algemeen.

De toepassing van informatieverwerkende apparatuur ter sturing van het productieproces heeft het leren oplossen van technische problemen een nieuwe dimensie gegeven. De metaalbewerker die thans werkt met numeriek bestuurbare apparatuur zal, indien hij niet enkel als machinebediener is opgeleid, een technisch *en* een programmeerprobleem moeten oplossen. Het probleemoplossingsproces kent hierdoor dan ook twee stadia van betrekkelijk hoge moeilijkheidsgraad.

Het eerste stadium van het probleemoplossingsproces is het vertalen van de praktische vraag (het produceren van een werkstuk) naar een wiskundig ruimtelijk model, dat in een werktekening wordt vastgelegd.

Het tweede stadium betaamt uit, ten eerste, het formuleren (vanuit de werktekening) van een programma van stappen, de bewerkingsvolgorde, die achtereenvolgens door de machine moeten worden uitgevoerd en, ten tweede, het vastleggen van deze stappen in een CNC-computerprogramma.

Mentaal Model.

Onderzoek dat uitgaat van de opvatting dat mensen zich een mentaal model vormen van de interne structuur en wijze van gegevensverwerking in een computer is betrekkelijk nieuw (Carroll & Mack, 1985; Carroll & Thomas, 1982; Gentner & Stevens, 1983; Johnson-Laird, 1980, 1983; Mayer, 1981; Rouse & Morris, 1986).

Hoewel deze opvatting in de cognitieve wetenschappen aan invloed wint, vindt er op het gebied van de cybernetische ergonomie (Stassen, 1976) al geruime tijd onderzoek plaats waarbij er van uit wordt gegaan dat mensen zich een mentaal model vormen van een (dynamisch) systeem (Cohen & Ferrell, 1969; Jagacinski & Miller, 1978; Rouse & Morris, 1986; Smallwood, 1967; Veldhuyzen & Stassen, 1977). Verondersteld wordt, dat 'de menselijke regelaar' een mentaal model gebruikt om daarmee de huidige toestand van het systeem te kunnen bepalen, en mogelijke toekomstige toestanden af te leiden en te voorspellen (van Bussel, 1980; van Heusden, 1980; Rouse & Morris, 1986; Smallwood, 1967; Stassen, 1976; Veldhuyzen & Stassen, 1977). Meer recentelijk, vanaf het begin van de jaren tachtig, is men zich in de cognitieve en computerwetenschappen met het begrip *denkbeeld* gaan bezighouden. Integenstelling tot de cybernetische ergonomie ging de onderzoeks aandacht hierbij voornamelijk uit naar de wijze waarop mensen omgaan en leren omgaan met computers en computersystemen.

Uit het hiervoor aangehaalde onderzoek kunnen twee conclusies worden getrokken. (a) Het hebben van een mentaal model, denkbeeld, van een dynamisch systeem of een computersysteem helpt zowel de computerbediener als de menselijke regelaar bij het verrichten van specifieke handelingen. (b) Deze handelingen worden afgeleid uit een algemeen mentaal model dat men zich van het systeem heeft gevormd en hebben betrekking op een gespecificeerde, opgedragen, taak.

Zoals hierboven al is opgemerkt, dient een metaalbewerker een ruimtelijk en een programmeerprobleem op te lossen. Bij de mens - CNC-machine-interactie zijn derhalve twee aspecten te onderscheiden:

(A) *een ruimtelijk aspect*, dat betrekking heeft op de concrete, tastbare, werkelijkheid; met als belangrijkste eenheid het te vervaardigen, direkt observeerbare, CNC-produkt, en

(B) *een verbaal aspect*, dat betrekking heeft op de abstracte, talige, werkelijkheid; met als belangrijkste eenheid het voor het te vervaardigen werkstuk te schrijven CNC-programma.

Hieronder zullen deze twee aspecten nader worden uitgewerkt. De kern van ons betoog zal daarbij zijn, dat aan het (leren) programmeren van CNC-machines *een* mentaal model ten grondslag ligt. Dit mentale model uit zich echter in *twee* verschijningsvormen: Het produktvervaardigingsproces en het CNC-computerprogramma voor het vervaardigen van dat produkt. We veronderstellen verder, dat beide vormen of representaties één-op-één (als een bijectie) samenhangen en door de CNC-programmeur mentaal op elkaar afgebeeld dienen te worden. In de hierna volgende afbeeldingstheorie zal met name het tweede stadium in het probleemoplossingsproces, het formuleren van een CNC-programma worden besproken.

Besturingssystemen.

In een CNC-computerprogramma staat stapsgewijs exact de benodigde informatie vermeld voor het besturen van een gereedschapbeweging (de baan van de beitel, frees, boor, etc.). Deze bewegingen kunnen, afhankelijk van de soort CNC-machine (noot 5), langsdraaibewegingen, aanzetbewegingen en dwarsbewegingen zijn. Hierbij kan er een onderscheid worden gemaakt tussen (a) 'Point-To-Point Systems', en (b) 'Contouring Systems' (Koren, 1983). Bij Puntbesturingssystemen wordt het werkstuk, ten opzichte van het gereedschap, langs de machine-assen voortbewogen. Bij Contouringssystemen is dat het gereedschap (zoals, bijvoorbeeld, bij een draaibank de beitel) (zie Figuur 1-a en 1-b). Een bijzonder geval is het Lijnbesturingssysteem. Bij Puntbesturingssystemen wordt het begin- en eindpunt van de beweging opgegeven. De tussenliggende verplaatsing is daarbij van ondergeschikt belang. Bij lijnbesturing vindt de beweging eerst in rechte lijn langs een as plaats, en vervolgens in rechte lijn langs een andere as van de machine. Hierbij is de rechte, afgelegde, baan wel van belang. Bij Lijnbesturingssystemen worden echter geen cirkelbogen, of elliptische banen afgelegd. Alle bewegingen dienen

standaard vastgelegd te worden in een coördinatenstelsel. Afhankelijk van de soort CNC-machine is dit een coördinatenstelsel met twee assen (2-D; twee dimensionaal) of een stelsel met drie assen (3-D; drie dimensionaal). Een CNC-machine kan echter van minimaal twee, zoals bij een draaibank, evengoed wel zes assen bezitten, zoals bij een bewerkingscentrum, waar frees- en boorbewerkingen in een machine zijn gerealiseerd (Carlier, 1986a, 1986b; de Korte, 1986b). In dit geval is er dus sprake van assen van de *machine* (zie Figuur 1-a en 1-b).

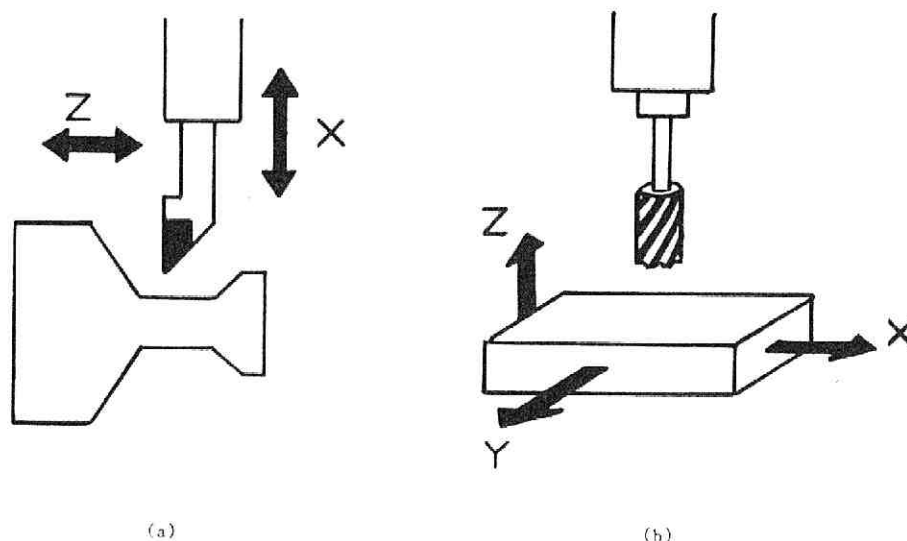


Fig. 1A-B: Voorbeeld van een Contouringsysteem (a) en een Puntbesturingssysteem (b).

Een Afbeeldingstheorie voor CNC-Programmeren.

Johnson-Laird (1983) levert met zijn boek '*Mental Models*' een belangrijke bijdrage aan de beantwoording van de vraag op welke wijze mensen zich een voorstelling vormen van de hen omringende werkelijkheid en op welke wijze deze voorstelling wordt vastgelegd en weergegeven in de taal. Johnson-Laird maakt een onderscheid tussen drie typen van mentale voorstellingen: beweringen (*propositions*), beelden (*images*) en mentale modellen (*mental models*). Een propositionele representatie is een mentale voorstelling van een (talige) bewering. Een mentaal model is een structurele analogie van de [buiten]wereld. Beelden zijn de perceptuele correlaten van een mentaal model, die het mogelijk maken een mentaal model vanuit een bepaald blikpunt (point of view) te gebruiken. Aangezien een bewering waar of niet-waar kan zijn, met betrekking tot de toestand in de wereld waar de bewering naar verwijst, is een propositionele representatie de mentale voorstelling van een functie van een toestand in de buitenwereld naar (een) waarheidswaarde(n) (*truth value*) (Johnson-Laird, 1983). Zo zorgt het mentale model in de zin 'het kopje staat naast het bord' ervoor, dat wij, op grond van onze, minder of meer specifieke, kennis van de buitenwereld, het kopje in onze voorstelling rechts naast het bord plaatsen. De beelden (*images*) zijn dan de perceptuele instanties van het mentale model die maken dat wij het kopje en bord 'zien' en vanuit verschillende blikpunten kunnen 'bezien'. De propositionele representatie geeft aan of de zin, indien deze zin vergeleken wordt met een daadwerkelijke toestand in de buitenwereld, waar of niet-waar is.

Johnson-Laird maakt in het verlengde hiervan een onderscheid tussen fysische modellen (fysische modellen verwijzen rechtstreeks naar de ons omringende fysische wereld), en conceptuele modellen (conceptuele modellen stellen meer abstracte zaken voor). Taaluitingen ('discourse') vereisen, volgens Johnson-Laird, een conceptueel model.

Voor ons betoog is het van belang hier reeds op te merken dat ons CNC-proces-model *zowel* een fysisch model als een conceptueel model belichaamt. De fysische kant komt tot uitdrukking

in de voorstelling van het produktvervaardigingsproces (de wijze waarop het werkstuk wordt gevormd uit een stuk metaal of kunststof). Beelden (de perceptuele correlaten) stellen ons in staat het werkstuk en de vervaardiging ervan vanuit verschillende gezichtspunten waar te nemen (het ruimtelijke aspect). De conceptuele kant speelt een rol bij het opstellen van het computerprogramma voor dat produktvervaardigingsproces (het verbale aspect). Zowel voor het voorstellen van het werkstuk en de bewerkingsvolgorde als voor het opstellen van het CNC-computerprogramma moet gebruik worden gemaakt van *dezelfde kennisbasis*, namelijk het mentale model. Noodzakelijk voor een juiste uitvoering van een CNC-programma is, dat het mentale model (het CNC-proces-model) weet te bewerkstelligen dat de conceptuele representatie en de perceptueel-fysische representatie één-op-één op elkaar worden afgebeeld. Of nu het één-op-één afbeelden op juiste wijze is geschied kan gecontroleerd worden aan de stand van de waarheidswaarde ('waar', of 'niet-waar'). Hetgeen tot uitdrukking komt in de vragen: 'Heeft het CNC-programma geleid tot het gewenste produkt?' en 'is de juiste programmacode in het CNC-programma gebruikt voor het laten uitvoeren van een bepaalde bewerking?'

Het boek 'Mental Models' van Johnson-Laird levert een belangrijke bijdrage aan het verschaffen van inzicht in de wijze waarop beelden en gedachten kunnen worden omgezet in beweringen in de taal. Toch is het niet de eerste stelselmatige verhandeling over de wijze waarop beelden, en gedachten, worden omgezet in beweringen in de taal. Treffend is de overeenkomst met het werk van de Oostenrijkse taalfilosoof Ludwig Wittgenstein. In 1921 verscheen er van zijn hand 'Die Logisch-Philosophische Abhandlung'. Een jaar later, in 1922, met een Engelse vertaling verschenen onder de nieuwe, door Moore voorgestelde, pompeuze titel: '*Tractatus Logico-Philosophicus*'. En in de *Tractatus* kan men al lezen: 2.1 Wij maken ons beelden van de feiten; 2.1.2 Het beeld is een model van de werkelijkheid; 3.1 In de bewering drukt de gedachte zich zintuiglijk waarneembaar uit; 4.0.1 De bewering is een beeld van de werkelijkheid. De bewering is een model [een toonbeeld] van de werkelijkheid, zoals wij ons die [werkelijkheid] denken [voorstellen]; 4.0.5 De werkelijkheid wordt met de bewering vergeleken; 5 De bewering is een waarheidsfunctie van de elementaire beweringen. (De elementaire bewering is een waarheidsfunctie van zichzelf).

De algemene strekking van de *Tractatus* zou men, naar onze mening, als volgt kunnen samenvatten: Wij maken ons beelden van de beelden. Hetgeen tot de, door Wittgenstein ontdekte, filosofische conclusie leidt, dat de stellingen (beweringen) van de Logika tautologieën zijn. En hoezeer dit ook een, filosofisch gezien, onthutsende vaststelling mag zijn, vanuit een cognitief psychologisch standpunt gezien blijft het, juist van wege deze tautologie, van belang te onderzoeken op welke wijze nu beelden uit een bepaald domein (zoals de taal) worden vergeleken met beelden uit een bepaald bereik (zoals de concrete realiteit). Cognitief psychologisch gezien is namelijk het (als een bijectie, en daardoor dus tautologisch) afbeelden van domein op bereik (en bereik op domein, want tautologisch) een interessant onderzoeksuitgangspunt.

Wittgenstein stelt (*filosofisch*), dat hetgeen wat het beeld van de werkelijkheid en het beeld van de bewering overeenkomstig moet hebben om met elkaar vergeleken te kunnen worden het logische beeld is. Dat wil zeggen:

Werkelijkheid → Beeld → (Logische Beeld) ↔ Beeld ← Bewering

Johnson-Laird betoogt (*taalpsychologisch*), dat de propositionele representatie de mentale voorstelling is van een functie van een toestand in de buitenwereld naar (een) waarheidswaarde(n).

Wat wij hier (*cognitief psychologisch*) willen betogen is, dat deze twee voorgaande uitspraken voor wat het CNC-programmeren betreft belichaamd en gereguleerd wordt door één onderliggend mentaal model: Het CNC-proces-model.

Dit brengt ons nu bij het volgende: In de cognitief psychologische literatuur bestaat er een oude, maar zeer levendige, onderzoekslijn waarbinnen men zich bezig houdt met de vraag of informatie over objecten in de buitenwereld in het semantische geheugen amodaal wordt

opgeslagen en gelijkelijk toegankelijk is voor zowel woorden als beelden (de 'common-code' theorie). Of dat deze informatie voor woorden en beelden in het semantische geheugen gescheiden ligt op geslagen (de 'dual-code' theorie).

Pylyshyn (1973), een criticus van de dual-code opvatting, heeft betoogd dat mensen abstracte mentale structuren bezitten, die niet bewust toegankelijk zijn, en die conceptueel en propositioneel van aard zijn. Voorstanders van het dual-code model van representatie (zoals Kolers & Brison, 1984; Paivio, 1986) voeren als belangrijkste argument tegen het common-code model aan, dat het speuren naar een overeenkomstig patroon in de verkregen resultaten voor woorden en beelden een veel te simpele wijze van wetenschap bedrijven is. En, dat de gebruikte methodieken, zoals het uitvoeren van onderzoek bij tweetaligen, te beperkt zijn om er zoveel veronderstellingen op te baseren en algemeen geldende conclusies uit te trekken (Kolers & Brison, 1984). Dat het omgekeerde, op grond van dezelfde resultaten dan maar te kiezen voor de dual-code theorie, eveneens te simpel is, is duidelijk geworden uit onderzoek naar het identificeren en benoemen van voorwerpen op het zogenaamde 'basic level' (Jolicoeur, Gluck & Kosslyn, 1984). Een en ander gaf Snodgrass (1984) aanleiding tot het formuleren van een informatie verwerkingsmodel voor het verwerken van beelden en woorden, dat zowel de dual-code opvatting als de common-code opvatting in zich verenigt.

Bij het programmeren van CNC-machines hebben wij te maken met (a) het produktvervaardigingsproces en (b) het CNC-computerprogramma voor het vervaardigen van dat produkt. Bij '(a)' dient de programmeur zich nauwkeurig de bewerkingstappen die leiden tot het eindprodukt voor te stellen. Bij '(b)' dient elke concrete stap vertaald, afgebeeld, te worden in de juiste, abstracte, CNC-computerprogrammacode.

Bovenstaande brengt ons tot een aantal uitspraken over CNC-programmeren, die te zamen de grondslag vormen van onze afbeeldingstheorie:

1. Onze *eerste aanneming* is: Hetgeen beweerd is voor de taal in het algemeen, geldt evenzeer voor het bijzondere geval van een kunstmatige taal, zoals een computerprogramma.
2. Onze *tweede aanneming* is: In het CNC-computerprogramma drukken de gedachten van de programmeur zich zintuiglijk waarneembaar uit en het CNC-computerprogramma drukt zich zintuiglijk waarneembaar uit in het produktvervaardigingsproces. Aldus bestaat er een één-op-één relatie tussen de stapsgewijze voortgang van het abstract talige CNC-computerprogramma voor de vervaardiging van het werkstuk en het concreet ruimtelijke vervaardigingsproces van dat betreffende werkstuk.
3. Uit een en twee volgt nu onze *derde aanneming*.
Namelijk: Dat het verbinden en vergelijken van taal en werkelijkheid evenzeer geldt voor het verbinden en vergelijken van (algemeen) CNC-programmeerkennis met de vereiste geometrie van het te vervaardigen werkstuk; en (specifiek) het schrijven van een CNC-programma voor dat betreffende werkstuk.
4. Uitgaande van een common-code model, onze *vierde aanneming*: Wat beide probleemoplossingsprocessen gemeenschappelijk hebben is een onderliggend mentaal model, het CNC-proces-model. Dit CNC-proces-model is een algemeen, zintuiglijke waarnemingsvorm vrij, model. Het model is de gemeenschappelijke kennisbasis voor beide probleemoplossingsprocessen.
5. Tenslotte, onze *vijfde aanneming*: (cf. Johnson-Laird, 1983; Wittgenstein, 1921) De waarheidswaarde ('waar' of 'niet-waar') geeft aan of het afbeelden van taal en werkelijkheid op juiste wijze is geschied.

In het hierna volgende gedeelte willen we nu het informatieverwerkingsmodel voor CNC-programmeren nader uitwerken. Twee onderwerpen zullen daarbij in ons betoog de nodige aandacht krijgen: (1) De procesgang van idee tot produkt, en (2) De ontwikkelingsgang van novice tot expert.

HET INFORMATIEVERWERKINGSMODEL.

Algemeen.

Door de invoering van computergestuurde gereedschapswerktuigen ter verbetering van het productieproces heeft het vak van machinebediener in korte tijd een grote verandering ondergaan van eerst vooral *handvaardigheid* (Skill) naar thans voornamelijk *geestvaardigheid* (Cognitive Skill). Laur-Ernst (1982) heeft er op gewezen, dat de machine bediener die thans werkt met informatieverwerkende apparatuur voor alles een 'abstract cognitieve vaardigheid' dient te bezitten. Want het vertalen van het technische probleem in een door de CNC-machine 'begrepen' programma is een cognitieve prestatie, dat een handeling is, die verricht wordt in het hoofd van de CNC-machinebediener.

Lag vroeger het accent op de nauwkeurigheid van de bewerking, op de kundigheid die men als vakman had, thans ligt het accent voornamelijk op het hebben van kennis en inzicht in bewerkingscondities. Kundigheid, met de hand, is verschoven naar deskundigheid, met het hoofd.

Mentaal Model.

Onderzoek heeft aangetoond, dat het hebben van kennis, in de vorm van een mentaal model, van de wijze waarop een computer werkt in termen van de veronderstelde interne structuur en interne gegevensverwerking:

(A) het leren bedienen van de computer vergemakkelijkt (Bayman & Mayer, 1984; du Boulay e.a., 1981; Carroll & Mack, 1985; Carroll & Thomas, 1982; Kieras, 1987; Kieras & Bovair, 1984; Mayer, 1975, 1976a, 1981; Norman, 1983; Polson & Kieras, 1984; Young, 1981)

(B) de mens-machine-interactie in het algemeen vergemakkelijkt (Carroll & Thomas, 1982; Kieras, 1987; Kieras & Bovair, 1984; Mayer, 1979; Rouse & Morris, 1986; Young, 1983).

Dit vergemakkelijken heeft met name betrekking op het volgende. In de kontekst van het aanleren van specifieke procedures voor het bedienen van het apparaat ligt het directe nut van de Hoe-Het-Werkt kennis, het denkbeeld, in het voordeel dat de gebruiker hiervan heeft bij het *exact afleiden* van hoe het apparaat bediend met worden met betrekking tot een bepaalde, opgedragen, taak (Kieras, 1987; Kieras & Bovair, 1984).

Bij eerdere gelegenheid (Rakers, 1987) hebben wij erop gewezen, dat wat een computergestuurd gereedschapswerktuig wezenlijk onderscheidt van andere computersystemen de direkt observeerbare werking (de uitgevoerde bewerkingen op het opgespannen materiaal) is. Anders geformuleerd: Het hiervoor ter sprake gebrachte mentale model is voor CNC-machines niet slechts een denkbeeld met betrekking tot de veronderstelde *interne* structuur en werking ('*a picture of its inner workings*', Young, 1981). Integendeel. Het is met name een mentaal model met betrekking tot de *externe*, direkt observeerbare, werking. Dat wat het CNC-programma doet (zie Figuur 2), het produceren van, bijvoorbeeld, een frees-, draai-, of boorprodukt, kan rechtstreeks worden waargenomen aan de machine.

Een veronderstelling luidt dan ook, dat deze direkt te observeren werking van het CNC-computerprogramma een aanspraak zal doen op bepaalde, specifiek, voor deze apparatuur noodzakelijke cognitieve vaardigheden. Beschikken over voldoende *ruimtelijk inzicht*, *analytisch* kunnen *denken*, en in staat zijn zich op adequate wijze *een beeld te vormen* van het produktvervaardigingsproces, zijn naar onze mening de drie basiselementen die nodig zijn voor het uitvoeren van de CNC-programmeertaak.

Het Ruimtelijke Aspect.

Het vervaardigen van een werkstuk op een CNC-machine is een proces, dat rechtstreeks aan de machine kan worden waargenomen.

Het proces voltrekt zich in de concrete realiteit. Dit is eveneens aan de orde, indien er voorafgaande aan het produktvervaardigingsproces eerst simulatie op een beeldscherm plaatsvindt (de Korte, 1986a; Prischow & Viefhaus, 1987; Weule e.a., 1987).

Simon (1973) onderscheidt vanaf het idee, het ontwerpen van het produkt, tot en met het

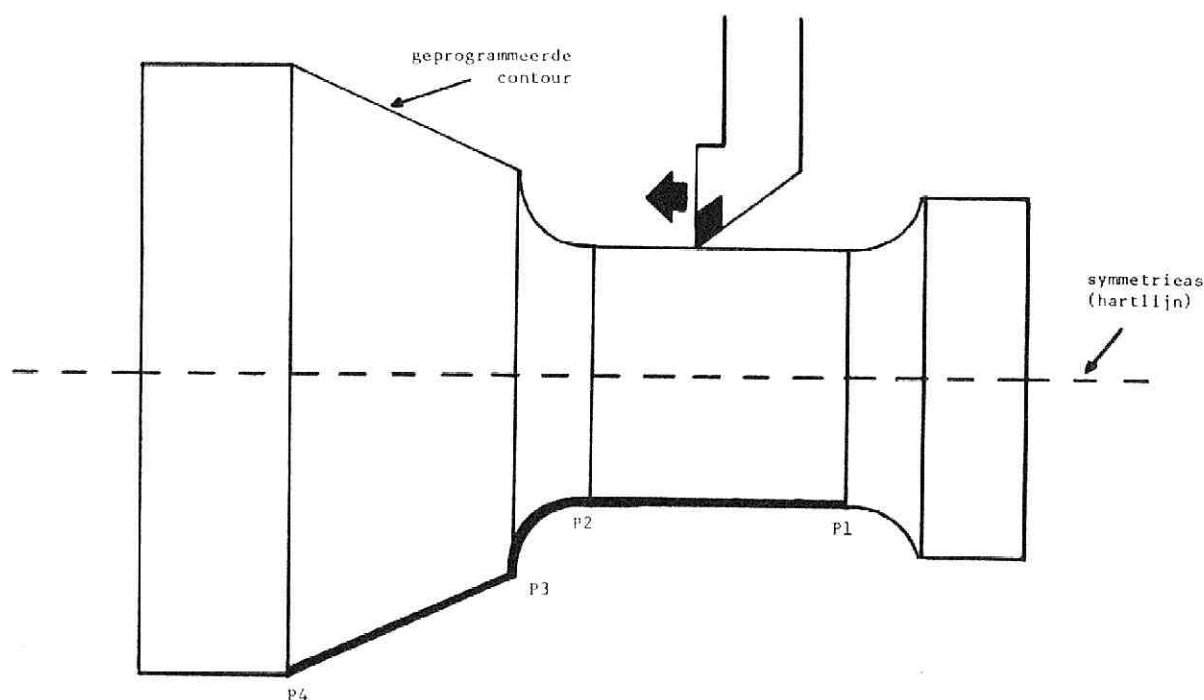


Fig. 2: Contour van een eenvoudig draaiproduct.

Tabel 1: Twee programmaregels voor een deel van het werkstuk uit figuur 2.

Punten	Regelnummer	G.-Functie	Diameter (X-as)	Sledeverplaatsing (z-as)	Aanzetsnelheid (g)
P2	N 16	G G01	X 1800	Z —5000	F 40
P2	17	G02	2500		40

inspecteren na de vervaardiging twee soorten van werktuigbouwkundige gegevensverwerkings-systemen. *Interne* gegevensverwerking staat voor de wijze waarop (a) technologische informatie, zoals signalen, stroomtoevoer, etc., door de machine wordt verwerkt, en waardoor (b) sledebewegingen ten behoeve van de werkstukvervaardiging plaatsvinden. *Externe* gegevensverwerking verwijst naar de wijze waarop (a) kennis van het specifieke produkt, en (b) technische informatie over het productieproces (inclusief logistieke zaken, kwaliteitseisen en veiligheidsaspecten) worden aangewend bij het schrijven van het computerprogramma.

Hoezeer deze werktuigbouwkundige tweedeling ook ter zake doende mag zijn, een cognitief psychologische opdeling, gericht op de mens-machine-interactie, dient naar onze mening de accenten anders te leggen. Enerzijds is daar de zuiver werktuigbouwkundige kant, de CNC-machine-kant, alsmede de elektronische gegevensverwerking door de machine. Anderzijds, de menszijde, de gegevensverwerking met betrekking tot een bepaalde machine in termen van het mentale model, het 'Weten-Hoe-het-werktuig-Werkt'. Dus, voor wat dit laatste, de menszijde, betreft: Een samengaan van externe gegevens verwerking (cf. Simon, 1973) *plus* de observaties, veronderstellingen en voorspellingen met betrekking tot '(b)' sledebewegingen, in een drie dimensionale ruimte, ten behoeve van de werkstukvervaardiging (cf. Kieras, 1987; Kieras & Bovair, 1984; Rouse & Morris, 1986).

Het op de machine vervaardigde CNC-eindprodukt is een voorwerp in een driedimensionale ruimte. Het is zichtbaar, tastbaar, meetbaar, etc.

Maar het CNC-werkstuk ontstaat echter *in de tijd*, in de vierde dimensie.

Dit heeft als noodzakelijke, cognitief psychologische, consequentie, dat de CNC-programmeur beslissingen dient te nemen over welke bewerkingen, stappen, hij achtereenvolgens door de machine wil laten uitvoeren. Het CNC-denkbeeld is daarom dan ook een proces model. Dat wil zeggen: Een mentaal model van het vervaardigingsproces van een CNC-produkt.

Het Verbale Aspect.

In het eerste stadium van het probleemoplossingsproces is de praktische vraag (het idee) vertaald in een wiskundig ruimtelijk model, en vastgelegd in een werktekening. Met het bepalen van de bewerkingsvolgorde, een fase voorafgaande aan het schrijven van het programma, zijn we in het tweede stadium van het probleemoplossingsproces beland, en daarmee bij het verbale aspect.

Een bewerkingsplan voor een CNC-produkt bestaat uit een lijst (het programma van stappen) waarin de uit te voeren bewerkingen stap na stap worden vastgelegd en beschreven in de gewone dagelijkse omgangstaal (eventueel met toegevoegd commentaar). Tevens worden materiaalgegevens, gereedschapkeuzes, de technologische instelgrootheden, enzovoorts vast gelegd. Nadat het bewerkingsplan is opgesteld kan het CNC-programma worden geschreven. Het programmeren bestaat uit het exact vastleggen van de gereedschapsbaan, voorzien van de bijbehorende instelgrootheden (zoals aanzetsnelheid, snedediepte en toerental).

Wat wij hiervoor voor het ruimtelijke aspect hebben opgemerkt, dat een drie dimensionaal voorwerp (het CNC eindprodukt) ontstaat in de tijd, geldt evenzeer voor het verbale aspect. Een CNC-computerprogramma loopt van programma-begin tot programma-einde eveneens in de tijd. En aangezien elke programmaregel een stadium vertegenwoordigt in het produktvervaardigingsproces kan er worden aangenomen, dat dit ruimtelijke vervaardigingsproces stap na stap samenhangt met de wijze waarop het CNC-programma, regel na regel, door de computer wordt afgewerkt. Anders geformuleerd: Er bestaat een één-op-één relatie tussen de talige voortgang van het computerprogramma in de tijd en het ruimtelijke vervaardigingsproces in de tijd. Het abstracttalige CNC-programma wordt één-op-één afgebeeld in het concreet-ruimtelijke vervaardigingsproces. Wij veronderstellen dat deze afbeelding een bijectie is, hetgeen wil zeggen, dat naar elk element (beeld) van de verzameling elementen uit het ruimtelijke bereik precies een element (beeld) uit de verzameling van elementen uit het talige domein wijst, en omgekeerd. Dit één-op-één afbeelden geschiedt, als mentale prestatie, eerst in het hoofd van de programmeur. Cognitief psychologisch gezien betekent dit, dat er verondersteld mag worden dat ook het programmeren van computergestuurde gereedschap-werktuigen onderhevig zal zijn aan de informatieverwerkingsbeperkingen van de mens.

Figuur 3 geeft globaal de procesgang weer van werktekening naar werkstuk, met daarbij aangegeven of dit geschiedt in het ruimtelijke of talige informatieverwerkingsgebied, opgedeeld naar mens- en machinezijde.

Computerprogrammeren: Van Novice naar Expert.

Bij het leren programmeren van computers valt er een ontwikkelingsvoortgang waar te nemen van nieuweling (*novice*) via gevorderde (*intermediate*) naar deskundige (*expert*). Voor elk stadium in deze ontwikkelingsgang kan een specifiek patroon worden onderscheiden, voor wat betreft de wijze waarop de programmeerkennis mentaal wordt voorgesteld en bij het probleemoplossen wordt aangewend. Hoewel er tot op heden geen exacte definities van de begrippen 'novice' en 'expert' zijn gegeven (zie Shneiderman, 1976, 1978 en 1980), hebben studies met betrekking tot novice- en expertgedrag het volgende kunnen aantonen.

Beginnende computerprogrammeurs:

(A) onthouden aangeboden computerprogramma's op grond van een syntax georiënteerde geheugenorganisatie (Adelson, 1981);

(B) vormen zich een concrete, meer op details gerichte, mentale voorstelling van het

programma (d.w.z. een voorstelling uitgaande van hoe een programma funktioneert: algoritmen) (Adelson, 1984);

(C) vertonen een rijke variëteit aan programmeervergissingen, voortkomende uit het verbinden van betekenissen van begrippen uit de gewone dagelijkse omgangstaal met betekenissen van programmataalbegrippen (McKeithen, Reitman, Rueter & Hirtle, 1981);

(D) presteren minder goed dan experts, daar zij veel minder ervaring hebben dan deskundigen met het schrijven van dezelfde soort procedure(s) in verschillende andere programmeertalen (Sheppard e.a., 1979).

Mayer (1975, 1976a, 1976c, 1979, 1981) heeft aangetoond, dat het leren programmeren gemakkelijker verloopt als (a) de interne gegevensverwerking van de computer wordt uitgelegd aan de hand van een *concreet model*, en (b) als nieuwelingen aangemoedigd worden (eerst) de te leren technische begrippen in *eigen woorden* weer te geven.

Experts, in tegenstelling tot nieuwelingen:

(A) vormen een abstracte geheugenorganisatie van het aangeboden computerprogramma (waar het betreffende programma toe dient: 'output specification') (Adelson, 1984);

(B) trachten aangeboden programma's te herinneren uitgaande van de abstracte en hiërarchisch, van algemeen naar specifiek, georganiseerde mentale voorstelling die zij zich van het programma hebben gevormd (een mentale voorstelling uitgaande van de functie die het programma als geheel dient te vervullen) (Adelson, 1981);

(C) vertonen een geheugenorganisatie opgebouwd op grond van hun programmeerkennis met betrekking tot de taal waarin het aangeboden programma was geschreven (McKeithen e.a., 1981); en

(D) hoewel voor die programmeurs die minder dan drie jaar programmeerervaring hadden, het aantal talen waarin men kon programmeren en de score op de lijst van bij hen bekende Fortran begrippen, uitstekende voorspellers bleken te zijn voor de wijze waarop zij zouden presteren, bleken deze voorspellingen niet uit te komen voor programmeurs met meer dan drie jaar ervaring (Sheppard e.a., 1979).

Overeenkomstige resultaten zijn er ook gerapporteerd voor expert-novicegedrag ten aanzien van het onthouden en uit het hoofd natekenen van werktekeningen van elektronische schakelingen (Egan & Schwartz, 1979), het oplossen van natuurkunde vraagstukken, en wiskunde problemen (Larkin, McDermott, Simon & Simon, 1980; Schoenfeld, 1979), waterbouwkunde (Bhaskar & Simon, 1977) en schaken en GO (Charness, 1976; Chase & Simon, 1973; Frey & Adelman, 1976; Goldin, 1978a, 1978b, 1979; de Groot, 1966; Reitman, 1976).

Moran (1981) heeft er voor gepleit, dat, juist voor wat de toepassing van onderzoeksresultaten betreft, onderzoek zich met name dient te richten op de verschillen in het verloop van het leerproces bij computergebruikers. Hierbij dient echter wel te worden aangetekend, dat de door Moran en hiervoor door ons aangeduide onderzoeksresultaten tot nu toe overwegend verkregen zijn bij *hoog begaafde, hoger opgeleide* en *beter getrainde* personen: Studenten aan universiteiten en experts met een jarenlange beroepservaring. De verkregen resultaten dienen dan ook in die, beperkte, kontekst geplaatst en beoordeeld te worden.

Desondanks willen we hier een resultaat nadrukkelijk vermelden. Polson & Kieras (1984) rapporteerden, dat novices expliciet aan elke mededeling die er op het beeldscherm verschijnt aandacht besteden en op juistheid controleren. En eveneens controleren zij expliciet of elke procedure ook op juiste wijze is uitgevoerd.

Juist bij CNC-machines is dit stapsgewijs en logisch na elkaar laten uitvoeren van bewerkingen zo belangrijk, aangezien bijna elke stap (dat wil zeggen: elke CNC-programmaregel) in het CNC-programma een bepaalde bewerking met het gereedschap in het te bewerken materiaal betekent (voor een voorbeeld van een draaiprodukt, zie Tabel 1 en Figuur 2). Een eenmaal uitgevoerde, verkeerde, bewerking in het ruwe materiaal kan niet meer ongedaan worden gemaakt. Materiaal dat verwijderd is (verspaand) is weg.

Zoals hiervoor is beschreven onderscheiden experts zich van nieuwelingen, doordat zij uitgaan van een abstracte en hiërarchisch van algemeen naar specifieke kenmerken georganiseerde mentale voorstelling. Dat dit onderscheid eveneens voor CNC-programmeren van belang is zullen wij nu hierna nader uitwerken.

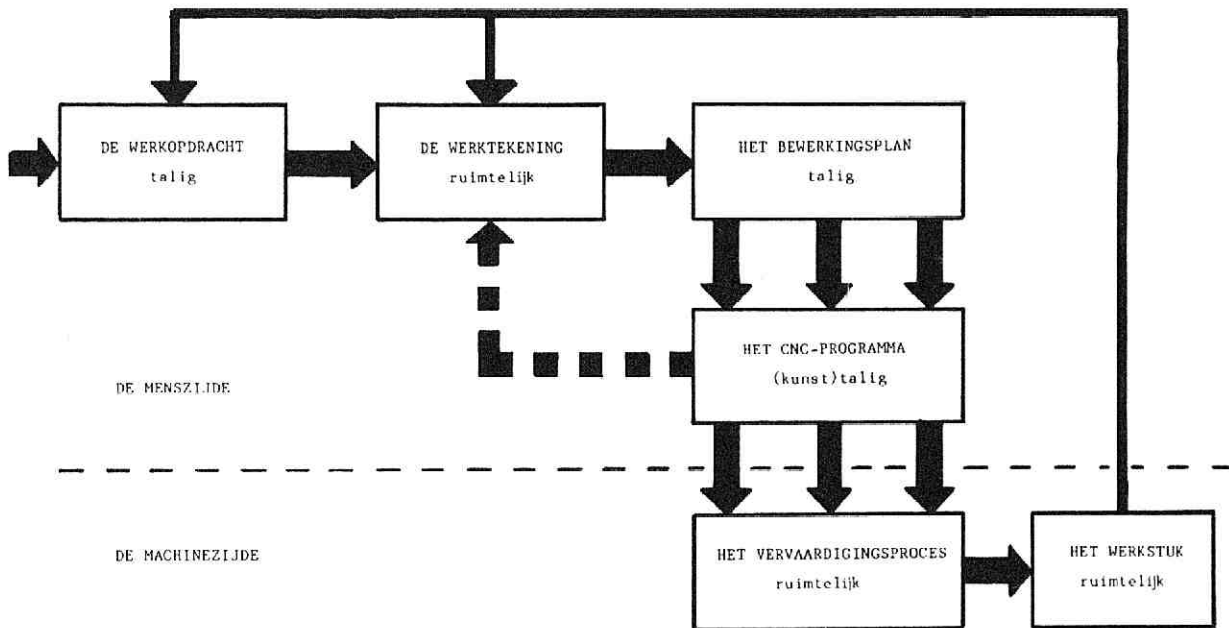


Fig. 3. Handelingsverloop van Werkopdracht naar Werkstuk.

CNC-machines programmeren.

Programmeren wil zeggen: Het adequaat in een programma kunnen toepassen van programmeerkennis. Maar, zoals alle kennis, programmeerkennis dient te worden aangeleerd.

Allereerst, declaratieve kennis (*Weten Wat*) van de geometrie van het te vervaardigen werkstuk en van de technologie voor het vervaardigen van een werkstuk, voornamelijk ruimtelijke aspecten. Alsmede (*Weten Wat*) van de programmeertaal, de verbale aspecten.

Ten tweede, procedurele kennis (*Weten Hoe*) van de variabelen en functies van de programmeertaal en de werking van het gereedschapswerktuig.

Maar declaratieve en procedurele kennis alleen is niet voldoende. *Weten Wat* en *Weten Hoe* zijn voorwaarden voor:

Ten derde, *Weten Wanneer* (en *Waar* in het CNC-programma) er *Wat* en *Hoe* geprogrammeerd moet worden (*criterium kennis*). Bij NC- en CNC-programmeren ook wel aangeduid met 'sequence control' (Koren, 1983). Hetgeen wil zeggen: 'Het besturen van werktuigen door een in een programma vastgelegde reeks van instructies' (Simons, 1982).

Het CNC-proces-model (zie Figuur 4) is dus een voorstelling van het gehele produktvervaardigingsproces van een bepaald produkt op een bepaalde CNC-machine. Met behulp van het proces-model dient de programmeur een ruimtelijke reeks beelden (de uit de werktekening afgeleide baan van het gereedschap) af te beelden op een talige reeks beelden (het CNC-programma). Het (abstracte) CNC-programma wordt ten slotte, middels de machine, weer één-op-één afgebeeld in de (concrete) realiteit.

We kunnen nu het ruimtelijke beeld definiëren als een toestand in het vervaardigingsproces ('the current state'): een momentopname. En het talige beeld laat zich aldus definiëren als een 'toestand', een met het ruimtelijke beeld corresponderend moment in de programma-afwerkingsfase, in het CNC-programma. Afbeelden wil dan ook zeggen: Een beeld na beeld transformatie van de *vorm* van de ene representatie in de *vorm* van de andere representatie. *Afbeelden is het transformeren van vormkenmerken*. Geometrische kenmerken in taalkenmerken (de menszijde) en van taalkenmerken in geometrische kenmerken (de machinezijde).

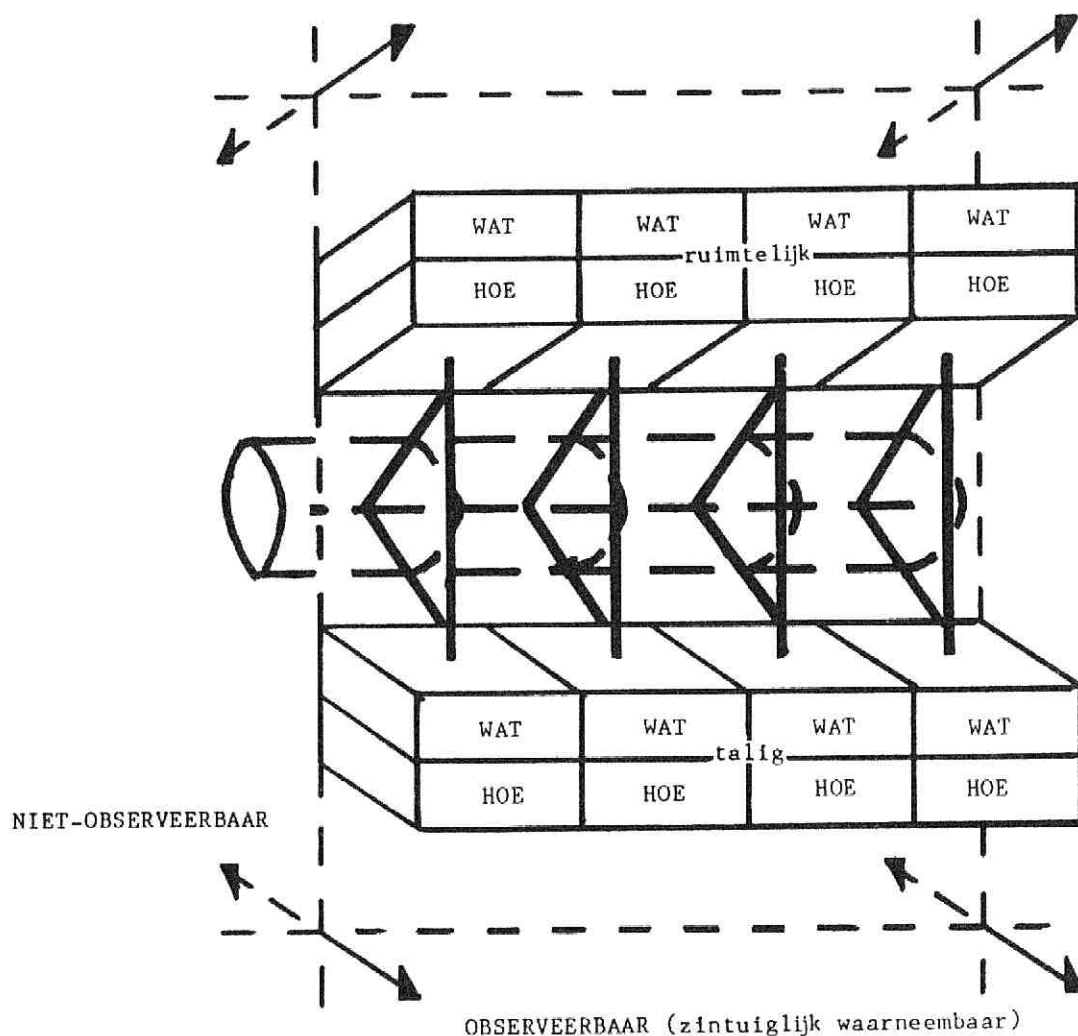


Fig. 4. De twee zintuiglijk waarneembare verschijningsvormen van het ene onderliggende CNC-proces model.

Nu zijn echter de werktekening zowel als het vervaardigde werkstuk statische beginen eindobjecten. Statisch, doordat zij eindprodukten zijn van een reeks handelingen. Tevens bevinden deze eindprodukten zich als voorwerp in ten hoogste een drie dimensionale ruimte. Het produktvervaardigingsproces vindt echter plaats in de tijd, in een vier dimensionale ruimte. Dit betekent, dat het CNC-programma als dynamisch proces een *uitgebreide afbeelding* is. Dat wil zeggen: De werktekening is uitgebreid met de door de programmeur gemaakte beslissingen Waar (en Wanneer) in het programma bepaalde bewerkingen uitgevoerd dienen te worden. Dit betekent, dat verschillende programma's, die onderling sterk afwijken in een vakbekwame aanpak zoals efficiëntie, etc., aldus desondanks hetzelfde eindprodukt kunnen opleveren.

CNC-Programmeren: Van Novice naar Expert.

In het hierna volgende gedeelte willen we het CNC-procesmodel (zie Figuur 4) en de expert-noviceverschillen plaatsen in het algehele informatieverwerkingsmodel voor CNC-programmeren. Daarbij zullen we de vijf uitspraken nader uitwerken tot specifieke hypothesen.

Er is aangetoond, dat er bij het leren een clustering van informatie in bepaalde patronen optreedt die een bijzondere psychologische betekenis bezitten (Buschke, 1976; Furukawa, 1972; Miller, 1956; Simon, 1974). Dit groeperen van informatie in betekenisvolle eenheden heeft, in het Engels, de naam 'chunking' gekregen (Miller, 1956). Leren kenmerkt zich in hoge mate door chunking en het herkennen van aangeleerde 'chunks' (Buschke, 1976). Tevens is er aangetoond,

dat experts 'een vocabulair' (Simon, 1974) bezitten van overbekende patronen ('familiar chunks') en novices niet of nauwelijks.

Het programmeren van CNC-machines wil zeggen, het vastleggen van bepaalde bewerkingen die stapsgewijs, na elkaar, door de machine uitgevoerd dienen te worden. Daarom dient de CNC-programmeur te beschikken over voldoende ruimtelijk inzicht, analytisch kunnen denken en in staat te zijn zich op adequate wijze een denkbeeld te vormen van het gehele produktvervaardigingsproces.

Uit onderzoek is naar voren gekomen (Rouse & Morris, 1986), dat het al dan niet op adequate wijze hebben, of kunnen vormen, van een mentaal model afhankelijk is van (a) de mate van deskundigheid (expertise) die men met betrekking tot de taak bezit, (b) de mate waarin er gebruik wordt gemaakt van aanwijzingen die in de taak zelf gelegen zijn, en (c) de context waarin het mentale model gebruikt dient te worden.

Zoals reeds ter sprake is gebracht, valt er (a) een ontwikkelingsvoortgang te constateren van nieuwelingen naar deskundigen, (b) gaan experts uit van een algemeen en abstract hiërarchisch georganiseerde mentale voorstelling van het programma en nieuwelingen van een concrete mentale voorstelling, en (c) besteden nieuwelingen expliciet aan elke mededeling die er op het beeldscherm verschijnt aandacht.

Een mogelijke bron van fouten bij het leren programmeren van CNC-machines zou dit typische, punt voor punt alles willen controleren, gedrag van novices kunnen zijn. Nieuwelingen houden zich in tegenstelling tot experts te veel bezig met het verifiëren van details (Egan & Schwartz, 1979). Met hierbij de opmerking, dat de menselijke geest tijdens dit controleren eenvoudig niet in staat is om alle relevante informatie tegelijkertijd vast te houden of te verwerken (Miller, 1956). Tevens kan er hier verondersteld worden, dat uitgaan van uitsluitend een *concrete* mentale voorstelling, zoals nieuwelingen plegen te doen, nog eens extra belastend zal zijn bij het afbeelden van geometrische kenmerken op (abstracte) taalkenmerken.

Experts daarentegen kunnen op grond van een andere geheugenorganisatie, ten eerste, wel grotere eenheden relevante informatie (chunks) aan (Adelson, 1981, 1984; Buschke, 1976; Egan & Schwartz, 1979; McKeithen e.a., 1981; Miller, 1956; Simon, 1974; Wiedenbeck, 1985). En, ten tweede, het afbeelden van geometrische kenmerken op (abstracte) taalkenmerken zal voor experts minder belastend zijn, daar zij als experts al uit gaan van een *abstract* georganiseerde mentale voorstelling.

Van experts kan echter eveneens worden verwacht dat zij andere, hun ontwikkelingsstadium kenmerkende, programmeerfouten zullen maken (Adelson, 1984; Boies & Gould, 1974; Rakers, 1987). Er kan verondersteld worden, dat deze expertfouten zich bijna uitsluitend zullen vertonen op het Weten Wanneer (criterium) niveau. Een (eventueel) minder adequaat mentaal model zal in het geval van de experts dan ook (zintuiglijk) zijn waar te nemen in het geval van het bewerkingsplan.

Bij nieuwelingen zal een inferieur mentaal model echter zijn vast te stellen uit de fouten die er gemaakt worden met betrekking tot de geometrie van het werkstuk. Dit tengevolge van het feit, dat nieuwelingen (a) minder goed op de hoogte zijn van het Weten Wat (feiten) en Weten Hoe (procedures) van de programmeertaal en werking van het gereedschapswerktuig, en (b) ten gevolge daarvan voortdurend zullen moeten controleren of de uitgevoerde procedure overeenstemt met de geldende criteria. Nieuwelingen dienen voortdurend de 'stand' van de waarheidswaarde (waar, of niet-waar) te controleren. Hetgeen betekent, dat de één-op-één relatie tussen het CNC-programma en de uitwerking van dat programma in het ruimtelijke bereik (het vervaardigingsproces) voor nieuwelingen niet een chunk na chunk, maar een programmaregel na programmaregel is (cf. Polson & Kieras, 1984). De reden hebben we hiervoor al aangegeven: Het afbeelden van beelden uit het concrete domein op het abstracte bereik kost nieuwelingen, uitgaande van een concrete mentale voorstelling, extra veel inspanning.

Zal het expertgedrag zich kenmerken door een algemene 'top-down' aanpak, nieuwelingen zitten, bij wijze van spreken, met hun neus in de punt van het gereedschap.

In positieve zin betekent het een en ander ook, dat programma's voor complexe werkstukken, geschreven door experts (a) kwantitatief minder fouten zullen vertonen, dan novice programmas, (b) compacter zullen zijn (minder regels), en (c) efficiënter zullen zijn (geen onnodige sledebewegingen, gereedschapwisselingen, etc.).

Maar het interessantste hebben we voor het laatst bewaard.

Mayer (1975, 1976a, 1976b, 1979, 1981) heeft aangetoond, dat het leren programmeren gemakkelijker gaat als: (a) de interne gegevensverwerking door de computer wordt uitgelegd aan de hand van een concreet model, en (b) als leerlingen worden aangemoedigd de te leren technische begrippen eerst in eigen woorden weer te geven. Carroll & Thomas (1982) en Carroll & Mack (1985) hebben die reeds bij nieuwelingen bestaande verbale structuur (het in eigen woorden zeggen) benadrukt, door te wijzen op het feit dat nieuwelingen in staat zijn spontaan metaforen te genereren. Metaforen, die *tijdelijk* als mentaal model dienen bij het *actief leren* bedienen van een computer. McKeithen e.a. (1981) rapporteerden echter, dat nieuwelingen een rijke variëteit aan programmeervergissingen vertoonden, voortgekomen uit het verbinden van betekenissen van begrippen uit de dagelijkse omgangstaal met betekenissen van programma-taalbegrippen. Een bevinding die kan worden opgevat als pleitend *tegen* het gebruik van metaforen en het in eigen woorden weergeven van te leren technische begrippen.

Het interessante voor wat betreft het CNC-programmeren is nu, dat het concrete model het produktvervaardigingsproces is, en het in eigen woorden weergeven is het bewerkingsplan, met eventueel toegevoegd commentaar. Daar er bij CNC-programmeren het (concreet) ruimtelijke aspect en het (abstract) verbale aspect allebei even belangrijk zijn en rechtstreeks, één-op-één, samenhangen kan er het volgende verondersteld worden: Nieuwelingen zullen *gehinderd* worden bij het leren programmeren als zij met gelijkenissen werken. Want een metafoor voegt een extra afbeeldingsstap toe aan het transformeren van beelden uit een ruimtelijk domein naar beelden in een talig bereik.

Analoog geredeneerd zullen novices eveneens *gehinderd* worden als zij een teveel aan criteria-informatie (Weten Wanneer) krijgen aangeboden. Dat wil zeggen, aanwijzingen waar zij bij het vastleggen van de gereedschapbaan allemaal rekening mee dienen te houden. In het verlengde van wat Polson & Kieras (1984) rapporteerden is namelijk het volgende van belang. Volgens du Boulay e.a. (1981) hebben nieuwelingen baat van het op het scherm laten zien van twee soorten informatie: (a) sequentiële informatie, d.w.z. een toelichting bij en verduidelijking van de opeenvolgende stappen, en (b) toestand informatie, een verduidelijking van hoe het is gekomen dat een programma zich thans in een bepaalde 'toestand' van gegevensverwerking bevindt. Sheppard e.a. (1979) vonden echter, dat ter verheldering toegevoegd commentaar alleen hielp in korte programma's. De *context* (zie hierover ook Rouse & Morris, 1986) is veel belangrijker voor het programmeren. Het is in dit verband het vermelden waard, dat een medewerker van Johnson-Laird heeft gewezen op de cognitief psychologische mogelijkheden die de latere Wittgenstein, die van de Philosophische Untersuchungen (PU), biedt voor wat betreft de toepassing van de betekenis van begrippen in een bepaalde context. Dat wat de betekenis en het gebruik van begrippen in een context bepaalt zijn de spelregels die bij die context horen: De spelregels van het taalspel (zie bijvoorbeeld: L. Wittgenstein, PU, nrs 69-71). Garnham (1980) heeft hier gewezen op de *criteria* die het gebruik vanuit een bepaalde context gezien bepalen. Daarmee wordt een bepaald begrip dus opgevat als een invulbare, context (= 'Sprachspiel') afhankelijke, grootheid (als een *default value*) (Garnham, 1980). Een opvatting die interessante aanknopingspunten biedt voor het bestuderen van het al dan niet weten te gebruiken van begrippen met een kunstmatig bepaalde betekenis, zoals in het geval van een programeertaal. Voor wat betreft het CNC-programmeren betekent dit, dat de programmeur de criteria van het CNC-taalspel (Wat en Hoe) dient te vergelijken met de criteria van het CNC-ruimtespel (Wat en Hoe van de gereedschapbaan). Kennis en inzicht in de criteria (in de verspaningscondities) zal daarom van fundamenteel belang zijn voor het Weten Wanneer programmeerkennis toegepast dient te worden.

De CNC-criteria gegeven vanuit de context het werkkader (zie Figuur 5) zullen volgens ons dan ook het denkkader (de 'Frame of Mind' waarbinnen de programmeur dient te werken, cf.

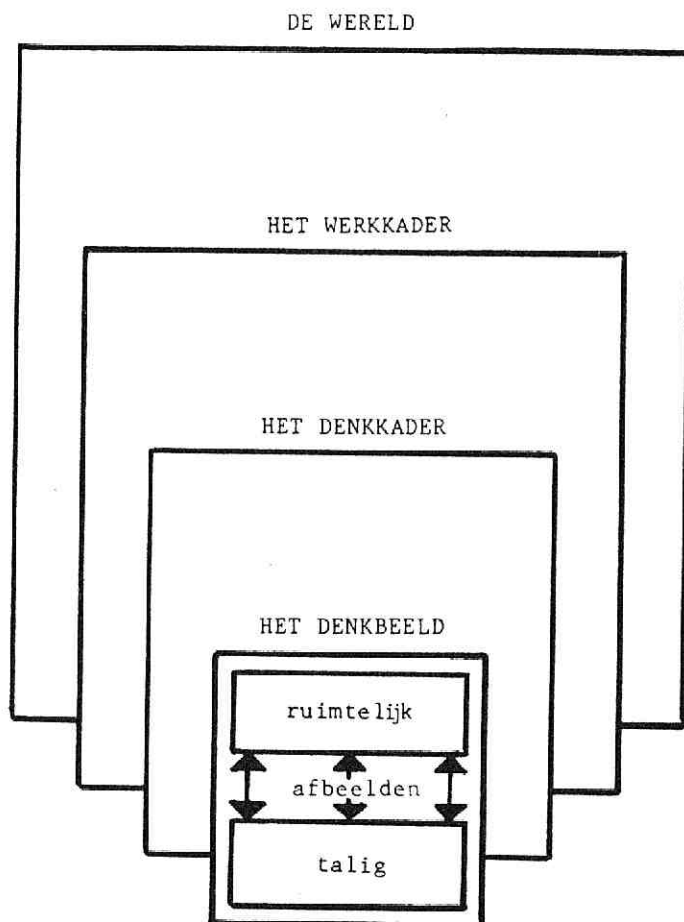


Fig. 5. Kennis-, Werk- en Denkkaders.

Rakers, 1987; Ryle, 1949/1980) van de novice belasten en daarmee een toch al minder goed denkbeeld van het produkt vervaardigingsproces nog eens extra ten negatieve beïnvloeden. Uit inmiddels uitgevoerde experimenten, bij studenten werktuigbouwkunde, toegepaste onderwijskunde, en programmeerexperts met een jarenlange beroepservaring, van de Universiteit Twente, is ons gebleken, dat deze laatste hypothese inderdaad experimenteel was vast te stellen (Rakers, Pieters, & de Bruijn, in voorbereiding).

Figuur 6 toont nu het informatieverwerkingsmodel voor het programmeren van computergestuurde gereedschapswerktuigen.

RELEVANTE ONDERZOEKSVRAAGSTELLINGEN.

Het is niet verwonderlijk, dat verschillende onderzoekers reeds hebben gewezen op het belang van het bestuderen van Hoe en Wanneer proefpersonen hun (programmeer)kennis toepassen (Adelson, 1981; du Boulay e.a. 1981; Larkin, McDermott, Simon & Simon, 1980; Lesgold, 1984; Mayer, 1981; Schoenfeld, 1979; Sheppard e.a., 1979). Want, om Schoenfeld (1979) te citeren:

‘Dat studenten een bepaalde probleemoplossingstechniek *kunnen* leren is geen garantie dat dat zij die ook *zullen* gebruiken. () Zij dienen niet alleen onderwezen te worden *hoe* [probleemoplossings]strategieën te gebruiken, maar ook *wanneer*’.

Voor wat betreft de stand van de huidige CNC-technologie heeft deze opmerking van Schoenfeld een bijzondere betekenis. Door de invoering van geavanceerde bewerkingscentra,

Senker, 1984; Thijsen & Mulder, 1987a, 1987b; Weule e.a., 1987; noten 1, 2, 3, 4).

Een veel gehoorde klacht van het bedrijfsleven in Nederland blijkt echter te zijn, dat de vooropleiding van de afgestudeerde MTSer-Werktuigbouwkunde niet probleemloos aansluit bij de verlangens van de industrie (Heesters, 1986; Nijhof & Mulder, 1986; Reijmerink & Mulder, 1987; Thijsen & Mulder, 1987a, 1987b; noten 1 en 4). Als gevolg daarvan krijgen de pas afgestudeerde MTSer-Werktuigbouwkunde en -Produktietechniek in hun startfunctie veelal eerst een 'on the job' inleerperiode (Heesters, 1986; noot 1).

Cognitief psychologisch onderzoek zou zich naar onze mening met name kunnen richten op de vraag op welke wijze *Weten Wat*, *Weten Hoe* en *Weten Wanneer* kennis wordt voorgesteld en bij het programmeren wordt aangewend. De kernvraag hierbij luidt: Op welke wijze vindt het afbeelden (transformeren) van beelden uit het domein op beelden uit het bereik plaats? En wat is de rol daarbij van de voor dat domein en bereik geldende *criteria*? Instruktietechnologisch onderzoek zou zich met name kunnen richten op de wijze in hoeverre het mogelijk is het leerproces van novice naar expert te beïnvloeden en te *bespoedigen*. Vooral dit bespoedigen is van belang, gezien het geringe aantal lesuren dat er, noodgedwongen, op een school ter beschikking staat voor het aanleren van kennis en vaardigheden met betrekking tot de moderne produktietechnieken.

REFERENTIES.

- Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & Cognition*, 9(4), 422-433.
- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(3), 483-495.
- Bayman, P. & Mayer, R.E. (1984). Instructional manipulation of users mental models for electronic calculators. *International Journal of Man-Machine Studies*, 20, 189-199.
- Bhaskar, R. & Simon, H.A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science*, 1, 193-215.
- Boies, S.J. & Gould, J.D. (1974). Syntactic errors in computer programming. *Human Factors*, 16(3), 253-257.
- du Boulay, B., O'Shea, T., & Monk, J. (1981). The black box inside the glass box: Presenting computing concepts to novices. *International Journal of Man-Machine Studies*, 14, 237-249.
- Bullinger, H.J., Fähnrich, K.-P. & Raether, C. (1985). *Task and User Adequate Design of Man-Computer Interfaces in Production*. In: B. Shackel (Ed.) *Human-Computer Interaction INTERACT 84*. Amsterdam: Elsevier Science Publishers B.V.
- Bullinger, H.J., Fähnrich, K.-P. & Sprenger, M. (1984). *User Oriented and Task-Consistent Programming Interfaces for CNC-Machines*. In: T. Lupton (Ed.) *Human Factors in Manufacturing*. Amsterdam: North-Holland.
- Burnes, B. (1984). *Factors Affecting the Introduction and Use of Computer Numerically Controlled Machine Tools*. In: T. Lupton (Ed.) *Human Factors in Manufacturing*. Amsterdam: North-Holland.
- Buschke, H. (1976). Learning is organized by chunking. *Journal of Verbal Learning and Verbal Behavior*, 15, 313-324.
- van Bussel, F.J.J. (1980). Human prediction of time series. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(7), 410-414.
- Card, S.K., Moran, T.P. & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum.
- Carroll, J.M. & Mack, R.L. (1985). Metaphor, computing systems, and active learning. *International Journal of Man-Machine Studies*, 22, 39-57.
- Carroll, J.M. & Thomas, J.C. (1982). Metaphor and the cognitive representation of computing systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 12(2), 107-116.
- Carlier, J. (1986a). *Computerondersteunde werkvoorbereiding voor mechanische werkstukken (1)*. MB Produktietechniek, 52(21), 576-581.
- Carlier, J. (1986b). *Computerondersteunde werkvoorbereiding voor mechanische werkstukken (2)*. MB Produktietechniek, 52(22), 596-603.
- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2(6), 641-653.

- Chase, W.G. & Simon, H.A. (1973). *The mind's eye in chess*. In: W.G. Chase (Ed.), Visual Information Processing. New York: Academic Press.
- Cohen, H.S. & Ferrell, W.R. (1969). Human operator decision making in manual control. *IEEE Transactions on Man-Machine Systems*, 10(2), 41-47.
- Das, B. (1987). An ergonomic approach to designing a manufacturing work system. *International Journal of Industrial Ergonomics*, 1, 231-240.
- Dirkzwager, A., Fokkema, S.D., van der Veer, G.C., & Beishuizen, J.J (Red) (1984). *Leren met Computers in het onderwijs*. 's-Gravenhage: SVO.
- Egan, D.E. & Schwartz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, 7(2), 149-158.
- Elsässer, B. & Lindvall, J. (1984). *Impact of NC Machine Tools on Employment and Skill Levels*. In: T. Lupton (Ed.) Human Factors in Manufacturing. Amsterdam: North-Holland.
- Frey, P.W. & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition*, 4(5), 541-547.
- Furukawa, J.M. (1972). Critical chunk, physical length, and number of words per chunk in programmed instruction. *The Journal of Educational Research*, 65(8), 361-364.
- Garnham, A. (1980). Default values, criteria and constructivism. *Cognitive Science*, 4, 427-433.
- Gentner, D. & Stevens, A.L. (Eds.) (1983). *Mental Models*. Hillsdale, NJ: Erlbaum.
- Goldin, S.E. (1978a). Memory for the ordinary: Typicality effects in chess memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 605-616.
- Goldin, S.E. (1978b). Effects of orienting tasks on recognition of chess positions. *American Journal of Psychology*, 91(4), 659-671.
- Goldin, S.E. (1979). Recognition for chess positions: Some preliminary research. *American Journal of Psychology*, 92(1), 19-31.
- de Groot, A.D. (1966). *Perception and memory versus thought: Some old ideas and recent findings*. In: Kleinmuntz (Ed.), Problem solving: Research, Method, and Theory. New York: John Wiley & Sons.
- Heesters, J (1986). *Onderzoek Naar De Startfuncties Van MTS-ers Werktuigbouwkunde*. Enschede: Sociale Bedrijfskunde, TH Twente.
- van Heusden, A.R. (1980). Human prediction of third-order autoregressive time series. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(1), 38-43.
- Jagacinski, R.J. & Miller, R.A. (1978). Describing the human operators internal model of a dynamic system. *Human Factors*, 20(4), 425-433.
- Johnson-Laird, P.N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 71-115.
- Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.
- Jolicoeur, P. Gluck, M.A. & Kosslyn, S.M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, 16, 243-275.
- Kemp, N.J. & Clegg, C.W. (1987). Information technology and job design: A case study on computerized numerically controlled machine tool working. *Behaviour and Information Technology*, 6(2), 109-124.
- Kieras, D.E. (1987). *What mental model should be taught: Choosing instructional content for complex engineered systems*. Technical report No. 24. Ann Arbor: University of Michigan.
- Kieras, D.E. & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, 8, 255-273.
- Kolers, P.A. & Brison, S.J. (1984). Commentary: On pictures, words, and their mental representation. *Journal of Verbal Learning and Verbal Behavior*, 23, 105-113.
- Koren, Y. (1983). *Computer Control of Manufacturing Systems*. New York: McGraw-Hill Book Company.
- de Korte, D.J. (1986a). *Gebruik van grafische simulatie in Nederland*. *MB Produktietechniek*, 52(7), 206-208.
- de Korte, D.J. (1986b). Principes en ontwikkelingen bij bewerkingscentra. *MB Produktietechniek*, 52(21), 562-574.
- Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Laur-Ernst, U. (1982). *Qualifizierungskonzept für das Arbeiten mit CNC-Maschinen im Rahmen der Erstausbildung für Metallberufe*. Berlin: Bundesinstitut für Berufsbildung.
- Lesgold, A.M. (1984). Human skill in a computerized society: Complex skills and their acquisition. *Behavior Research Methods, Instruments, & Computers*, 16(2), 79-87.
- Mayer, R.E. (1975). Different problem-solving competencies established in learning computer programming with and without meaningful models. *Journal of Educational Psychology*, 67(6), 725-734.
- Mayer, R.E. (1976a). Some conditions of meaningful learning for computer programming: Advance

- organizers and subject control of frame order. *Journal of Educational Psychology*, 68(2), 143-150.
- Mayer, R.E. (1976b). Comprehension as affected by structure of problem representation. *Memory & Cognition*, 4(3), 249-255.
- Mayer, R.E. (1976c). Integration of information during problem solving due to a meaningful context of learning. *Memory & Cognition*, 4(5), 603-608.
- Mayer, R.E. (1979). Can advance organizers influence meaningful learning? *Review of Educational Research*, 49(2), 371-383.
- Mayer, R.E. (1981). The psychology of how novice learn computer programming. *Computing Surveys*, 13(1), 121-141.
- McKeithen, K., Reitman, J.S., Rueter, H. & Hirtle, S.C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Moran, T.P. (1981). An applied psychology of the user. *Computing Surveys*, 13(1), 1-11.
- Mulder, M. & Nijhof, W.J. (1987). Hoe werkt de automatisering door op de MTS en LEAO? *Didaktief*, 17(2), 31-33.
- Nijhof, W.J. & Mulder, M. (Red.) (1986). *Basisvaardigheden in het Beroepsonderwijs*. 's-Gravenhage: SVO.
- Norman, D.A. (1983). *Some Observations on Mental Models*. In: Gentner, D. & Stevens, A.L. (Ed.), *Mental Models*. Hillsdale, New Jersey: Laurence Erlbaum Associates.
- Norman, D.A. (1984). Stages and levels in human-machine interaction. *International Journal of Man-Machine Studies*, 21, 365-375.
- Oliver, M.V.L. (1984). *The Management and Training Implications of Innovation and Change Due to Advanced Manufacturing Technologies*. In: T. Lupton (Ed.) *Human Factors in Manufacturing*. Amsterdam: North-Holland.
- Pavio, A. (1986). *Mental Representations: A dual coding approach*. New York: Oxford University Press.
- Polson, P.G. & Kieras, D.E. (1984). A formal description of users knowledge of how to operate a device and user complexity. *Behavior Research Methods, Instruments, and Computers*, 16(2), 249-255.
- Pritschow, G. & Vieffhaus, R. (1987). *Workshop Programming of Numerical Controls*. In: H.-J. Bullinger and B. Shackel (Eds.) *Human-Computer Interaction - INTERACT 87*. Amsterdam: Elsevier Science Publishers B.V.
- Pylyshyn, Z.W. (1973). What the minds eye tells the minds brain: A critique of mental imagery. *Psychological Bulletin*, 80(1), 1-24.
- Rakers, G.G.H. (1987). *The theory of 'Frame of Mind'*. IST-MEMO-87-01. Enschede: Vakgroep Instructie-technologie, Universiteit Twente.
- Rakers, G.G.H., Pieters, J.M. & de Bruijn, I. Programming computerized numerical controlled manufacturing systems for discrete products: A theory, a processing model, and two experiments. (in voorbereiding)
- Reitman, J.S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, 8, 336-356.
- Reijmerink, C.H.L. & Mulder, M. (1987). *Het Profiel van de MTS-er in een Startfunctie*. Enschede: Onderzoek centrum voor Toegepaste Onderwijskunde.
- Reijmerink, C.H.L., Thijsen, J.A., Mulder, M. & Pieters, J.M. (1987). *Regionale Voorzieningen in de Werktuigbouwkunde: Een structurele oplossing?* Enschede: Onderzoekcentrum voor Toegepaste Onderwijskunde.
- Reijmerink, C.H.L. & Pieters, J.M. (1987). *Apparatuur in het MTO: Ideaal en werkelijkheid*. Enschede: Onderzoekcentrum voor Toegepaste Onderwijskunde.
- Reuterswärd, A. (1984). *Skill Structures and Job Learning in Manufacturing*. In: T. Lupton (Ed.) *Human Factors in Manufacturing*. Amsterdam: North-Holland.
- Rouse, W.B. & Morris, N.M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349-363.
- Ryle, G. (1949/1980). *The Concept of Mind*. Harmondsworth, Middlesex: Penquin Books Ltd.
- van der Sanden, J. (1986). *Het Leren van Technische Vaardigheden: individuele verschillen bij het uitvoeren van praktijk opdrachten in het lager technisch onderwijs*. Den Haag: Stichting voor Onderzoek van het Onderwijs.
- Schoenfeld, A.H. (1979). Explicit heuristic training as a variable in problem-solving performance. *Journal for Research in Mathematics Education*, 10, 173-187.
- Senker, P.J. (1984). *Coping with New Technology: The Need for Training*. In: T. Lupton (Ed.) *Human Factors in Manufacturing*. Amsterdam: North-Holland.
- Sharit, J. (1985). Supervisory control of a flexible manufacturing system. *Human Factors*, 27(1), 47-59.

- Sheppard, S.B., Curtis, B., Milliman, P., & Love, T. (1979). Modern coding practices and programmer performance. *Computer*, 12(12), 41-49.
- Shneiderman, B. (1976). Exploratory experiments in programmer behavior. *International Journal of Computer and Information Sciences*, 5(2), 123-143.
- Shneiderman, B. (1977). Measuring computer program quality and comprehension. *International Journal of Man-Machine Studies*, 9, 465-478.
- Shneiderman, B. (1978). Improving the human factors aspect of database interactions. *ACM Transactions on Database Systems*, 3(4), 417-439.
- Shneiderman, B. (1980). *Software Psychology: Human factors in computer and information systems*. Boston: Little, Brown and Company Inc.
- Sime, M.E., Arblaster, A.T. & Green, T.R.G. (1977a). Reducing programming errors in nested conditionals by prescribing a writing procedure. *International Journal of Man-Machine Studies*, 9, 119-126.
- Sime, M.E., Arblaster, A.T. & Green, T.R.G. (1977b). Structuring the programmers task. *Journal of Occupational Psychology*, 50, 205-216.
- Simon, H.A. (1974). How big is a chunk? *Science*, 183, 482-488.
- Simon, W. (1973). *The Numerical Control of Machine Tools: Basic principles, systems analysis and industrial applications*. London: Edward Arnold (Publishers) Ltd.
- Simons, G.L. (1982). *Computers in Engineering and Manufacture*. Manchester: National Computing Centre Publications.
- Smallwood, R.D. (1967). Internal models and the human instrument monitor. *IEEE Transactions on Human Factors in Electronics*, 8(3), 181-187.
- Snoddgrass, J.G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, 23, 3-22.
- Stassen, H.G. (1976). *Mens-Machine-Systemen: regel-, stuur en bewakingstaken*. In: J.A. Michon, E.G.J. Eijkman & L.F.W. de Klerk (Red.) *Handboek der Psychonomie*. Deventer: Van Loghum Slaterus.
- Thijssen, J.A. & Mulder, M. (1987a). *Trends in de Productie-automatisering en Gevolgen voor het Beroepsonderwijs*. Enschede: Onderzoekcentrum voor Toegepaste Onderwijskunde.
- Thijssen, J.A. & Mulder, M. (1987b). *Aansluiting MTS-Werktuigbouwkunde - Arbeidsmarkt*. Enschede: Onderzoekcentrum voor Toegepaste Onderwijskunde.
- Veldhuyzen, W. & Stassen, H.G. (1977). The internal model concept: An application to modeling human control of large ships. *Human Factors*, 19(4), 367-380.
- Weule, H., Löffler, L. & Selinger, Th. (1987). *Factory Workers and the Language barrier - Making Computers a Tool Rather than a Nuisance*. In: H.-J. Bullinger and B. Shackel (Eds.) *Human-Computer Interaction - INTERACT 87*. Amsterdam: Elsevier Science Publishers B.V.
- Wiedenbeck, S. (1985). Novice/Expert differences in programming skills. *International Journal of Man-Machine Studies*, 23, 383-390.
- Wittgenstein, L. (1921/1984). *Tractatus Logico-Philosophicus*. Frankfurt am Main: Suhrkamp.
- Wittgenstein, L. (1984). *Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp.
- Young, R.M. (1981). The machine inside the machine: User's models of pocket calculators. *International Journal of Man-Machine Studies*, 15, 51-85.
- Young, R.M. (1983). *Surrogates and Mapping: Two kinds of conceptual models for interactive devices*. In: Gentner, D. & Stevens, A.L. (Eds.) *Mental Models*. Hillsdale, NJ: Laurence Erlbaum Associates.

Noten.

1. Nieuwe Technologieën in de Industriële Sector, veranderingen in Bedrijf en Onderwijs Basisstudie 2. 's-Gravenhage: Commissie voor Ontwikkelingsproblematiek van Bedrijven van de Sociaal-Economische Raad (COB/SER) (1986).
2. De Technologische Verandering, de Werkgelegenheid, de Beroepskwalificaties en de Opleiding. Berlin: Europees Centrum voor de Ontwikkeling van de Beroepsopleiding (1985).
3. Nationaal Plan voor Innovatie in de Afdeling Werktuigbouwkunde van het Middelbaar Technisch Onderwijs: Verslag van de werkzaamheden en uitgebrachte adviezen. (1984). Verslag van de werkgroep, onder voorzitterschap van A.M.W. v.d. Linden (MTO), ingesteld naar aanleiding van de brief van de Staatssecretaris van Onderwijs en Wetenschappen en van de Minister van

Organisatorische verschillen tussen basisscholen en hun effect op leerlingprestaties.*

Hennie P. Brandsma en Anja W.M. Knuver

RION, instituut voor onderwijsresearch,

ABSTRACT

In this paper the first results will be discussed of an empirical study on Dutch primary schools in which the research question regarding the influence of school characteristics on individual learning is central. This study makes use of the theoretical insights regarding 'effective schools'. The design is set up in such a way that it can control for most of the criticisms on the effective school research and makes use of recent developments in multi level analysis. A cluster sampling procedure was followed: a random sample of 250 primary schools was drawn. Pupil-level data were gathered through the pupils of grades five and seven of these schools. Information regarding the school level and grade level was collected by the school principal and the teachers of grades five and seven. The analyses show that about 15 - 25% of the between school variance in pupil achievement can be explained by several school (organizational) factors.

1. INLEIDING

Onderzoek naar het verband tussen schoolkenmerken en het functioneren van leerlingen heeft een duidelijke ontwikkeling doorgemaakt.

De eerste onderzoeken kenmerkten zich door een pessimistische visie. Verschillen tussen scholen in (gemiddelde) leerprestaties werden vrijwel volledig wegverklaard door verschillen in de instroomkenmerken (aanleg en sociaal milieu) van de leerlingen. Bekende representanten uit dit tijdperk zijn Coleman e.a. (1966) en Jencks (1972). Deze pessimistische geluiden werden overstemd door o.a. publicaties van Weber (1971), Edmonds e.a. (1978), Brookover e.a. (1979), Rutter e.a. (1979) die op basis van andere inzichten en met behulp van andere methoden tot positievere resultaten kwamen. Veelal werkend met zeer selecte en kleine steekproeven en door middel van meer kwalitatieve technieken van dataverzameling komen zij tot de conclusie dat 'school matters'. In deze periode, het tijdperk van de 'effectieve scholenbeweging', worden een vijftal 'effectief' gebleken schoolkenmerken steeds opnieuw 'ontdekt'.

Kritiek van zowel conceptuele als methodologische aard (vgl. Rowan e.a., 1983; Blom e.a., 1986) zette vraagtekens bij de houdbaarheid en generaliseerbaarheid van de gevonden verbanden. Deze kritiek had voornamelijk te maken met:

- de beperkte omvang en representativiteit van de steekproef van scholen (case-studies);
- geen of weinig rekening houden met de achtergrondgegevens van leerlingen (SES, IQ, ethniciteit);
- de beperkte reikwijdte van de uitkomstvariabelen (alle cognitieve uitkomsten);
- geen rekening houden met het multi-niveau karakter (leerlingen in klassen, klassen in scholen);
- een beperkte definitie van het concept schooleffectiviteit.

De effectieve scholenbeweging heeft echter de aanleiding gegeven tot een verdere uitdieping van de theorievorming over, en empirisch onderzoek naar de effecten van schoolkenmerken op het functioneren van leerlingen. De methodologische en statistische problemen in de onderzoeken van

*De uitvoering van dit onderzoek was mogelijk dankzij een subsidie van de Stichting voor Onderzoek van het Onderwijs (SVOnr. 6009)

van de effectieve scholenbeweging lijken langzamerhand te worden overwonnen. De meest recente ontwikkelingen die hebben plaatsgevonden op het gebied van statistische multi-variate modeltoetsing op meerdere niveau's tegelijk (vgl. Van den Eeden e.a., 1984; De Leeuw & Kreft, 1985; Aitkin & Longford, 1986; Raudenbush & Bryck, 1986, 1987) en het ontwikkelen van programmatuur daarvoor maken het mogelijk om preciesere schattingen te maken van de invloed van schoolkenmerken op het individuele leerlingfunctioneren.

In dit artikel worden de eerste resultaten besproken van een grootschalig onderzoek op Nederlandse basisscholen, waarin de vraag naar de invloed van schoolorganisatiekenmerken op de individuele leerprestaties centraal staat. In dit onderzoek is getracht zoveel mogelijk aan de kritiek op het effectieve scholenonderzoek tegemoet te komen. Met name de multiniveau-analyse krijgt in dit artikel de aandacht. In het eerste deel wordt de probleemstelling theoretisch verder uitgewerkt en wordt ingegaan op het multi-niveau-karakter ervan. In par. 3 wordt in het kort het onderzoeksdesign en het gebruikte instrumentarium besproken. De resultaten komen in par. 4 aan de orde.

2. THEORETISCH KADER EN PROBLEEMSTELLING

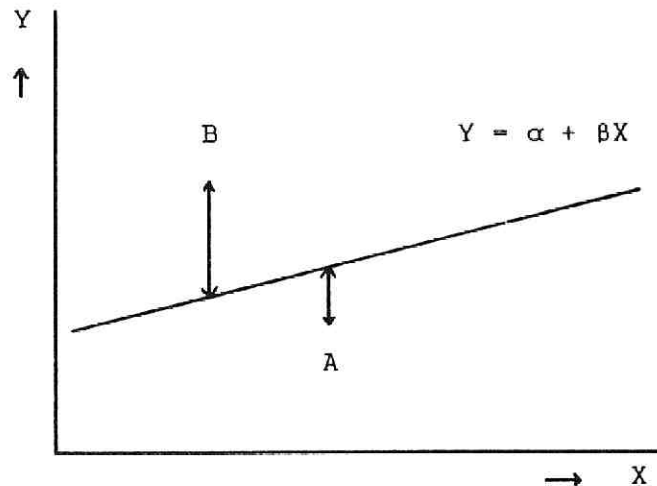
2.1 Schooleffectiviteit

Een van de belangrijkste doelstellingen van het onderwijsbeleid is ongetwijfeld het oplossen van de achterstand in leerprestaties van kinderen uit de lagere sociale milieus. De resultaten van voornamelijk noord-amerikaans effectieve-scholen onderzoek wijzen erop dat er tussen scholen (met vergelijkbare (lage) leerling-instroomkenmerken verschillen kunnen bestaan in leerprestaties en dat schoolkenmerken een verklaring kunnen geven voor dit verschil in effectiviteit.

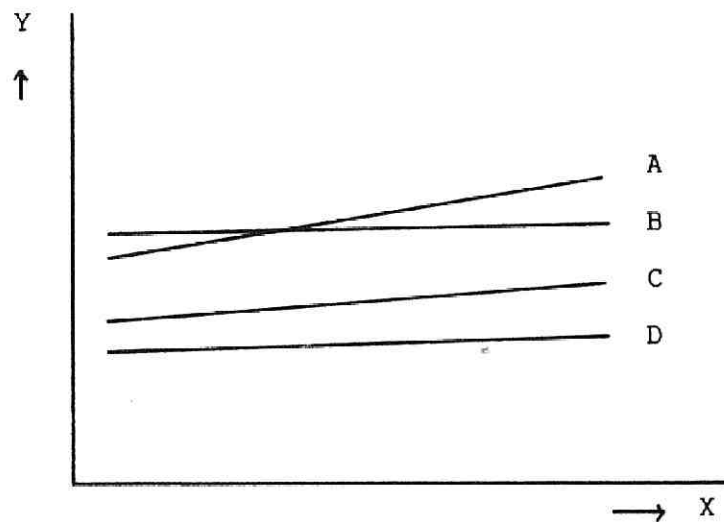
Centraal in de literatuur over effectieve scholen staat het begrip 'effectiviteit'; een relatief begrip. Sommige scholen zijn effectiever dan andere in het scheppen van betere onderwijskansen voor leerlingen uit achterstandssituaties.

De bepaling van de effectiviteit van scholen heeft op verschillende manieren plaatsgevonden. De meest simpele wijze is het afzetten van de gemiddelde score van een school (meestal met betrekking tot schoolprestaties) tegen een standaard (meestal de gemiddelde score over alle scholen). De scholen die boven het gemiddelde scoren zijn effectief, de scholen eronder niet effectief. Dit 'standaard-model' corrigeert echter niet voor verschillen in instroomkenmerken tussen leerlingen (b.v. SES en IQ) en voor het selectiemechanisme (tendens van scholen om leerlingen met dezelfde achtergrond te selecteren). Een model dat wel corrigeert voor (gemiddelde) instroomkenmerken op leerprestaties, staat bekende als het 'residu-model'. In dit model worden de afwijkingen van scholen ten opzichte van de algemene regressielijn (regressie van prestaties op IQ en/of SES) als indicatie gezien voor de effectiviteit van scholen (zie figuur 1).

School B is in dit model effectief en school A ineffectief en de afstanden van A en B tot de regressielijn kunnen gezien worden als een maat voor de effectiviteit van de scholen. Hoewel dit model een verbetering inhoudt ten opzichte van het standaardmodel, wordt geen rekening gehouden met de variantie binnen scholen; het is gebaseerd op verschillen tussen gemiddelden van scholen. Inmiddels is zowel conceptueel als statistisch vooruitgang geboekt op dit gebied. Door Cuttance (1987) wordt het begrip schooleffectiviteit gesplitst in twee dimensies, nl. kwaliteit (quality) en compenserend vermogen (equity). De kwaliteitsdimensie heeft betrekking op de gemiddelde score van een school en wordt grafisch gerepresenteerd door het intercept α . In de dimensie 'compenserend vermogen' is de capaciteit van de school vervat om in meer of mindere mate te kunnen compenseren voor verschillen in instroom. Deze dimensie wordt grafisch gerepresenteerd door de grootte van de hellingshoek β . In deze opzet wordt dus expliciet rekening gehouden met de individuele variantie binnen scholen. Per school wordt een schatting gemaakt van de kwaliteit (het intercept) en het compenserend vermogen (de hellingshoek). In figuur 2 worden de regressielijnen van vier verschillende scholen (A t/m D) weergegeven.



Figuur 1:



Figuur 2:

Alle leerlingen in school A en B presteren beter dan de leerlingen in de scholen C en D, onafhankelijk van SES en IQ. De schatting van de gemiddelde scores geeft inzicht in de kwaliteitsverschillen tussen scholen en kan met een (co)variantie-analyse getoetst worden. Het compenserend vermogen van scholen komt tot uitdrukking in de grootte van de hellingshoek. Een vergelijking tussen deze scholen laat zien dat in de scholen A en C de prestaties van kinderen met laag IQ en SES relatief lager zijn dan in school B en D, en van kinderen met hoog IQ en SES de prestaties in school A en C relatief hoger dan in school B en D. De scholen A en C differentiëren sterk en compenseren weinig naar IQ en SES; ze kunnen getypeerd worden als elitair. Het compenserend vermogen van de scholen B en D is daarentegen relatief hoog en zij kunnen getypeerd worden als egalitaire scholen.

Nu het begrip 'effectiviteit' is gedefinieerd in termen van kwaliteit en compenserend vermogen is het interessant verschillen in effectiviteit tussen scholen te verklaren met behulp van schoolvariabelen. In de volgende paragraaf zullen verschillende school-effect modellen besproken worden die tot nu toe gebruikt zijn in schooleffectiviteitsonderzoek en zal onze modelkeuze worden beargumenteerd.

2.2 Schooleffectiviteitsmodellen

Modellen om school-effecten te toetsen zijn onder te brengen in een drietal sub-groepen.

- modellen met constant intercept en constante helling (standaard regressiemodel);
- modellen met variabele intercepten en constante helling (ANCOVA);
- modellen met variabele intercepten en variabele hellingen (Aptitude-Treatment-Interaction).

De modellen onder a. komen overeen met het effectiviteitsconcept zoals weergegeven in figuur 1 in de vorige paragraaf. Bij de modellen onder b. worden de verschillen tussen scholen wat betreft kwaliteit gemodelleerd. In de modellen onder c. worden zowel kwaliteitsverschillen als verschillen in compenserend vermogen gemodelleerd. Aitkin en Longford (1986) hebben de verschillende modellen vergeleken met betrekking tot hun vooronderstellingen, overeenkomsten en verschillen. De subtypes onder a. en b. zijn homogene binnengroeps regressiemodellen, die geen variantie tussen scholen toestaan wat betreft de relatie tussen y- en x-variabelen (compenserend vermogen). Dat de relatie tussen y- en x-variabelen mag variëren is een essentiële assumptie in het effectieve scholenonderzoek, zoals aangegeven in de vorige paragraaf. Voor de analyse van de data is dan ook gekozen voor een model waarin zowel intercepten als hellingen mogen variëren. De analyse vindt dan ook plaats op twee niveaus tegelijk: de school en de leerling.

De coëfficiënten voor kwaliteit en compenserend vermogen kunnen geschat worden als fixed of random regressie-coëfficiënten. Fixed effects modellen geven een verkeerde schatting van de hoeveelheid variantie die verklaard kan worden door schoolfactoren, zowel in de kwaliteitsdimensie als in het compenserend vermogen. Dit komt doordat alle variantie in de afhankelijke variabele beschouwd wordt als systematische variantie. Een gedeelte van deze variantie is echter steekproefvariantie (Cuttance, 1987; Kreft, 1987).

Recentelijk zijn voorstellen gedaan schooleffecten te verklaren met behulp van random effects modellen (Raudenbush & Bryk, 1986; Longford, 1985). In deze modellen worden de schattingen voor kwaliteit (α) en compenserend vermogen (β) beschouwd als random variabelen met een proportie ware en een proportie steekproefvariantie. Met deze modellen kunnen schooleffecten efficiënter geschat worden (grotere power en nauwkeuriger schattingen) en het wordt mogelijk resultaten te generaliseren naar de totale populatie van scholen.

Uiteindelijk stellen Aitkin en Longford het volgende random coëfficiënt model voor (met random intercepten en random hellingen).

$$y_{ij} = (\alpha + \alpha_j) + (\beta + \beta_j) x_{ij} + \delta_{zj} + e_{ij}$$

waarbij:

y_{ij} : afhankelijke variabele (leerling i in school j)

α : intercept

α_j : random deel van het intercept

β, δ : regressie-coëfficiënten

β_j : random deel van de regressie-coëfficiënt

x_{ij} : leerlingsscore op een covariaat

z_j : score op een schoolvariabele (op schoolniveau)

e_{ij} : individuele errorterm

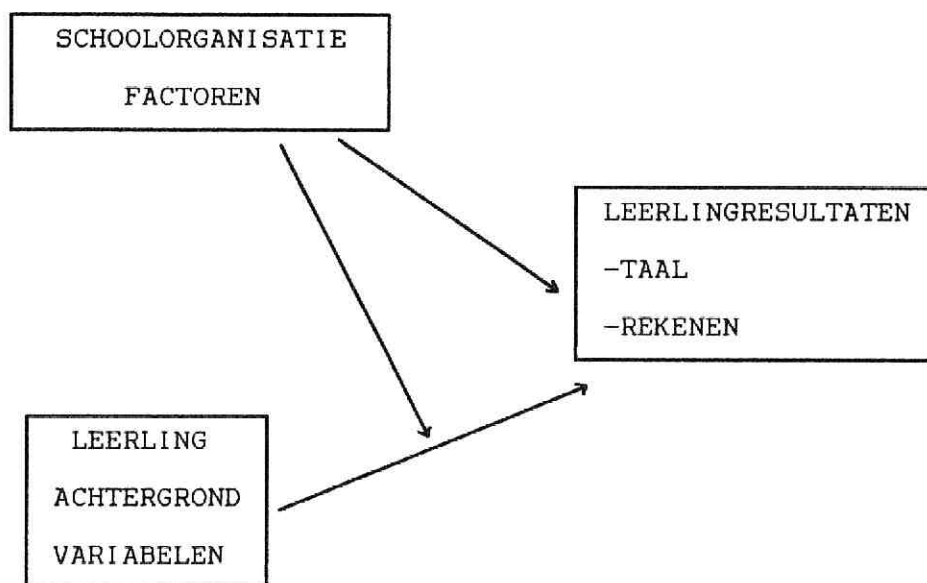
Dit model is gebruikt voor de analyses van het hier beschreven onderzoeksmateriaal.

2.3 Probleemstelling

Op basis van genoemde theoretische inzichten op het gebied van schooleffectiviteit (m.n. de verschillen in kwaliteit en compenserend vermogen) en de ontwikkelingen in het modelleren van schooleffecten, kan de probleemstelling als volgt geformuleerd worden:

Welke organisatiekenmerken van scholen bepalen het verschil in effectiviteit van scholen, zowel in kwaliteit als in compenserend vermogen, wat betreft leerlingprestaties in taal en rekenen.

Het algemene schema van de onderzoeksopzet kan geschetst worden met behulp van het volgende model:



Dit model is het algemene kader voor het verklaren van verschillen in schoolprestaties in taal en rekenen. De leerlingvariabelen SES en IQ worden gezien als belangrijke voorspellers van prestaties, maar ze kunnen niet of nauwelijks beïnvloed worden. Ze worden beschouwd als covariabelen op leerlingniveau. Uit het model valt te concluderen dat er een direct effect verwacht wordt van schoolkenmerken op leerlingprestaties; dit zijn de algemene schooleffecten, de kwaliteitsdimensie. Er zijn ook interacties tussen leerling-achtergrond en schoolfactoren in het model opgenomen. Deze interacties kunnen een verklaring vormen voor eventuele heterogene binnengroepsregressies.

3. ONDERZOEKSOPZET

3.1 Steekproef

Er is een random steekproef getrokken van 250 Nederlandse basisscholen uit een populatie van ongeveer 8000 scholen. Op deze scholen is informatie verzameld van alle leerlingen uit de groepen 5 en 7. Er zijn gegevens verzameld op drie verschillende niveaus: de school, de klas en de leerling. De informatie over de school is gegeven door de directeur, de informatie over de klas door de klasseleerkrachten van de groepen 5 en 7. De gegevens op leerlingniveau zijn verkregen door verschillende toetsen en vragenlijsten af te nemen aan de leerlingen. In dit rapport zal alleen aandacht besteed worden aan enkele variabelen op schoolniveau en de leerlingvariabelen van groep 7.

3.2 Variabelen

Voor het hier besproken deel van het onderzoek zijn drie groepen variabelen meegenomen: schoolfactoren als predictorvariabelen, taal- en rekenscores als afhankelijke variabelen en leerlingachtergrondvariabelen als covariaten.

Voor het meten van taal- en rekenvaardigheid zijn in samenwerking met het CITO twee tests

ontwikkeld. Beide tests hebben een redelijke betrouwbaarheid: KR-20-taal = .81; KR-20-rekenen = .70.

De covariabelen zijn ook gemeten op leerlingniveau. In de hier besproken analyses zijn de covariabelen sociaal economische status (SES), intelligentie (ISI) en sexe meegenomen. Voor de variabele SES is informatie verzameld over het beroeps- en opleidingsniveau van de vader en de moeder. Voor het bepalen van het IQ is gebruik gemaakt van het performale deel van de ISI-test (Van Boxtel e.a., 1982).

Voor de eerste analyses is een selectie gemaakt uit de schoolorganisatiekenmerken die in het onderzoek zijn gemeten. In Blom e.a. (1986) is uitgebreid beschreven wat verwacht wordt van de relaties tussen deze variabelen en leerprestaties van kinderen. Hier volstaan we met een overzicht waarin ook de psychometrische gegevens van de variabelen zijn vermeld. Voor een uitgebreide beschrijving van de instrumentering verwijzen we naar Brandsma en Knuver (1988).

Overzicht opgenomen schoolorganisatie-variabelen:

	verwacht verband met ll.prestaties	aantal items	alpha
1. teamstabiliteit	positief	1	
2. taakdelegatie	positief	10	
3. besluitvorming	positief	8	.71
4. planmatigheid	positief	15	.83
5. intensiteit ll.evaluatie	positief	3	
6. aantal innovaties	?	4	
7. onderwijsk. leiderschap	positief	8	.81

De correlaties tussen de schoolorganisatie-variabelen onderling zijn relatief laag (de hoogst voorkomende correlatie is .26), zodat het probleem van multicollineariteit zich slechts in geringe mate voordoet. Bovendien wordt in het VARCL-programma, waarmee de analyses zijn uitgevoerd, aan het multicollineariteitsprobleem tegemoet gekomen. De coëfficiënten worden simultaan geschat, wat betekent dat de variabelen die overblijven in het model onafhankelijk van elkaar een bepaalde hoeveelheid variantie verklaren. Alle continue variabelen zijn getransformeerd naar z-scores, om de interpretatie van de β -schattingen te vergemakkelijken.

3.3 Analyse

Bij de analyses van deze eerste gegevens is gebruik gemaakt van het VARCL-pakket. Vanwege het niet kunnen hanteren van missing values door dit pakket zijn een aantal scholen uit het databestand verwijderd. De analyses zijn uiteindelijk uitgevoerd op een dataset van 187 scholen en 3657 leerlingen uit groep 7. Omdat er sprake is van een clustersteekproef (scholen zijn random getrokken, daarbinnen klassen en daarbinnen leerlingen, zodat de kans voor leerlingen om in de steekproef terecht te komen niet voor alle leerlingen gelijk is), is gebruik gemaakt van de wegingsprocedure zoals beschikbaar in het VARCL-programma.

4. RESULTATEN

Bij het analyseren van de data is de volgende strategie gehanteerd.

Er zijn afzonderlijke analyses uitgevoerd voor de output-variabelen taal en rekenen. Dit maakt het mogelijk de eventueel verschillende effecten van co-variabelen en schoolfactoren voor taal en rekenen te onderscheiden.

Voor zowel de taal- als de rekenanalyses is eerst nagegaan of wat betreft de co-variabelen op leerlingniveau er verschillen in helling tussen scholen worden gevonden. Hiermee wordt nagegaan

of de hellingscoëfficiënt β als random (in het geval er significante verschillen in hellingen tussen een covariaat en de afhankelijke variabele worden gevonden) of fixed effect (in het geval geen verschillen in hellingen tussen scholen gevonden worden) gemodelleerd dient te worden. In tabel 1 is een overzicht gegeven van de verschillende stappen die hiertoe zijn ondernomen.

Tabel 1: Vergelijking random en fixed effecten voor de covariabelen sexe, SES en ISI.

	Taal	Rekenen
Model 1		
LEEG		
dev.	10048.5	9933
SEXE F(ixed)		
SES F		
ISI F		
dev.	9065.7*	8891.6*
SEXE R(andom)		hellingen: s= .17 se = .04**
SES R		
ISI F		
dev.	9063.1	8882.0*
SEXE R(andom)		hellingen: s= .17 se = .04**
SES R		hellingen: s= .06 se = .03**
ISI F		
dev.	9059.7	8874.4*
SEXE R		hellingen: s=.17 se=.04**
SES R		hellingen: s=.06 se=.03**
ISI R		hellingen: s=.03 se=.05**
dev	9057.9	8872.8

dev. = deviance. Verschillen in deviance zijn chi-kwadraat verdeeld.

* = significant verschil in vergelijking met het vorige model

** = significante waarde, $\alpha = .05$

Bij taal wordt de modelfit niet verbeterd als de covariaten random gemodelleerd worden in plaats van fixed. De verbetering van fit van model tot model wordt uitgedrukt als een verschil in deviance, welke chi-kwadraat verdeeld is, waarbij elke toevoeging van een random covariaat twee extra vrijheidsgraden betekent.

Voor taal lijken de hellingen voor intelligentie, SES en sexe niet te verschillen van school tot school. Dat betekent dat wat taal betreft niet aangenomen mag worden dat scholen verschillen in hun compenserend vermogen. Alleen kwalitatieve verschillen tussen scholen wat taal betreft kunnen verklaard worden met behulp van schoolkenmerken.

Voor rekenen wordt de beste modelfit verkregen als sexe en SES random, en ISI fixed worden gemodelleerd. We kunnen aannemen dat het effect van zowel sexe als SES op de rekenresultaten tussen scholen verschilt.

De volgende stap in de analyse bestaat uit het aanpassen van de modellen voor taal en rekenen door ook schoolfactoren in te bouwen. De volgende strategie is daarbij ondernomen. Het model waarin alleen de covariaten zijn opgenomen (model II) is vergeleken met het model waarin tevens alle schoolfactoren opgenomen zijn (model IV). Deze schoolfactoren zijn: teamstabiliteit, taakdelegatie, besluitvorming, planmatigheid, intensiteit leerlingevaluatie, aantal innovaties en onderwijskundig leiderschap.

In het model voor taal zijn deze schoolorganisatievariabelen opgenomen als fixed effects. Bij het model voor rekenen zijn bovendien de interacties opgenomen van deze schoolfactoren met de random covariabelen SES en sexe. Hierdoor wordt het mogelijk de gevonden hellingsverschillen tussen scholen weg te verklaren met behulp van deze interactietermen.

Bij het vergelijken van het model met alleen covariaten (model II) met het model waarin ook alle schoolkenmerken en eventuele interacties zijn opgenomen (model IV), kan een verschil in modelfit worden waargenomen (een verschil in deviance). Zowel voor taal als voor rekenen is de modelverbetering significant.

In de volgende stap is getracht de modellen zó aan te passen, dat het uiteindelijke model zo veel mogelijk variantie verklaart, met behulp van zo weinig mogelijk schoolfactoren (streven naar parsimonie). Zo ontstaat het meest economische model. Om dit laatste te bereiken is een stapsgewijze procedure uitgevoerd. De schoolvariabelen die het minst belangrijk leken (kleine β -schattingen en grote standaardfouten) zijn één voor één uit de analyse verwijderd. Daarbij is telkens gecontroleerd of het verschil in deviance tussen de modellen niet significant was. Was dit wel het geval, dan is de betreffende variabele in het model opgenomen. In de tabellen 2 en 3 worden de uiteindelijke resultaten voor het taal- respectievelijk het rekenmodel gepresenteerd, samen met het lege model (model I), het model waarin alleen covariaten zijn opgenomen (model II), en het

Tabel 2: variantie-component-analyse taal

	modellen			
	model I leeg	model II covariabelen	model III uiteindelijk	model IV volledig
		schatt. se	schatt. se	
Schattingen				
Sexe (vrouwelijk)		.23 .03	.23 .03	
SES		.20 .02	.20 .02	
ISI		.40 .01	.40 .01	
evaluatie innovaties			— .09 .02 — .09 .02	
Variantie-componenten				
Leerlingen niveau	.860	.665	.665	.664
Schoolniveau	.119	.065	.048	.046
Deviance	10048.5	9065.7	9034.0	9028.6
Afname in deviance in vergelijking met het vorige model		$X^2_d=982.8$ df=3 p<.000	$X^2_d=31.7$ df=2 p<.000	$X^2_d=5.4$ df=5 p>.10

Tabel 3: variantie-component-analyse rekenen

	modellen			
	model I leeg	model II covariabelen	model III uiteindelijk	model IV volledig
		schat. se	schat. se	
Schattingen				
Sexe (vrouwelijk)		.01 .03	.01 .03	
SES		.15 .02	.15 .02	
ISI		.44 .01	.44 .01	
teamstabiliteit			.063 .03	
besluitvorming			— .054 .028	
evaluatie			.037 .027	
innovaties			— .081 .028	
interactie SES- schoolleiderschap			.026 .015	
Variantie-componenten				
Leerlingen niveau	.824	.613	.614	.613
Schoolniveau	.163	.103	.088	.104
hellingen sexe		.029	.029	.027
hellingen SES		.00361	.00091	.001
Deviance	9933.5	8874.4	8851.4	8842.4
Afname in deviance in vergelijking met het vorige model		$X^2_d=59.1$ df=5 p<.000	$X^2_d=23$ df=5 p<.000	$X^2_d=9$ df=16 p>.10

Uit tabel twee kan geconcludeerd worden dat voor taal het uiteindelijke model (model III) het meest economische model is. De fit van het model waarin alle schoolorganisatievariabelen zijn opgenomen is niet beter dan die van het uiteindelijke model. Ongeveer 26% van de totale variantie in taalscores kan verklaard worden door de leerling-covariabelen sexe, SES en intelligentie.

Ongeveer 26% van de overblijvende interceptvariantie in de taalscores kan verklaard worden door de twee schoolorganisatie-variabelen evaluatie en innovaties (dit is 1,7% van de totale variantie), waarbij het aantal innovaties een negatief verband en de intensiteit van leerlingevaluatie een positief verband laat zien met de output-variabele.

Uit tabel 3 kan afgelezen worden dat ook voor rekenen het uiteindelijke model (model III) het meest economische model is. De afname in deviance in vergelijking met het volledige model is niet significant.

Ongeveer 24% van de totale variantie in rekenscores wordt verklaard door de drie leerling-covariabelen. De overblijvende interceptvariantie kan voor ongeveer 15% verklaard worden door schoolorganisatorische factoren (dit is 1,5% van de totale variantie). Deze schoolfactoren zijn teamstabiliteit, besluitvorming, evaluatie van leerlingresultaten, het aantal innovaties en de interactie van SES en schoolleiderschap. De variantie in hellingen voor SES kan bijna volledig

verklaard worden door deze interactie. Het gaat hier echter om een zeer klein percentage variantie; er mogen dus geen conclusies aan deze bevinding verbonden worden. De variantie in hellingen voor de variabele *sexe*, die veel groter is, kan niet verklaard worden met behulp van de schoolfactoren in dit model.

5. DISCUSSIE

Voordat de resultaten van deze studie besproken worden, moeten nog enkele opmerkingen worden gemaakt. Het is van belang te wijzen op het voorlopige karakter van deze analyses. Een selectie van de door ons gemeten schoolkenmerken is opgenomen in de hier besproken analyses. Waarschijnlijk zijn de hier beschouwde modellen dus niet uitputtend gespecificeerd. Dit lijkt ook zeer waarschijnlijk als bedacht wordt dat slechts een gering deel van de interceptvariantie verklaard kan worden door enkele schoolfactoren (voor taal 25% en voor rekenen 15%).

Er zullen nieuwe analyses worden uitgevoerd, waarin naast schoolorganisatorische factoren ook klasse-organisatiefactoren worden meegenomen. In de nabije toekomst (herfst 1988) zullen nieuwe data beschikbaar zijn voor dezelfde leerlingcohorten, zodat tevens analyses kunnen worden uitgevoerd met 'leerwinst' als afhankelijke variabele.

In het voorgaande zijn de resultaten besproken van een studie waarin gezocht is naar organisatorische kenmerken van scholen die verschillen tussen scholen wat betreft leerlingprestaties kunnen verklaren. Bij de analyses is gebruik gemaakt van het VARCL-pakket, waarin multi-niveau analyse mogelijk is. Het analysemodel is zo ontworpen dat zowel verschillen in kwaliteit als verschillen in compenserend vermogen (verschillen in hellingen van de regressielijnen) gemodelleerd zijn. Een opmerkelijk resultaat is, dat er geen verschillen in compenserend vermogen van scholen gevonden zijn wat betreft de taalscores. Deze verschillen zijn wel gevonden bij rekenen, maar daar zijn de verschillen erg klein. Alleen *sexe* geeft een redelijke variantie in hellingen te zien (ongeveer 3% van de totale variantie); er zijn echter geen schoolorganisatorische variabelen gevonden die deze variantie kunnen verklaren. Zowel voor taal als voor rekenen zijn, in de tot nu toe uitgevoerde analyses, slechts verschillen in kwaliteit gedeeltelijk te verklaren door schoolfactoren.

De interceptvariantie voor taal kan voor ongeveer 25% verklaard worden door twee schoolorganisatorische variabelen; de intensiteit van de evaluatie van leerlingresultaten hangt positief, en het aantal innovaties in de school negatief samen met taalvaardigheid. Het eerste resultaat is in overeenstemming met de conclusies uit kwalitatief onderzoek met betrekking tot de kenmerken van effectieve scholen.

Het verzamelen van informatie over het vaardigheidsniveau van leerlingen en de evaluatie daarvan kan een belangrijke signaalfunctie hebben, zowel voor de leerkracht als voor de school. Het stelt leerkrachten in staat te ontdekken aan welke kennis het leerlingen ontbreekt, zodat een geschikte aanpassing van de curriculuminhoud kan plaatsvinden (Edmonds, 1979; Levine & Stark, 1982). Het aantal innovaties, waarmee een school bezig is, laat een negatief verband zien met taalvaardigheid. Deze bevinding kan op verschillende manieren verklaard worden. In de eerste plaats is het mogelijk dat de causaliteit van het verband in de omgekeerde richting wijst; het is niet het aantal innovaties dat leidt tot lagere prestaties, maar lagere prestaties die leiden tot het invoeren van innovaties. Het is tevens mogelijk dat het tegelijkertijd bezig zijn met meerdere innovaties een indicatie is voor onrust in de school, hetgeen een verklaring kan zijn voor de lagere prestaties. Een ordelijk schoolklimaat wordt vaak gezien als voorwaarde voor het bereiken van goede leerlingresultaten (Brookover, 1978).

Voor rekenen kan ongeveer 15% van de totale variantie tussen scholen verklaard worden door vier schoolorganisatie-variabelen. Teamstabiliteit en de intensiteit van evaluatie van leerlingresultaten hangen positief samen met rekenvaardigheid. Het aantal innovaties en de manier waarop belangrijke beslissingen worden genomen (autocratisch versus democratisch) hangen negatief samen met rekenvaardigheid. De verschillen in de hellingen voor *SES* bij rekenen kunnen bijna volledig verklaard worden door het interactie-effect tussen *SES* en onderwijskundig leiderschap.

Deze variantie is echter zo klein dat hieruit geen theoretische conclusies getrokken kunnen worden.

Bij het vergelijken van de resultaten van taal en rekenen valt op dat er een zekere consistentie is. Het aantal innovaties in een school en evaluatie van leerlingresultaten kunnen een deel van de variantie in taal en een deel van de variantie in reken-scores verklaren. De positieve relatie van teamstabiliteit en rekenscores kan ook verklaard worden met behulp van het concept 'ordelijk schoolklimaat'. Samen met het gevonden negatieve verband tussen het aantal innovaties en leerlingresultaten kan geconcludeerd worden dat leerlingen in een stabiele schoolorganisatie, zowel wat betreft personeel als aantal innovaties, betere resultaten zullen behalen op taal- en rekentoetsen.

Voor besluitvorming kan geconcludeerd worden dat autocratisch leiderschap, dus een schoolleider die belangrijke beslissingen vaker zelf neemt dan samen met de stafleden, een positief verband vertoont met rekenresultaten. Dit resultaat is tegengesteld aan de verwachtingen. Dat de variabelen taakdelegatie, onderwijskundig leiderschap en planmatigheid geen verbanden te zien geven met leerlingprestaties is niet volgens de verwachting. De betrouwbaarheid van de meting van deze variabelen lijkt hiervoor niet de oorzaak, omdat deze voor onderwijskundig leiderschap en planmatigheid juist hoog is. Bij planmatigheid is het mogelijk dat geen verband gevonden wordt door het hoge gemiddelde en de relatief kleine standaarddeviatie ($m=3.23$, $sd=0.46$ bij $max=4$ en $min=1$). Bij onderwijskundig leiderschap hebben we te maken met een betrouwbare schaal met een goede spreiding. Een andere mogelijke verklaring voor het ontbreken van gevonden verbanden is dat de bestaande verbanden niet rechtlijnig zijn, en dus in een variantie-componentanalyse niet gevonden worden. Deze mogelijkheid moet nog nader onderzocht worden.

Samenvattend kan geconcludeerd worden dat in de hier besproken analyses de verschillen tussen scholen in kwaliteit veel belangrijker zijn dan de verschillen in compenserend vermogen. Als de verschillen tussen scholen voor taal en rekenen met elkaar vergeleken worden, wordt duidelijk dat de mogelijke effecten van schoolfactoren voor rekenen groter lijken dan voor taal. Deze bevinding is in overeenstemming met gegevens uit andere onderzoeken (b.v. Mortimore e.a., 1988).

LITERATUUR

- Abalos, J., Jolly, J. & Johnson, R. (1985). *Statistical methods for selecting merit-schools*. Chicago: AERA-paper.
- Aitkin, M. & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. In *The Journal of the Royal Statistical Society, Series A (General)* (p. 1-43), 149, Part 1.
- Blom, M.J.T., Brandsma, H.P. & Stoel, W.G.R. (1986). *Effectieve scholen in het basisonderwijs*. Groningen: RION.
- Boxtel, H.W. van, Snijders, J.Th. & Welten, V.J. (1982). *ISI Vorm III. Verantwoording van het prestatiegedeelte & Handleiding voor de gehele testreeks*. Groningen: Wolters-Noordhoff.
- Brandsma, H.P. & Knuver, J.W.M. (1988). *Basisschoolkenmerken en leerlingfunctioneren. Deelrapport II. Instrumentontwikkeling*. Interne publicatie, concept. Groningen: RION.
- Brookover, W., Beady, C., Flood, P., Schweitzer, J. & Wisenbaker, J. (1979). *School social systems and student achievement: schools can make a difference*. New York: Bergin Publishers Book.
- Brookover, W.B., Schweitzer, J.H., Schneider, J.M., Beady, C.J., Flood, P.L. & Wisenbaker, J.M. (1978). Elementary school climate and school achievement. *American Educational Research Journal*, 151, 301-318.
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. & York, R. (1966). *Equality of educational opportunity*. Washington D.C.: U.S. Government Printing Office.
- Cuttance, P. (1987). *Modelling variation in the effectiveness of schooling* (draft).
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15-24.
- Eeden, P. van den & Saris, W.E. (1984). Empirisch onderzoek naar multilevel uitspraken. *Mens en Maatschappij*, 59, 165-178.
- Frechtling, J.A. (1982). *Alternative methods for determining effectiveness: Convergence and divergence*. New York: AERA-paper.
- Jencks, C.S. (1972). *Inequality: a reassessment of the effect of family and schooling in America*. New York: Basic Books.

- Kreft, G.G. (1987). *Models and methods for the measurement of schooleffects*. Dissertatie, Universiteit van Amsterdam.
- Levine, D.U. & Stark, J. (1982). Instructional and organizational arrangements that improve achievement in innercity schools. *Educational Leadership*, 41-46.
- Leeuw, J. de & Kreft, I. (1985). *Random coefficient models for multilevel analysis*. Leiden.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988). *School matters: the junior years*. Somerset, Open Books.
- Purkey, S.C. & Smith, M.S. (1983). Effective schools: a review. *The elementary School Journal*, 83, 427-452.
- Raudenbush, S.W. & Bryk, A.S. (1986). A hierarchical model for studying schooleffects. *Sociology of Education*, 1-17.
- Raudenbush, S.W. & Bryk, A.S. (1987). *Quantitative models for estimating teacher and school effectiveness*. Paper for the invitational Conference of the Center for student testing, evaluation and standards. Princeton (NY).
- Rowan, B., Bossart, S.T. & Dwyer, D.C. (1983). Research on effective schools. A cautionary note. *Educational researcher*, 24-31.
- Rutter, M., Maughan, B. Mortimore, P., Ouston, J. & Smith, A. (1979). *Fifteen thousand hours. Secondary schools and their effects on children*. London: Open Books, Publishers Ltd.
- Weber, G. (1971). *Inner-city children can be taught to read: four successful schools*. Washington D.C.: Council for Basic Education.

Manuscript ontvangen 17-2-1988

Definitieve versie ontvangen 14-6-1988

Betrouwbaarheid van observatietoetsen voor praktische vaardigheden in het medisch onderwijs

C.P.M. van der Vleuten¹ & S.J. van Luyk²

¹ Vakgroep Onderwijsontwikkeling & -research

² Skillslab Rijksuniversiteit Limburg

ABSTRACT

Tests for observations of practical skills are increasingly popular in many educational settings. Despite this popularity their psychometric characteristics are fairly unknown. This study reports on the reliability of an observation test which is used at the medical school of the University of Limburg. Every year all medical students show their ability in practical medical skills by means of a so-called skills-test, consisting of a series of stations in which students have to demonstrate their skills apprehended.

Medical students from all classes in the 1984/1985, 1985/1986, and 1986/1987 academic years were included in the study. Using methods derived from generalizability theory, analyses investigated interrater reliability and reproducibility of scores.

Overall interrater reliability appeared to be sufficient across all classes. On the other hand, variation in the quality of examinee performance from station to station proved to be a large source of measurement error. Decision studies indicated that, depending on the interpretation perspective taken, a minimum of three to five hours of testing time is required to obtain reasonably reproducible scores, regardless of year of training, and despite differences in test content. Some strategies for improvement of the reliability of observational tests are discussed.

INLEIDING

Praktische vaardigheden spelen in vele sectoren van het onderwijs een belangrijke rol, met name en vooral in de meer beroepsgeoriënteerde disciplines. Hoewel er in onderwijskundige zin een groeiende aandacht bestaat voor praktische vaardigheden (b.v. De Klerk, 1980; Pieters, 1984; Houtman & Schinkelshoek, 1986) is over de toetsen van praktische vaardigheden veel minder bekend. De aandacht voor praktische vaardigheden en voor toepassingen daarvan in het onderwijs is van betrekkelijk recente datum (Sanders, 1980a en 1980b).

Ook binnen de medische opleidingen, het onderwerp van deze studie, neemt de belangstelling voor toetsen van praktische vaardigheden sterk toe. Na de introductie van praktijktoetsen in de vorm van Objective Structured Clinical Examinations in Engeland (OSCE's; Harden & Gleeson, 1979) en soortgelijke toetsen als 'Patient Based Instruments' (Stillman e.a., 1976) in de Verenigde Staten, hebben geleidelijk meer medische opleidingen dergelijke toetsvormen geïntroduceerd. De standaard vorm bestaat uit een circuit, dat wil zeggen een reeks van z.g. stations. Bij ieder station - een apart ingerichte ruimte met relevante hulpmiddelen en materiaal - wordt aan de student gevraagd een vaardigheid te tonen. Alle kandidaten starten gelijktijdig, ieder bij een verschillend station en na een vastgestelde periode (variërend tussen vijf en dertig minuten) gaan alle kandidaten naar een volgend station, totdat het gehele circuit is doorlopen. De kandidaten worden direct geobserveerd en beoordeeld door een aanwezige beoordelaar aan de hand van tevoren opgestelde gestandaardiseerde beoordelingsformulieren. De beoordelingsformulieren kunnen verschillende vormen hebben, variërend van globale oordelen door middel

van rating scales, tot gedetailleerde checklists van het gewenste handelen.

Het doel van het toepassen van toetsen voor praktische vaardigheden kan nogal verschillen. Vaak wordt meer dan praktische vaardigheden alleen beoogd en worden ze gebruikt om het klinisch denken en handelen of probleemoplossend vermogen van studenten te meten. Hierbij krijgen zij vaak de vorm van arts-patiënt simulaties. In alle gevallen is echter het direct observeren en zo objectief mogelijk registreren van het gedrag van kandidaten onder standaard instructies in standaard situaties kenmerkend.

Ondanks de toenemende belangstelling voor directe observatie als toetsvorm, is er slechts weinig bekend over de psychometrische eigenschappen ervan. Over toetsen voor praktische vaardigheden zoals toegepast in het lager en middelbaar beroepsonderwijs zijn spaarzame gegevens bekend (Sanders, 1980a en 1980b). Met betrekking tot toepassingen in medische opleidingen zijn er recentelijk enkele studies naar betrouwbaarheid en validiteit verricht.

Psychometrische bevindingen

Petrusa e.a. (1986) onderwierpen 99 studenten geneeskunde aan 17 stations met een totale toetstijd van ongeveer 2 uur en 15 minuten. Bij de afzonderlijke stations werden patiëntenproblemen aan de studenten voorgelegd door gebruikmaking van simulatiepatiënten. De beoordeling van de studenten werd door de simulatiepatiënten verricht aan de hand van betrekkelijk globale beoordelingslijsten. De bereikte betrouwbaarheidscoëfficiënt (α) bedroeg .57. Voor het bereiken van een betrouwbaarheid van .80 - een waarde die wel als minimaal wordt beschouwd - zouden 51 stations nodig zijn met een totale toetsduur van ongeveer zeven uur. De interbeoordelaarsovereenstemming bleek gemiddeld aanvaardbaar (κ .77). De correlaties, gecorrigeerd voor attenuatie, met stagebeoordelingen en een multiple choice test (NBME, nationale Amerikaanse algemeen medische kennistoetsen) waren resp. .73 en .64. Het verband met kennistoetsen wordt als positief opgevat, omdat, naast het feit dat kennis wordt gezien als een noodzakelijke voorwaarde voor het goed kunnen beoefenen van praktische vaardigheden, de te meten concepten niet onafhankelijk van elkaar zijn: betere studenten zullen op de verschillende competenties beter zijn; de slechtere studenten scoren op elk gebied slechter. Afgezien van de benodigde toetslengte zijn dit aanvaardbare onderzoeksresultaten.

Newble & Swanson (ter perse), integreerden en analyseerden het toetsmateriaal van een aantal afnames van verschillende cohorten van in totaal 336 studenten geneeskunde. De schriftelijke stations buiten beschouwing latend, resulteerde een toets bestaande uit 5 stations met een totale gemiddelde toetsduur van een half uur in een generaliseerbaarheidscoëfficiënt van .31. Een aanvaardbare betrouwbaarheid zou pas bereikt worden bij ongeveer 6 uur toetsen. Beoordelingen vonden plaats door facultaire medewerkers en de correlaties met co-beoordelaars varieerden van middelmatig tot hoog (gemiddelde intraclass correlatie .75). Bij correlatie van de toets met een soortgelijke multiple choice toets werd een hoog verband gevonden (.77; gecorrigeerd voor attenuatie).

Williams e.a. (1987), construeerden een zeer lange toets bestaande uit 18 stations van ongeveer 40 minuten per stuk (totale duur 15 uren). Deze bestond voor de ene helft uit een uitvoerend deel (anamnese en lichamelijk onderzoek) en de andere helft uit schriftelijke vragen over het betreffende patiëntenprobleem. De beoordelingen in het uitvoerende deel werden verricht door simulatiepatiënten die voor dit doel waren getraind. De toets werd afgenomen bij 70 studenten in hun examenjaar. De gerapporteerde betrouwbaarheidscoëfficiënt was .75. Er werd echter geen uitsplitsing gegeven naar uitvoerend en schriftelijk deel, noch werden beoordelaarsovereenstemmingen vermeld. Het verband met stagebeoordelingen was .65 en de correlaties met NBME part I en II resp. .53 en .51.

Stillman e.a. (1987) analyseerden een toets van 14 stations met een totale toetsduur van 3,5 uur. De opdrachten hadden betrekking op sociale vaardigheden (zoals interviewstijl en counseling) en beperkt fysisch diagnostisch onderzoek. Beoordelingen vonden plaats door simulatiepatiënten op basis van checklists en rating scales. De auteurs rapporteren een betrouwbaarheid van .78. De beoordelaarsovereenstemming bedroeg .66 voor de rating scales

en .93 voor de checklists (Pearson correlaties). Correlatie met de NBME part I en II valt hier lager uit dan bij de andere studies met resp. .23 en .37; het verband met de stagebeoordelingen bleek .50 te zijn.

Hoewel er ook in Nederland initiatieven worden genomen voor de toetsing van medische vaardigheden (Hiemstra e.a., 1986; Van Rossum, 1985; Metz, 1986) zijn psychometrische gegevens zeer schaars voorhanden (Metz, 1984).

Met betrekking tot de aanwezige gegevens concludeert Swanson (1987), dat het zwakste punt van deze toetsvorm de toetslengte betreft. Iedere station kan opgevat worden als een enkel test-item en diens gevolg is al snel een groot aantal stations noodzakelijk voor het verkrijgen van een voldoende generaliseerbaar resultaat. Dat maakt uiteraard de praktische uitvoerbaarheid van deze toetsvorm extra bezwaarlijk, te meer omdat er een zware logistieke belasting van de onderwijsorganisatie wordt geleverd. Interbeoordelaarsbetrouwbaarheid blijkt in het algemeen een geringer probleem dan inter-item (station) betrouwbaarheid. De validiteit (na correctie voor attenuatie) en de acceptabiliteit leiden er echter toe, dat de hier besproken toetsen over het algemeen gunstig worden beoordeeld.

De Maastrichtse vaardigheidstoets

Voor de onderwijskundige opzet van de Rijksuniversiteit Limburg - het probleemgestuurde leren (Schmidt, 1983) - is de integratie van theorie en praktijk een belangrijke doelstelling. Gepoogd wordt om praktische vaardigheden vanaf het begin van de studie onderdeel te laten zijn van de theoretische kennisvergarig. Daartoe is onder andere een onderwijskundige voorziening gecreëerd in de vorm van een vaardigheidslaboratorium, het zogenaamde Skillslab. Studenten kunnen in het Skillslab met alle mogelijke hulpmiddelen en voorzieningen in een veilige 'laboratorium-omgeving' medisch praktische vaardigheden aanleren en oefenen. De vaardigheden variëren van simpele motorische handelingen zoals het aanleggen van een verband of het meten van de bloeddruk, tot complexe fysisch diagnostische handelingen en interpretaties zoals die bijvoorbeeld bij het onderzoek van de thorax nodig zijn. Ook complexe sociale vaardigheden zoals het voeren van een slecht nieuws gesprek komen aan de orde. Gemiddeld besteedt een student 2 à 3 uur per week in het Skillslab. Voor een uitgebreide uiteenzetting van doelstellingen en uitwerkingen van het vaardigheidsonderwijs wordt verwezen naar Dochy & Van Luyk (1987).

Aan het einde van elk studiejaar wordt elke jaargroep (ongeveer 150 studenten) getoetst op hun vaardigheidsbeheersing door middel van een zogenaamde vaardigheidstoets. Ook in het vijfde en zesde studiejaar, wanneer klinische stages worden doorlopen, worden de studenten op deze wijze getoetst. De toets heeft consequenties voor de besluitvorming over de studievoortgang en vormt een geïntegreerd deel van het totale toetsingssysteem (cf. Verwijnen e.a., 1982).

Een vaardigheidstoets bestaat voor een individuele student uit een twee uren durend circuit van een variërend aantal stations (tussen de 6 en 11). Afhankelijk van de opdracht varieert de stationsduur van 10 tot 30 minuten. De keuze van de te toetsen vaardigheden is gebaseerd op een blauwdruk van 12 zogenaamde vaardigheids categorieën. Afhankelijk van het genoten onderwijs wordt per categorie een vaardigheid geselecteerd en ondergebracht in een station. Bij een meerderheid van de stations wordt gebruik gemaakt van (simulatie-)patiënten, die zo nodig speciaal getraind zijn voor de te spelen rol.

In elk station is tenminste één beoordelaar aanwezig, de z.g. observator. De observator is een deskundige op het gebied van de getoetste vaardigheid, en is tevoren getraind en voorbereid op zijn taak. In ongeveer 20 procent van de gevallen is ook een co-observator aanwezig, die onafhankelijk van de observator dezelfde beoordelingslijsten scoort. Voor het toetsen van één complete jaargroep van 150 studenten zijn twee werkdagen nodig.

Om het doorgeven van informatie over de inhoud van de toets te voorkomen wordt elke twee uur een ander, maar (op basis van de blauwdruk) inhoudelijk parallel circuit ingericht. De observaties worden gescoord aan de hand van criterialijsten. Dit zijn gedetailleerde lijsten waarop de te toetsen vaardigheden zijn ontleed in operationele onderdelen. Een lijst kan variëren van 8 tot 90 items (gemiddeld ongeveer 30). De meeste criterialijst-items zijn

dichotoom (uitgevoerd/niet uitgevoerd), maar bij sommige is een intermediaire codering mogelijk voor onvolledig uitgevoerde handelingen. Belangrijke items kunnen zwaarder worden gewogen. De score op een station wordt gevormd door de som van de itemscores; de score op een totaal circuit is de somscore van alle stationsscores en vormt de uitslag voor een individuele student. Scores worden uitgedrukt in percentages van het totaal mogelijk te behalen punten.

Resultaten op de vaardigheidstoets tellen mee in de beoordeling van de studievoortgang. Uitgangspunt bij de bepaling van de caesuur is een relatieve normering volgens de methode Wijnen (Wijnen, 1971).¹

Enkele validiteitsstudies hebben aangetoond, dat de vaardigheidstoets (vooral in hogere jaargroepen) sterk correleert met kennismetingen (Van der Vleuten e.a., *ter perse*), predictieve waarde heeft ten opzichte van het functioneren in de klinische stages (Van der Vleuten e.a., 1987) en convergeert met oordelen van deskundigen op een aantal verschillende validiteitsaspecten (Van der Vleuten & Van Luyk, 1986). De onderhavige studie zal zich beperken tot de betrouwbaarheid van de vaardigheidstoets.

Vraagstelling en verwachtingen

De stabiliteit van de toets zal worden onderzocht door analyse van een drietal variatiebronnen: verschillen tussen beoordelaars, verschillen in moeilijkheidsgraad tussen vaardigheden (stations) en verschillen in moeilijkheidsgraad tussen parallelle circuits binnen een toets. Een vierde variatiebron, verschillen ontstaan door gebruik van verschillende (simulatie)patiënten met eenzelfde rol, zal hier niet nader worden onderzocht omdat de benodigde gegevens (vooralsnog) ontbreken. Uit recentelijk onderzoek blijkt, dat weliswaar op stationsniveau statistisch significante verschillen tussen simulatiepatiënten worden gevonden (Hiemstra e.a., 1987; Dawson-Saunders e.a., 1987), maar dat de variatie over de gehele test heen aanvaardbaar klein is (Swanson & Norcini, in voorbereiding). Tot slot betreft een mogelijke vijfde variatiebron de stabiliteit van de afzonderlijke stations. Het ligt voor de hand dat stations met korte beoordelingslijsten minder stabiele resultaten opleveren dan beoordelingen op grond van langere lijsten. Dit kan weer repercussies hebben voor de totale toetsscore. Psychometrisch is dit verband echter nog niet eenduidig te beschrijven (Van der Vleuten, 1987) en nader onderzoek dient eerst nog plaats te vinden.

Gegeven de bevindingen uit de studies die in de inleiding werden genoemd is te verwachten, dat ook hier het aantal te toetsen vaardigheden (stations) voor het verkrijgen van een betrouwbaar resultaat een probleem zal zijn. Voortvloeiend uit de te verwachten verschillen in moeilijkheidsgraad van stations, zullen moeilijkheidsgraadverschillen tussen parallelle circuits wellicht ook optreden.

De beoordelaarsovereenstemming zou voor de vaardigheidstoets gunstig moeten uitvallen, te meer daar gewerkt wordt met getrainde en deskundige beoordelaars, werkend met gedetailleerde operationele criterialijsten.

METHODE

Subjecten

In totaal waren 18 toetsen (3 academische jaren x 6 toetsen) voor de analyse beschikbaar. Binnen iedere toets kan er sprake zijn van een verschillend aantal circuits, terwijl ieder circuit een verschillend aantal stations en een verschillend aantal studenten kan bevatten. Tabel 1 bevat hiervan een overzicht. De toetsen in het eerste jaar bestonden uit twee delen die op afzonderlijke tijdstippen werden afgenomen en waaraan alle eerstejaars deelnamen, maar waarover een gecombineerde uitslag werd gegeven. In 1986/1987 werd de zesdejaars-toets eveneens twee maal afgenomen. In dit geval echter namen studenten deel aan hetzij de ene hetzij de andere afname, afhankelijk van hun vorderingen in de stages. Recentelijk zijn aan de regel absolute grenzen toegevoegd, waarboven of beneden altijd een (on)voldoende wordt behaald.

1. Recentelijk zijn aan de regel absolute grenzen toegevoegd, waarboven of beneden altijd een (on)voldoende wordt behaald.

Tabel 1: Beschrijving van vaardigheidstoetsen over de periode 1984-1987.

Ac. jr.	Jaar	Aantal circuits	Aantal stations	Totale toetsduur in min.	Aantal studenten
84/85	1	5	3	30	143
	1	3	4	60	141
	2	4	9	120	141
	3	3	10	120	126
	4	9	6/7	120	116
	5	4	7	120	82
	6	4	8/9	120	51
85/86	1	4	3	30	149
	1	6	4	60	151
	2	3	9	120	121
	3	4	10	120	140
	4	4	6/7	120	128
	5	3	7	120	110
	6	4	8/9	120	83
86/87	1	6	3	30	150
	1	4	4	90	153
	2	4	9	120	143
	3	3	10	120	125
	4	3	6/7	120	135
	5	3	8	120	117
	6	2	8/9	120	47
	6	2	8	120	55

Sommige toetsen bestaan uit nogal veel circuits met weinig studenten (minder dan 20). Soms bleek dit grote aantal circuits noodzakelijk in verband met logistieke en organisatorische problemen (bijvoorbeeld uitval van apparaten of fantomen).

Analyseprocedure beoordelaarsovereenkomst

In totaal waren 3402 paren observator/co-observator scores beschikbaar voor analyse. Deze werden op twee manieren geanalyseerd.

Op stationsscore-niveau werden de gepaarde scores van observator/co-observator geanalyseerd met een 'random effects beoordelaars genest binnen stations' variantie-analyse (ANOVA), geen rekening houdend met student-identiteit: het i:p design in termen van de generaliseerbaarheidstheorie (Cronbach e.a., 1982; Brennan, 1983). Variantiecomponenten werden vervolgens geschat op basis van de ANOVA mean squares volgens standaard procedures (Brennan, 1983) en een generaliseerbaarheidscoefficiënt werd bepaald (voor één beoordeling), waarbij de gemiddelde verschillen tussen observatoren in de error term werden betrokken.

Op het itemniveau van de criterialijsten binnen stations werd een percentage overeenkomst tussen observator en coöbservator scores berekend. Discrepancies tussen itemovereenkomst en stationsscore-overeenkomst kunnen kwalitatieve aanduidingen geven over de aard van de beoordeling.

Analyseprocedure reproduceerbaarheid van scores en beslissingen

De beoordelingen per getoetste vaardigheid, d.w.z. de stationsscores, worden opgevat als items en vormen de basis voor de betrouwbaarheidsschatting van de totale toets. Voor ieder circuit binnen een jaargroep werd een 'random effects personen x stations' ANOVA uitgevoerd (p x i design) en werden variantiecomponenten geschat voor personen, stations en error (voor de gesplitse eerstejaarstoetsen werden de variantiecomponenten van beide delen gesommeerd). De variantiecomponenten werden vervolgens gemiddeld per en over jaargroepen (gewogen naar

steekproefgrootte) om een schatting van de totale effecten te verkrijgen.

De berekening van de betrouwbaarheidscoëfficiënten zal op verschillende wijze worden uitgevoerd, ter illustratie van verschillende mogelijke interpretaties. Uitgaande van een norm-georiënteerd standpunt zullen generaliseerbaarheidscoëfficiënten worden berekend, waarbij geen rekening wordt gehouden met moeilijkheidsgraadverschillen tussen stations. Daarnaast zal een berekening plaatsvinden volgens een domein-georiënteerd perspectief, enerzijds door verschillen in moeilijkheidsgraad in de error term op te nemen (dependability coefficient) en anderzijds rekening houdend met verschillende zak/slaag grenzen (adjusted phi coefficient; Brennan & Kane, 1977). Feitelijke berekeningen vonden plaats met het programma GENOVA (Crick & Brennan, 1983).

RESULTATEN

Beoordelaarsovereenstemming

Tabel 2 bevat beoordelaarsovereenstemmingen uitgesplitst naar jaargroep.

Tabel 2: Beoordelaarsovereenstemmingen naar jaargroep en totaal.

Jaar	Aantal paren	Intraclass correlatie	Percentage overeenkomst
1	581	.85	87
2	682	.86	84
3	699	.84	84
4	527	.86	85
5	443	.75	84
6	470	.82	87
1-6	3402	.84	85

De intraclass correlaties zijn redelijk hoog en vertonen weinig verschillen tussen jaargroepen. De waarde in jaar 5 vormt hierop een uitzondering. Wellicht kan dit worden toegeschreven aan de geringere ware score variantie in jaar 5 (zie variantiecomponenten tabel 4). Het percentage overeenstemming op itemniveau van de criterialijsten van jaar vijf is overeenkomstig de overige jaren. Voor het overige laten de overeenstemmingspercentages eenzelfde beeld zien. Hoewel zij niet corrigeren voor toeval en gevoelig zijn voor de frekwenties in de randtotalen, blijken de percentages een betrekkelijk aanvaardbaar en stabiel resultaat over de verschillende jaren heen op te leveren. Concluderend kan worden gesteld, dat de totale error die kan worden toegeschreven aan beoordelaarsfouten laag is. Op het niveau van de totale toets zullen de verschillen een nog kleinere rol spelen, omdat iedere student door een meervoud van observatoren wordt beoordeeld en aangenomen mag worden dat beoordelaarsverschillen zullen uitmiddelen.

Een tweede analyse van beoordelaarsovereenstemming betreft uitsplitsing naar de verschillende inhoudsgebieden. Het ligt voor de hand dat sommige vaardigheden moeilijker te beoordelen zijn dan andere. Tabel 3 bevat een uitsplitsing naar (blauwdruk)categorieën van de vaardigheidstoets over alle jaren heen. Voor de categorie sociale vaardigheden blijken beide overeenkomstmaten laag. Gegeven de subjectiviteit in de beoordeling van sociale vaardigheden is dit geen opmerkelijk resultaat. De criterialijsten die gebruikt worden bevatten allerlei proces- en uitkomstenmaten, die interpretaties van de beoordelaar vergen. Ondanks de zorgvuldige samenstelling van de beoordelingslijsten voor sociale vaardigheden (cf. Kraan & Crijnen, 1987) en de specifieke voorbereidende trainingen aan de hand van een uitgebreide handleiding, blijft de beoordeling een meer subjectieve aangelegenheid, leidend tot verschillen tussen beoorde-

Tabel 3: Beoordelaarsovereenstemmingen uitgesplitst naar vaardigheidscategorieën over alle jaargroepen.

Categorie	Aantal paren	Intraclass correlatie	Percentage overeenkomst
Neurologische vaardigheden	122	.90	88
Therapeutische vaardigheden	476	.84	89
Abdomen	219	.82	89
Thorax/perifere circulatie	306	.89	85
Oogheelkundige vaardigheden	168	.73	87
Gynaecol./Obst. vaardigheden	216	.87	82
Bewegingsapparaat	245	.86	85
KNO vaardigheden	139	.77	87
Pediatrische vaardigheden	41	.82	87
Sociale vaardigheden	511	.66	75
Laboratorium vaardigheden	799	.84	87
Receptuur vaardigheden	45	.71	94
Diversen ¹	115	.84	83

¹ Experimentele stations met combinaties tussen verschillende categorieën

laars. De discrepantie voor 'receptuur' tussen de relatief lage intraclass correlatiecoëfficiënt en het hoge overeenkomstpercentage kan worden verklaard uit de eenvoud van deze vaardigheden en het geringe aantal items op de criterialijst. Als gevolg daarvan wordt algemeen zeer hoog gescoord resulterend in een geringe ware score variantie, terwijl anderzijds de overeenkomst tussen beoordelaars op itemniveau hoog is.

Variantiecomponenten

Tabel 4 bevat de geschatte variantiecomponenten uitgesplitst per jaargroep en over de totale dataset.

Tabel 4: Geschatte variantiecomponenten per jaargroep en over alle jaargroepen.

Jaar	Geschatte variantiecomponenten			Percentage van totale variantie		
	Personen	Stations	Error	Personen	Stations	Error
1	67.1406	208.2347	284.5977	11.99	37.19	50.82
2	68.2282	81.7080	212.3640	18.83	22.55	58.62
3	47.0295	159.2826	190.5516	11.85	40.14	48.01
4	51.7329	72.5922	140.8575	19.51	27.37	53.12
5	18.5279	89.0927	150.5470	7.18	34.515	8.31
6	68.1309	92.0852	151.9589	21.82	29.50	48.68
1-6	54.1622	122.0340	195.0798	14.59	32.87	52.54

Zoals in het algemeen bij toetsen gebruikelijk is bestaat ook hier de grootste variantiecomponent uit de error. Gerelateerd aan de variantiecomponent van personen is de variantiecomponent voor stations eveneens aanzienlijk. Blijkbaar zijn er aanmerkelijke verschillen in de moeilijkheidsgraad van de verschillende stations.

Bij vergelijking van de variantiecomponenten over de verschillende jaren valt de lage persoonscomponent in jaar 5 op. Het vijfde studiejaar is het eerste stagejaar en de studenten doorlopen allerlei verschillende klinische programma's. In de keuze van de chronologie van de stages bestaat enige vrijheid. Hierdoor zouden verschillen in achtergrond kunnen bestaan die een verklaring kunnen geven voor de lage persoonscomponent. In geringe mate blijkt dat ook

uit de toename van de error component, waarin de persoon x station interactie term besloten zit. Een tweede verklaring ligt mogelijk in de verschillen in zwaarte van de verschillende toetsen. Beslissingen over de studievoortgang vinden plaats volgens de twee-fasenstructuur in jaren 1, 4 en 6. In de besluitvorming hebben vaardigheidstoetsen in de overige jaren een minder zwaar gewicht. De gemiddeld behaalde scores in de tussenliggende jaren liggen dan ook doorgaans iets lager. Voor jaar vijf geldt daarenboven dat er nauwelijks ruimte is voor voorbereiding op de toets in verband met drukte van de stage-routine.

Een mogelijke invloed van het feit, dat er beslissingen worden genomen over de studievoortgang is ook terug te vinden in de jaren 4 en 6 waar de persoonscomponenten relatief hoog zijn, maar minder in het eerste jaar. Deze lage persoonscomponent lijkt ook niet verklaard te kunnen worden door de betrekkelijke eenvoud van de vaardigheden in het eerste jaar. Gecombineerd met een goede voorbereiding zou eenvoud van vaardigheden kunnen resulteren in algemeen hoge scores, leidend tot een lage persoonscomponent, maar ook in een lage stationscomponent. In feite is deze laatste echter een van de grootste van alle jaargroepen. Deze bevinding zou erop kunnen wijzen, dat er in het eerste jaar één of enkele stations zitten, die als uitbijter functioneren. De alternatieve verklaring kan natuurlijk ook gelegen zijn in de kwaliteit van de gebruikte beoordelingslijsten in het eerste jaar. Nadere (inhoudelijke) analyse zal dit moeten uitwijzen.

Samenvattend kan gesteld worden dat de variantiebronnen behorend bij stations en error relatief groot zijn, waardoor noodzakelijkerwijs het aantal te toetsen vaardigheden eveneens groot zal moeten zijn voor het verkrijgen van een reproduceerbaar resultaat. Het betekent ook dat parallelle toetsvormen, zoals die bijvoorbeeld noodzakelijk zijn in de Maastrichtse situatie, zullen variëren in moeilijkheid, waardoor bij een absolute interpretatie sommige studenten benadeeld kunnen worden. De reproduceerbaarheid van scores zal worden geanalyseerd in de navolgende paragraaf.

Reproduceerbaarheid van scores

De schattingen van de variantiecomponenten uit tabel 4 vormden de basis voor de schatting van de tussen stations betrouwbaarheid en standaardmeetfout, geprojecteerd naar verschillende testlengtes en score-interpretaties. De berekeningen zijn alleen gebaseerd op de gecombineerde gewogen gemiddelde resultaten over jaar 1 tot en met 6. Tabel 5 bevat een overzicht.

Tabel 5: Reproduceerbaarheid van scores bij verschillende test-lengtes voor norm-georiënteerde interpretatie (generaliseerbaarheidscoëfficiënten: G) en domein-georiënteerde interpretatie (dependency coëfficiënten: D) en bijbehorende standaardmeetfout (SEM).

Toetstijd in uren	Aantal stations	Norm-georiënteerde interpretatie		Domein-georiënteerde interpretatie	
		G	SEM	D	SEM
1	4	.53	6.98	.41	8.90
2	8	.69	4.94	.58	6.30
3	12	.77	4.03	.67	5.14
4	16	.81	3.49	.73	4.45
5	20	.85	3.12	.77	3.98
10	40	.92	2.21	.87	2.82

Uitgaande van een toetsbenadering waarin alleen de relatieve posities van personen van belang zijn en niet het absolute scoreniveau, is de generaliseerbaarheidscoëfficiënt de aangewezen index. Deze is (hier) vergelijkbaar met Cronbach's alpha. Het is duidelijk dat nogal wat tijd wordt vereist voor het verkrijgen van een adequate betrouwbaarheid. Pas na ongeveer drie uur wordt een minimaal aanvaardbaar niveau bereikt.

In het domein-georiënteerde perspectief wordt aan de scores een absolute betekenis gehecht en dienen moeilijkheidsgraadverschillen als error te worden opgevat. Uit tabel 5 blijkt dat in dat geval meer dan vijf uur toetstijd noodzakelijk wordt.

De standaardmeetfouten zijn weergegeven in de oorspronkelijke score-schaal (percentages). Bij een reproduceerbaarheidscoëfficiënt van ongeveer .80 is de standaardmeetfout nog altijd van respectabele grootte (ongeveer 4 procent).

Een alternatieve domein-georiënteerde benadering betreft de verschuiving van de aandacht van de reproduceerbaarheid van scores naar de reproduceerbaarheid van zak/slaag beslissingen. Bij examens is men geïnteresseerd in het scheiden van voldoende en onvoldoende prestaties, vaak meer dan in de mate van niveaubeheersing. Dat kan, afhankelijk van de caesuur en het (gemiddeld) behaalde examenresultaat, een heel ander beeld opleveren van de reproduceerbaarheid. De gemiddelde score behaald over alle vaardigheidstoetsen was 77.09 procent. Tabel 6 bevat betrouwbaarheden berekend voor daarbijbehorende verschillende zak/slaag grenzen.

Tabel 6: Reproduceerbaarheid (adjusted phi coëfficiënten) bij verschillende zak/slaag percentages.

Toetstijd in uren	Aantal stations	Zak/slaag percentage ¹			
		60%	70%	80%	90%
1	4	.80	.48	.29	.71
2	8	.89	.69	.54	.84
3	12	.93	.78	.66	.89
4	16	.94	.83	.74	.91
5	20	.96	.86	.78	.93
10	40	.98	.93	.88	.96

¹ Behaalde gemiddelde score 77.09%

Naarmate de caesuur verder verwijderd ligt van de gemiddeld behaalde score verandert de betrouwbaarheid van de beslissing aanzienlijk. Afhankelijk van de toetssituatie (bijvoorbeeld de mate van voorbereiding) en het vereiste niveau kunnen daarmee nogal grote verschillen in de benodigde toetstijd ontstaan. In dit geval levert een wat lager vereist niveau van 60 procent slechts één uur benodigde toetstijd op, hetgeen aanmerkelijk korter is dan bij de hierboven gevonden betrouwbaarheidscoëfficiënten. Voor een norm dichter bij het behaalde gemiddelde zijn aanzienlijk langere toetstijden nodig. Overigens moet er rekening mee worden gehouden, dat de zak/slaag grens en gemiddeld behaalde score niet onafhankelijk zijn van elkaar. Het is denkbaar dat de gemiddeld behaalde score zich aanpast aan de caesuur.

Circuitverschillen

De circuits binnen één vaardigheidstoets van één jaargroep kunnen worden opgevat als paralleltoetsen. Om benadeling van studenten te voorkomen zouden verschillende circuits van gelijke moeilijkheid behoren te zijn. Door verschillen in moeilijkheidsgraad in aanmerking te nemen bij de berekening van de betrouwbaarheid, wordt reeds een schatting gegeven van de betrouwbaarheid volgens een norm-gerichte interpretatie voor een situatie waarin niet iedereen dezelfde items krijgt aangeboden. De D-coëfficiënten uit tabel 5 zijn aldus berekend en vormen in dat geval de juiste norm-georiënteerde interpretatie voor de betrouwbaarheid. De studenten

krijgen in de vaardigheidstoets echter geen willekeurige set van stations voorgelegd, maar zij worden groepsgewijs aan dezelfde stations onderworpen. Directe vergelijking van de grootte van het effect van circuitverschillen kan worden bekeken door variaties in circuitgemiddelden te vergelijken.

Studenten worden random toegewezen aan de verschillende circuits. Toetsing van de verschillen vond plaats aan de hand van een one-way variantie-analyse over de scores van studenten, met circuitnummer als groepsvariabele. Tabel 7 bevat de resultaten.

Tabel 7: Variantie-analyses over circuitgemiddeldes binnen een toets.

Ac. jaar	jr	Gemiddeld behaalde procentuele score op circuit									F	P
		1	2	3	4	5	6	7	8	9		
84/85	1	80	80	80	82	81					.119	.98
	1	69	79	80							16.485	.00*
	2	72	76	74	73						1.060	.37
	3	64	66	68							2.344	.10
	4	76	66	71	73	72	71	72	78	73	2.296	.03*
	5	76	76	79	79						1.034	.38
	6	72	75	70	71						1.487	.23
85/86	1	80	88	84	85						3.918	.01*
	1	80	84	83	80	81	78				.585	.71
	2	71	71	77							5.291	.01*
	3	74	74	76	74						.694	.56
	4	79	83	81	81						1.453	.23
	5	71	70	73							1.811	.17
	6	82	86	82	81						1.229	.30
86/87	1	79	79	77	84	75	77				1.101	.41
	1	80	84	78	81						3.792	.01*
	2	76	78	78	69						2.072	.11
	3	73	76	74							2.767	.07
	4	83	82	83							.258	.77
	5	77	76	77							.685	.51
	6	80	79								.192	.66
	6	83	85								.411	.52

* $p < .05$

Van de vijf statistisch significante verschillen worden er drie in het eerste jaar geconstateerd. Hier is echter sprake van korte deelttoetsen met een gering aantal stations (3 of 4). Het verschil in jaar 4 van 1984/1985 is evenmin erg representatief, aangezien hier sprake was van een (noodgedwongen) opdeling in veel circuits met weinig personen. Een en ander wijst erop dat de circuits niet al te sterk van elkaar verschillen in moeilijkheidsgraad. In de praktijk hebben deze gegevens nooit aanleiding gegeven tot correctie van individuele scores van studenten.

DISCUSSIE

De resultaten uit deze studie komen sterk overeen met de bevindingen uit de literatuur. De verschillen tussen beoordelaars zijn niet erg groot en de resulterende betrouwbaarheden zijn aanvaardbaar. Bovendien blijkt dit betrekkelijk invariant over de verschillende jaargroepen. Wel kwam naar voren dat de verschillen tussen beoordelaars varieerden met de inhoud van de

vaardigheden. Met name bij de beoordeling van sociale vaardigheden bleek de overeenkomst relatief laag. Geconcludeerd kan echter worden dat beoordelaarsverschillen geen ernstige bedreiging vormen voor de toetsing van praktische vaardigheden, te meer niet wanneer elke kandidaat door verschillende beoordelaars wordt gezien, waardoor beoordelaarskenmerken elkaar zullen neutraliseren door uitmiddeling.

Veel groter en belemmerender is de variatiebron tussen de verschillende vaardigheden. Conform de bevindingen uit de literatuur bleken de stations nogal in moeilijkheidsgraad te verschillen. De bijbehorende variantiecomponent bleek relatief groot in vergelijking tot de variatie tussen personen. De algemene error variantie bleek (zoals gewoonlijk) de grootste foutenbron. In verhouding tot de geobserveerde variantie is daarmee de ware score variantie betrekkelijk laag. Om tot een adequate totale betrouwbaarheid te komen dient als gevolg daarvan het aantal te toetsen vaardigheden groot te zijn waardoor lange toetstijden nodig worden.

Dit werd duidelijk door de uitkomsten van de D-studies waarin de betrouwbaarheid werd geprojecteerd naar verschillende toetslengtes. Een adequate betrouwbaarheid bleek pas verkregen te kunnen worden na tenminste enkele uren toetstijd, afhankelijk van de interpretatie en het perspectief voor het gebruik van de toets. Bij een norm-georiënteerde benadering is men slechts geïnteresseerd in de relatieve posities van de kandidaten en worden hogere betrouwbaarheden geboekt dan bij de domein-georiënteerde benadering. Bij deze laatste is men geïnteresseerd in het absoluut behaalde niveau en worden verschillen in moeilijkheidsgraad tussen de verschillende vaardigheden als meetfout beschouwd, resulterend in lagere betrouwbaarheids-schattingen. Voor een norm-georiënteerde interpretatie was hier tenminste drie uren toetstijd noodzakelijk, terwijl voor de domein-georiënteerde interpretatie minimaal vijf uren nodig bleek. De standaardmeetfout lag in dit geval rond de 4 procent van de oorspronkelijke schaalscores.

De in de literatuur vermelde betrouwbaarheidscoëfficiënten zijn allen berekend volgens een norm-georiënteerde benadering. De betrouwbaarheidscoëfficiënten uit het hier gerapporteerde onderzoek blijken iets gunstiger te zijn dan in de literatuur vermelde resultaten. Meerdere verklaringen zijn hiervoor mogelijk. De meest waarschijnlijke verklaring is wellicht de invloed van de inhoud van de toets. Zoals aangegeven beogen de meeste toetsen uit de vermelde studies een bredere competentie dan de vaardigheidstoets in het kader van de Maastrichtse situatie. Deze beperkt zich meer tot medisch praktische vaardigheden, terwijl men in de andere studies vaker (hogere) cognitieve vaardigheden tracht te meten, zoals probleemoplossen en kennis. Beperking tot toetsing van praktische vaardigheden betekent dat een kleiner domein wordt beoogd, waardoor het generalisatiedomein makkelijker wordt gedekt bij steekproeftrekking en samenstelling van de toets. Een tweede verklaring ligt mogelijk in de inmiddels opgebouwde routine van vaardigheidstoetsing. Kenmerkend voor de Maastrichtse situatie is de grootschalige werkwijze, welke in geen van de andere studies voorkomt. De samenstelling, organisatie en afname is routine geworden, zowel docenten als studenten weten wat van hen verwacht wordt en wellicht dat deze exogene kenmerken de psychometrische kwaliteit van de vaardigheidstoets ten goede komen.

Ter voorkoming van lange toetstijden is recentelijk gesuggereerd om de aandacht te verschuiven naar de kwaliteit van de genomen beslissing (Swanson & Norcini, in voorbereiding) in plaats van naar de kwaliteit van de scores. In praktijksituaties waar de toetsen gehanteerd worden in de besluitvorming is dit een zinvolle strategie. In deze studie zijn betrouwbaarheidscoëfficiënten berekend voor enkele verschillende zak/slaag grenzen en werden aanmerkelijke verschillen in betrouwbaarheid en benodigde toetslengte gevonden. Naarmate de caesuur verder afligt van het algemene gemiddelde neemt de betrouwbaarheid van de beslissing sterk toe er kan worden volstaan met aanmerkelijk kortere toetstijden. Hierbij valt echter aan te tekenen dat de zak/slaag grens en de toetsprestatie niet onafhankelijk zijn van elkaar. In de praktijk blijken prestaties van studenten zich vaak aan te passen aan het gewenste niveau voor een voldoende beoordeling.

Samenvattend kan echter geconcludeerd worden dat de toetslengte van observatietoetsen

voor praktische vaardigheden een zwakke schakel vormt voor de betrouwbaarheid. Op zich is dat een wat ontmoedigend gegeven, zeker als de logistieke belasting van dit soort toetsen in aanmerking wordt genomen. Voor toetsconstructeurs betekent dit dat zuinigheid betracht moet worden bij de samenstelling. Er zal extra nauwkeurig op gelet moeten worden dat geen vaardigheden worden getoetst die ook, maar vooral ook makkelijker, met andere meetinstrumenten gemeten kunnen worden.

Uiteraard bestaat de rechtvaardiging van observatietoetsen niet uitsluitend uit de kwaliteit van de psychometrische kenmerken. Vanuit een onderwijskundig standpunt bezien kunnen deze toetsen grote invloed hebben op het onderwijs zelf (Newble & Jaeger, 1983; Stillman & Swanson, 1987). Toetsen voor praktische vaardigheden stimuleren het aanleren van vaardigheden, kunnen motiverend werken voor docenten en kunnen aanleiding geven tot standaardisatie (consensus) over wat vaardigheden zijn en hoe ze moeten worden aangeleerd (Bouhuijs e.a., 1987).

Niettemin zou de techniek aan gewicht winnen als ook voldoende betrouwbare informatie zou kunnen worden verkregen. Daarbij zijn enkele mogelijkheden denkbaar.

In de eerste plaats zou men lange toetsen kunnen gebruiken, zij het dat dit een zware logistieke inspanning zal vergen. Wel kan in dit verband worden opgemerkt, dat het uitbreiden van de toets met meer stations een gunstiger effect zal hebben op de betrouwbaarheid dan het inzetten van meerdere beoordelaars bij dezelfde stations. Aangezien de variatie tussen vaardigheden groter is dan de variatie tussen beoordelaars, zal het verlengen van de toets de betrouwbaarheid sterker beïnvloeden dan het inzetten van meer beoordelaars.

In de tweede plaats kan worden aangeraden om toetsen van praktische vaardigheden te combineren met schriftelijke toetsen (Swanson, 1987). De efficiëntie van de schriftelijke toetsvorm kan worden gecombineerd met de adequate predictieve waarde ervan (Van der Vleuten e.a., *ter perse*). De gecombineerde score zal een betrouwbaarder resultaat opleveren.

Een derde optie betreft de hiervoor genoemde verschuiving van reproduceerbaarheid van toetsscores naar de reproduceerbaarheid van zak/slaag beslissingen. Wellicht zou dit kunnen worden gecombineerd met een sequentiële testprocedure, waarin per individu onderscheid wordt gemaakt in toetstijd. Bij hoger scorende kandidaten kan worden volstaan met een korte toets, terwijl bij lager scorende studenten langere toetstijden nodig zullen zijn.

Tot slot is een vierde mogelijkheid, voor zover inhoudelijk mogelijk, het verkorten van de stationsduur en de gewonnen tijd te compenseren door meer vaardigheden te toetsen. Voorlopige aanwijzingen duiden op een per saldo grotere winst in betrouwbaarheid door het verkorten van de stationslengte, zelfs wanneer dat resulteert in een sterke reductie van de precisie van de meting binnen een enkel station (Van der Vleuten, 1987).

Bezien zal moeten worden of de conclusies uit deze studie uitsluitend beperkt zijn tot praktische vaardigheden in het medische domein. Voor meer simpele praktische vaardigheden, of meer pure psychomotorische vaardigheden zoals bijvoorbeeld technische vaardigheden uit het lager beroepsonderwijs, kan een hele andere situatie gelden. Verder psychometrisch onderzoek zal dat moeten uitwijzen. Wat betreft meer eenvoudige handelingsgerichte verpleegkundige vaardigheden hebben de eerste resultaten niettemin soortgelijke conclusies opgeleverd (Van der Vleuten e.a., in voorbereiding).

Referenties

- Bouhuijs, P.A.J., Vleuten, C.P.M. van der & Luyk, S.J. van, (1987), The OSCE as a part of a systematic skillstraining approach. *Medical Teacher*, 9, 183-191.
- Brennan, R.L., (1983), *Elements of Generalizability Theory*. Iowa: American College Testing Program.
- Brennan, R.L. & Kane, M.T., (1977), An index of dependability for mastery tests. *Journal of Educational Measurement*, 14, 277-289.
- Crick, J.E. & Brennan, R.L., (1983), *Manual for GENOVA: A Generalized Analysis of Variance System*. Iowa: American College Testing Program.

- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N., (1972), *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley.
- Dawson-Saunders, B., Verhulst, S.J., Marcy, M. & Steward, D.E., (1987), Variability in standardized patients and its effect on student performance. In: I.R. Hart & R.M. Harden, (Eds.), *Further Developments in Assessing Clinical Competence*. Montreal: Can-Heal.
- Dochy, F.J.R.C. & Luyk, S.J. van, (Eds.), (1987), *Handboek Vaardigheidsonderwijs*. Lisse: Swets & Zeitlinger.
- Harden, R.M. & Gleeson, F.A., (1979), ASME Medical Education Booklet No. 8. Assessment of clinical competence using an objective structured clinical examination (OSCE), *Medical Education*.
- Hiemstra, R.J., Bender, W., Scherpbier, A., Lunsen, H.W., Vries, P.G.M. de & Soeters, D., (1986), An objective structured clinical examination (OSCE) in Groningen, The Netherlands. In: I.R. Hart, R.M. Harden, R.M. & J.H. Walton, (Eds.), *Newer Developments in Assessing Clinical Competence*, Montreal, Heal Publications.
- Hiemstra, R.J., Scherpbier, A.J.J.A. & Roze, B.J., (1987), Assessing history-taking skills or ... simulated patients' peculiarities. In: I.R. Hart & R.M. Harden, (Eds.), *Further Developments in Assessing Clinical Competence*. Montreal: Can-Heal.
- Houtman, I. & Schinkelshoek, D., (1986), *Toetsen van praktische vaardigheden*. IFLO docentenopleiding en Afdeling Onderwijs Research, Vrije Universiteit, Amsterdam.
- Klerk, L.F.W. de, (1980), *Het leren van psychomotorische vaardigheden: een onderwijspsychologische benadering*. Deventer: Van Loghum Slaterus.
- Kraan, H. & Crijnen, A., (1987), *The Maastricht History Taking and Advice Checklist: Studies of instrumental utility*. Academisch proefschrift Rijksuniversiteit Limburg.
- Metz, J.C.M., (1984), *Medische competentie, Een onderzoek naar de betrouwbaarheid en de validiteit van het Gestructureerd Klinisch Examen*. Dissertatie, Nijmegen.
- Metz, J.C.M., (1986), Het gestructureerd klinisch examen. *Nederlands Tijdschrift voor de Geneeskunde*, 130, 2091-2093.
- Newble, D.I. & Jeager, K., (1983), The effect of assessments and examinations on the learning of medical students. *Medical Education*, 17, 165-171.
- Newble, D.I. & Swanson, D.B., (Ter perse), Psychometric characteristics of the Objective Structured Clinical Examination. *Medical Education*.
- Petrusa, E.R., Blackwell, T., Parcel, S. & Saydjari, C., (1986), Psychometric properties of the Objective Clinical Exam as an instrument for final evaluation. In: I.R. Hart, R.M. Harden, R.M. & J.H. Walton, (Eds.), *Newer Developments in Assessing Clinical Competence*, Montreal: Heal Publications.
- Pieters, J.M., (1984), Praktische vaardigheden in het voortgezet onderwijs. In: L.F.W. De Klerk & A.M.P. Knoers, (Eds.), *Onderwijspsychologisch Onderzoek*. Lisse: Swets & Zeitlinger.
- Rossum, H.J.M. van, (1985), Honderd groepen Alco. In: Th. W.M. Hoeks, E. van der Putte & H.J.M. van Rossum, (Eds.), *Vijf jaar Alcoschap, een geslaagd experiment*. Faculteit der Geneeskunde, Leiden.
- Sanders, P.F., (1980a), *Een inleiding tot toetsen praktische vaardigheden*. Specialistisch Bulletin, nr. 5, Cito Arnhem.
- Sanders, P.F., (1980b), *Een procedure voor de constructie toetsen praktische vaardigheden*. Specialistisch Bulletin, nr. 9, Cito Arnhem.
- Schmidt, H.G., (1983), Problem based learning: Rational and description. *Medical Education*, 17, 11-16.
- Stillman, P.L., Sabers, D. & Redfield, D., (1976), The use of paraprofessionals to teach and evaluate interviewing skills in medical students. *Pediatrics*, 57, 769-774.
- Stillman, P.L. & Swanson, D.B., (1987), Ensuring the clinical competence of medical school graduates through standardized patients. *Archives of Internal Medicine*, 147, 1049-1052.
- Stillman, P.L., Regan, M.B. & Swanson, D.B., (1987), A diagnostic fourth year performance assessment. *Archives of Internal Medicine*, 105, 762-771.
- Swanson, D.B., (1987), A measurement framework for performance based tests. In: I.R. Hart & R.M. Harden (Eds.), *Further Developments In Assessing Clinical Competence*. Montreal: Can-Heal.
- Swanson, D.B. & Norcini, J.J., (In voorbereiding), Factors influencing the reproducibility of tests using standardized patients.
- Verwijnen, G.M., Imbos, Tj., Snellen, H., Stalenhoef, B., Pollemans, M., Luyk, S. van, Sprooten, M., Leeuwen, Y. van & Vleuten, C.P.M. van der, (1982), The evaluation system of the medical school of Maastricht. *Assessment and Evaluation in Higher Education*, 3, 225-244.
- Vleuten, C.P.M., van der, (1987), Intracase versus intercase reliability: A trade-off? *PES-Publ. nr. 178, Interne publicatie Rijksuniversiteit Limburg*.
- Vleuten, C.P.M. van der & Luyk, S.J. van, (1986), A validity study of a test for clinical and technical medical

- skills. In: I.R. Hart, Harden, R.M. & Walton, H.J. (Eds.), *Newer Developments in Assessing Clinical Competence*, Montreal: Heal Publications.
- Vleuten, C.P.M. van der, Luyk, S.J. van & l'Espoir, N.E.J.C. (1987), Effecten van vaardigheidsonderwijs. In: F.J.R.C. Dochy & S.J. van Luyk, (Eds.), *Handboek Vaardigheidsonderwijs*. Lisse: Swets & Zeitlinger.
- Vleuten, C.P.M. van der, Luyk, S.J. van & Beckers, A.J.M., (Ter perse), A written test as an alternative to performance testing. *Medical Education*.
- Vleuten, C.P.M. van der, Robroek, W.C.L. & Luyk, S.J. van, (In voorbereiding), Assessing practical nursing skills: Three methods compared.
- Williams, R.G., Barrows, H.S., Vu, N.V., Verhulst, S.J., Colliver, J.A., Marcy, M. & Steward, D., (1987), Direct, Standardized Assessment of Clinical Competence. *Medical Education*, 21, 482-489.
- Wijnen, W.H.F.W., (1971), *Onder of Boven de Maat*. Academisch proefschrift Rijksuniversiteit Groningen. Amsterdam: Swets & Zeitlinger.

Manuscript ontvangen 15-2-1988

Definitieve versie ontvangen 15-6-1988

Boekbesprekingen

B.H.A.M. Van Hout Wolters: *'Het markeren van kerngedeelten in studieteksten'*. Een 'proces-product'-benadering. (Proefschrift K.U. Brabant). Lisse: Swets & Zeitlinger 1986.

Op de eerste bladzijde van het proefschrift van B. van Hout Wolters staat:

'Het proefschrift heeft vooral tot doel een bijdrage te leveren aan het verbeteren van onderzoek op dit terrein'. Met 'dit terrein' wordt dan bedoeld, het terrein van het lezen en verwerken van teksten, waarbij hier vooral gelet wordt op een effectieve samenstelling van het studiemateriaal alsmede op de vereiste studievaardigheden van de leerlingen. De keuze van zo'n doelstelling getuigt van moed en vertrouwen in eigen kunnen. Ik kan me voorstellen, dat je op twee manieren tot zo'n uitspraak komt, nl., omdat je vindt dat dergelijk onderzoek in het verleden onzorgvuldig is uitgevoerd, of omdat je de hele tot nu toe gehanteerde onderzoeksaanpak afwijst en dan daar een nieuwe onderzoeksaanpak tegenover wilt stellen.

Van Hout Wolters heeft zich tot doel gesteld om duidelijk te laten zien, dat de huidige onderzoeksaanpak aanzienlijk verbeterd kan worden, door veel zorgvuldiger om te gaan met de opzet, de materiaalkeuze en de analyse van de gegevens. Ik ben het met haar eens, dat tot nu toe op dit gebied verricht onderzoek vaak erg te wensen overlaat. Vaak zijn het slecht gecontroleerde experimenten, die slordig waren opgezet en met geen enkele theoretische pretentie. Deze onzorgvuldigheid heeft tot gevolgd, dat ook de bruikbaarheid voor de praktijk, de praktische opbrengst, vrijwel nihil geweest is. Van Hout Wolters heeft willen laten zien, dat het ook anders kan: Dit heeft zij gedaan, door een onderzoek uit te voeren, dat als 'ideaalmodel' gehanteerd kan worden. Op voorhand wil ik hier stellen, dat ze mijns inziens die pretentie heeft waargemaakt. Het is een heel goed onderzoek, waarover zorgvuldig gerapporteerd is. Het aantrekkelijke van haar verslag is, dat ze alle keuzes die ze maakt en de compromissen die daarbij gesloten zijn, expliciet verwoordt. Dat is uniek. Het proefschrift krijgt daardoor didactische waarde, omdat juist die verwoording van dit soort keuzes, die vaak zeer essentieel zijn voor de kwaliteit van de onderzoeksopzet, vrijwel nooit terug te vinden zijn in onderzoeksrapporten. Bij de bespreking is als uitgangspunt genomen of Van Hout Wolters dit modelkarakter bij alle onderdelen van haar onderzoek ook in dezelfde mate gerealiseerd heeft.

Het proefschrift bestaat uit tien hoofdstukken, die zijn opgesplitst in drie delen. In het eerste deel, dat betrekking heeft op theoretische achtergronden, wordt onderzoek op dit gebied geordend en besproken. In het tweede deel wordt de constructie van de studieteksten, alsmede de meetinstrumenten beschreven. In het derde deel wordt gerapporteerd over de product- en procesgegevens.

Deel I:

Hierin bespreekt Van Hout Wolters het onderzoek naar het markeren van studieteksten. Dit gebeurt op een grondige en systematische manier. De hele procedure, die aan de selectie van de besproken onderzoeksartikelen ten grondslag ligt wordt beschreven. Eerst wordt uiteengezet hoe het literatuur-onderzoek is opgezet, waar gezocht is en wanneer. Dit levert meteen informatie op over het soort onderzoek wat plaatsgevonden heeft, welke variabelen daarbij gekozen zijn en wie erin participeerden. Het resultaat van deze analyse wordt door Van Hout Wolters gekoppeld aan theorieën over informatieverwerking. Het door haar gevonden hogere rendement van leerling-markering boven de docent-markering, wordt vanuit die theorie geïnterpreteerd. In het eerste deel wordt de modelfunctie betrekking hebbend op: 'Hoe kom je tot een onderzoekstopic?', 'Wat is er over bekend?', 'In welk theoretisch kader past het?', zonder meer waar gemaakt.

Deel 2:

In deel twee komt de constructie van de toetsen en de teksten aan de orde. De verantwoording van de kwaliteit en vergelijkbaarheid van de teksten vind ik zonder meer klassiek. Vrijwel alle mogelijke beoordelingsmaten worden gebruikt om greep te krijgen op de onderlinge vergelijkbaarheid. Iets minder enthousiast ben ik over de keuze van de leerling-variabelen, waarin zich toch wel een zekere overdaad manifesteert. Gezien de hoeveelheid proefpersonen, 24 Havo-leerlingen die zorgvuldig geselecteerd zijn uit een aantal van 95, en de repercussies, die uit dit aantal voortvloeien voor de mogelijke celvullingen, acht ik het meenemen van de variabelen, zoals cognitieve stijl, prestatiemotivatie, verbale intelligentie, naast variabelen zoals interesse voor het onderwerp, geoefendheid en voorkennis, e.d. te veel van het goede. Een selectie hiervan, gebaseerd op de te verwachten invloed, zou mijns inziens zeker op zijn plaats geweest zijn. De gevolgde procedure voor de constructie van de toetsing, is daarentegen weer uiterst zorgvuldig en correct. In het onderzoek, een within-subject design, worden drie condities gecreëerd: een markeringsconditie, waarin de leerlingen delen van de tekst markeerden, een conditie waarbij de markeringen door de docent waren aangegeven en een conditie, waarbij er niet gemarkeerd mocht worden. Iedere leerling participeerde in alle condities. Dit is uiteraard alleen maar mogelijk, als de teksten onderling optimaal vergelijkbaar zijn, omdat anders teksteffecten voor de nodige interferentie kunnen zorgen. Naast productgegevens, zijn er ook procesgegevens verzameld, waarbij gebruik gemaakt is van de zgn. 'XY'-methode. Een elektronische pen, waarmee de onderstrepingen plaatsvinden, registreert hoeveel tijd besteed wordt per regel. Daarmee is ook te bepalen waar gelezen wordt, doordat de proefpersonen aanstippen op welke regel ze lezen. Het nagaan van de validiteit van de XY-methode is door Van Hout Wolters in zorgvuldig opgezette vooronderzoekjes uitgebreid getest. Daarnaast is er gebruik gemaakt van de 'dual-task' methode waarbij de inspanning door proefpersonen gemeten is met behulp van een neventaak, nl. 'Het zo snel mogelijk indrukken van een knop bij het horen van een bepaalde toon'. Tenslotte hebben leerlingen lijsten ingevuld en zijn ze na afloop geïnterviewd. Bij de keuze van het design, vraag ik mij af of deze keuze voor het relatief ingewikkelde within-subject design, gezien alle eisen van vergelijkbaarheid die daaraan gesteld worden, zoveel beter is dan een between-subject design. Weliswaar betekent dit meer proefpersonen, maar het kost per proefpersoon minder tijd. Ook al zijn de teksten goed vergelijkbaar dan blijft nog altijd de vraag of de toetsen die bij de teksten horen op dezelfde manier vergelijkbaar zijn. Volgens mij blijft 't een kwestie van afweging of de winst die geboekt wordt met de keuze van het within-design, in termen van reductie van individuele verschillen, opweegt tegen de inspanning die je moet getroosten om strikt vergelijkbare teksten en toetsen te construeren.

Deel 3:

In dit deel worden de product-proces-gegevens gepresenteerd en aan elkaar gerelateerd. Hierin worden een groot aantal analyses uitgevoerd. Niet alleen worden de groepen sec vergeleken, maar ook wordt gekeken of de effecten vooral te vinden zijn in de gemarkeerde delen van de tekst. Dit laatste blijkt het geval. De enige kritiek die op dit deel denkbaar is, is dat de instructie die aan de leerlingen werd gegeven in de verschillende condities, tot diverse interpretaties kan leiden. Sommige leerlingen zullen pas na herlezing de hoofdzaken zijn gaan markeren, anderen zijn wellicht al direct gaan markeren bij de eerste keer lezen. De instructie, wat dit betreft, is te vaag en te vrijblijvend. Toch is het belangrijkste van dit proefschrift niet in eerste instantie het resultaat, maar de presentatie van het gehele onderzoek. Van Hout Wolters laat zien, waar belangrijke beslismomenten zitten. Inplaats van het maskeren van subjectieve momenten, laat ze door een expliciete presentatie de lezer mee in de keuken kijken van het onderzoek, iets wat wellicht alleen goede koks aandurven.

J.H. Boonman

Vakgroep Onderwijskunde
Universiteit Utrecht

Heidelberglaan 2
3508 TC Utrecht

Mart-Jan de Jong: *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs*. Lisse: Swets en Zeitlinger, 1987, Proefschrift Erasmus Universiteit, Rotterdam. f 49,90.

De laatste jaren verschijnen er regelmatig proefschriften over allochtone kinderen in het Nederlandse onderwijs. Tot nu toe betroffen het vooral onderzoeken over specifieke aspecten van de onderwijsproblematiek van allochtone kinderen, zoals biculturele opvang- of onderwijsmodellen, of tweede-taal verwerving. Het proefschrift van Mart-Jan de Jong handelt over een zeer relevant aspect, zeker gezien tegen de achtergrond van het vigerende onderwijsvoorrrangsbeleid, namelijk de onderwijsprestaties van allochtone en autochtone kinderen.

Kern van het boek wordt gevormd door het verslag van een longitudinaal onderzoek naar het overgangsproces van het lager naar het voortgezet onderwijs bij 1000 scholieren van 57 Rotterdamse scholen. Hiertoe zijn deze leerlingen gedurende vier jaar gevolgd vanaf de 5e klas van het lager onderwijs (groep 7 van het basisonderwijs) tot het einde van de tweede klas van het voortgezet onderwijs. Van deze 1000 leerlingen behoren 251 (25%) tot de etnisch-culturele groepen (exclusief 55 met een etnisch-gemengd ouderpaar). Van hen hebben 140 een Surinaamse en 61 een Turkse of Marokkaanse achtergrond. Daarnaast bevinden zich 15 Zuid-Europese, 16 Chinese en 19 Kaap-Verdiaanse leerlingen in de onderzochte groep.

Het onderzoek concentreert zich op de volgende vragen: hoe groot zijn de onderlinge verschillen in schoolprestaties tussen de autochtone en de diverse groepen allochtone leerlingen en hoe liggen de verschillen binnen de resp. etnisch-culturele groepen? Wijzigen deze verschillen zich tijdens en na de doorstroming naar het voortgezet onderwijs? Hoe kunnen deze verschillen worden verklaard?

Om deze verschillen in kaart te brengen, te analyseren en te verklaren is gekozen voor een gewijzigd Wisconsin-model. De wijzigingen in dit model waren noodzakelijk, omdat in het originele model geen rekening wordt gehouden met verschillen in etnisch-culturele herkomst. Het analyse model onderscheidt drie soorten variabelen: achtergrondvariabelen (socio-econ. status, etnisch-culturele herkomst, sekse, verblijfsduur in Nederland), interveniërende variabelen (structurele gezinskenmerken, opvoedingskenmerken, onderwijskenmerken, percentage allochtonen in de klas, IQ, studiegedrag en -houding, waarden en verwachtingen, leerproblemen) en de afhankelijke variabele: bereikt onderwijsniveau. De veronderstelling achter dit model is duidelijk: de achtergrond- en interveniërende variabelen zijn verantwoordelijk voor het bereikte onderwijsniveau. De operationalisatie van deze variabelen geschiedde vooral op basis van literatuurstudie. De keuze van de onderzoeksinstrumenten en de wijze waarop deze zijn gebruikt worden vooral bepaald door praktische overwegingen. Zo noodzaakten de randvoorwaarden van het onderzoek tot klassikaal af te nemen IQ tests (de GALO-intelligentietest) en zo besloot de onderzoeker de vragen over structurele en gezinsopvoedingskenmerken (o.a. die naar de opvoedingswaarden van ouders) rechtstreeks aan de kinderen te stellen. Dit betekent onvermijdelijk een bias in de uitkomsten.

Wat betreft de resultaten van dit onderzoek, zowel het correleren van de diverse variabelen alsook ook de beantwoording van de vragen rond de afzonderlijke variabelen leveren interessante uitkomsten op. Allochtone leerlingen gaan bijvoorbeeld minder vaak gebukt onder een gebrek aan zelfvertrouwen dan men geneigd is te veronderstellen en hun prestatie-motivatie is significant sterker dan bij autochtone leerlingen.

Verontrustend is de achterstand van de allochtone leerlingen op het gebied van de beheersing van de Nederlandse taal aan het einde van het tweede jaar in het voortgezet onderwijs. In schooljaren uitgedrukt bedraagt deze voor de Chinese leerlingen 0,5 jaar, voor de Surinaamse, Zuid-Europese en Kaap-Verdiaanse leerlingen 1,5-2 jaar en voor de Turkse en Marokkaanse leerlingen 2,5-3 jaar. De ontwikkeling van het taalbeheersingsniveau blijkt geen significant verband te vertonen met de verblijfsduur in Nederland. Een belangrijke conclusie die dan ook in hoofdstuk 11 leidt tot de aanbeveling om in het onderwijs meer en efficiënter aandacht te besteden aan het leren van de Nederlandse taal. Merkwaardigerwijs komt deze aanbeveling op een moment dat een aantal grote steden juist overweegt om de jonge allochtone leerlingen meer

onderwijs te geven met de eigen (welke?) taal als instructietaal, omdat zij het Nederlands nog zo weinig beheersen.

Tot slot wil ik nog enkele kritische kanttekeningen maken. Deze betreffen het onderzoek als zodanig en de opzet van het boek. Een van de voorwaarden voor een goede verslaglegging is het inzichtelijk maken van de relevante aspecten van het onderzoek. Dit doet de onderzoeker wel wanneer het gaat om de operationalisatie van de variabelen; echter niet wanneer het de gevolgde methode van onderzoek betreft. Nergens wordt in een paragraaf of een hoofdstuk een overzicht gegeven van de gevolgde methode van onderzoek, de hierbij gehanteerde instrumenten en de jaren waarin deze werden afgenomen of gebruikt. De lezer moet de schaarse informatie hierover uit de diverse hoofdstukken bijeenprokkelen.

Een tweede kritische kanttekening betreft het feit dat de auteur onvoldoende aangeeft dat de onderzochte groep allochtone leerlingen niet representatief genoemd kan worden voor de etnisch-culturele schoolpopulatie in Nederland, hetgeen de titel van het boek wel suggereert. De percentages allochtone leerlingen uit de onderzochte groep (60% van Surinaamse en 25% van Turkse en Marokkaanse herkomst), zijn noch representatief voor de Rotterdamse scholen, noch voor die in geheel Nederland (25% van Surinaamse en > 50% van Turkse en Marokkaanse herkomst). Een omissie lijkt mij dat er binnen de categorie Surinaamse leerlingen geen onderscheid gemaakt is naar etnische groep (bijv. naar leerlingen behorend tot de Hindoe-staanse, dan wel Creoolse groepering). Dat tussen deze groepen grote verschillen kunnen bestaan wordt wel erkend (bijvoorbeeld bij de bespreking van de invloed van het opgroeien in een één-ouder gezin), maar niet uitgewerkt.

De opzet van het boek vertoont een zekere onevenwichtigheid. Er wordt relatief veel ruimte in beslag genomen door de bespreking van literatuur over onderwerpen die geen directe relevantie voor het onderzoek en het analysemodel hebben (de hoofdstukken 1 en 2). De probleemstelling en het analysemodel van het onderzoek - de kern van het boek - worden pas in hoofdstuk 3 gepresenteerd temidden van een bespreking over ongelijke onderwijskansen en het schoolloopbaanonderzoek. Vanaf hoofdstuk 4 volgt dan het verslag van het onderzoek. Het geheel maakt de indruk alsof de auteur de gehele onderwijsproblematiek van allochtone leerlingen heeft willen bespreken. Hij haalt heel wat van stal. Hij begint bijvoorbeeld een discussie over het onderwijs in de eigen taal en cultuur en over het intercultureel onderwijs. De ruimte (of tijd) ontbreekt om deze discussie af te ronden. Soms blijft de auteur steken in opiniërende opmerkingen. Zo vermeld hij bijvoorbeeld dat zich in zijn onderzoekspopulatie een groot aantal leerlingen uit etnisch-gemengde huwelijken bevindt (hij noemt dit 'Mix-categorie'). Deze groep blijkt groter te zijn dan zijn Turkse en Marokkaanse groep van resp. 36 en 25 leerlingen. Dit gegeven stimuleert hem echter niet tot reflectie over de samenstelling van zijn onderzoekspopulatie, maar wel tot een uithaal naar sociale wetenschappers die het moeilijk zouden hebben met het begrip assimilatie (pag. 75 en 76).

Kortom dit boek is zeer zeker de moeite waard om te lezen, te bestuderen en te bediscussiëren. Voor mensen die nog weinig kennis en ervaring hebben met allochtone groepen lijkt het mij minder geschikt.

L. van den Berg-Eldering

R.U. Leiden

Sectie interculturele pedagogiek.

Vakgroep Orthopedagogiek

Henri Dors: *Vriendschap en sociale relaties in multi-etnisch samengestelde schoolklassen*. Lisse; Swets & Zeitlinger, 1987. Proefschrift Universiteit van Amsterdam. f 36,30.

Een tweede proefschrift dat in 1987 over het onderwerp etnische groepen en het onderwijs is verschenen, is dat van Henri Dors. Het onderzoek waarvan in het boek 'Vriendschap en sociale relaties in multi-etnische samengestelde schoolklassen' verslag wordt gedaan wil nagaan hoe de interetnische relaties tussen kinderen liggen en hoe deze samenhangen met een aantal schoolinterne factoren.

In de introductie zegt Dors de relatie te willen blootleggen tussen ideologieën die samenhangen met de etnische factor en de aard van interetnische relaties binnen de basisschool. De directe aanleiding tot het opzetten van dit onderzoek is het overheidsbeleid met betrekking tot het intercultureel onderwijs. De doelstelling van dit intercultureel onderwijs is mede onder invloed van de stellingname van Dors verschoven in de richting van het bevorderen van interetnische relaties, althans volgens de officiële beleidsnotitie van 1986. In deze notitie wordt echter niet aangegeven hoe die interetnische relaties in positieve zin kunnen worden beïnvloed. Dors stelt zich dan ook voor om door middel van zijn onderzoek hierin meer inzicht te verschaffen en zodoende bij te dragen aan de oplossing van problemen in de praktijk van beleid en onderwijsveld.

Meer in concreto wil hij nagaan hoe het verband ligt tussen de etnische beleving van de leerkracht en de leerstof die hij aanbiedt (de onafhankelijke variabelen) en de vriendschappen en interetnische relaties tussen de kinderen in zijn klas (afhankelijke variabelen). Het theoretisch kader voor dit onderzoek wordt gevormd door de multi-etnische theorie van Banks, aangevuld met elementen uit de anti-racisme theorie. Banks onderscheidt een drietal benaderingen van etnische vraagstukken: assimilationisme, cultureel pluralisme en de multi-etnische benadering. De eerste benadering ziet de minderheidsculturen als deficiënt, gedeprimeerd en pathologisch, de tweede gaat ervan uit dat de openbare instituties in de samenleving de belangen van etnische groepen in beleid en praktijk dienen te representeren. De multi-etnische theorie van Banks zelf neemt een middenpositie in. Zij gaat ervan uit dat alle etnische groepen worden beïnvloed door de 'mainstream' cultuur en dat het onderwijs tot taak heeft leerlingen te helpen bij het ontwikkelen van hun etnische identiteit en het bewust maken van etnische en culturele alternatieven, zodat zij cross-cultureel kunnen functioneren in een 'open society'.

Deze multi-etnische theorie schiet volgens Dors op één punt tekort: zij geeft namelijk niet aan welke instrumenten en mechanismen de realisering van deze open samenleving in de weg staan. De anti-racismetheorie doet dit wel. Deze theorie gaat ervan uit dat de structuren van de samenleving doortrokken zijn van institutioneel racisme. De leerkracht maakt deel uit van deze structuren en zal dus in meerdere of mindere mate racistisch zijn in zijn houding naar de leerling. Dit geldt ook voor de leerstof die hij aanbiedt.

De concrete onderzoeksvraagstelling wordt dan of er verband bestaat tussen het racisme van de leerkracht en diens leerstof en de relaties tussen de kinderen in zijn klas. Dit is in het onderhavige onderzoek in meer neutrale termen geformuleerd. De onderzoeker spreekt over de etnische beleving van de leerkracht of over diens SOE (Stage Of Ethnicity). Voor het meten van de leerkracht- en leerstofvariabele gaat de onderzoeker weer te rade bij Banks. Deze heeft een aantal fasen onderscheiden in de etnische beleving van de leerkracht en richtlijnen gegeven waaraan leerstof zou moeten voldoen. Voor het meten van de relaties tussen de kinderen heeft het onderzoek van Bartel model gestaan. Bartel heeft, in tegenstelling tot vele anderen, niet slechts gekeken naar de etnische voorkeuren van kinderen, maar ook een samenhang gezocht met de intellectuele en sociale vaardigheden van kinderen (welke kinderen kies je uit om mee te rekenen of om mee te spelen en welke juist niet?).

Het hoofdonderzoek heeft plaats gevonden in de maanden mei en juni 1986 onder 1549 leerlingen uit 68 groepen (36 behorend tot groep 3 en 32 tot groep 6) van 42 basisscholen in onderwijsvoorrangsgebieden in Amsterdam en onder 164 onderwijsgegenden.

Voorwaarden voor deelname aan het onderzoek waren de multi-etnische samenstelling van de klassen (groep 3 en 6) en voldoende aantallen behorend tot de volgende raciale/etnische

categorieën: blank Nederlands, Surinaams en Antilliaans-Nederlands, Marokkaans en Turks. Van de leerkrachten was 63% vrouw en 34% man, 86% behoorde tot de blank Nederlandse categorie en 8% tot de Surinaamse of Antilliaanse categorie.

De onderzoeksinstrumenten bestonden uit vragenlijsten (etnische beleving, aangeboden leerstof, sociometrisch interview, gegevens leerlingen, lijst met inschatting door leerkracht van sociale relaties tussen de leerlingen).

De belangrijkste conclusies van dit onderzoek zijn -) dat sekse-voorkeur een duidelijke rol speelt en relatief meer gewicht in de schaal legt dan etnische voorkeur bij de kinderen, -) dat er geen algemene samenhang is tussen de leeftijdsgroep (5⁺ en 9⁺) en de aard en sterkte van de etnische voorkeur bij de kinderen, -) dat er geen algemene samenhang kan worden geconstateerd tussen de etnische beleving van de leerkracht en de etnische voorkeur van de kinderen uit zijn klas en dat dit evenmin het geval is tussen de aangeboden leerstof en de interetnische relaties.

Tussen de onderscheiden etnische groepen doen zich wel verschillen voor wat betreft etnische voorkeur. Bij de blanke Nederlandse kinderen bestaat geen specifieke etnische voorkeur, bij de Surinaamse kinderen daarentegen wel. De laatsten hebben niet alleen een duidelijke voorkeur voor de eigen groep, maar wijzen ook andere categorieën af. Marokkaanse en Turkse kinderen worden vooral afgewezen om hun intellectuele (on)vaardigheden.

Wat betekenen deze uitkomsten nu voor de doelstellingen van de onderzoeker: het blootleggen van verbanden tussen ideologie en etnische voorkeur in de klas en het geven van adviezen voor intercultureel onderwijs? Gezien de beperkte ruimte ga ik slechts in op één van de uitkomsten: het verband tussen de etnische beleving van de leerkracht en de etnische voorkeur van de kinderen uit zijn klas. Het onderzoek toont aan dat er geen algemeen verband tussen beide variabelen bestaat. Voor deze uitkomst zijn in principe twee verklaringen mogelijk: óf de theorie, in casu de anti-racisme theorie klopt niet óf het onderzoek, in casu het instrumentarium, deugt niet. Als aanhanger van de anti-racisme theorie neigt Dors meer naar de tweede verklaring, hoewel hij niet met zoveel woorden afstand neemt van het fasenmodel van Banks.

Volgens dit fasenmodel zouden de meeste leerkrachten zich in een bi-etnische fase bevinden, hetgeen betekent dat zij een gezond besef van etnische identiteit hebben en de psychologische karakteristieken en vaardigheden bezitten die nodig zijn om op succesvolle wijze te participeren in de eigen (blanke) als in één andere etnische gemeenschap, aldus de auteur op p. 136. De interviews met de leerkrachten bevestigen deze uitkomsten nog eens. Toch meent Dors zijn eigen onderzoeksresultaten in twijfel te moeten trekken. De leerkrachten hebben immers geen vrienden in een andere etnische gemeenschap. Dit moet wel invloed hebben op de relatie die zij met de leerlingen hebben, met als gevolg een deficiet benadering van kinderen uit minderheidsgroepen en een sterke doorstroming naar lagere vormen van voortgezet onderwijs. De werkelijke oorzaak hiervoor moet gezocht worden in het feit dat leerkrachten onbewust in het systeem van instituties, regels en gewoonten participeren die institutioneel etnische ongelijkheid produceren en reproduceren. Wat leerkrachten nodig hebben, aldus Dors, is een heroriëntatie met het oog op een daadwerkelijke strijd tegen het structureel ingekaderde racisme à la Mullard. Dit lijkt een cirkelredenering.

Het onderzoek van Dors bevat een interessante vraagstelling. Een vraagstelling die directe relevantie heeft voor het beleid en de praktijk van het intercultureel onderwijs. De uitkomsten geven echter niet veel houvast voor dit onderwijs. De aanbevelingen van Dors stoelen meer op morele gronden en op de anti-racisme theorie dan op de uitkomsten van het onderzoek. Het theoretische kader is nogal gekunsteld. Het onderzoek steunt vooral op de theorie van Banks, bij de bespreking van het onderzoek van Bartel komen de intellectuele en sociale vaardigheden van leerlingen als variabelen om de hoek kijken. Deze worden zonder nadere bespreking in het onderzoeksdesign opgenomen. De mogelijke relevantie van sociale klasse en van concentratie van etnische groepen bij de etnische voorkeur van kinderen wordt wel even genoemd, maar niet meegenomen in het onderzoek. Het onderzoek roept een aantal vragen op die niet worden beantwoord zoals de vraag naar de verklaring van de geringe etnische voorkeur bij blanke Nederlandse kinderen en de uitgesproken etnische voorkeur bij Surinaams-Nederlandse

kinderen. En de vraag in hoeverre de etnische beleving van de leerkrachten uit de diverse etnische groepen verschilt en of dit ook invloed had op de voorkeur van de kinderen.

Tot slot wil ik nog een opmerking maken over de medewerking van scholen en leerkrachten aan onderzoek. Dors vermeldt een vrij hoog percentage negatieve respons (43%) ondanks de zachte dwang van het ABC en het bevoegd gezag om mee te werken aan dit onderzoek. Een veelgehoorde reden van weigering was dat scholen al zo vaak worden benaderd voor onderzoek en dat zij zo weinig met de resultaten kunnen. Van de leerkrachten die wel meewerkten aan dit onderzoek werd veel gevraagd. Zij moesten de gegevens verzamelen door het (laten) invullen van de vragenlijsten. De onderzoekers leverden het pakket onderzoeksmateriaal bij de school af en kwamen het later weer ophalen. Bij de leerkrachten die wel meededen wekte de vragenlijst naar de etnische beleving weerstanden op, daar zij de indruk hadden getest te worden op hun racisme, hetgeen klopte.

Ik vraag mij af hoe groot de medewerking op scholen bij een volgend onderzoek zal zijn, wanneer de leerkrachten vernemen dat zij, ondanks de positieve scores op de schaal van etnische beleving, toch verantwoordelijk gesteld blijven voor de productie en reproductie van de etnische ongelijkheid in het onderwijs.

L. van den Berg-Eldering
R.U. Leiden
Sectie interculturele pedagogiek
Vakgroep Orthopedagogiek,

Th. H. L. Compennolle, *Stress op de middelbare school*. Amersfoort/Leuven, Acco 1987.

Met het onderzoek dat als ondertitel meekreeg "een literatuuroverzicht van het stressonderzoek en een onderzoek naar stressoren en stressresponsen bij leerlingen in het Vlaams middelbaar onderwijs", promoveerde Theo Compennolle, thans werkzaam als kinder- en jeugdpsychiater aan het kinderpsychiatrisch centrum Tulpenburg te Amstelveen, tot doctor in de geneeskunde. Als "motto" kreeg het proefschrift een citaat van T.S. Eliot mee, waarmee de exploratieve aard van het onderzoek wordt weergegeven:

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

Het proefschrift omvat zeven hoofdstukken plus een Engelse samenvatting, een uitgebreide lijst met literatuurreferenties en bijlagen. *Hoofdstuk 1* geeft een algemene inleiding tot het onderzoek. De aanleiding tot het onderzoek ligt in de dagelijkse kinder- en jeugdpsychiatrische praktijk, waar men sinds de jaren '50 steeds meer aandacht is gaan schenken aan het belang van de sociale context voor het veroorzaken en in stand houden van psychiatrische dysfuncties bij kinderen. De school, die op het gezin na de belangrijkste context voor kinderen is, heeft binnen het onderzoek tot nu toe relatief weinig aandacht gekregen.

In *Hoofdstuk 2* wordt een overzicht gegeven van de "beperkte en versnipperde literatuur over stress bij leerlingen". Op basis hiervan wordt een theoretisch kader gepresenteerd in de vorm van een geïntegreerd theoretisch stressmodel. Een viertal centrale begrippen worden achtereenvolgens besproken: stressoren, stressresponsen, beoordelingsprocessen en weerbaarheid. Stressoren of stress-situaties zijn stimuli die het "eco-psycho-somatisch evenwicht van een individu objectief verstoren of bedreigen, op zo'n wijze dat het individu de stimulus niet zomaar kan negeren of routinematig kan verwerken". Deze stressoren zijn niet per definitie slecht: zij kunnen ook een uitdaging betekenen, die tot beter presteren leidt. De fysiologische, psychologische en gedragsmatige reacties waarmee het organisme de verstoring van het evenwicht zoals veroorzaakt door de stressoren poogt te herstellen, worden stressresponsen

genoemd. Verschillende habituele stressresponsen worden besproken (o.a. aangeleerde hulpeloosheid, habituele doeloriëntatie waarmee o.a. bedoeld wordt prestatie-motivatie, habituele angst en examenangst). Cognitieve beoordelingsprocessen spelen een belangrijke mediërende rol tussen stressoren en stressreponsen: of een individu een situatie al dan niet als bedreigend ervaart is afhankelijk van zijn interpretatie en beoordeling van de situatie. Deze beoordelingsprocessen moeten worden opgevat als een continu feedback proces: de situatie wordt steeds opnieuw beoordeeld op basis van de dan beschikbare informatie. De veerkracht van een individu dient te worden opgevat als een resultante van individuele en contextuele mediërende factoren, en is als zodanig van invloed op de cognitieve beoordeling van de situatie. Met name goede sociale steun (of het geloof erover te beschikken) blijkt van invloed te zijn op fysische en psychische gezondheid. De vier centrale begrippen worden samengevoegd in een model dat het stressproces opvat als een dynamisch gebeuren, waarin rekening dient te worden gehouden met de volgende factoren: 1) tijd en ontwikkeling, 2) de context waarin de stressfenomenen zijn ingebed, en 3) het niveau van organisatie waarop stress een rol speelt (o.a. het niveau van de individuele leerling en het niveau van de schoolorganisatie). Het hoofdstuk wordt afgesloten met een kritische beschouwing van preventieve en curatieve behandelingsstrategieën en een heldere vergelijking van het Vlaamse en het Nederlandse onderwijssysteem.

De hoofdstukken 3 en 4 beschrijven de doelstellingen en de procedure van het onderzoek. Het eerste doel van dit onderzoek was het ontwerpen van een instrument waarmee de factoren die in het schoolmilieu stress veroorzaken kunnen worden opgespoord. Met behulp van een dergelijk instrument kunnen relevante subgroepen vergeleken worden zowel wat betreft hun stressresponsen, als de schooleigenschappen die ze als stresserend ervaren. Ook kan de school als stresserend milieu beschreven worden, en kan worden nagegaan welke verbanden er bestaan tussen door leerlingen gerapporteerde stressoren en hun stressresponsen. Aan het onderzoek werd deelgenomen door 1180 leerlingen uit eerste, tweede, derde en zesde klassen van veertien Vlaamse middelbare scholen (in leeftijd variërend van 12 tot 18 jaar). Middels vragenlijsten werd nagegaan welke aspecten van de school als stresserend worden ervaren en welke de habituele stressresponsen van leerlingen zijn. De stress-situaties op school werden geëvalueerd met behulp van een "Subjectieve Schoolstressoren Vragenlijst", die samengesteld werd op basis van de "Classroom Environment Scale" van Moos en zijn medewerkers (o.a. Moos, 1979), en die in een voorstudie (N=657) was aangepast voor de Vlaamse onderwijssituatie. De items (94 in totaal) zijn zodanig geformuleerd dat de "gemeenschappelijke perceptie van de klassituatie" (p. 14) wordt gemeten. Een voorbeeld van een item uit de lijst kan dit illustreren:

- | | |
|---------------------------------|---|
| De leerkrachten zijn erg streng | ☐ bijna nooit
☐ soms
☐ dikwijls
☐ bijna altijd |
|---------------------------------|---|

Als stressresponsen werden examenangst en habituele angst gemeten met respectievelijk de Nederlandse aanpassing van Spielberger's "Test Anxiety Inventory" (Examen/toets Attitude Vragenlijst; Van der Ploeg, 1980), en de "Zelf Beoordelings Vragenlijst" (ZBV) (Van der Ploeg, Defares & Spielberger, 1979). Daarnaast werden ter exploratie enige vragen gesteld omtrent de rol van het sociaal steunend netwerk, alsook over de prestatiegerichtheid van de ouders.

De resultaten van het onderzoek worden besproken en samengevat in de hoofdstukken 5 en 6. De belangrijkste bevindingen zullen we hieronder kort weergeven:

1. **Analyse van de vragenlijsten.** Om na te gaan of de tien theoretische subschalen van de "Subjectieve Schoolstressoren Vragenlijst" tot uitdrukking kwamen werden hoofdcomponenten analyses met varimax-rotatie uitgevoerd. De theoretisch te verwachten tien factoren van de "School Environment Scale" konden niet uit de data worden geëxtraheerd. De analyse met zes factoren (totaal verklaarde variantie = 34.7%) beantwoordde het beste aan de theoretisch te verwachten categorieën. Deze factoren werden benoemd als "Gebrek aan persoonlijke inzet van de leerkrachten" (32 items), "Wanorde en passiviteit van de leerlingen" (15 items),

"Prestatiegerichtheid" (10 items), "Individualisme van de leerlingen" (12 items), "Werkomstandigheden" (14 items), en "Gebrek aan didactische attitude van de leerkrachten" (11 items). De vragenlijst bleek goed te differentiëren tussen verschillende scholen.

2. Beschrijving van het schoolmilieu. Niet alleen de totale groep, maar ook relevante subgroepen werden geanalyseerd m.b.t. hun appreciatie van de school (m.a.w. aspecten van het schoolgebeuren waarop leerlingen kritiek hebben). Uit de bevindingen blijkt o.a. dat het Vernieuwd Secundair Onderwijs in vergelijking tot het Traditionele Onderwijs meer kritiek van de leerlingen krijgt (zie ook van Oudshoorn, 1987). Gemengde scholen blijken er eveneens slechter vanaf te komen dan niet gemengde scholen. In tegenstelling tot de gesignaleerde verschillen tussen de schooltypes is er een vrij grote overeenstemming tussen jongens en meisjes.

3. Verbanden tussen stressoren, stressresponsen en leerlingkenmerken op het niveau van de leerling. De twee factoren die het hoogst correleren met de stressresponsen van leerlingen waren de persoonlijke attitude van de leerkracht en het gebrek aan cohesie tussen de leerlingen (correlaties variërend van .13 tot .30). Ook de materiële omstandigheden spelen een belangrijke rol. Prestatiedruk was de enige factor van de "Schoolstressoren Vragenlijst", die niet met habituele angst correleerde. De prestatiedruk zoals die werd gemeten blijkt eerder als positieve kwaliteit van de school te worden ervaren en moet dus eerder als steun biedende "prestatiegerichtheid" worden geïnterpreteerd dan als stressor. Compennolle zegt hierover "Het lijkt erop dat de voor onderwijs en prestaties belangrijke factor prestatiegerichtheid ook een welzijnsbevorderende factor is (p. 169)".

Geconcludeerd wordt in *hoofdstuk 7* dat de "Schoolstressoren Vragenlijst" een goed instrument lijkt te zijn om stress in scholen op te sporen, en dat deze als zodanig bruikbaar is voor de onderwijspraktijk. Gezien de grote verschillen die tussen scholen geconstateerd konden worden, doen scholen er volgens Compennolle goed aan zelf stress op hun school te onderzoeken (het maken van een schoolstress-profiel) en op basis daarvan de zwakke punten te verbeteren. Twee factoren uit de schoolomgeving blijken het duidelijkst samen te hangen met stressresponsen bij leerlingen: de cohesie tussen leerlingen en de persoonlijke relatie met de leerkrachten. Het bevorderen van de cohesie tussen leerlingen onderling en tussen leerlingen en leerkrachten lijkt dan ook een belangrijke raadgeving naar de onderwijspraktijk toe.

De belangrijkste waarde van deze studie ligt enerzijds in het in kaart brengen van factoren uit het sociaal- en onderwijskundig klimaat die door de leerlingen als stressor worden waargenomen, en anderzijds in het relateren van deze factoren aan een tweetal met stress samenhangende responsen (angstdispositie en examenangst). Het is de eerste studie binnen het Nederlandse taalgebied die dit verschijnsel als zodanig bestudeert. Ze mag dan ook van grote waarde worden geacht. Theorieën en onderzoek naar stress worden in dit boek systematisch besproken. In hoeverre de gerapporteerde bevindingen ook gelden voor het Nederlandse onderwijssysteem dient nader onderzocht te worden.

Een tweetal zaken met betrekking tot dit onderzoek wil ik tot slot nader aan de orde stellen. Het eerste heeft betrekking op de gehanteerde vragenlijst, en het tweede op het belang van copinggedrag binnen het stress-proces waaraan door Compennolle helaas weinig aandacht wordt besteed in zijn literatuuroverzicht.

De "Subjectieve Schoolstressoren Vragenlijst" is in wezen een instrument dat peilt naar de mate van aanwezigheid van potentieel bedreigende schoolsituaties. De subjectieve perceptie van leerlingen ten aanzien van schoolsituaties wordt in feite niet gemeten: met andere woorden de mate waarin dergelijke situaties door de leerlingen als een bedreiging voor het welbevinden worden waargenomen blijft een open vraag. Het eerder genoemde item zou dan bij wijze van illustratie als volgt kunnen luiden:

Als de leerkrachten streng zijn vind ik dat

- ☐ helemaal niet vervelend
- ☐ een beetje vervelend
- ☐ wel vervelend
- ☐ heel erg vervelend

Bij een dergelijke formulering van items wordt wel gepeild naar de subjectieve inschatting van schoolsituaties die een potentiële bedreiging voor het welbevinden kunnen vormen (zie Hendriksen en Boekaerts, 1988). Dit zou beter aansluiten bij de veelvuldig geciteerde definitie van stress zoals gegeven door Lazarus en Folkman (1984): "Psychological stress is a particular relationship between the person and the environment, that is appraised by the person as taxing or exceeding his resources and endangering his or her well-being" (p. 19). In de recente (onderzoeks)literatuur naar stress wordt veel aandacht besteed aan het copinggedrag als belangrijke factor van invloed op het welbevinden (o.a. Lazarus en Folkman, 1984). Coping, of omgaan met stress (in het nederlandse taalgebruik bestaat hier geen goed equivalent voor) verwijst naar cognitieve of gedragsmatige pogingen van het individu om stress te reduceren of te controleren. Op basis van recent onderzoek is de overtuiging gegroeid, dat de manier waarop mensen stress hanteren, belangrijker zou zijn dan de stress-situatie zelf. Dit gedrag bepaalt immers de gevolgen van stress-situaties voor het welbevinden.

REFERENTIES

- Hendriksen, J. en Boekaerts, M. (1988). Stress-situaties op school: een onderzoek naar alledaagse schoolsituaties die door leerlingen als belastend worden waargenomen. *Tijdschrift voor Onderwijsresearch*, 13, (3), 161-171.
- Lazarus, R.S. & Folkman, S. (1984) *Stress, Appraisal and Coping*. Springer Publishing Company, New York.
- Moos, R.H. (1979) *Evaluating Educational Environments*. San Francisco, Jossey Bass.
- Oudshoorn, S. (1987) Een gezonde prestatiedruk op school bevordert het welzijn en de resultaten. *Didaktief*, 17 (8), 9-11.

J. Hendriksen
 Facultaire vakgroep ISO/Onderwijspsychologie
 Katholieke Universiteit Nijmegen.

Akker, J.J.H. van den (1988)
 Ontwerp en implementatie van natuuronderwijs.
Amsterdam/Lisse: Swets en Zeitlinger (proefschrift)

Over de ontwikkeling en implementatie van curricula bestaan uiteenlopende opvattingen. Vooral over de mate waarin het curriculum het handelen van de leerkracht kan voorstructuren, lopen de meningen uiteen. Volgens de ene opvatting dienen schoolwerkplan en leerboeken duidelijk gestructureerd en uitgewerkt te zijn zodat ze gemakkelijk gevolgd kunnen worden door de leerkrachten in de voorbereiding en uitvoering van het onderwijs. Anderen menen dat het curriculum de leerkracht vooral ideeën aan de hand moet doen, aan de hand waarvan hij zelf het onderwijs kan plannen en uitvoeren. Deze laatste opvatting sluit goed aan bij de statutaire opdracht van de Stichting voor Leerplanontwikkeling. De SLO stelt dat zij geen concrete schoolwerkplannen over onderwijsleerpakketten mag maken. Bovendien is de opvatting van de SLO dat deze halffabricaten (ideeën, technieken, voorbeelden) makkelijker door leerkrachten zullen worden overgenomen dan concreet uitgewerkte onderwijsleerpakketten. Immers, leerkrachten zouden niet bereid zijn dergelijke concrete aanwijzingen op te volgen. Hoewel al enige empirische evidentie aanwezig is om op z'n minst deze opvatting in twijfel te trekken, heeft een vergelijking tussen een SLO-variant en een meer gestructureerde variant op een zelfde scholingsgebied niet plaatsgevonden. Van den Akker is in een veldexperiment bij natuuronderwijs in de basisschool nagegaan welke kenmerken van het onderwijsleerpakket bevorderend werken op het gebruik ervan door leerkrachten in de eerste fase na de invoering. Een analyse van het SLO-materiaal leert dat de rol van de leerkracht nog weinig helder en concreet is

uitgewerkt. Er zijn dan ook een aantal kernproblemen bij het functioneren van leerkrachten, zoals het moeizaam wijzigen van de rol van de leerkracht, onzekerheden en gebreken bij leerkrachten ten aanzien van de gewenste vakinhoudelijke achtergrond, gecompliceerde en tijdrovende lesvoorbereiding en onvoldoende realisering van effecten bij leerlingen. Op grond van deze resultaten en empirisch onderzoek naar de relatie tussen curriculumkenmerken en het gebruik van materiaal door leerkrachten komt Van den Akker tot de conclusie dat het curriculummateriaal slechts dan een effectieve functie kan vervullen, als het gekenmerkt wordt door een grote mate van eenduidigheid, structurering en detaillering, met name wanneer het veel procedurele specificaties bevat, dat wil zeggen zo nauwkeurig mogelijke aanwijzingen voor het handelen van de leerkracht geeft. Deze aanbeveling is omgezet in een omvangrijke lijst van ontwerp-(ontwikkelings)specificaties. Op basis hiervan is curriculummateriaal ontwikkeld voor één les. In een veldexperiment is vervolgens het SLO-materiaal en dit experimentele materiaal met elkaar vergeleken. De effecten werden vanuit het perspectief van zowel de ontwikkelaars als de leerkrachten beschouwd. Bij het perspectief van de ontwikkelaars gaat het om de mate waarin de lesuitvoering (vooral van leerkrachtengedrag) in overeenstemming is met de oorspronkelijke bedoelingen van de SLO-ontwikkelaars. Dit heeft plaatsgevonden via klasse-observatie door middel van videoregistratie en handnotities. Bij het perspectief van de leerkracht gaat het om de wijze waarop de leerkracht na afloop van de les reflectie pleegt op het lesvoorstel, lesverloop en zijn eigen rol daarin. Dit is gemeten via een gestructureerd interview, aangevuld met een korte schriftelijke vragenlijst. Ten aanzien van de eerste onderzoeksvraag, blijkt dat de experimentele variant in alle opzichten superieur is aan het oorspronkelijke SLO-materiaal. Opvallend in dit verband is dat wat de tijdsbesteding betreft, de experimentele variant langer duurt dan de SLO-variant, die minder tijd dan gesuggereerd aan het lesthema besteedt. Nog belangrijker is echter dat de lesonderdelen waarvan verwacht mag worden dat ze van belang zijn voor leerlingeffecten, duidelijk in het voordeel van de experimentele variant uitvallen. De SLO-variant blijkt een opvallend hoog aantal niet-taakgerichte activiteiten te omvatten. Wat het perspectief van de leerkracht betreft, blijkt dat het SLO-materiaal aanleiding geeft tot veel logistiek-organisatorische activiteiten, terwijl het experimentele materiaal aanleiding geeft tot inhoudelijke coördinatie en planning van het onderwijsleerproces. Leerkrachten blijken het materiaal complementair te beoordelen: wat bij de ene materiaalversie als sterk punt wordt beschouwd, wordt bij de andere materiaalversie juist als gemis of negatief beoordeeld. De SLO-versie heeft veel mogelijkheden voor keuzes en eigen inbreng, maar weinig houvast voor de leerkracht. De experimentele versie geeft veel houvast maar weinig keuzemogelijkheden. Wat de voorkeur voor de versies betreft, blijken toch nogal wat leerkrachten van versie te wisselen.

Conclusie: het lijkt dan ook terecht dat gestructureerd materiaal meer en duidelijker de intenties van ontwikkelaars kan weerspiegelen en dat er geen grote weerstand van leerkrachten tegen sterke structurering van lesmateriaal bestaat. Wat de ontwikkeling van curricula betreft, is het onderzoek een pleidooi voor het verschuiven van de strategie van de SLO, die nu vooral gericht is op schoolwerkplannen, model-leerplannen e.d., naar het micro-niveau: de ontwikkeling van onderwijsleerpakketten met voorbeeld lesmateriaal conform de in het experimentele materiaal vervatte ontwerp-specificaties. De resultaten van het onderzoek komen overeen met die van een aantal andere implementatiestudies en de conclusies sporen met een aantal aanbevelingen die in het kader van de evaluatie van het werk van SLO in het verleden door anderen zijn gemaakt. Belangrijk is de empirische ondersteuning via dit helder geschreven, duidelijke proefschrift. Op de wijze waarop deze empirische ondersteuning heeft plaatsgevonden, is echter wel kritiek uit te oefenen.

1. Het is de vraag of met één les van twee uren wel zo veel gezegd kan worden over effecten van een curriculum of methode. Zijn leerkrachten wel bereid om gedurende langere tijd met voorgestructureerd curriculummateriaal te werken? Curriculummateriaal waarvan als zwak punt wordt opgemerkt dat het een lange lestijd vergt.

2. De experimentele versie kost meer tijd. Dat wordt overigens als een positief punt, in relatie tot vermoedelijke leerresultaten van leerlingen, naar voren gehaald. Gesuggereerd is om voor beide versies ongeveer twee uren te besteden, maar de aanwijzingen in de experimentele versie waren zodanig dat ze uitdrukkelijk twee uren impliceerden. Gezien de resultaten van andere curriculumprojecten, bijvoorbeeld het project Onderwijs Sociaal Milieu, blijft de vraag of lestijd, voorgestructureerd door het curriculummateriaal een voldoende indicatie is voor leerresultaten. Daarom is meting van leerresultaten noodzakelijk en is deze naar mijn mening ten onrechte in dit onderzoek achterwege gelaten.

3. Een belangrijk punt van kritiek betreft de onderzoeksmethode. Die is onvoldoende onpartijdig. Allerlei belangrijke keuzes en beoordelingen van leerkrachtgedrag, observaties en ontwikkeling van de schalen, zijn allemaal gemaakt door de onderzoekers die een duidelijk belang hadden bij de experimentele versies. Hoe graag ik ook de resultaten en de conclusies van het onderzoek zou willen onderschrijven, het is de vraag of de veldstudie wel helemaal fair is opgezet en uitgevoerd.

Bert P.M. Creemers
RION
Postbus 1286
9701 BG Groningen

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 OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 12, nr. 4

September 1988

Inhoudsopgave

VOR-Bulletin, Jrg 12, nr. 4

Oproep Prof. Dr. H.C.J. Duijkerfonds	240
Congressen	241
Publicaties	244

Oproep Prof. Dr. H.C.J. Duijkerfonds

Het Prof. Duijkerfonds is door de Universiteit van Amsterdam ingesteld om door het doen van uitkeringen de beoefening van de toegepaste psychologie te bevorderen.

Voor het jaar 1988-1989 is een bedrag in de orde van f 25.000,- beschikbaar ten behoeve van één of meer onderzoeksprojecten. De bedoeling is kwalitatief goed onderzoek mogelijk te maken, dat anders niet gerealiseerd zou kunnen worden omdat de middelen ontbreken.

Bij voorkeur komen voor subsidiëring in aanmerking de kosten van een niet omvangrijk onderzoek of deelonderzoek. Ook is te denken aan kosten van een voorstudie of vooronderzoek, danwel aan de uitwerking van een onderzoeksopzet.

Bij gegeven beperkte ruimte komen bijvoorbeeld ook in aanmerking de kosten van inschakeling van tijdelijke hulpkrachten bij onderzoek en/of de aanschafkosten van onmisbare apparatuur, literatuur enz.

Voorwaarden

1. Het voorstel moet liggen op het terrein van de toegepaste psychologie.
2. Aanvragen moeten gemotiveerd worden. Vermeld dient te worden wat de doelstelling van het project is, de opzet en de wijze van uitvoering, de duur, de wijze van verslaglegging, de medewerkers en degene die verantwoordelijk is voor de uitvoering, en voorts een nauwkeurige begroting van de kosten.
3. Aanvragen zullen door de Adviescommissie van het Prof. Duijkerfonds beoordeeld worden op:
 - de wetenschappelijke kwaliteit van de opzet
 - de maatschappelijke relevantie van de opzet
 - realiseerbaarheid van het project
 - het aangevraagde bedrag (eventueel in het verband van de totale projectbegroting).

De Adviescommissie wordt gevormd door de leden van het bestuur van de Stichting Instituut voor Sociale en Bedrijfspsychologie aan de Universiteit van Amsterdam (ISBP), aangevuld met een lid van de Wetenschapscommissie van de Subfaculteit Psychologie.

4. Subsidie-aanvragen moeten worden ingediend op een aanvraagformulier, dat door het secretariaat van het ISBP op verzoek wordt toegestuurd.

Aanvragen moeten **vóór 1 oktober 1988 in 15-voud** worden gericht aan:

De Adviescommissie van het Prof. Duijkerfonds

t.a.v. Dr. W.H.C. Kerkhoff

p/a Stichting ISBP

Weesperplein 8

1012 XM Amsterdam

(tel. 020-525.3773)

Congressen

INTERNATIONAL CONGRES FOR SCHOOL EFFECTIVENESS 1989

JANUARY 4TH - JANUARY 6TH, 1989
ROTTERDAM.

The purpose of the 1989 congress is (as it was in London) to bring together researchers, policy makers and practitioners in order to exchange information and to develop new insights on the important issue of the relation between pupil achievement and the organisation of schools and classrooms.

Furthermore, it also happens that ICSE '89 will coincide with the conclusion (December 31, 1988) of the evaluation of the largest project fighting educational deprivation in The Netherlands: the "Project O.S.M." (the Educational and Social Environment Project).

We hope to be able to present this project's concluding report at the ICSE '89.

The Programme/Paper Committee invites practitioners, researchers and the policy making community to submit proposals for papers on effective schools. Topics on which papers are especially welcome are:

- research on effective schools and effective instruction (e.g. factors related to effective schools and instruction, or research problems);
- development of effective schools (e.g. development of programmes, teacher training);
- experiences in school improvement (implementation strategies);
- applications of results in policy-making with regard to effective schools, school improvement;
- evaluation of projects in effective school development (e.g. description of large scale programmes in effective schools).

As noted before, the programme will consist primarily of papers. Paper sessions are generally scheduled for a 1½ hour period, containing 3 paper presentations, a prepared critique and an audience discussion. Individual paper presentations will therefore be of approximately 20 minutes.

Paper proposals should

- not exceed three pages (A 4-format);
- be written in the English language (this goes also for the abstracts);
- be accompanied by an abstract of about 100 words;
- include whether the paper is about research, practice or policy-making on effective schools and a clear indication of the specific topic the paper will be about;
- include the author's name, address, affiliation and telephone number;
- be received before **September 15, 1988**.

Notifications of acceptance will be sent before October 15. Final versions of papers should be received before December 1.

The paper proposals should be sent to the ICSE secretariat. Further distribution to the members of the Programme/Paper Committee will be taken care of by the ICSE secretariat.

HET INTERUNIVERSITAIR CENTRUM VOOR ONDERWIJSEVALUATIE

In het Interuniversitair Centrum voor Onderwijsevaluatie (ICO) werken evaluatiedeskundigen en onderwijsonderzoekers van de universiteiten van Groningen en Twente samen om op het gebied van onderwijsevaluatie hoogwaardig onderzoek te bevorderen en een post-doctorale opleiding in te richten. Om deze doelstellingen te verwezenlijken, richt het ICO een netwerk in voor geconcentreerde a.i.o.-begeleiding, met een gestructureerd opleidingsprogramma en worden regelmatig conferenties en symposia georganiseerd voor een breder publiek.

Het opleidingsprogramma is niet alleen bedoeld voor a.i.o.'s die participeren in het netwerk, maar tevens voor afgestudeerde onderzoekers die zich willen ontwikkelen tot vakbekwaam evaluatie-onderzoeker, en voor al diegenen die vanuit hun beroep te maken hebben met evaluatie van het onderwijs (evaluatie van beleidsprogramma's, leerplanevaluatie, toetsontwikkeling, evaluatie van vernieuwingsprogramma's, onderwijsassessment).

Onderwijsevaluatie wint aan betekenis

Binnen het Nederlands onderwijs wint evaluatie aan betekenis. Dat blijkt uit op stapel staande programma-evaluaties, zoals de evaluatie van het onderwijsvoorrangsbeleid, de evaluatie van de invoering van de basisvorming in de eerste fase van het voortgezet onderwijs, de evaluatie van het informaticastimuleringsplan, de evaluatie van de onderwijsverzorging. Dat blijkt ook uit de wat langere traditie van leerplanevaluatie en evaluatie van onderwijsvernieuwingen.

Daarnaast worden nieuwe activiteiten ontwikkeld die niet als evaluatie-onderzoek in strikte zin kunnen worden benoemd, maar wel duidelijk evaluatiedeskundigheid gebruiken. Gewezen kan worden op het periodiek peilingsonderzoek, systemen voor zelfevaluatie van scholen en onderwijsinstellingen, onderzoek naar schoolorganisatorische effectiviteit en toetsservice-systemen.

In het ICO samenwerkende onderzoekers nemen een vooraanstaande plaats in bij veel van deze activiteiten en zijn betrokken bij de uitvoering van op gang gebracht evaluatie-onderzoek.

Evaluatie en wetenschap

Wetenschappelijke redenen voor een expertisecentrum op het gebied van onderwijsevaluatie kunnen worden gevonden in de bijzondere wijze van "toetsing" van "wat werkt" in het onderwijs, die in de genoemde evaluatieprogramma's aan de orde komt. Daarbij kan worden gedacht aan fundamentele evaluatievraagstukken, zoals het vaststellen van evaluatiecriteria en het bereiken van overeenstemming over evaluatie-standaarden of -maatstaven. Ook het ontwerpen van evaluatiedesigns die recht doen aan belangen van betrokkenen en aan wetenschappelijke criteria, en de benutting van (in theorieën) beschikbare kennis bij ex-ante en ex-post evaluatie-onderzoek krijgen aandacht. De beïnvloeding van evaluatie-onderzoek door de politieke en organisatorische context is daarbij van belang. Methodologische ontwikkelingen worden in verband gebracht met ontwikkelingen in de theorievorming over het onderwijs, met name met factoren die effectiviteit van het onderwijs beïnvloeden.

Evaluatie: om de kwaliteit van het onderwijs

Onder bovenstaande titel zal op vrijdag 30 september 1988 een symposium worden georganiseerd bij gelegenheid van de officiële opening van het ICO. Tijdens het symposium zal aandacht worden besteed aan problemen van standaarden van onderwijsevaluatie, met name gerelateerd aan leertoetsen, aan het gebruik van beschikbare empirische kennis bij evaluatie ex-ante en ex-post van onderwijsprogramma's, en aan de voor kwaliteit en benutting van evaluatie-onderzoek vaak ongunstige invloeden van de politieke en organisatorische context. Als het programma en de namen van de gastsprekers definitief bekend zijn, zal nog uitgebreid aandacht worden geschonken aan dit symposium.

Inlichtingen

Prof. dr. J. Scheerens (voorzitter),
Faculteit Toegepaste Onderwijskunde van de Universiteit Twente,
Postbus 217,
7500 AE Enschede,
tel. 053-893597

Dr. W.Th.J.G. Hoebe (secretaris),
RION, Instituut voor Onderwijsonderzoek van de Rijksuniversiteit Groningen,
Postbus 1286,
9701 BG Groningen
tel. 050-636645

Publicaties

- Franssen, H.A.M., Jager, A. & Rinkema-Kohlman, V.W. (1987) Denken over doelen, een literatuurstudie naar het inventariseren, formuleren en classificeren in onderwijsdoelstellingen als bijdrage aan de eindtermendiscussie. Utrecht: ISOR, Afdeling Onderwijs-onderzoek
- Van de Par, I. (1988). Stages in Zeeland. Goes: Contactcentrum Onderwijs-Arbeid.
- Koetsier, C.P. & Ten Brinke, J. S. (1988). Experiment individuele eindstage-1 op de afdeling didactiek van de biologie: analyse in het rondetafelgesprek met instituutsbegeleiders. Utrecht: Afd. Didaktiek in de Biologie, R.U.U.
- Koetsier, C.P. & Ten Brinke, J.S. (1988). Analyse in de interviews met (con)rectoren. Utrecht: Afd. Didaktiek in de Biologie, R.U.U.
- DOZ (1988). Praktische handleiding voor efficient onderwijs. Den Haag: De Som/DOZ.
- Tunissen, C. & Knapen, M. (1988) Begeleiding van lichamelijke gehandicapten in het voortgezet onderwijs. Nijmegen: ITS.
- EBN (1988) Europa: basis om te beginnen. Rapport van een werkgroep ad hoc van de Europese Beweging in Nederland over het Europese perspectief van de basisvorming in het Nederlandse onderwijs.

Reflecteren in systematische samenspraak

J.M.C. Nelissen

S.A.C. Utrecht:

ABSTRACT

This article is a report of a research on the effects of a new mathematics curriculum on problem solving and reflection of children. First the protocol method which is used in the present study, called 'systematic dialogue', is discussed.

The results of a first (1983) and a second (1986) experiment show that there is only a slight difference between the experimental group and the control group in problem solving under the condition that the children in both groups were assisted by the experimenter. However, the differences between both groups are significant regarding:

1. speed and accuracy of performance and
2. flexibility of thinking and capacity for strategy change.

The reflection of the experimental group was found to be superior to that of the control group. These findings were attributed mainly to the differences between the mathematics curricula used in the schools.

1. INLEIDING

Door middel van een aan het Schooladviescentrum (S.A.C.) te Utrecht ontwikkeld reken/wiskundeprogramma voor de basisschool werd getracht het reflecteren en het probleemoplossen van de kinderen op het gebied van de wiskunde te stimuleren. In een tweetal onderzoeken is nagegaan of, en zo ja, in welke mate dit streven succesvol is geweest. Voor een uitvoerige analyse van de theoretische achtergronden van het ontwikkelde programma en voor een concrete beschrijving van de inhoud, de bedoeling en de opbouw van het programma wordt verwezen naar Nelissen (1987).

Reflectie wordt gezien als een menselijke mogelijkheid (functie) het eigen handelen te plannen, te overwegen, bij te sturen en te veranderen. Het reflecteren wordt niet beschouwd als een algemeen, maar als een vak- of domeinspecifiek handelen, een handelen dat nauw verbonden is met een vakgebied. Dat wil zeggen dat het reflecteren alleen tot stand kan komen als een persoon begrippen en procedures tot zijn beschikking heeft die karakteristiek zijn voor een bepaald vakgebied. De gedachte achter deze zienswijze is als volgt.

Indien een persoon niet of onvoldoende over begrippen en procedures beschikt, is het voor die persoon ook niet goed mogelijk over het hanteren van die begrippen en procedures te reflecteren. In de onderhavige studie wordt er een nauw verband gezien tussen reflecteren en wiskundig denken of mathematiseren. Kenmerkend voor het (leren) mathematiseren zijn eigen constructies, dat wil zeggen dat de (lerende) persoon eigen vondsten en procedures ontwerpt. Deze moeten vervolgens op hun bruikbaarheid worden getoetst: is deze vondst (constructie) bruikbaar en efficiënt?

Met andere woorden, constructie impliceert reflectie. En omgekeerd kan reflectie leiden tot nieuwe en betere constructies op hoger niveau. Reflectie ontstaat of wordt opgeroepen wanneer zich onverwachte problemen voordoen: wat is dit voor probleem, kan ik het oplossen, welke strategieën of procedures zou ik kunnen gebruiken?

In deze studie is het reflecteren onderzocht in de context van het oplossen van wiskundige problemen.

2. ONDERZOEKMETHODE EN ONDERZOEKINSTRUMENT

Welke onderzoeksmethode is nu het meest geëigend om inzicht te krijgen in dit constructief handelen en reflecteren?

Volgens veel onderzoekers vraagt onderzoek op dit terrein om een kwalitatieve benadering waarbij gebruik wordt gemaakt van technieken als interview, introspectie of protocolanalyse (De Corte en Verschaffel 1983, 1985, De Klerk 1982, Breuker 1982, Knoers 1982, Van Parreren c.s. 1983, Ginsburg 1981, 1983, Ericsson en Simon 1984, Kruteckij 1976, Meichenbaum e.a. 1985, Borkowski 1985, Kirby 1984, Stepanov en Semenov 1985, e.a.).

Het gebruik van kwalitatieve methoden is evenwel niet geheel zonder bezwaren. Er kunnen dan ook in de onderzoekliteratuur suggesties gevonden worden om de kwaliteit van het gebruik van zulke methoden te verhogen (De Corte en Verschaffel 1985, Kruteckij 1986, Ericsson en Simon 1984). We gaan op deze kwestie, vanwege ruimtegebrek, nu niet in.

De onderzoeksmethode waarvoor werd gekozen, is een protocolmethode. Het onderzoeksinstrument dat werd ontwikkeld, is gebaseerd op het idee (zie Borkowski 1985, Forrest-Pressley en Gillier 1983, Kilpatrick 1985, Stepanov en Semenov 1985) dat veranderen van strategie, volgend op een conflict, inclusief de motivering van die verandering, als een belangrijke indicatie voor reflectie kan gelden. Het onderzoeksinstrument, genoemd 'Systematische samenspraak', is in vier achtereenvolgende versies ontwikkeld. De eerste versie bestond uit 41 items, de tweede en derde uit 31 resp., 14 items en de definitieve versie uit 8 items. Deze toetsversies zijn met ruim 70 kinderen uitgeprobeerd.

De kinderen worden tijdens de samenspraak met 'open' wiskundige problemen geconfronteerd. Vijf problemen moeten ze direct oplossen, voor drie problemen moeten ze een adequate aanpak bedenken. Wanneer een kind geconfronteerd wordt met een probleem, is het mogelijk dat het kind dat probleem direct oplost, c.q. dat direct een adequate aanpak wordt bedacht.

De proefleider vraagt in het eerste geval waarom het probleem juist op die manier is opgelost en waar het kind aan heeft gedacht. In het tweede geval wordt gevraagd waarom juist die handelwijze is verkozen (hier is sprake van reflectie vooraf). Het is ook mogelijk dat het probleem zodanig wordt opgelost, dat de oplossing - die op zich niet fout hoeft te zijn - aan te vechten is. De items zijn nl. zo geconstrueerd dat de kinderen geneigd zijn voor de meest in het oog springende (en op zich ook juiste) oplossing c.q. aanpak te kiezen. Indien dit het geval is, worden ze met een dilemma geconfronteerd. 'Wat je zei dat kan natuurlijk, dat is niet fout, maar kijk eens, het is anders. Zou je dit kunnen verklaren?'

Voorbeeld

'Gerard en Jasper hebben allebei een half uur in hun eigen tuin gewerkt. In die tijd heeft Gerard de helft van zijn grasveld gemaaid en in diezelfde tijd heeft Jasper een kwart van het zijne gemaaid. Wie heeft het meeste gedaan?'

'De meeste kinderen zijn spontaan geneigd 'Gerard' te antwoorden. De proefleider reageert in dit geval als volgt: 'Ja zeker, dat kan! Maar weet je wat het geval is, Jasper heeft het meeste gedaan. Hoe zou dat nu kunnen, hoe los je dit probleem op?' De kinderen worden voor een zogenoemd 'socio-cognitief' conflict geplaatst. Deze term vraagt om nadere uitleg. Doise en Mugny (1984) onderscheiden twee conflicttypen. Er is sprake van een individueel conflict indien niet gebeurt wat een persoon had verwacht; zijn handelen levert een onvermoed of géén resultaat op.

Het tweede type conflict speelt zich af in de communicatie tijdens een sociale interactie. Een persoon wordt door een ander met een dilemma of probleem geconfronteerd. De persoon ontdekt dat er een ander gezichtspunt, een andere oplossing in het geding is. Er is daarom sprake van een sociaal conflict en vervolgens van een cognitief conflict.

Nadat het kind voor een (socio-cognitief) conflict is geplaatst, zijn er twee mogelijkheden. De eerste is dat het kind direct van strategie verandert en het probleem transformeert in (zie voorbeeld) een verhoudingsprobleem. De proefleider vraagt nu wat het kind in eerste instantie heeft gedaan en gedacht en wat het verschil is met wat het nú denkt. Dit is een essentieel moment tijdens het onderzoek: het veranderen van strategie en het overwegen van twee oplossingsmethoden in relatie tot elkaar.

Een tweede mogelijkheid is dat het kind niet verandert, maar vasthoudt aan de eerst gekozen oplossing (bij de items 2,4,5,6 en 8) of aanpak (bij de items 1,3 en 7). In dit geval biedt de proefleider *hulp*; deze hulp is standaard en dus voor elk kind hetzelfde.

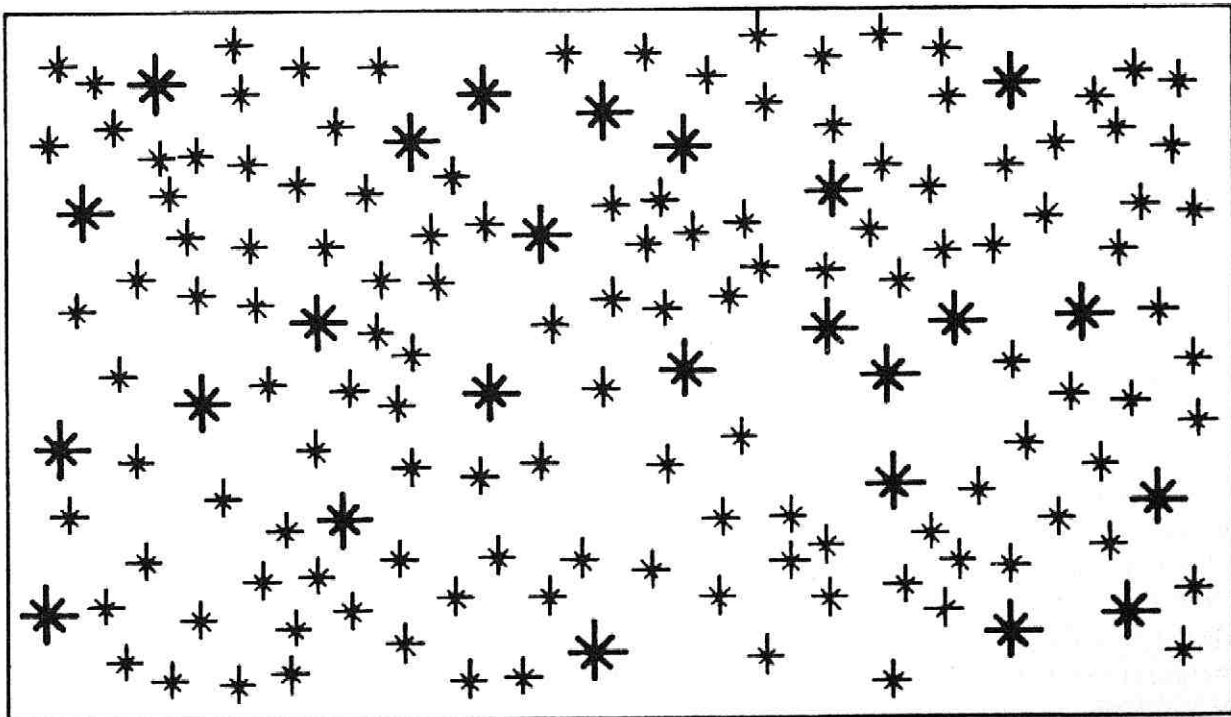
Er doen zich nu weer twee mogelijkheden voor: òf het kind verandert van strategie, kiest een adequate strategie en dan loopt de procedure zoals boven beschreven. Of het kind houdt nog vast aan zijn eigen oplossing of strategie, het kiest een andere maar inadequate strategie, het vergeet de strategie, het gaat gokken of raden. In dit geval biedt de proefleider weer de hulp die voorgeschreven is en zo gaat het verder. Voor elk item is een afbreekregel bepaald, waarna het volgende probleem wordt geïntroduceerd.

Bij de problemen 1, 3 en 7 moeten de kinderen vooraf een adequate aanpak bedenken én motiveren. Op basis van de antwoorden van de kinderen wordt nagegaan of de kinderen een plan van handelen kunnen maken. Wij noemen dit reflectie vooraf.

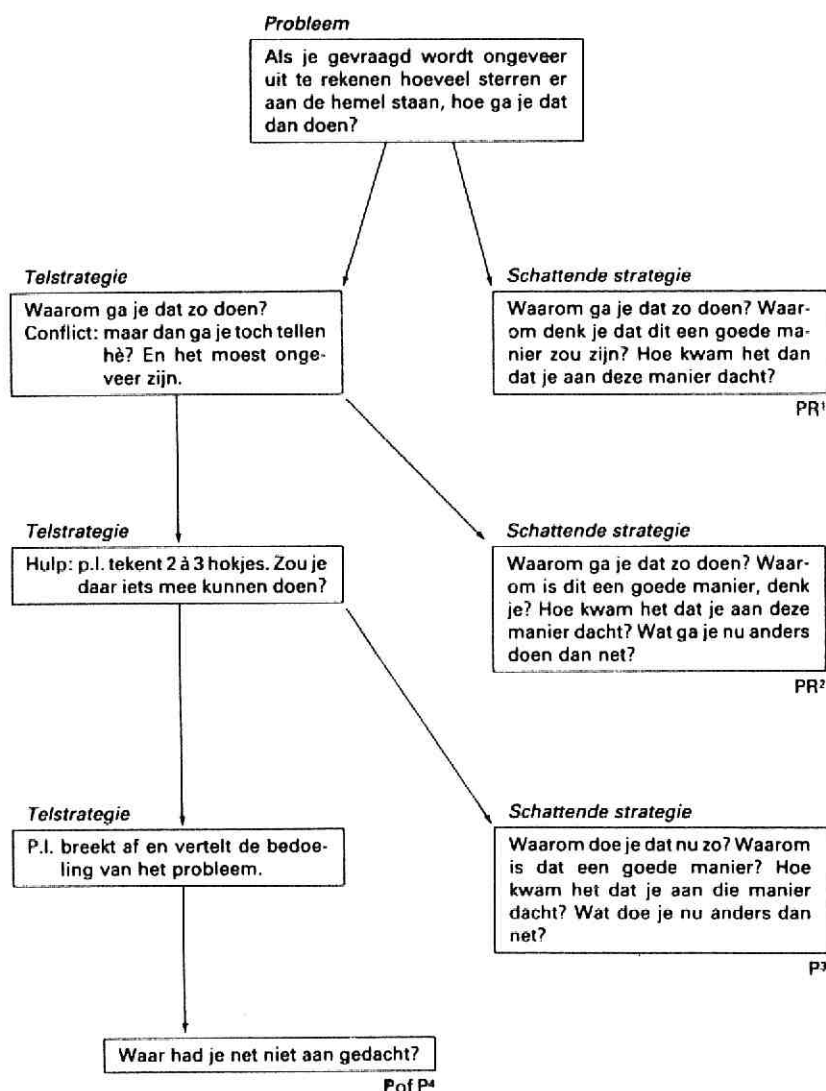
De items 2,4,5,6 en 8 bestaan uit problemen die de kinderen moeten oplossen. Indien een kind direct - dus niet na een conflict - een juiste oplossing voor een probleem vond, wordt het gevraagd de gehanteerde strategie te formuleren en te motiveren. Dit wordt beschouwd als een indicatie van reflectie achteraf. Het is echter mogelijk dat kinderen, die direct een probleem oplossen al tijdens dit proces reflecteerden; zelfs is het mogelijk dat het probleem juist *dóór* te reflecteren opgelost werd. Dit is echter niet onderzocht. Wat onderzocht werd, is of ze achteraf *in staat zijn* tot reflectie. Bij elk item hoort een instructieschema.

Dit schema bevat de regels voor de samenspraak, die tijdens het onderzoek werden gevolgd. Het werd bovendien ná het onderzoek gebruikt om de bandopnamen te beluisteren en te analyseren. Het kan daarom ook opgevat worden als een protocolanalyzeschema. In het protocolanalyse-schema is het theoretisch meest waarschijnlijke verloop van het oplossingsproces gegeven. Het schema is opgesteld op grond van een 'logische analyse'.

Hieronder volgen de 8 items. Alleen bij de eerste en het vierde item geven we, vanwege gebrek aan ruimte, het volledige instructie- c.q. protocolanalyzeschema. De items 3 en 7 verlopen zoals het eerste item en de items 2,5,6 en 8 verlopen zoals het vierde item.



Figuur 1. Item 1



Figuur 2. Toetsinstructies bij item 1

Rob is in 1961 geboren.
Nu is het 1983.
Hoe oud is Rob?

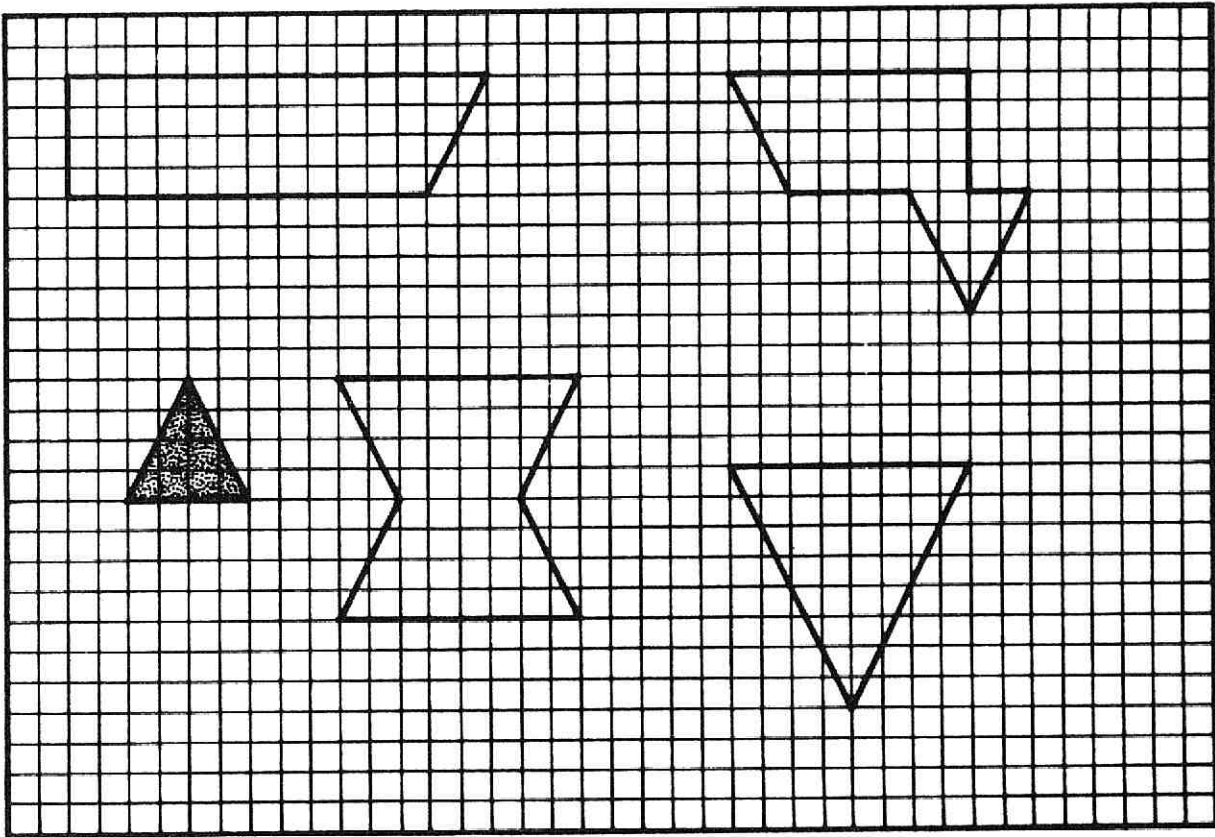
Figuur 3. Item 2

3. INDICATIES VAN REFLECTIEF HANDELEN

In het onderzoek golden verschillende typen handelingen van kinderen als indicaties van reflectie. Er wordt over indicaties van reflectie gesproken, omdat reflectie als zodanig uit niet direct zichtbaar en waarneembaar handelen bestaat, reflecteren is een mentaal handelen. Door middel van bijvoorbeeld verbale handelingen kan reflectie zichtbaar worden gemaakt.

Als *eerste* indicatie van reflectie werd gezien het formuleren én motiveren van een plan van handeling (Doblaev 1984, Kirby 1984, Sternberg 1985, Borkowski 1985).

De kinderen wordt gevraagd (item 1,3 en 7) hoe ze een probleem denken aan te pakken. De proefleider voegt hieraan toe dat ze het probleem dus niet direct hoeven op te lossen, maar dat ze alleen moeten aangeven *hoe* ze het willen oplossen. Als ze een adequate aanpak kunnen



Probleem

Je ziet hier een aantal figuren. Als je moet weten welke figuur het grootst is, hoe ga je dat dan berekenen?

Figuur 4. Item 3

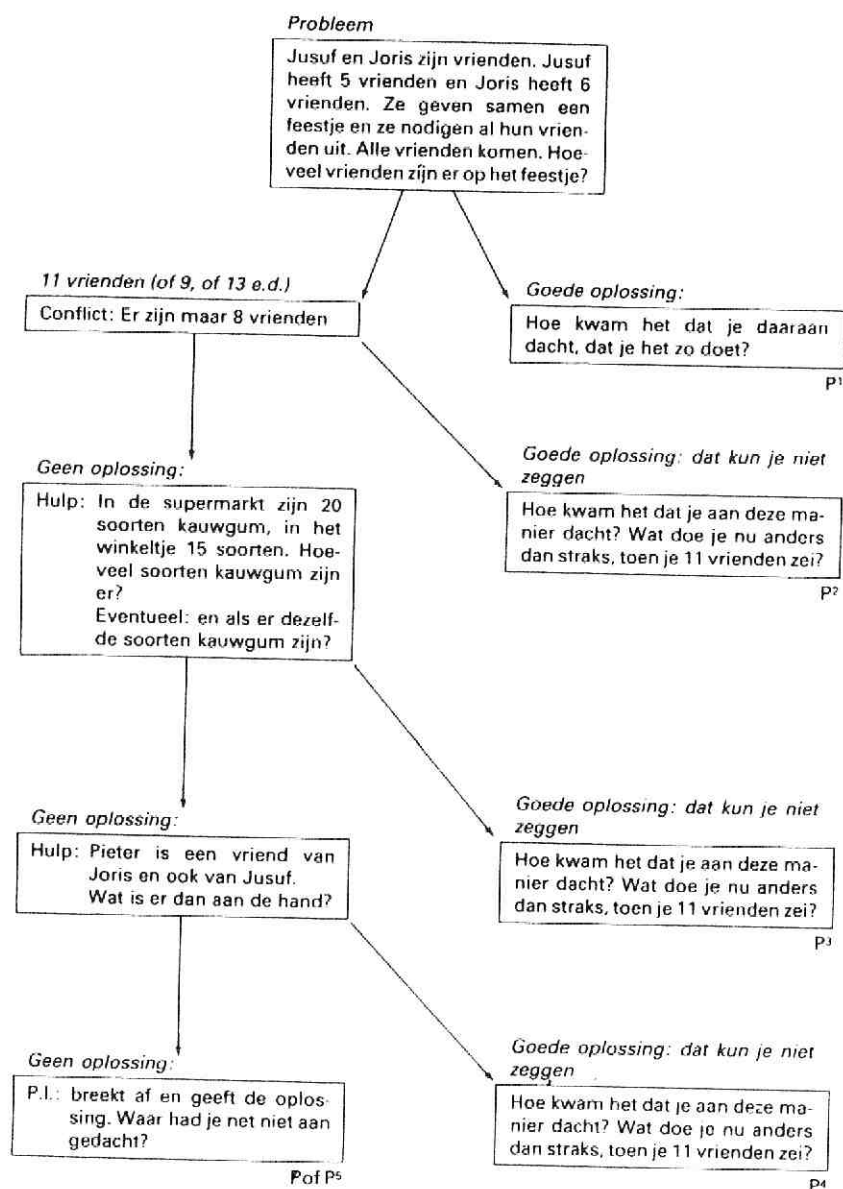
formuleren, vraagt de proefleider *waarom* ze het juist zó willen doen, waarom ze denken dat hun manier een goede manier is (zie de protocolanalyseschema's in §2). De kinderen moeten een relatie kunnen aangeven tussen de gekozen strategie en het op te lossen probleem, anders is er geen sprake van reflectie vooraf. Als indicatie van reflectie geldt dus het formuleren en het adequaat motiveren van de strategie.

Als *tweede* indicatie van reflectie geldt het formuleren en motiveren van een reeds gehanteerde oplossingsstrategie (Polivanova 1978, Ginsburg 1981, Borkowski 1985). Deze wijze van reflecteren kan zich voordoen nadat een kind direct een juiste oplossing voor een probleem vond (item 2,4,5,6 en 8). De proefleider vraagt in dit geval de gevolgde strategie te formuleren en te motiveren. Zoals gezegd, werd niet onderzocht of reflectie leidt tot probleemoplossen maar of de kinderen achteraf *in staat zijn* tot reflectie.

Laten we dit nog nader toelichten.

De proefleider stelt de vraag 'Hoe heb je gedacht? Waarom heb je zo gedacht?' Er zijn nu in principe twee antwoordmogelijkheden. Of een kind overweegt naar aanleiding van deze vraag zijn handelen: reflectie wordt in dit geval door de vraag van de proefleider opgeroepen. Of het kind rapporteert (retrospectief) over het reflecteren dat voorafgaande aan of tijdens het oplossingsproces plaatsvond.

In beide gevallen is sprake van reflectie, al kan de proefleider moeilijk uitmaken of er sprake is



Figuur 5. Toetsinstructies bij item 4

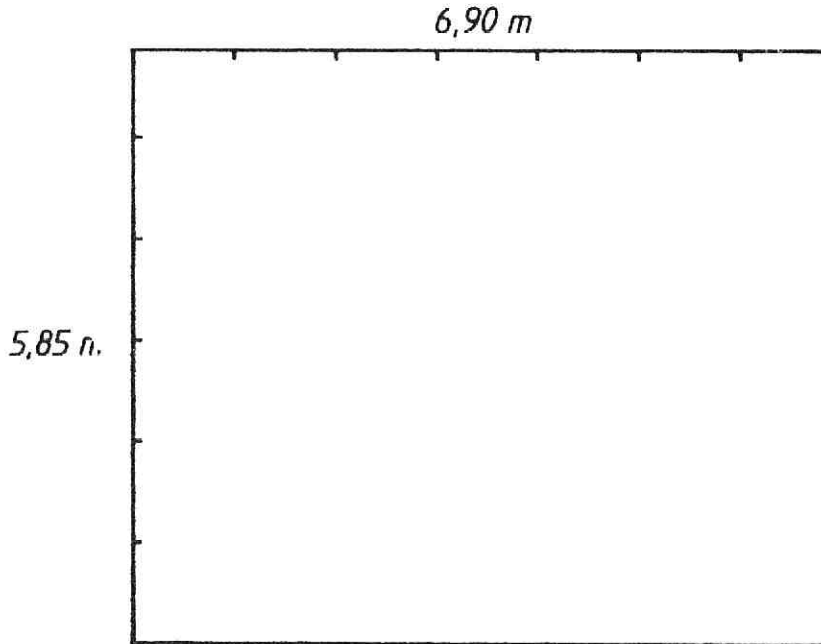
van opgeroepen dan wel spontane reflectie. De afspraak is dat in beide gevallen gesproken wordt van reflectie achteraf.

Als *derde* indicatie van reflectie wordt gezien het veranderen van strategie en het motiveren en vergelijken van twee strategieën. Borkowski (1985) spreekt in dit verband van reflectieve beslissing.

Deze vorm van reflectie kan als volgt gezien worden.

Indien een kind niet direct (vooraf) een juiste - dat wil hier zeggen: niet aan te vechten - *aanpak* voor een probleem vindt (item 1,3 en 7) en indien een kind niet direct en juiste - niet aan te vechten - *oplossing* voor een probleem vindt (item 2,4,5,6 en 8), dan plaatst de proefleider het kind voor een dilemma of conflict. Door dit conflict wordt een koerswijziging opgeroepen.

Dit is de kamer bij jullie thuis. Hij is 6 meter 90 lang en 5 meter 85 breed. Vlak voordat de winkels sluiten, besluiten ze bij jullie thuis dat ze nieuwe vloerbedekking in die kamer willen. Ze vragen jou heel vlug ongeveer uit te rekenen hoeveel vierkante meter vloerbedekking ze nodig hebben. Doe dat eens.



Figuur 6. Item 5

Indien die koerswijziging inderdaad optreedt en het kind gaat over op een andere, adequate handelingswijze (strategie) dan vraagt de proefleider aan het kind de eerste en de tweede handelingswijze te vergelijken. Pas als een kind deze vergelijking kan maken, is er sprake van reflectie

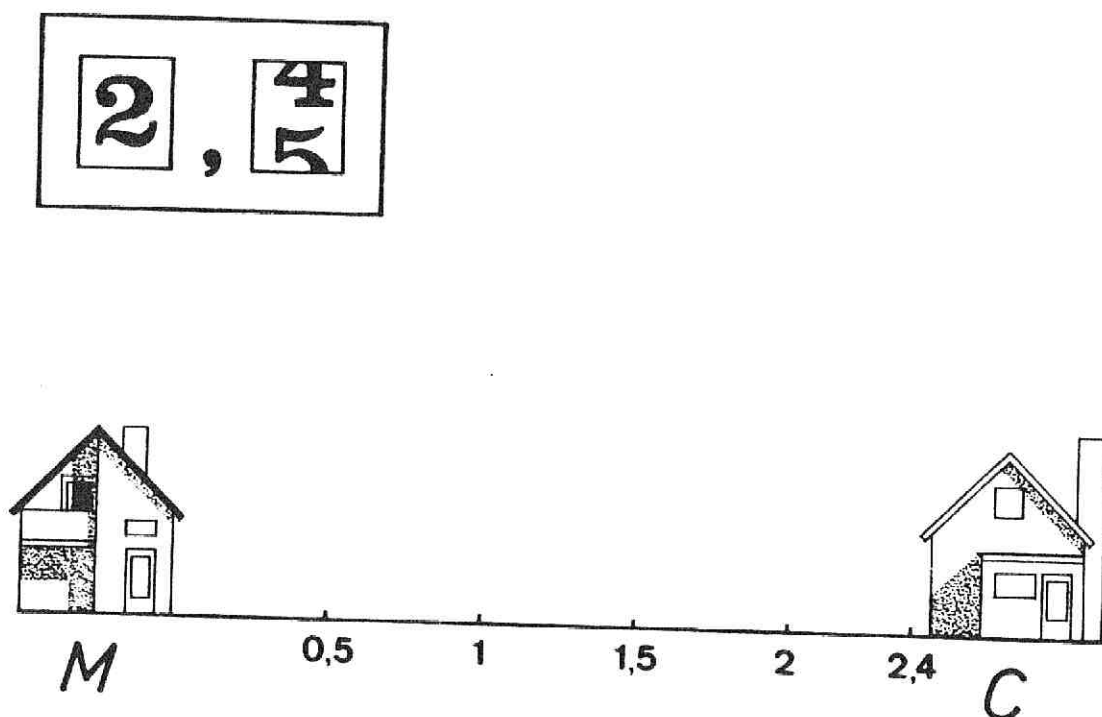
De *vierde* indicatie van reflectie luidt: een kind dat na hulp van de proefleider een probleem uiteindelijk niet oplost, komt alsnog tot beoordeling van de gevolgde handelingswijze en wel nadat de onderzoeker de juiste handelingswijze of de oplossing aan het kind heeft meegedeeld.

De *vijfde* indicatie luidt dat de kinderen een andere handelingsmogelijkheid overwegen. Dit overwegen kan worden opgeroepen door de kinderen te vragen hoe ze als eersteklasser een bepaald probleem aangepakt zouden hebben. Aangezien de kinderen in de eerste klas zo'n probleem nooit hebben gezien, moeten ze nu een beeld, een hypothese vormen van een mogelijke handelingswijze.

De *zesde* indicatie kan omschreven worden als het controleren van het eigen handelen tijdens of ná het oplossingsproces (Polivanova 1978, Matis 1982, Sacharowa 1982, Van Parreren c.s. 1983, Forrest-Pressley en Gillier 1983, Lawson 1984, Kirby 1984).

De *zevende* indicatie kan omschreven worden als het anticiperen op de kritiek of het commentaar van de proefleider, waardoor de keuze voor een strategie wordt beïnvloed. Deze dialogische benadering van reflectie staat centraal in het werk van Vygotskij 1977 (cf. Wertsch c.s. 1980, Matis 1982).

Als *achtste* indicatie geldt dat het kind op de overeenkomsten wijst in structuur tussen verschillende problemen, en inzicht vertoont in de gespreksstructuur. Op grond daarvan kiest of motiveert het kind zijn handelingswijze. Er is in zo'n geval sprake van generalisatie van handelen, van strategieën (Zak 1980, Borkowski 1985).

*Probleem*

Dit is het huis van Marijke en dit is het huis van Carla. Op een woensdagmiddag fietst Marijke van haar eigen huis naar het huis van Carla. Als ze vertrekt, staat haar (fiets)kilometerteller op 0,0 en als ze bij Carla is, staat hij op ... (aarzelend uitgesproken) ... 2,4. Op hoeveel staat de kilometerteller als Marijke weer terug is?

Figuur 7. Item 6

4. REFLECTEREN IN SYSTEMATISCHE SAMENSPRAAK: HET ONDERZOEK

4.1 Opzet van het onderzoek.

Het onderzoek vond plaats in vier proefscholen en drie controlescholen in de stad Utrecht. Na drie jaar volgde een kort herhalingsonderzoek in één proef- en één controleschool.

Het is helaas niet mogelijk om in het bestek van dit artikel het reken-/wiskundeprogramma in de proef- en controle-scholen uitvoerig te omschrijven en daarom volsta ik met enkele typerende voorbeelden en een korte karakterisering.

Het programma dat in de proefscholen wordt gebruikt, is op het schooladviescentrum (S.A.C.) ontwikkeld en bestaat uit handleidingen en materialen voor het onderwijs aan vier tot twaalfjarigen.

De dagelijkse praktijk wordt gekenmerkt door een afwisseling in probleemsituaties, toepassingssituaties en oefensituaties. Omdat het onderscheid in deze drie soorten onderwijssituaties essentieel is, zal ik ze hier kort toelichten.

Probleemsituaties zijn georganiseerd rond een context of thema. Het doel is de kinderen te leren een probleem aan te pakken op eigen niveau. De eigen constructies zijn essentieel, maar ook de gezamenlijke discussie over ieders aanpak.

Je ziet hier een bladzijde van een boek. Deze bladzijde ziet er net zo uit als alle andere bladzijden van dit boek. Het boek heeft honderd bladzijden.

Bovendien werd het zeevisserijbedrijf wat socialer. De harde en zeer harde kanten van het beroep gingen eraf. De arbeidsovereenkomsten werden beter. De voortschrijdende techniek maakte het mogelijk veel handenarbeid te verlichten of te vervangen door mechanische hulpmiddelen. De lonen stegen flink, terwijl ook de verlofregelingen verbeterd werden.

Werken in de zeevisvaart is een uitdaging voor mensen die hoofd en handen weten te gebruiken en niet bang zijn voor extra werk.

3. Waarom in de zeevisvaart?

Waarom wil iemand in de zeevisvaart werken? Dat is een moeilijk te beantwoorden vraag. De tijd is voorbij dat men zei: „Jij deugt nergens voor, ga maar varen bij de visserij.”

Het beroep is vanouds overgegaan van vader op zoon en dat is in veel gevallen nog zo. De kustbevolking van Zeeuws-Vlaanderen tot Schiermonnikoog en Zoutkamp levert sinds onheugelijke tijden de nodige bekwame mensen voor de zeevisvaart. Uit andere delen van het land komen maar weinig mensen, die de zeevisvaart kiezen als hun toekomstig werkterrein. Als er één vak is, dat zo duidelijk door één bepaalde bevolkingsgroep wordt uitgeoefend, dan is dat wel het beroep van visser.

Hoe sterk die aangeboren drang naar de zeevisvaart kan zijn, bewijzen onder andere de inwoners van Urk. Die zijn, ondanks de afsluiting van de Zuiderzee, toch de zeevisserij blijven beoefenen. In het algemeen vormt de vissersbevolking een tamelijk besloten gemeenschap. Voor buitenstaanders is het moeilijk daarbinnen te komen. Daarom zal voorlopig het grootste deel van het varende personeel wel uit die kustbevolking blijven komen.

Opmerkelijk is, dat er maar zelden sprake is van een weloverwogen keuze. In die kringen gaat men gewoon varen in de visserij en vaak ook in de koopvaardij.

Er bestaat onder vissers veel liefde voor het vak en er heerst bij hen een duidelijke beroepstrots. Dat het vak overgaat van vader op zoon is dan ook niet verwonderlijk.

Probleem

Als je gevraagd wordt ongeveer te berekenen hoeveel woorden er in dit boek staan, hoe ga je dat dan doen?

Figuur 8. Item 7

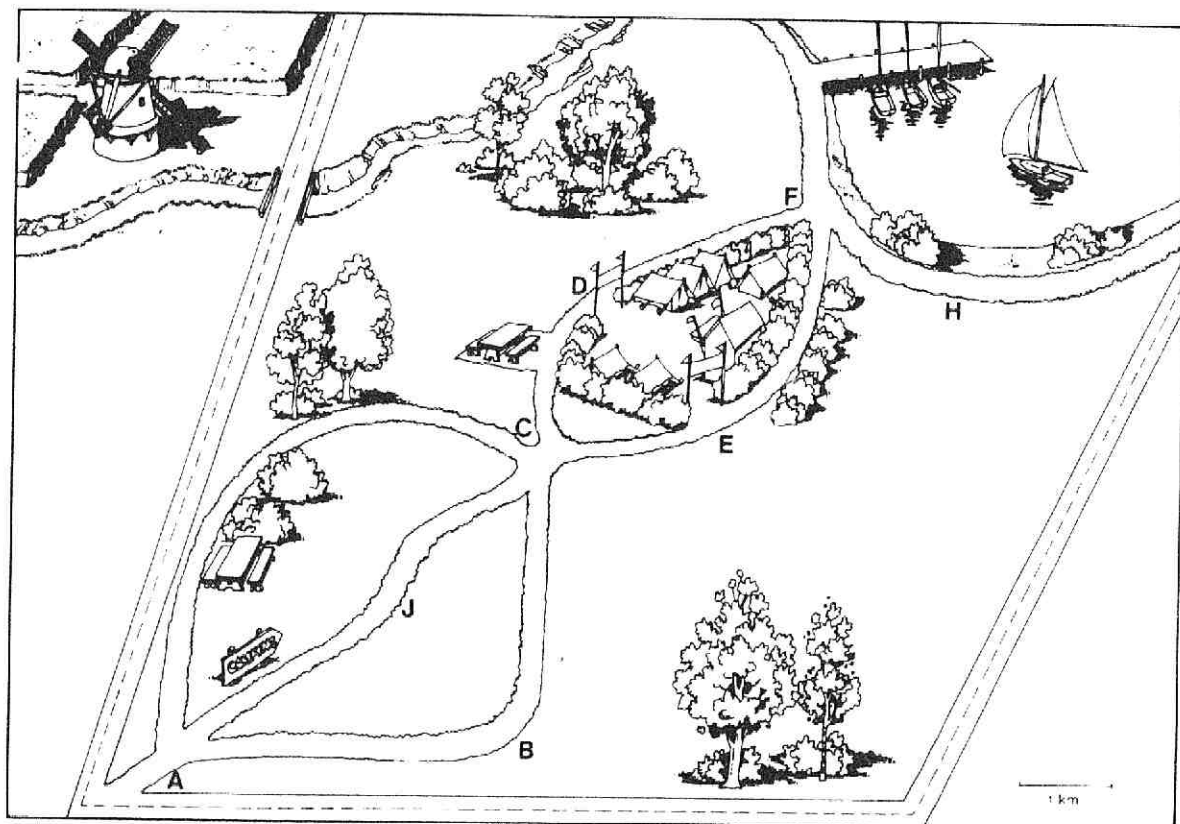
Gerard en Jasper hebben allebei een half uur in hun eigen tuin gewerkt. In die tijd heeft Gerard de helft van zijn grasveld gemaaid en in diezelfde tijd heeft Jasper een kwart van het zijne gemaaid. Wie heeft het meest gedaan?

Figuur 9. Item 8

Dit moet leiden tot reflectie over de eigen aanpak waardoor verhoging van het niveau van probleemoplossen kan worden bereikt.

In de toepassingssituatie gaat het erom de verworven kennis verder uit te bouwen en te verdiepen door deze in gevarieerde contexten te gebruiken. Toepassen moet niet opgevat worden als het oefenen en het inslijpen van vaardigheden door het maken van rijtjes sommen, maar als het beoefenen van een aanpak, een denkwijze, een oplossingsprocedure in telkens nieuwe contexten. In de oefensituatie gaat het wél om oefenen, maar dit is geen 'blind' oefenen. Er is steeds aandacht voor de aanpak, voor het gebruik van een inzichtelijke rekenregel; bijvoorbeeld: $9 \times 7 = (10 \times 7) - 7$ of $29 + 9 = 29 + 10 - 1$.

De volgende situatie is een probleemsituatie uit Rekenwerk.



Figuur 10. Een 'probleemsituatie' uit Rekenwerk.

De volgende situatie is een probleemsituatie uit Rekenwerk.

Wat is de bedoeling van deze onderwijssituatie?

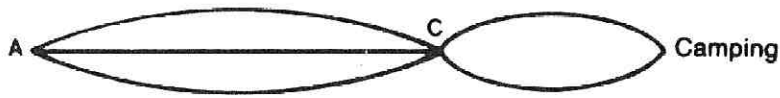
De leraar introduceert het werkblad: een groep wandelaars houdt een trektocht te voet van enkele dagen. Ze bekijken samen de kaart en zoeken uit hoe ze naar de camping moeten lopen, hoeveel routes er zijn en welke de kortste route is. De eerste vraag luidt dus om na te gaan hoeveel mogelijke routes er zijn.

De tweede vraag luidt: hoe bepaal je welke route het kortste is.

Hoeveel routes zijn er? Iemand zegt zes, een ander vijf, wie heeft gelijk? De kinderen moeten - eventueel in tweetallen - zonder hulp dit probleem maar eens proberen op te lossen. Daarna volgt een klassikale bespreking van de verschillende aanpakken. De leraar wordt aanbevolen het volgende schema te introduceren:

Hoeveel routes zijn er van A naar C?

Bij elke van de drie routes, horen weer twee routes.



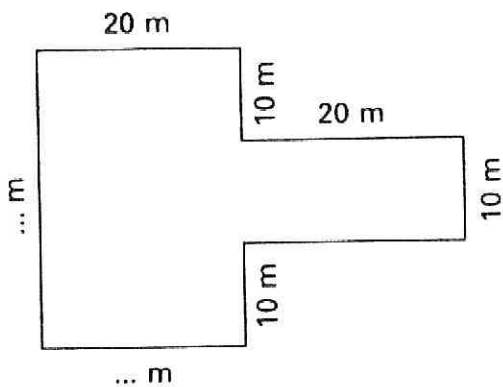
Figuur 11. Een combinatieprobleem uit Rekenwerk.

In de controlescholen werden traditionele ('mechanistische') programma's gebruikt.

- 16** Teken deze figuur in je schrift.
Vul de ontbrekende maten in.

Bij omtrek denk ik aan een touw.

Wijs de omtrek van deze figuur eens aan.



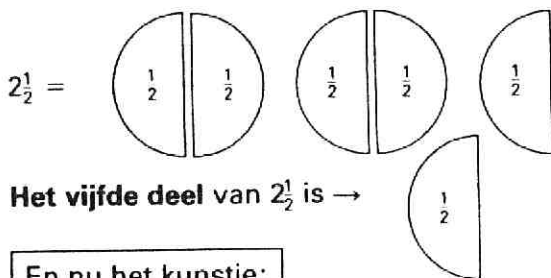
De omtrek is

$$..m + ..m + ..m + ..m + ..m + ..m + ..m + ..m = ..m$$

14 $4\frac{1}{3} = \frac{..}{..}$ $3\frac{1}{7} = \frac{..}{..}$ $5\frac{1}{2} = \frac{..}{..}$ $8\frac{2}{3} = \frac{..}{..}$

$3\frac{1}{5} = \frac{..}{..}$ $9\frac{1}{3} = \frac{..}{..}$ $7\frac{1}{5} = \frac{..}{..}$ $4\frac{1}{6} = \frac{..}{..}$

$\frac{1}{5} \times 2\frac{1}{2}$ betekent eigenlijk het vijfde deel van $2\frac{1}{2}$.



$\frac{1}{5} \times 2\frac{1}{2} = \frac{1}{5} \times \frac{5}{2} = \frac{1}{2}$ Je hebt de vijven tegen elkaar weggestreept.

Maak er eerst breuken van:

$3\frac{1}{5} \times 3\frac{3}{4} = \frac{..}{..} \times \frac{..}{..} = ..$ $9\frac{1}{3} \times 2\frac{1}{7} = .. \times .. = ..$

Figuur 12. Opgaven uit 'Naar Zelfstandig Rekenen'.

Zoals men ziet, gaat het bij deze opgaven telkens om het leren, inprenten en oefenen van standaardregels. De kinderen doen geen eigen onderzoek, en moeten de geleerde regels direct (kunnen) toepassen. De opgaven worden niet gepresenteerd in contexten en ze moeten individueel (meestal schriftelijk) worden verwerkt.

4.2 Selectie en vergelijkbaarheid van de proef- en controlegroep.

Bij de *selectie* van de scholen hebben verschillende criteria een rol gespeeld: de proefscholen gebruiken het programma van het S.A.C. ('Rekenwerk'), het programma wordt tot en met het zesde (nu achtste) leerjaar door alle leerkrachten gebruikt, de betrokken schoolteams onderschrijven de vernieuwingsideeën achter het programma. Voorts worden in de controle-scholen traditionele rekenmethoden gebruikt, de schoolteams van de proef- en controle-scholen worden begeleid door het S.A.C., ze zijn bereid aan het onderzoek deel te nemen, de proef- en controle-scholen liggen verspreid over de verschillende wijken in Utrecht en er is één proef- en één controle-school die overwegend bezocht worden door buitenlandse kinderen.

De proefgroep die volgens deze criteria geselecteerd werd, bestond uit zeventig zesdeklassers, waarvan tweëndertig uit het milieu van handarbeiders. De controlegroep bestond uit zevenenveertig zesdeklassers, van wie vijftientig uit het milieu van handarbeiders.

Het is een belangrijke vraag of met de hantering van de bovenvermelde criteria de vergelijkbaarheid van de proef- en controlegroep voldoende is gegarandeerd.

Dit kon in mijn onderzoek niet door middel van een voortoets gecontroleerd worden. Zulk een toets had immers vóór de invoering van het programma - op de kleuterschool - afgenomen moeten worden. Wat op vier- of vijfjarige leeftijd gemeten kan worden, is echter moeilijk vergelijkbaar met de uitkomsten van metingen op twaalfjarige leeftijd. Gezien de criteria is er wel enige maar geen absolute garantie, dat de groepen vergelijkbaar zijn en bovendien vergelijkbaar waren toen de kinderen zes jaar waren. In verband met de vereiste vergelijkbaarheid van de beide groepen kan op nog twee zaken worden gewezen: Ten eerste hebben in het onderwijs op de betrokken proef- en controlescholen tijdens de implementatie van het programma, voor zover uit de begeleidingscontacten bekend, geen ingrijpende veranderingen plaatsgevonden, die de resultaten van het onderzoek merkbaar hadden kunnen beïnvloeden. Ten tweede is de vraag of de selectie van de groepen op juiste gronden heeft plaatsgevonden, onderzocht door de scores op het Algemeen Proefwerk (A.P.), de Utrechtse variant van de CITO-toets, van de proef- en controlegroep te vergelijken. In een analyse is nagegaan of de proef- en controle-groep aan elkaar gelijkgesteld mogen worden wat betreft het algemeen niveau van de kinderen, uitgedrukt in de variabelen A.P.-rekenen en A.P.-overigen (begrijpend lezen, taal en dergelijke). Uit een variantie-analyse is gebleken dat de proef- en de controlegroep op geen van deze variabelen van elkaar verschilden ($\alpha = .05$).

Het onderzoek werd voorbereid en uitgevoerd door vijf proefleiders. Aan de leerkrachten van de betrokken scholen werd vooraf de bedoeling van het onderzoek toegelicht en werden afspraken gemaakt over plaats, tijdstip van de afnamen e.d..

Het onderzoek werd uitgevoerd in de maanden oktober, november en december 1983.

Het vond altijd op school plaats, in een rustige ruimte en duurde ongeveer vijfenveertig minuten. Aan de kinderen werd vooraf de bedoeling van de samenspraak duidelijk gemaakt en er werd hun gevraagd zoveel als mogelijk hardop te denken. Benadrukt werd dat het om hun eigen manier van denken ging. De meeste kinderen vonden het de eerste minuten wel spannend en na korte tijd meestal ook leuk om te doen. Slechts één kind maakte aan de medewerking snel een einde.

4.3 Onderzoeksvragen.

1. Lossen de kinderen uit de proefgroep beter wiskundige problemen op dan de kinderen uit de controlegroep wanneer gelet wordt op de zelfstandigheid, snelheid en adequaatheid waarmee de problemen worden opgelost?
2. Lossen de kinderen uit de proefgroep beter wiskundige problemen op dan de kinderen uit de

controlegroep wanneer gelet wordt op het flexibel en zonder hulp van de proefleider veranderen van strategie, nadat de proefleider het kind voor een conflict had geplaatst? Aangezien de interesse vooral uitgaat naar het verschil in hulp tussen de proef- en controlegroep, worden categorie 1 en 2 (zie 4.4.1) samen genomen.

3. Lossen de kinderen uit de proefgroep de wiskundige problemen beter op dan de kinderen uit de controlegroep, terwijl in beide groepen hulp wordt geboden door de proefleider?

Deze 3 onderzoeksvragen worden beantwoord op basis van analyse van het feitelijk probleemoplossend handelen van de kinderen zoals vastgelegd in de protocollen. Er is echter een belangrijk *verschil* tussen het probleemoplossen waarvan sprake is in de eerste twee vragen en in de derde vraag. Zoals we nl. in § 2 zagen, wordt als dat nodig is een kind *hulp* geboden. Met de derde vraag is bedoeld na te gaan *hoeveel hulp* in beide groepen wordt geboden en in welke mate in beide groepen van die hulp werd geprofiteerd. Op basis van het aantal interventies (hulp) werden verdelingen verkregen (zie 4.4.1), die statistisch werden getoetst.

4. Reflecteren de kinderen uit de proefgroep 'vooraf' beter dan de kinderen uit de controlegroep?

Gespecificeerd luidt de vraag of meer kinderen in de proefgroep in staat zijn hun strategie vooraf te bepalen en te motiveren dan in de controlegroep. (zie 3)

5. Reflecteren meer kinderen uit de proefgroep op hoog niveau vergeleken met de kinderen uit de controlegroep? De niveaubepaling is gebaseerd op de indicaties zoals weergegeven in § 3.

Deze twee vragen worden beantwoord op basis van een analyse van het handelen van de kinderen zoals dat in §3 is weergegeven. Dat wil zeggen dat in de onderzoeksprotocollen wordt gezocht naar mentaal handelen dat als indicatie van reflectie kan worden geïnterpreteerd.

6. Bestaat er een samenhang tussen het oplossen van wiskundige problemen en reflectief handelen, en zo ja, is die samenhang in de proefgroep en in de controlegroep van gelijke aard?

7. Worden de onderzoeksbevindingen uit het eerste onderzoek bevestigd door de onderzoeksbevindingen uit het herhalingsonderzoek?

Onderzoeksvraag 6 wordt beantwoord via een aparte statistische analyse en onderzoeksvraag 7 via een herhalingsonderzoek met dezelfde vragen en analysemethoden als in het eerste onderzoek.

4.4 Onderzoeksresultaten.

4.4.1 Het oplossen van wiskundige problemen.

De resultaten van het onderzoek naar het probleemoplossen van alle kinderen zijn samengebracht in een totaaloverzicht, dat hieronder is weergegeven. Zoals uit dit overzicht is af te lezen, werden de resultaten van het onderzoek naar het probleemoplossen (en reflectie vooraf) verdeeld in 6 categorieën. Deze categorieën kunnen als volgt worden getypeerd.

1. De kinderen in categorie 1 hebben het betreffende probleem direct opgelost c.q. direct een goede - niet aan te vechten - strategie gekozen; (de termen 'probleem' en 'strategie' - ook wel aanpak genoemd zijn in §2 toegelicht).
2. In de tweede categorie zijn de kinderen ondergebracht die, nadat ze voor een conflict waren geplaatst, het probleem oplosten, c.q. een adequate strategie bedachten.
3. In de derde categorie zijn de kinderen ondergebracht die na éénmaal hulp van de proefleider het probleem oplosten c.q. een adequate strategie kozen.
4. In de vierde categorie zijn de kinderen ondergebracht die twee maal hulp ontvingen alvorens een correcte aanpak te bedenken c.q. oplossing te vinden.
5. Wanneer de kinderen alsnog een oplossingsstrategie adequaat konden formuleren en toelichten, nadat de proefleider de oplossing had gegeven, kwamen ze in de vijfde categorie. (Een kind zegt dan bijv. 'Oh ja natuurlijk, dat kan op die manier...' en licht die inval toe).
6. In de laatste categorie kwamen de kinderen die het probleem niet oplosten c.q. geen adequate aanpak of strategie bedachten.

Over de overeenstemming in beoordeling tussen de proefleiders zijn de volgende gegevens berekend. Er werden 104 items beoordeeld door 4 proefleiders in onafhankelijke tweetallen en 30 items door 4 proefleiders gezamenlijk. De beoordeling had betrekking op de wijze en het moment van probleemoplossen.

	categorie	meer programmegebonden items								minder programmegebonden items							
		1		3		7		5		2		4		6		8	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
PROEFGROEP N = 70	1	16	23	12	19	45	65	20	29	6	9	2	3	12	18	29	42
	2	13	18,5	11	16	3	5	22	32	39	56	1	1,5	20	29	13	19
	3	24	34	23	37	5	7	7	10	11	15	27	38,5	22	32	3	4
	4	1	1,5	8	13	7	10	3	5	5	7	12	17	3	4,5	6	9
	5			3	5	1	1	4	6	4	6	12	17	5	7,5	11	16
	6	16	23	5	8	8	12	13	18	5	7	16	23	6	9	7	10
CONTROLE- GROEP N = 47	1	6	13	2	4	19	41	7	15					3	7	12	26
	2	5	11			11	24	12	26	26	55	2	4	12	26	8	17,5
	3	20	43	15	32	4	9	11	23	6	13	21	45	13	28	4	8,5
	4	3	6	21	45	10	22	6	13	6	13	14	29,5	8	17	5	11
	5	1	2	5	11			2	4	5	10,5	6	13	4	8	12	26
	6	12	25	4	8	2	4	9	19	4	8,5	4	8,5	6	13	5	11

-  = PR probleemoplossen, reflectie vooraf
 = P probleemoplossen
 = P probleem niet opgelost

Tabel 1 Overzicht van de toetsresultaten.

Meer programmegebonden items bestaan uit problemen die op soortgelijke manier voorkomen in het SAC-programma. Minder programmegebonden betekent dat het algemene, inhoudelijke problemen betreft, die niet in het programma voorkomen.

De proportieovereenstemming in beoordeling (Cohen 1960) was .90 resp. .97.

De betrouwbaarheid van de toets, zo blijkt uit berekening, is bevredigend (KRV = .93).

Alvorens de toetsing van deze tabel op item-niveau te bespreken, wordt eerst onderzoeksvraag 1 en 2 beantwoord.

Tabel 2 Probleem 'direct goed' opgelost.

Frequentietabel				
score	proef (n=70)	(2n)	contr. (n=47)	(3n)
8	0	(0)	0	(0)
7	0	(0)	0	(0)
6	2	(4)	0	(0)
5	4	(8)	0	(0)
4	8	(16)	1	(3)
3	8	(15)	5	(15)
2	18	(38)	8	(24)
1	14	(28)	14	(42)
0	16	(32)	19	(57)
$\bar{x}$	2.0		1.0	

De eerste onderzoeksvraag luidde of de kinderen in de proefgroep de wiskundige problemen sneller en adequater *zonder hulp* van de proefleider oplossen dan de kinderen in de controlegroep. In die adequaatheid en snelheid van probleemoplossen komen de volgende karakteristieken van probleemoplossen tot uiting: het probleem wordt als zodanig herkend en er vindt een adequate probleemrepresentatie plaats.

Die karakteristieken worden samengevat als: probleem 'direct goed' opgelost (categorie 1). 'Direct goed' betekent dat het kind direct een niet aanvechtbare oplossing voor een probleem vond (item 2,4,5,6 en 8) of dat direct een niet aanvechtbare aanpak voor een probleem werd gevonden (item 1, 3 en 7). In dit verband wordt het vinden van een juiste aanpak beschouwd als probleemoplossen; er wordt nl. een antwoord gegeven op de vraag: hoe zou je dit probleem oplossen?

De score is het aantal malen dat per kind 'direct goed' wordt gecategoriseerd (maximaal 8). De

Tabel 3 Probleem 'flexibel zonder hulp' opgelost.

Score verdeling				
score 'flexibel zonder hulp'	proef (n=70)	(2n)	contr. (n=47)	(3n)
8	1	(21)	0	(0)
7	5	(10)	0	(0)
6	11	(22)	2	(6)
5	10	(20)	2	(6)
4	14	(28)	8	(24)
3	10	(20)	12	(36)
2	10	(20)	9	(27)
1	5	(10)	5	(15)
0	4	(8)	9	(27)
$\bar{x}$	3.9		2.4	

vraag of er een verschil is tussen de proef- en controlegroep is nagegaan door de t-toets op de gemiddelde score op deze variabele.

Door het aantal kinderen in de proefgroep en de controlegroep met respectievelijk de factor 2 en 3 te vermenigvuldigen, kunnen de resultaten direct worden vergeleken. Uit de berekening blijkt dat de gemiddelde score tussen de proef- en de controlegroep verschilt in het voordeel van de proefgroep. Dit verschil is significant ($p < .001$). De onderzoeksvraag 1, is hiermee positief beantwoord.

In het onderzoek werd ook nagegaan of een kind flexibel zijn strategie kon veranderen, makkelijk en soepel een nieuwe 'mentale handeling' construeerde (Kruteckij 1976), nadat het (door de proefleider) voor een conflict was geplaatst.

Wij vatten die karakteristieken samen als: probleem 'flexibel zonder hulp' opgelost. Wij spreken van 'flexibel zonder hulp' probleemoplossen bij de items 2,4,5, 6 en 8 indien een kind direct (1e categorie) of nadat het is geplaatst voor een conflict (2e categorie) een niet-aanvechtbare oplossing voor het probleem vindt (5 punten).

'Flexibel zonder hulp' betekent bij de items 1,3 en 7 dat direct een niet aanvechtbare strategie voor een probleem werd bedacht (1e categorie) of - nadat het kind voor een conflict is geplaatst (2e categorie) - een niet aanvechtbare strategie voor een probleem werd gevonden (3 punten). Maximaal zijn per kind dus weer 8 punten te behalen, maar de kans op 8 punten is nu groter, omdat ook de tweede categorie ('na conflict') mee wordt berekend. De scores op de variabele 'flexibel zonder hulp' hangen deels samen met de scores op de variabele 'direct goed' opgelost, maar bij de scores op de variabele 'flexibel zonder hulp' wordt dus tevens naar de tweede categorie gekeken. De vraag of er een verschil is tussen de proef- en de controlegroep is nagegaan door de t-toets op de gemiddelde score op de variabele probleem 'flexibel zonder hulp' opgelost (zie tabel 3).

Uit de berekening blijkt dat de gemiddelde score tussen de proefgroep en de controlegroep verschilt in het voordeel van de proefgroep. Dit verschil is significant ($p < .001$). De onderzoeksvraag 2 is hiermee positief beantwoord.

De derde onderzoeksvraag. Deze luidde of de kinderen uit de proefgroep beter wiskundige problemen oplossen dan de controlekinderen. Met deze vraag wordt het probleemoplossen nagegaan dat tot stand kwam *met hulp* van de proefleider.

Er is *per item* een χ^2 -toets op de verdeling over de zes categorieën toegepast. Hierbij wordt getoetst of de verdeling over de categorieën voor de proefgroep en de controlegroep (per item) hetzelfde is.

Het resultaat van die toetsing ziet er als volgt uit :

Programmagebonden items: Niet programma gebonden items:

item 1 $p < .20$	item 2 $p < .15$
item 3 $p < .00002$	item 4 $p < .10$
item 5 $p < .12$	item 6 $p < .11$
item 7 $p < .005$	item 8 $p < .15$

Bij twee items levert de berekening - uitgaande van een significantieniveau van .05 - een verschil op, maar met name bij de niet-programma gebonden items is het verschil in de verdeling tussen beide groepen gering. Op basis van deze vergelijking van de verdeling kan *onderzoeksvraag 3* slechts in gematigd positieve zin worden beantwoord. Er is een tendens dat de proefgroep iets positiever uit de verdeling naar voren komt dan de controlegroep. Met andere woorden: gelet op de resultaten van het probleemoplossen tot stand gekomen *met hulp* van de proefleider, is het verschil in het voordeel van de proefgroep bescheiden.

Zónder hulp lossen de proefgroepkinderen beter wiskundige problemen op dan de controlekinderen (zie vraag 1 en 2), maar de laatste groep blijkt van de hulp, wanneer die verstrekt wordt, bijna evenveel te profiteren als de kinderen uit de proefgroep.

Uitsplitsing naar geslacht laat het verdelingspatroon ongewijzigd. Uitsplitsing naar rekenniveau (zoals vastgesteld door de leerkracht) en naar milieu is van enige invloed op het eerder verkregen patroon.

4.4.2 Het reflecteren.

4.4.2.1 Reflectie vooraf.

Met behulp van 3 items nl. 1, 3 en 7 (zie §2) werd nagegaan of de kinderen vooraf een aanpak, of plan van handelen, konden bedenken en die aanpak konden motiveren. We noemden dit reflectie vooraf.

Er is nagegaan hoeveel kinderen uit de proefgroep en de controlegroep bij alle drie deze items vooraf reflecteerden, hoeveel kinderen bij een of twee items en hoeveel kinderen bij geen van die items reflecteerden. Uit de χ^2 -toets blijkt dat de verdeling van de proefgroep en de controlegroep significant ($p < .01$) verschilt.

Uit tabel 4 is af te lezen dat dit komt doordat relatief meer kinderen uit de proefgroep vooraf reflecteren dan uit de controlegroep. Onderzoeksvraag 4 is hiermee in positieve zin beantwoord.

Tabel 4 Reflectie vooraf.

score	p-groep		c-groep	
alle items (3)	6	(12)	0	(0)
1 of 2 items	46	(92)	22	(66)
geen item	18	(36)	25	(75)
	n=70	(2n)	n=47	(3n)

4.4.2.2 Niveaus van reflectief handelen.

In aansluiting bij het werk van Zareckij, Semenov en Stepanov (1980) wordt onderscheid gemaakt tussen drie niveaus van reflectie: een hoog (productief), een redelijk (deels productief, deels reproductief) en een laag (reproductief) niveau van reflectie. Bij de bepaling van het niveau van reflectie van een kind werd gewerkt met het kwalitatief *totaalbeeld* dat de studie van het protocol opleverde, maar er werden ook kwantitatieve normen opgesteld. Het reflecteren van een kind werd als *hoog* gekwalificeerd als de indicaties in totaal 12 maal of meer in het protocol gesignaleerd werden, *redelijk* als ze 7 t/m 11 maal en *laag* als ze maximaal 6 maal werden gesignaleerd (deze indeling is arbitrair).

In principe waren deze kwantitatieve normen bepalend voor de niveau indeling. In sommige gevallen echter, wanneer een kind als een grensgeval beschouwd werd, bijvoorbeeld tussen laag en redelijk niveau, dan waren de karakteristieke reflectiemomenten voor de niveau-indeling van doorslaggevende betekenis.

Bijvoorbeeld :

Bram (controlegroep) heeft veel hulp nodig. Na conflict komt hij niet tot een betere aanpak (geen koerswijziging), hij houdt vast aan een algoritme (item 1 en 2). Bij verschillende items wisselt hij lukraak van algoritme (item 3 en 5), overweegt wel een andere handelingsmogelijkheid, maar licht deze matig toe: 'Nu heb ik dat allemaal geleerd.' Soms komt na veel hulp een koerswijziging tot stand (item 4). Bij de laatste drie items wordt hij opeens alerter. Bij het achtste item is er sprake van zelfcontrole. Er werd besloten de reflectie van dit kind als 'redelijk' te beschouwen. Doorslaggevend werd geacht de wijze waarop dit kind gedurende het gespreksverloop steeds alerter werd en

steeds beter ging letten op zijn eigen handelen.

De niveau-indeling is niet tot stand gekomen door tussen de reflectie-indicaties kwalitatieve niveauverschillen te onderscheiden (hoewel dit wellicht mogelijk zou zijn geweest), maar op basis van summaties van het aantal malen dat de indicaties in het protocol werden gesignaleerd. Deze niveaubepaling is nog enigszins grofmazig. Een hoog niveau kan in principe immers bereikt worden zowel op basis van een hoge frequentie van slechts een of twee indicatietypen, als op basis van een lage frequentie van alle acht de indicatie- typen.

In het onderzoek bleek dit echter niet op die manier voor te komen. Indien de kinderen scoorden dan was dit voornamelijk op de indicaties 1,2,3 en 6. De 4e indicatie werd al veel minder frequent, de 7e en 8e sporadisch en de 5e indicatie werd niet één maal gesignaleerd.

Per saldo blijkt de niveau-indeling vooral op de indicaties 1,2,3 en 6 gebaseerd te kunnen worden, hetgeen kan leiden tot de overweging af te zien van de overige indicaties, die het onderzoek alleen maar ingewikkelder en arbeidsintensiever maakten (indicatie 5 hebben we in het herhalingsonderzoek al laten vervallen).

De resultaten van de niveaubepaling op basis van de analyse van de protocollen zijn samengevat in de onderstaande tabel (5 protocollen in de proefgroep bleken onbruikbaar).

Tabel 5. Verdeling in niveaus van reflectie.

	p-groep	c-groep
hoog	31	8
redelijk	21	17
laag	13	22
	65	47

Op deze verdeling werd een χ^2 -toets toegepast. Het verschil is significant ($p < .025$). Uit de tabel blijkt dat er een verschil is in het voordeel van de proefgroep. De onderzoeksvraag 5 - 'Reflecteren meer kinderen uit de proefgroep op hoog niveau vergeleken met de kinderen uit de controle-groep?' - is hiermee in positieve zin beantwoord.

4.4.2.3. Samenhang tussen probleemoplossen en reflectie.

Uit een kwalitatieve analyse van de protocolgegevens werd de indruk verkregen dat er een positieve samenhang bestaat tussen probleemoplossen en reflecteren. In de voorgaande paragraaf werd toegelicht welke niveaus van reflectie we in dit onderzoek hebben onderscheiden. Eveneens hebben we een indeling gemaakt in niveaus van probleemoplossen.

Als goede probleemoplossers worden beschouwd de kinderen die minstens 4 maal een probleem 'flexibel zonder hulp' oplosten (deze kinderen scoren dus 4 maal of meer in categorie 1 of 2 (zie §4.4.1)).

Redelijke probleemoplossers zijn kinderen die 1, 2 of 3 maal een probleem 'flexibel zonder hulp' oplosten en zwakke probleemoplossers zijn kinderen die niet één maal een probleem 'flexibel zonder hulp' oplosten, doch uitsluitend met hulp of zelfs met hulp het probleem nog niet oplosten.

Tabel 6. Verdeling in niveaus van probleemoplossen en reflectie.

	p-groep (n=65)		c-groep (n=47)	
probl. opl reflecteren	goed	redelijk/ zwak	goed	redelijk/ zwak
hoog	28	3	5	3
redelijk	8	13	6	11
laag	2	11	3	19
	38	27	14	33

De tabel biedt een overzicht van de verdeling in niveaus van reflectie en van probleemoplossen (de redelijke en zwakke probleemoplossers zijn samen genomen). Uit analyse (met behulp van het ‘log-lineaire model’) blijkt dat probleemoplossen en reflectie condities zijn die nauw met elkaar samenhangen ($p<.005$), zij het dat die samenhang in de proefgroep - zoals al uit de gegevens in de tabel is af te lezen - sterker is dan in de controlegroep ($p<.01$). Het antwoord op onderzoeksvraag 6 luidt dus dat er een positieve samenhang is tussen probleemoplossen en reflectie zowel in proef- als controlegroep, maar deze samenhang is in de proefgroep sterker.

4.4.2.4. *Interbeoordelaarsbetrouwbaarheid.*

Op grond van gegevens, verzameld in het herhalings- onderzoek (1986), is nagegaan in welke mate 6 externe deskundigen en 3 proefleiders consistent oordelen over het niveau van reflectie. (In het onderzoek van 1983 was dit ten onrechte niet geschied). Zes externe deskundigen beoordeelden een viertal gespreksprotocollen op grond van de indicaties die in het onderzoek als operationalisaties van reflectie golden. Daarbij hanteerden ze de criteria volgens welke de kinderen geclassificeerd werden in een hoog, redelijk of laag niveau van reflectie.

De externe deskundigen werden niet getraind in het afnemen van de toets, noch in het analyseren en interpreteren - via bandopnamen - van de protocolgegevens. De interne deskundigen waren de drie proefleiders - inclusief de auteur - die het herhalingsonderzoek hadden uitgevoerd. Zowel de externe als de interne deskundigen hebben dezelfde vier onderzoeksprotocollen beoordeeld.

In de onderstaande tabellen is weergegeven in welke niveaus van reflectie de beoordelaars de vier kinderen indeelden. De getallen in de tabel geven steeds het aantal beoordelaars weer.

Tabel 7. Reflectie niveau-indeling van de zes deskundigen

niveau	leerling			
	a	b	c	d
hoog	1	1	6	5
redelijk	3	0	0	1
laag	2	5	0	0

Tabel 8 Reflectie niveau-indeling van de drie proefleiders.

niveau	leerling			
	a	b	c	d
hoog	0	0	3	3
redelijk	2	0	0	0
laag	1	3	0	0

Op basis van deze tabellen kan geconcludeerd worden dat de externe deskundigen redelijk consistent over het protocolmateriaal oordeelden (al is die beoordeling bij leerling a minder consistent). Deze conclusie geldt in nog iets sterkere mate voor de getrainde proefleiders. De overeenstemming tussen de externe beoordelaars bevat overigens ook een valideringsaspect. De overeenstemming wijst er immers op dat de indicaties van reflectie bij deze beoordelaars eenzelfde beeld oproepen, een beeld dat herkend wordt als reflectie zoals bedoeld in het onderzoek. Van de overeenstemming tussen de externe deskundigen onderling en de proefleiders onderling zijn vervolgens de proportie-overeenstemming en de kappawaarden berekend (zie tabel 9).

Tabel 9. Overeenstemming tussen de beoordelaars uitgedrukt in kappa's en proportie-overeenstemming (p).

<i>leerling</i>	<i>deskundigen</i>		<i>proefleiders</i>	
	p	kappa	p	kappa
a	.86	.33	.91	.44
b	.78	.41	.77	.41
c	.89	.39	.95	.55
d	.84	.49	.81	.42

De overeenstemming tussen de beoordelaars uitgedrukt in proportie-overeenstemming en in kappa-waarden mag bevredigend worden genoemd, maar is niet zeer hoog. Dit wordt waarschijnlijk mede veroorzaakt doordat één externe beoordelaar een zeer optimistisch beeld had van het reflecteren van alle kinderen en zelfs leerling B. als hoge reflecteerder beoordeelde. Aangezien voor kappa moeilijk normen zijn vast te stellen, zou voor acceptatie van de resultaten tevens op de proportie-overeenstemming gelet kunnen worden.

4.4.2.5 Herhalingsonderzoek (1986).

Het herhalingsonderzoek was in verschillende opzichten beperkter van omvang dan het eerste onderzoek. Het vond plaats in twee Utrechtse basisscholen. Beide scholen werden al jaren begeleid door het S.A.C.. De proefschool gebruikte het programma van het S.A.C. (Rekenwerk). In de controleschool werd een traditionele methode (Naar Zelfstandig Rekenen) gebruikt. De onderzoeksgesprekken met de kinderen werden uitgevoerd door drie proefleiders die ook aan het eerste onderzoek hadden deelgenomen.

De *onderzoeksvragen* die op grond van het materiaal dat verzameld werd in het herhalingsonderzoek beantwoord moesten worden, waren dezelfde als de vragen uit het eerste onderzoek. In het bijzonder staat de vraag centraal of en in welke mate de resultaten uit het eerste onderzoek door de resultaten uit het herhalingsonderzoek worden bevestigd. Over de resultaten van het herhalingsonderzoek (1986) kunnen we kort zijn en wel omdat de resultaten van alle toetsingen *dezelfde trends* vertonen als in het eerste onderzoek (1983). En hiermee is de zevende en laatste onderzoeksvraag beantwoord namelijk dat de bevindingen uit het eerste onderzoek worden bevestigd door de bevindingen van het herhalingsonderzoek.

5. DISCUSSIE

5.1 Interpretatie van de onderzoeksresultaten

Uit de resultaten van de beide onderzoeken kan afgeleid worden dat de kinderen uit de proefgroep, vergeleken met de kinderen uit de controlegroep:

- de wiskundige problemen sneller oplossen;
- minder hulp nodig hadden bij het oplossen van die problemen;
- hun handelen vooraf vaker planden en formuleerden;
- vaker reflectieve beslissingen (achteraf) namen, van strategie veranderden, twee strategieën vergeleken en zichzelf controleerden.

Hoe kunnen die verschillen worden geïnterpreteerd? Niet uit een verschil in algemeen niveau, zo lijkt het, omdat uit analyse is gebleken dat de proef- en de controlegroep gelijkgesteld mogen worden wat betreft het algemeen niveau van de kinderen, uitgedrukt in de variabelen A.P. rekenen en A.P. overigen ($\alpha = .05$).

Deze berekening kan echter de vraag oproepen of het werken met het S.A.C.-programma dan geen enkele invloed heeft uitgeoefend op de resultaten van de rekenonderdelen van het A.P.. Deze vraag kan als volgt worden beantwoord.

Ten eerste krijgen de kinderen op basis van het S.A.C.-programma aanzienlijk minder training in de traditionele A.P.-onderdelen hoofdrekenen, cijferen en vraagstukjes, dan de kinderen in de controlegroep. Ten tweede komen belangrijke onderdelen van het S.A.C.-programma in het A.P. in het geheel niet voor (zoals waarschijnlijkheid, meetkunde, eigenschapsrekenen, en dergelijke). Ten derde bestaat het A.P. voor een groot deel uit opgaven die in het S.A.C.-programma nauwelijks voorkomen.

Dat het werken met het S.A.C.-programma geen direct positieve invloed had op de resultaten van het A.P.rekenen, hoeft dus niet te verbazen, omdat met het S.A.C.-programma de kinderen niet getraind worden in het maken van zulke opgaven. Het gegeven dat de kinderen uit de proefgroep gelijkwaardige prestaties behaalden op de rekenonderdelen van een toets, die in feite voor hen niet 'fair' was, hoeft niet tot ontevredenheid te stemmen. Maar betekent dit niet dat de kinderen uit de proefgroep intelligenter waren? Ze zijn immers niet getraind en behalen toch gelijkwaardige prestaties op de rekenonderdelen van het A.P.. Als de kinderen uit de proefgroep inderdaad intelligenter zijn dan zou dat over de hele linie moeten blijken en dat is niet het geval. Op de niet-rekenonderdelen van het A.P., dat zijn de onderdelen Begrijpend Lezen, Taal-Allerlei, Wereldverkenning en Spelling, verschilden de beide groepen niet van elkaar ($\alpha = .05$). Omdat de programmavariabele niet weg te denken is uit rekenonderdelen van het A.P., kan dus eigenlijk beter gekeken worden naar de niet-rekenonderdelen van het A.P..

Uit onderzoek is bekend dat er een grote mate van samenhang is tussen schoolvorderingentoet-

sen, zoals het A.P., en intelligentietests (Zwarts 1983). Dat eventuele effecten van het werken met het S.A.C. programma toegeschreven kunnen worden aan verschillen in intelligentie is derhalve onwaarschijnlijk. De uitslagen op het A.P. tussen beide groepen verschillen immers niet.

Evenmin kunnen de verschillen geïnterpreteerd worden uit speciale trainingen van de proefgroepkinderen, want die zijn niet gegeven. Het lijkt plausibel de onderzoeksresultaten te verklaren uit een drietal factoren:

1. Verschillen tussen gebruikte curricula.
2. De gegroeide professionaliteit en gemotiveerdheid van de leerkrachten in de proefgroep ten opzichte van (de vernieuwing van) het reken-/wiskundeonderwijs.
3. De begeleiding van de vernieuwing van het rekenwiskundeonderwijs op de betrokken proefscholen.

De factor die de geconstateerde verschillen het meest verklaard, is waarschijnlijk de eerstgenoemde.

Het probleemoplossen en reflecteren van de kinderen uit de proefgroep, zo bleek uit de protocollen, weerspiegelden de karakteristieken van het gebruikte programma op duidelijke wijze. Het probleemoplossend en reflectief handelen van de kinderen uit de controlegroep toonde eveneens opvallende overeenkomst met de kenmerken van de in die groep gebruikte curricula. De kinderen hanteerden, zoals een mechanistische methode dat ook voorschrijft, één algoritme, ze weken daar niet gemakkelijk of pas met hulp van de proefleider van af, ze controleerden zichzelf weinig en ze neigden ertoe rechtstreeks op het resultaat af te stevenen.

5.2 Suggesties voor verbetering van het onderzoek

Voor de snelle probleemoplossers vormden sommige problemen geen enkele uitdaging. Spontane reflectie, bijvoorbeeld in de vorm van zelfcontrole (indicatie 6) vond in zo'n geval dan ook niet plaats. Maar bovendien lukte het de proefleider slechts met moeite bij deze kinderen reflectie op te roepen. Ze hadden het probleem toch al opgelost, maar ze wisten soms zelfs niet hoe (verg. expertgedrag).

We zien hier dat reflectie niet altijd een voorwaarde is voor probleemoplossen. Het omgekeerde echter kwam óók voor: hoewel sommige kinderen reflecteerden (zichzelf controleerden), geraakten ze niet tot de juiste oplossing van het probleem. Reflectie leidde dus niet altijd tot de oplossing van het probleem.

Een volgend probleem bleek te zijn dat niet alle items een voldoende conflict-karakter bezaten en daardoor bij de kinderen weinig reactie (en reflectie) uitlokten (bijv. het 2e, 3e en 6e item). Deze items zouden vervangen moeten worden door betere en meer uitdagende items.

Tot slot lijkt ons een belangrijke kwestie bij de interpretatie en analyse van de protocollen deze. De kinderen wordt gevraagd een gevonden oplossing of een strategie die ze bedachten te motiveren en toe te lichten. Bovendien worden ze gevraagd twee strategieën te vergelijken (na conflict) en te motiveren. Welnu, de *kwaliteit* van deze motiveringen kan erg uiteenlopen en deze kwaliteitsverschillen hebben we niet meegewogen bij de niveaubepaling van reflectie. De reden daarvoor was dat we geen hanteerbare onderscheidingscriteria konden vinden en in deze beginfase van het reflectieonderzoek ook niet al te veel hooi op onze vork wilden nemen. Het instrument wordt echter fijnmaziger en dus gevoeliger als in (verder) onderzoek rekening wordt gehouden met de kwaliteit van de gevraagde motiveringen.

5.3 Probleemoplossen en reflectie

In het eerste onderzoek en in het herhalingsonderzoek bleken probleemoplossen en reflectie samenhangende variabelen te zijn. Deze samenhang was in de proefgroep sterker dan in de controlegroep. Als verklaring voor deze sterke samenhang bij de kinderen uit de proefgroep kan op het volgende gewezen worden.

Doordat in de proefscholen - op basis van het programma - de kinderen werden gestimuleerd commentaar te leveren op elkaars ideeën en constructies, is bij deze kinderen wellicht op den duur een *reflectieve houding* ontstaan.

Die houding kan het probleemoplossen in positieve zin hebben beïnvloed, waardoor de

samenhang tussen beide variabelen kan zijn versterkt. Ten tweede leerden de kinderen in de proefgroep modellen, begrippen en procedures die noodzakelijk zijn om zowel met succes problemen te kunnen oplossen als om te kunnen reflecteren. Voorts kan het onderwijs dat gegeven wordt met Rekenwerk worden getypeerd als *interactief* onderwijs. Deze ervaring was tijdens de samenspraak met de proefleider (die de kinderen voor een *socio-cognitief* conflict plaatste) van positieve invloed, waarschijnlijk zowel op het probleemoplossen als op het reflecteren.

De problemen veroorzaakten bij de kinderen uit de proefgroep minder *emotionele* onzekerheid dan bij de controlekinderen, zo luidt de laatste veronderstelling.

De kinderen uit de *controlegroep* hebben immers vooral niet voor discussie vatbare standaardprocedures geleerd. Het gebruik van die procedures vraagt niet om reflectie en reflectie zou zelfs verstoring kunnen werken. Hierdoor én door het op individualisering gerichte onderwijs kan zich moeilijk een reflectieve houding ontwikkelen. Het non-interactieve onderwijssysteem heeft noch de motivatie voor discussie, noch de discussievaardigheid van de kinderen uit de controlegroep erg bevorderd. Dit heeft zowel het probleemoplossen als het reflecteren, tijdens de samenspraak met de proefleider, waarschijnlijk in *negatieve* zin beïnvloed.

De vragen van de proefleider hebben waarschijnlijk bij vele kinderen hun *zelfvertrouwen* ondermijnd en hun gevoel van zekerheid aangetast. Het sociale conflict veroorzaakt dan tevens een emotioneel conflict. Het *socio-cognitief* en *emotioneel* conflict kan zowel op het oplossen van problemen als op het reflecteren en negatieve invloed hebben uitgeoefend.

Tot slot van dit artikel willen wij nog op het volgende wijzen. In deze laatste paragraaf hebben we gewezen op een samenhang die er bestaat tussen probleemoplossen en reflectie. Daarmee is niet gezegd (of aangetoond) dat reflectie zonder meer zal leiden tot beter probleemoplossen. Dit is ook niet onderzocht.

Er is onderzocht, in de context van probleemoplossen, of de kinderen tot reflectie in staat zijn. Het onderzoek waarover wij rapporteerden is dan ook anders van opzet dan de meeste reflectie- of metacognitieonderzoeken. In deze onderzoeken vindt een training plaats van metacognitieve vaardigheden (vragen stellen, onderstrepen e.d.). Een controlegroep wordt niet getraind. Beide groepen worden (vóór en ná) getoetst, meestal niet op reflectievaardigheid, maar de toets bestaat uit een leertaak of een serie problemen. Indien de resultaten van de proefgroep beter zijn dan die van de controlegroep, dan wordt hieruit *afgeleid* dat dit verschil is veroorzaakt door de training die vooraf plaatsvond. Dit is verdedigbaar, maar niet geheel zeker. De resultaten kunnen ook veroorzaakt zijn door meer ervaring, meer inzicht e.d.. Wij hebben, in tegenstelling tot deze onderzoeken, niet het reflecteren als zodanig getraind. Wij hebben voorts uit de onderzoeksresultaten - met name uit het gegeven dat de kinderen in de proefgroep met minder hulp de problemen oplosten - *niet* afgeleid dat dit werd veroorzaakt door een hogere reflectievaardigheid.

We hebben het reflecteren *als zodanig* onderzocht en vastgesteld dat meer kinderen uit de proefgroep op hoog niveau reflecteren, vergeleken met de kinderen uit de controlegroep.

Om na te kunnen gaan of een hoog reflectieniveau van *directe* invloed is op de wijze en het niveau van probleemoplossen, zou men problemen moeten ontwerpen die *alleen* opgelost kunnen worden door te reflecteren.

Of zulke problemen bestaan, weten we niet, wij hebben ze in elk geval niet kunnen bedenken.

LITERATUUR

- Berg, W. van de, D. van Eerde (1983), Ontcijfering van het rekenen: diagnostiek van de Kwantiwijzer. *Tijdschrift voor orthopedagogiek*. 22 (1983) nr.2, 44-67
- Boonman, J.H. en W.A.M. Kok (1986), *Kennisverwerven uit teksten*. Vakgroep Onderwijskunde RU Utrecht 1986.
- Borkowski, J.G. (1985), Signs of Intelligence : Strategy Generalization and Metacognition. In : *The Growth of Reflection in Children*. S.R. Yussen (ed.), Academic Press Inc., 105-145
- Breuker, J.A. (1982), Hardop nadenken: Methodologie voor het analyseren van hardop denkprotocollen. In : Kamp, L. van der, M. van der Kamp, *Methodologie van Onderwijsresearch*. Lisse, 47-61, (Bijdrage tot de Onderwijsresearchdagen 1982, nr. 1)

- Cohen, J. (1960), Coefficient of agreement for nominal scales. In : *Education and psychological measurement*, 20, 37-46
- Confrey, J. (1985). Towards a Framework for Constructivists Instruction in : *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education* (volume 1). L.Streefland (ed.), P.M.E. Noordwijkerhout; 477-483
- Corte, E. de, L. Verschaffel (1983), Analyse en beïnvloeding van het probleemoplossend denken. *Tijdschrift voor orthopedagogiek* 22, 449-464
- Corte, E. de, L. Verschaffel (1985), Oplossingsstrategieën van eenvoudige redactieopgaven over optellen en aftrekken. *Pedagogische Studiën* 62, 125-138.
- Doblaev, L.P. (1982), *Smyslovaja struktura ucebnoego teksta i problemy ego ponimanija*. Moskou. *Studieteksten lezen en begrijpen* (vert. Z.Nelissen-Bradová). Apeldoorn 1984
- Doise, W. en G. Mugny (1984), *The Social Development of the Intellect*. Pergamon Press, Oxford.
- Donaldson, M. (1978), *Children's Mind*. London, Fontana
- Ericsson, K.A., H.A. Simon (1980), Verbal reports as data. *Psychological Review* 87, 215-251.
- Ericsson, K.A., H.A. Simon (1984), *Protocol Analyses; Verbal Reports as Data*. Cambridge, Massachusetts: MIT-Press.
- Forrest-Pressley, D.L., L.A. Gillier (1983), Children's Flexible Use of Strategies during Reading. In: *Cognitive Strategy Research*, Psychological Foundations, M.Pressley, J.R. Levin (ed.) New York, Berlin.
- Ginsburg, H. (1981). The Clinical Interview in Psychological Research on Mathematical Thinking : Aims, Rationales, Techniques. In: *For the learning of mathematics*. 1, 3 (march).
- Ginsburg, H. (1983) (ed.) *Development of mathematical thinking*. Acad.Press New York.
- Jong, R. de (1986). *Wiskobas in methoden*. Dissertatie Utrecht.
- Kilpatrick, J. (1985), *Reflection and recursion*. *Educational Studies in Mathematics*. 16.
- Kirby, J.R. (1984). *Cognitive Strategies and Educational Performance*. Academic Press Inc.
- Klerk, L.F.W. de (1982). Verslag van het AERA-congres 1982 te New York. *Pedagogische Studiën* 59, 519-536
- Knoers, A.M.P. (1983), 'Ik sta het dichtst bij de onderwijstechnologie'. Interview met Knoers. *Pedagogische Studiën* 60, 451-466.
- Krutetskii (Kruteckij), V.A. (1986). *The psychology of mathematical abilities in Schoolchildren*. Chicago.
- Lawson, M.J. (1984). Being Executive about Metacognition. In: *Cognitive strategies and Educational performance*. J.R. Kirby (ed.) Academic Press Inc.
- Lesh, R. (1985). Conceptual Analysis of Mathematical Ideas and Problem Solving Processes. In : Streefland (ed.) *Proceedings of the ninth international conference for the psychology of mathematics education* (vol. II); P.M.E., Noordwijkerhout; 73-97
- Matis, T.A. (1982). Wege zur Erforschung psychischer Neubildungen in der gemeinschaftlichen Lerntätigkeit. In.: *Ausbildung der Lerntätigkeit bei Schülern*. V.V. Davydov, J.Lompscher, A.K. Markova (ed.), Volk und Wissen Berlin.
- Meichenbaum, D., S.Burland, L.Gruson en R. Cameron, (1985). Metacognitive Assessment. In: *The Growth of Reflection in Children*. S.R. Yussen (ed.). Academic Press Inc. 3-31.
- Nelissen, J.M.C., A.C. Vuurmans (1983). *Aktiviteit en de ontwikkeling van het psychische*. SUA Amsterdam.
- Nelissen, J.M.C. (1987). *Kinderen leren wiskunde*. Een studie over constructie en reflectie in het basisonderwijs (dissertatie); Gorinchem.
- Parreren, C.F.van, E.M.H. Assink, J.W.M. Borghouts- Van Erp (1983). *Onderwijsleerprocessen in het kader van ontwikkelend onderwijs*. NIP konferentie, juni.
- Polivanova, K.N. (1978). Onderzoek naar controle als leerhandeling. In : *Issledovanija po problemam vozrastnoj i pedagogiceskoj psichologii*. Moskou. Vert. Z. Nelissen-Bradová.
- Sacharowa, A.W. (1982), Selbstkontrolle und Selbstbewertung in der Ausbildung der Lerntätigkeit. In: *Ausbildung der Lerntätigkeit bei Schülern*. V.V. Davydov, J. Lompscher, A.K. Markova (ed.), Volk und Wissen Berlin. *Pedagogika Moskou*; 149-164.
- Span, R., R.Overtoom-Corsmit (1986). Information Processing by Intellectually Gifted Pupils Solving Mathematical Problems. In: *Educational Studies in Mathematics*; special issue.
- Stepanov, S.J., I.N. Semenov (1982). De vorming van typen reflectie bij het oplossen van (creatieve) opgaven. In: *Voprosy psichologii*, nr. 1. Vert.: Z. Nelissen-Bradová.
- Stepanov, S.J., I.N. Semenov (1985), Psychologie van reflectie : problemen en onderzoeken In: *Voprosy psichologii*, 3, 31-40 (in het russisch).
- Sternberg, R.J. (1985), Review of Meichenbaum, Burland, Gruson and Cameron's metacognitive Assessment. In : *The Growth of Reflection in Children*. S.R. Yussen (ed.) Academic Press Inc. 31-37.
- Teyken, C. (1988). *Ontwerpen als proces*. Lisse, Swets & Zeitlinger
- Vygotskij, L.S. (1977). *Denken und Sprechen*. Fischer Taschenbuch Verlag.

- Wertsch, J.V., G.D. McNamee, J.B. McLane. N.A. Budwig (1980). The adult-child dyad as a problemsolving system. *Child development* 51, 1215-1221.
- IJzendoorn, M.H. van, S. van Vliet-Visser (1985). Gezinsuitbreiding. Een kwalitatieve voorstudie naar veranderingen in de relatie tussen eerstgeboren kind en zijn ouders. *Pedagogische Studiën* 62, 313-328.
- Zak, A.Z.(1977). Over de ontwikkelingen van het denken bij leerlingen. In: *Psychologische problemen van de leeractiviteiten bij leerlingen*. Red. V.V. Davydov, Moskou; vert. Z. Nelissen-Bradová.
- Zak, A.Z.(1980). Een experimenteel onderzoek naar de reflectie bij jongere basisschoolleerlingen. *Pedagogische Studiën* 57, 355-369, vert. Z. Nelissen-Bradová.
- Zareckij, V.K., I.N. Semenov, S.J. Stepanov (1980). Het reflectief-persoonlijk aspect van de vorming van oplossingen van creatieve opgaven. *Voprosy psichologii* 5, vert. Z. Nelissen-Bradová.
- Zwarts, M.A. (1983). *Criteriumtoetsen bij de aansluiting van primair en secundair onderwijs* (dissertatie). Utrecht.

Manuscript ontvangen 3-11-1987

Definitieve versie ontvangen 18-5-1988

Differentiële Effecten van Vraagvorm bij Aardrijkskunde- en Natuurkunde-examens¹

Huub van den Bergh, Mindert Eiting en Martha Otter²

Stichting Centrum voor Onderwijsresearch van de Universiteit van Amsterdam (SCO)

ABSTRACT

Differential effects of item format on geography and science examinations.

The effects of item format on achievement scores of examinees are still in debate. Many previous studies can be criticized, because of an inadequate design. In this study a design was used in which every possible combination of multiple-choice and open ended examinations was administered to random samples of examinees. The data were analysed with multi-group linear structural models. Geography exams differing in item format proved to measure different traits, whereas this was not found for the science exams. It was suggested that these differential results were not due to the subject of the exam, but due to the type of questions asked.

1. INLEIDING

De vraag of toetsen met een verschillende vraagvorm hetzelfde attribuut meten brengt de gemoederen nog steeds in beweging (zie o.a. Bonset, 1973, 1974, 1975; Bonset en Griffioen, 1972; Svinicki & Koch, 1983; Houtman, 1986). Sommigen menen dat open en meerkeuze-vragen volstrekt vergelijkbaar zijn, anderen menen dat beide vraagvormen een beroep doen op verschillende cognitieve vaardigheden.

Overeenstemming tussen de verschillende onderzoekers is er min of meer over twee zaken: de gemiddelde prestaties van leerlingen op een toets met meerkeuze-vragen zijn hoger dan die op dezelfde toets met open vragen, en de homogeniteitschatting valt vaak lager uit voor een toets met meerkeuze-vragen (zie o.a. Frary, 1985; Heim & Watts, 1967; Carcelli, 1980; Benson, 1981; Shohamy, 1984; Van den Bergh, 1987). Deze verschillen in gemiddelden en betrouwbaarheid zijn echter niet van belang voor de onderhavige vraagstelling. De vraag of twee toetsen met gelijke inhoud maar verschillende vraagvorm vergelijkbare attributen meten is een vraag naar de zogenaamde congenericiteit van de toetsen (Jöreskog, 1971). Toetsen met open en meerkeuze-vragen behoeven niet aan de strenge eisen van paralleliteit (gelijke betrouwbaarheden, varianties en gemiddelden) of tau-equivalentie (gelijke ware score varianties) te voldoen.

Interessanter is de al gestelde vraag of met verschillende vraagvormen hetzelfde attribuut gemeten wordt. De onderzoeksresultaten zijn niet eensluidend. In een aantal onderzoeken werd aangetoond dat de voor meetfouten gecorrigeerde prestatierangorde van leerlingen op toetsen met open en meerkeuze-vragen verschillend is (zie o.a. Vernon, 1962; Rowley, 1974; Williams e.a., 1978; Murphy, 1982; Ward, 1982; Kempa & L'Odiaga, 1984; Shohamy, 1984; Van den Bergh, 1987), of dat toetsen met open en toetsen met meerkeuze-vragen op verschillende wijze samenhangen met een criterium, vaak schoolprestaties of het docenten oordeel (Gates, 1921; Knight, 1922; Daniëls, 1954; Davis & Fiffer, 1957; Hopkins e.a., 1962, 1981; Kruglak, 1965; Nisbett, 1965; Bracht & Hopkins, 1970; Culpepper & Ramsdale, 1983; Koffler, 1986). Deze auteurs trekken de conclusie dat bij het maken van open en meerkeuze-vragen andere cognitieve vaardigheden een rol spelen. Anderen hebben gedemonstreerd dat de prestaties van leerlingen op

¹ Dit onderzoek kwam tot stand dankzij een subsidie van het ministerie van onderwijs en wetenschappen.

² De auteurs willen graag Jannemieke van de Gein, Gert Rijlaarsdam en Margaretha Vergeer bedanken voor hun uitvoerige commentaar.

toetsen met open en toetsen met meerkeuze-vragen vergelijkbaar zijn qua gemeten attribuut (zie o.a. Toops, 1921; Ruch & Stoddard, 1927; Shalson & Crawford, 1928; Correy, 1930; Carter & Crone, 1940; Hurlburt, 1954; Cook, 1965; Heim & Watts, 1967; Choppin, 1968; Van Westrhenen, 1973; Traub & Fisher, 1977) of qua criterium-validiteit (zie o.a. Peters & Martz, 1931; Emmett, 1954; Peel & Armstrong, 1956; Wiseman, 1956; Cheketts & Christensen, 1974; Weiss & Jackson, 1983).

Hierbij zij opgemerkt dat degenen die geen verschil vinden in gemeten attribuut tussen toetsen met verschillende vraagvormen daarvoor geen verklaring geven. De verklaringen van degenen die menen dat er verschillen zijn qua gemeten attribuut bevatten soms wel theoretische noties; enerzijds worden verschillen ten gevolge van formuleervaardigheid genoemd, anderzijds gaat men in op verschillen in gemeten attribuut ten gevolge van herkenings- of herinneringsprocessen.

Wij willen erop wijzen dat in de bovenstaande studies veelal geen deugdelijk onderzoeksontwerp is gebruikt. Er zijn drie oorzaken van foute conclusies denkbaar. We onderscheiden conceptuele onduidelijkheid, dubieuze psychometrie en een dubieus onderzoeksontwerp.

Wanneer men de vraag stelt of verschillende vraagvormen verschillende vaardigheden vereisen, is er sprake van conceptuele onduidelijkheid, wanneer niet tevens de mate van verschil wordt geïntroduceerd. In dit verband wijzen we op de recente aandacht voor de problematiek van de altijd verwerpelijke nulhypothese. Juist het voorbijgaan aan de mate van verschil vormt een motor van het voortdurende debat: zo wordt van een verschil gesproken wanneer een voor onbetrouwbaarheid gecorrigeerde correlatie tussen prestaties bij verschillende vraagvormen kleiner is dan 1.00 (Mellenbergh, 1971), kleiner dan .80 (Peterson, 1926; Hurd, 1932) of kleiner dan .70 (Van Westrhenen, 1973).

Een voorbeeld van dubieuze psychometrie is het berekenen van gecorrigeerde correlaties op basis van de coëfficiënt alpha omdat deze coëfficiënt een ondergrensschatting is van de feitelijke betrouwbaarheid, waardoor de gecorrigeerde correlatie veelal overschat wordt (zie voor dit punt: Van den Bergh, 1987). Ook het beoordelen van de criteriumvaliditeit van niet-tau-equivalente metingen, kan misleidend zijn, omdat verschillen in betrouwbaarheid kunnen leiden tot verschillen in de hoogte van de gevonden correlaties.

Wat onderzoeksontwerp betreft kan het meeste onderzoek de toets der kritiek niet doorstaan. Wat is het geval? De prestaties van leerlingen op een toets met open vragen moeten worden vergeleken met die op een toets met meerkeuze-vragen. De prestaties op de tweede toets worden dan systematisch beïnvloed door geheugeneffecten e.d., waardoor de effecten van toetsvorm en geheugen gecontamineerd worden. In het verleden werd vrijwel altijd eerst de toets met open vragen afgenomen gevolgd door de toets met meerkeuze-vragen. Een verbetering van het onderzoeksontwerp is mogelijk door de volgorde van de afname van de open-vraagversie en de meerkeuze-vraagversie van een toets te randomiseren (zie o.a. Hogan, 1981), maar ook deze procedure is niet ideaal. Immers, ook in dat geval kunnen geheugeneffecten een rol spelen, waardoor een overschatting van de bedoelde correlatie gevonden wordt. Daar staat tegenover dat men er niet van uit mag gaan dat in de periode tussen de beide afnamen, die niet al te kort mag zijn wegens ongewenste geheugen effecten, de vaardigheden van de leerlingen constant blijven of voor alle leerlingen in gelijke mate veranderen. Dit is, wanneer het onderwezen kennis en vaardigheden betreft, een minder plausibele aanname (vergelijk: Bloom, 1976; Anastasi, 1958). Dit verschijnsel kan leiden tot een onderschatting van de bedoelde correlatie.

De vraag of met verschillende vraagvormen hetzelfde attribuut gemeten wordt kan beter geherformuleerd worden, waarbij tevens een verschilgrootte wordt gedefinieerd. Centraal in onze formulering staat het begrip van een latente (of ware) score van een leerling, die kan worden opgevat als een voor meetfouten gecorrigeerde waargenomen score. Van belang is nu de proportie gemeenschappelijke variantie in een populatie van latente scores behorende bij twee toetsafnamen. Zelfs wanneer we twee keer een identieke toets afnemen kan deze proportie kleiner zijn dan één. De afwijking van één noemen we het proportioneel verlies bij identieke toetsen. Bij afname van niet-identieke toetsen, bij voorbeeld met verschillende vraagvormen, kan nog een extra verlies optreden.

We kunnen constateren dat de resultaten van de verschillende onderzoeken op dit terrein

moeilijk te interpreteren zijn. Dit geldt zowel de onderzoeken waarin geen als onderzoeken waarin wel verschillen gevonden zijn. Ons resten de op theorie gebaseerde verklaringen van degenen die verschillen hebben geconstateerd. Omdat er voor het verwerpen van een hypothese een grens getrokken moet worden, hebben wij de grens voor het extra verlies arbitrair vastgesteld op de helft of meer van het proportioneel verlies bij identieke toetsen. Onze hypothese luidt dat bij toepassing van open en gesloten vraagvormen, ongeacht het onderwerp, genoemd extra verlies zal optreden.

2. INSTRUMENTATIE EN STEEKPROEF

2.1. Instrumentatie

Dit onderzoek sluit aan bij de evaluatie van de zogenaamde 70/30 maatregel voor het Centraal Schriftelijk Eindexamen LBO/MAVO. Deze maatregel houdt in dat het examencijfer van LBO- en MAVO-examinandi voor 70% bepaald wordt door hun antwoorden op de gesloten en voor 30% bepaald wordt door hun antwoorden op de open vragen. Deze verandering van vraagvorm is onder andere doorgevoerd voor de vakken aardrijkskunde en natuurkunde.

In dit onderzoek is in eerste instantie een inventarisatie gemaakt van examens voor de vakgebieden aardrijkskunde en natuurkunde. Deze inventarisatie heeft betrekking op vijftien eindexamens LBO en MAVO uit de afgelopen jaren. Uit deze eindexamens is per onderwerp en subonderwerp, aselekt een groot aantal open en meerkeuze-vragen geselecteerd, minimaal vier keer het aantal vragen per onderwerp en subonderwerp dat normaliter in een eindexamen is terug te vinden. Per onderwerp en subonderwerp komt dus een groot aantal open en meerkeuze-vragen ter beschikking.

Voor de constructie van equivalente open en meerkeuze-vragen worden acht vakinhoudelijk-deskundigen aangetrokken. Deze deskundigen zijn (de afgelopen) vijf jaar betrokken geweest bij de constructie van eindexamens aardrijkskunde of natuurkunde voor LBO en MAVO. Aan elke deskundige wordt gevraagd om bij twintig gesloten vragen een open equivalent te construeren en bij twintig open vragen een gesloten equivalent te construeren, door formulering van drie foute alternatieven. Alle oorspronkelijk geselecteerde en later geconstrueerde open en vierkeuze-vragen worden vervolgens door alle specialisten gecontroleerd op gelijkwaardigheid. Uit de equivalent bevonden vragen worden per vakgebied aselekt twee examenversies geformeerd. Het aantal vragen per onderwerp en subonderwerp zal gelijk zijn aan het aantal vragen dat gemeenlijk in een eindexamen is terug te vinden.

2.2. Steekproef

Uit de populatie LBO- en MAVO-scholen zijn aselekt 180 scholen aangeschreven; 90 secties natuurkunde en 90 secties aardrijkskunde. Van deze 180 aangeschreven secties verklaarden 33 zich bereid tot medewerking. De 33 scholen die medewerking toezegden, bestonden uit twaalf LBO- en 21 MAVO-scholen. Van de twaalf LBO-scholen verklaarden elf leraren natuurkunde en twee leraren aardrijkskunde zich bereid tot deelname, van de 21 MAVO-scholen elf leraren natuurkunde en tien leraren aardrijkskunde. Een overzicht van het totaal aantal aangeschreven scholen, het totaal aantal deelnemende scholen en het totaal aantal deelnemende leerlingen wordt gegeven in Tabel 1.

Tabel 1. Totaal Aantal Aangeschreven Scholen, Totaal Aantal Deelnemende Scholen en het Totaal Aantal Deelnemende Leerlingen per Vakgebied (Natuurkunde en Aardrijkskunde) en per Schooltype (LBO en MAVO).

	Natuurkunde			Aardrijkskunde		
	LBO	MAVO	tot	LBO	MAVO	tot
Totaal aantal aangeschreven scholen	45	45	90	45	45	90
Totaal aantal deelnemende scholen	11	11	22	1	10	11
Totaal aantal deelnemende leerlingen	367	326	693	20	288	308

Uit Tabel 1 wordt duidelijk dat slechts één LBO-school voor het vak aardrijkskunde is geworven. De reden hiervoor is dat aardrijkskunde op het LBO slechts sporadisch gekozen wordt.

3. ONDERZOEKSONTWERP

Per onderwerp zijn twee vraagvormen en twee versies ontworpen zodat er vier examens zijn. Elke leerling maakt op twee tijdstippen van elk twee lesuren een examen. Daarbij is elke combinatie van twee examens mogelijk, dus ook combinaties van twee dezelfde examens. Omdat ook de volgorde bij ongelijke examens ertoe kan doen, betekent dit dat er zestien te onderscheiden combinaties zijn per vakgebied. Leerlingen worden aselekt toegewezen aan deze combinaties.

Het onderhavige onderzoek is een voorbeeld van een generaliseerbaarheidsstudie (zie o.a. Nußbaum, 1984), waarin natuurkunde- en aardrijkskunde-examens zijn gedefinieerd met behulp van een facet-ontwerp. Gezien de vraagstelling is het vraagvorm-facet met als niveaus open (O) en gesloten (G) vragen het belangrijkste. In termen van de variantie-analyse is dit facet gefixeerd. De twee vraagvormen zijn geen willekeurige keuze uit de verzameling van alle mogelijke vraagvormen, zodat de resultaten niet in statistische zin kunnen worden gegeneraliseerd naar deze verzameling. Omdat het tijdsinterval tussen de twee afnamen ook niet willekeurig is gekozen (variërend van 1 tot 3 maanden), is het tijdstip-facet met de niveaus T1 en T2 eveneens gefixeerd. Een aselekt facet vormt de examen-versie, omdat de niveaus V1 en V2 staan voor aselechte steekproeven uit de vragen-populatie. Een vierde aselekt facet is om model-technische redenen (zie hierna) ingevoerd. Op basis van vragen-paren met nagenoeg gelijke gemiddelden en betrouwbaarheden der scores zijn per versie aselekt parallelle subversies (P1 en P2) samengesteld volgens de methode van de 'matched random subtests' (zie Gulliksen, 1950). Dit facet is dus pas na afname gedefinieerd en genest in versie. Elke leerling maakt per tijdstip twee parallelle subversies van één der examens.

Het ontwerp voorziet in zestien groepen leerlingen voor zowel het natuurkunde- als het aardrijkskunde-examen. De zestien groepen zijn aselekt samengesteld uit de totale leerlingsteekproeven. Bijgevolg maken leerlingen uit dezelfde klas verschillende examens, bepaald door het lot. In Tabel 2 vindt men een weergave van het ontwerp. Daarin is tevens de grootte der groepen vermeld, zoals vastgesteld na afname. Leerlingen die op één of beide afnametijdstippen afwezig waren, zijn uit de steekproeven verwijderd.

In Tabel 2 is een experimenteel ontwerp weergegeven, waardoor (eventueel geattenueerde) correlaties in verschillende groepen vergeleken kunnen worden, dit in tegenstelling tot traditionele ontwerpen waarin het vergelijken van gemiddelden en varianties centraal staat. Met behulp van dit ontwerp kunnen effecten van vraagvorm worden gescheiden van niet-bedoelde effecten (zie: inleiding), waarbij hier onder effect een verandering in correlatie of rangorde van prestaties wordt verstaan.

Het ontwerp voorziet in een systematische toewijzing van leerlingen aan condities (groepen). In de eerste acht groepen maken de leerlingen twee examens met gelijke vraagvorm. Bij de volgende acht groepen is er sprake van examens met een verschillende vraagvorm. Per cluster van vier groepen maken twee groepen tweemaal eenzelfde versie van een examen en twee groepen verschillende versies van de examens.

4. VOORGENOMEN ANALYSES

Voor beide examens worden per vraagvorm, per tijdstip en per versie beschrijvende statistieken berekend in steeds vier groepen te zamen (zie Tabel 2). Bepaald worden gemiddelde, standaarddeviatie, scheefheid en gewelfdheid van de scoreverdeling en coëfficiënt alpha als ondergrensschatting van de score-betrouwbaarheid.

Tabel 2. Vier Facetten Ontwerp voor de Afname (A) van Natuurkunde en Aardrijkskunde (St = numer substeekproef; N_n = aantal leerlingen natuurkunde; N_a = aantal leerlingen aardrijkskunde).

Vraagvorm:			O																G			
Tijdstip:			T1				T2				T1				T2							
Versie:			V1		V2		V1		V2		V1		V2		V1		V2					
Subversie:			P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2				
St	N _n	N _a																				
1	25	6	A	A			A	A														
2	37	15			A	A			A	A												
3	37	16	A	A					A	A												
4	35	13			A	A	A	A														
5	22	8									A	A					A	A				
6	34	14											A	A						A	A	
7	32	11									A	A			A	A				A	A	
8	30	9											A	A			A	A				
9	43	14	A	A											A	A	A	A				
10	26	9			A	A											A	A				
11	42	17	A	A																A	A	
12	42	14			A	A											A	A				
13	37	14					A	A			A	A										
14	23	7							A	A			A	A								
15	41	16					A	A					A	A								
16	44	14							A	A	A	A										

4.1. Covariantie-structuur analyses

De tweede groep beschrijvende gegevens bestaat uit covariantie-matrices voor zestien groepen natuurkunde examinandi en zestien groepen aardrijkskunde examinandi (deze matrices worden niet integraal weergegeven). Elke covariantie matrix is van de orde vier en betreft de somscores op vragen uit twee parallelle subversies op het eerste tijdstip en twee parallelle subversies op het tweede tijdstip.

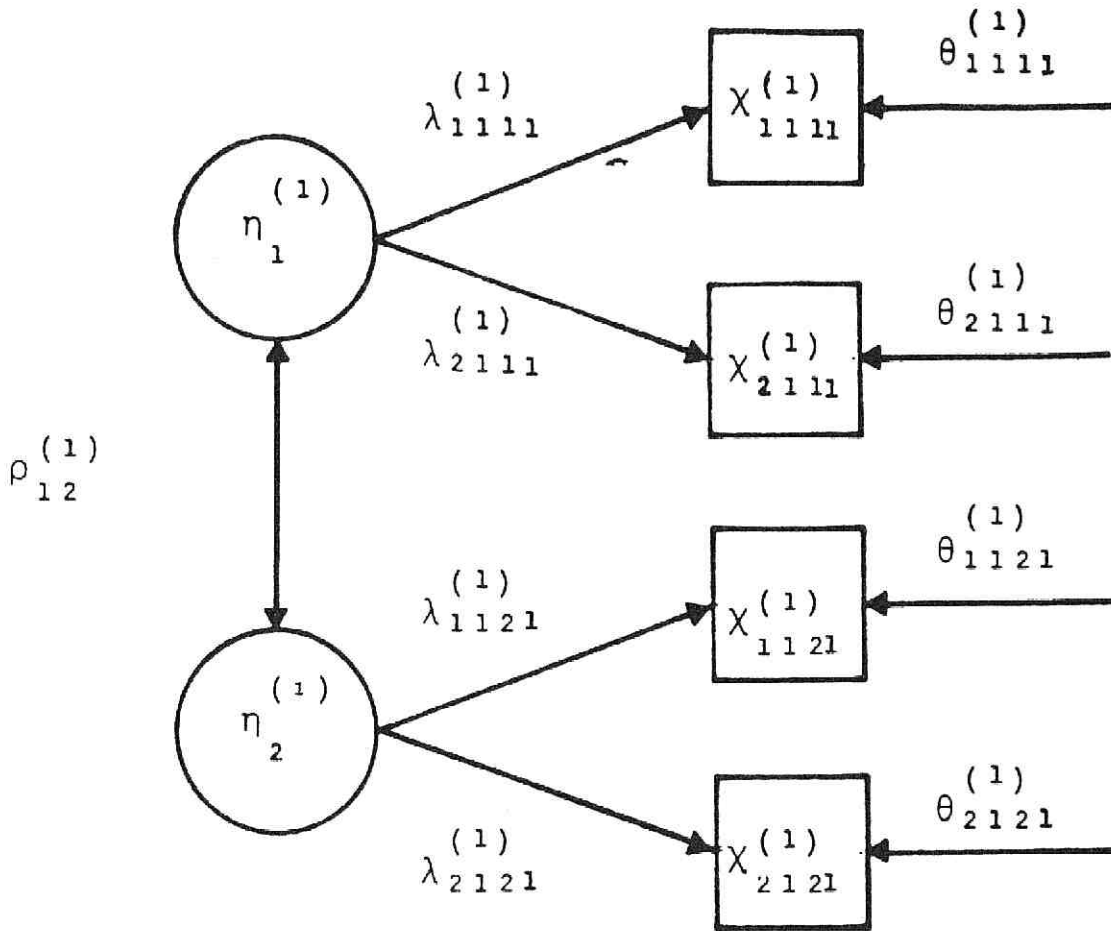
Zestien covariantie matrices zullen steeds simultaan worden geanalyseerd met behulp van meergroepen LISREL modellen (Jöreskog & Sörbom, 1986). Achtereenvolgens worden acht modellen geanalyseerd. Deze zijn in elkaar genest. In de reeks wordt van model tot model een restrictie opgeheven. Het eerste model is uitermate restrictief terwijl het achtste betrekkelijk veel parameters ongespecificeerd laat. Omdat de reeks volledig vastligt wordt het risico vermeden dat op basis van de resultaten modelaanpassingen worden verricht die onderhevig zijn aan toevalligheden.

We lichten de modellen toe met behulp van één van de zestien bestanddelen. Dit betreft de eerste groep (zie Tabel 2). Een getekende weergave staat in Figuur 1.

Het bestanddeel in Figuur 1 is een voorbeeld van een meetmodel, gebaseerd op de klassieke testtheorie (Jöreskog, 1971; Van den Bergh & Eiting, 1988). We schrijven de waargenomen somscore voor een subversie als x_{ijkl} , waarbij de index i ($i = 1, 2$) de parallelle subversie aangeeft, de index j ($j = 1, 2$) de versie, k ($k = 1, 2$) het afnametijdstip en l ($l = 1, 2$) de vraagvorm. De superindex g ($g = 1, 2, \dots, 16$) markeert de groep, of beter gezegd de subpopulatie, waarbinnen de somscore een stochastische variabele is. Door de toewijzing van groepen aan condities (zie: Tabel 2) zijn met de superindex g de indices j en l vastgelegd. Bijvoorbeeld: in de eerste groep (zie Figuur 1) hebben j en l altijd de waarde één, omdat op beide tijdstippen dezelfde versie met dezelfde vraagvorm afgenomen is.

Per tijdstip definiëren we een ware score $\eta_k^{(g)}$ met gemiddelde nul en variantie één. De correlatie van de ware scores schrijven we als $\rho_{12}^{(g)}$. De regressie van waargenomen op ware score duiden we aan met $\lambda_{ijkl}^{(g)}$. De restscore is $\epsilon_{ijkl}^{(g)}$ met variantie $\theta_{ijkl}^{(g)}$. De regressie-vergelijking in het meetmodel luidt:

$$x_{ijkl}^{(g)} = \lambda_{ijkl}^{(g)} * \eta_k^{(g)} + \epsilon_{ijkl}^{(g)}. \quad (1)$$



Figuur 1. Een van de Zestien Modelbestanddelen (weergegeven is het model voor groep 1).

Stellen we dat de restscores en ware scores ongecorrleerd zijn, dan kan de covariantiematrix van de geobserveerde scores ($\Sigma^{(g)}$; orde vier maal vier) geschreven worden als:

$$\Sigma^{(g)} = \Lambda^{(g)} P^{(g)} \Lambda'^{(g)} + \theta_{v^{(g)}} \quad (2)$$

waarin $\Lambda^{(g)}$ de matrixweergave van de regressie van de geobserveerde op de ware scores is (orde vier maal twee), $P^{(g)}$ de correlatiematrix van de ware scores is (orde twee maal twee) en $\theta_{v^{(g)}}$ de covariantiematrix van de restscores is (orde vier maal vier).

Vergelijking (1) is weinig restrictief. Alle te analyseren modellen bevatten meer restricties, die zowel binnen groepen als over groepen van toepassing zijn. Sommige daarvan zijn altijd van toepassing tot en met het achtste model, andere zijn voorwerp van onderzoek.

Invariante restricties

Parallele subtests worden gekenmerkt door gelijke betrouwbaarheden en varianties. Daarom zijn binnen groepen de λ -parameters van de parallelle subtests steeds gelijk en zo ook de θ -parameters. Hiermee zijn effecten van het aselechte facet van parallelle subversies uitgeschakeld. Ook over groepen zijn enkele restricties invariant. Het betreft de groepen die op hetzelfde tijdstip exact hetzelfde examen (gelijke versie en gelijke vraagvorm) maakten. Derhalve zijn voor telkens vier groepen de λ - en θ -parameters gelijk (zie Tabel 2).

Variabele restricties

We onderscheiden acht modellen: model Ia tot en met IVb. Het eerste model is een parallelliteitsmodel. We onderscheiden model Ia en Ib. In model Ia nemen we aan dat alle restscores ongecorrleerd zijn. De waargenomen scores worden uitsluitend verklaard door de ware scores; geheugeneffecten e.d. spelen geen rol. Verder nemen we aan dat alle regressies van waargenomen op ware scores gelijk zijn (binnen en over groepen). Hetzelfde geldt de variantie der restscores. Tenslotte nemen we aan dat de ware score correlatie over tijdstippen in alle groepen gelijk is aan één. Op beide tijdstippen meten we steeds hetzelfde attribuut, ongeacht de vraagvorm.

De volgende modellen kunnen beschreven worden door aan te geven welke restrictie uit het direct daaraan voorafgaande model wordt opgeheven. Er is sprake van een reeks geneste modellen.

In model Ib nemen we niet meer aan dat alle restscores ongecorrleerd zijn. Covariërende restscores verwachten we bij tweemaal afnemen van dezelfde examenversie, bij voorbeeld wegens geheugeneffecten. De variantie-covariantiematrix van de restscores (de symmetrische matrix $\theta(s)$ van de orde vier) bevat zes covarianties. Twee daarvan zijn covarianties van restscores op twee parallele subtests. Deze zijn per definitie nul. Van de overige vier nemen we aan dat ze wel gelijk zijn aan elkaar binnen elke groep maar niet noodzakelijk nul, wanneer tweemaal dezelfde examenversie is afgenomen. Voor steeds twee groepen veronderstellen we gelijke covarianties. Afhankelijk van de vraagvorm van het examen (open - open, gesloten - gesloten, open - gesloten en gesloten - open) mogen deze covarianties over groepen verschillen (zie: Tabel 2).

Model IIa en IIb noemen we tau-equivalentie modellen. In model IIa wordt de restrictie uit het voorgaande model (Ib) opgeheven dat de varianties van de restscores binnen groepen voor beide examens gelijk zijn. Wel blijft deze restrictie over groepen gehandhaafd. In feite wordt hiermee parallelliteit over tijdstippen als aanname opgeheven. Volstaan wordt met tau-equivalentie over tijdstippen. Opgemerkt moet worden dat in acht groepen vraagvorm en tijdstippen gecontamineerd zijn, zodat een eventueel tijdeffect niet duidelijk te onderscheiden is van een vraagvormeffect.

In model IIb wordt parallelliteit betreffende tijdstip en vraagvorm volledig opgeheven door ook toe te staan dat restscore-varianties van examens met verschillende vraagvorm mogen verschillen van groep tot groep. Zijn examens met gelijke vraagvorm op gelijke tijdstippen toegepast, dan blijven de restscore-varianties over groepen uiteraard gelijk aan elkaar.

Model IIIa en IIIb noemen we congenericiteitsmodellen. Analooq aan de voorafgaande twee worden hier restricties opgeheven, nu met betrekking tot de regressies van waargenomen op ware scores.

Model IVa en IVb noemen we ongelijke-attribuutmodellen. In model IVa nemen we niet langer aan dat de ware score correlatie tussen de prestaties op twee tijdstippen gelijk is aan één. Wel zal deze voor alle groepen dezelfde waarde hebben. In dit model specificeren we dus dat met het verstrijken van de tijd tussen twee afnamen het gemeten attribuut verandert. In model IVb tenslotte wordt de centrale hypothese van dit artikel getoetst. Hierin wordt toegestaan dat de ware score correlatie in groepen, die twee keer dezelfde vraagvorm kregen voorgelegd, een andere is dan in groepen met twee verschillende vraagvormen. Overigens wordt hier geen onderscheid gemaakt naar het soort vraagvorm: alleen gelijk óf verschillend telt.

De acht onderscheiden modellen kunnen worden geanalyseerd met een maximum likelihood (ML) methode. In dat geval kan een toetsingsgrootte worden bepaald (Lawley & Maxwell, 1971) die bij benadering chi-kwadraat is verdeeld indien (a) de steekproef voldoende groot is, (b) de waargenomen scores multivariaat normaal verdeeld zijn en (c) het model waar is. Is de overschrijdingskans van de grootte volgens de chi-kwadraatverdeling kleiner dan bijvoorbeeld vijf procent, dan besluiten we dat de toetsingsgrootte zeer waarschijnlijk niet chi-kwadraat verdeeld is. Indien omtrent de voorwaarde (a) en (b) enige zekerheid bestaat, volgt de conclusie dat het model verworpen moet worden. Bestaat met name omtrent voorwaarde (b) onzekerheid, dan zullen ook analyses worden uitgevoerd via de methode van ongewogen kleinste kwadraten (ULS).

Bij geneste modellen kan van model tot model het verschil van de resulterende toetsingsgroottes geëvalueerd worden. Dit verschil is bij benadering chi-kwadraat verdeeld, indien aan de

voorwaarden (a) en (b) is voldaan en, strikt genomen, het best passende model ook het ware model is. Om die reden wordt de onderzoekshypothese pas aan het einde van de reeks modellen getoetst. Wel geeft de verschil-grootte in het algemeen een goede aanwijzing voor het belang van een opheven restrictie.

4.2 Exploraties

De laatste analyses hebben tot doel inzicht te verkrijgen in de gehanteerde oplossingsprocessen bij open en meerkeuze-vragen. Als blijkt dat het oplossingsproces bij open en meerkeuze-vragen hetzelfde is dan verwachten we dat de antwoorden van leerlingen op de open vragen een zekere mate van samenhang vertonen met de gekozen antwoordalternatieven bij de vragen van het (equivalente) gesloten examen. Dit natuurlijk onder de aanname dat het door de leerlingen gegenereerde antwoord één van de aangeboden alternatieven is bij de gesloten vraag.

Om de samenhang tussen de antwoorden van leerlingen op de open en gesloten vragen te bepalen, worden de vier antwoordcategorieën van de gesloten vraagvorm gebruikt om de antwoorden op de open vraagvorm te categoriseren. Antwoorden op open vragen die niet op deze wijze gecategoriseerd kunnen worden, krijgen een aparte code. Deze code wordt ook toegekend aan open of gesloten vragen die door de leerlingen worden overgeslagen. Vervolgens wordt tussen elke open vraag en zijn gesloten equivalent de samenhang berekend met behulp van Cramers' V^2 , daar deze relatief ongevoelig is voor het aantal cellen in de kruistabel als ook voor de grootte van de steekproef (Hays, 1974, p. 745).

5. RESULTATEN EN DISCUSSIE

In Tabel 3 worden enkele beschrijvende maten weergegeven van de examens aardrijkskunde en natuurkunde. Per vakgebied worden per vraagvorm (open, gesloten), per tijdstip (T1, T2) en per versie (V1, V2) het aantal items, het gemiddelde, de standaarddeviatie, de scheefheid, de gewelfdheid en het aantal leerlingen gepresenteerd. Tevens wordt voor elk examen (per vraagvorm, per tijdstip en per versie) de coëfficiënt alpha weergegeven.

Tabel 3. Gemiddelde ($\bar{X}$), Standaardafwijking (sd), Scheefheid (sf), Gewelfdheid (gf), Steekproefgrootte (N), Coëfficiënt Alpha (α) en het Aantal Vragen van de Examens Natuurkunde en Aardrijkskunde per Vraagvorm (Open, Gesloten), per Tijdstip (T1, T2) en per Versie (V1, V2).

Vakgebied	Vraagvorm	Tijdstip	Versie	$\bar{X}$	sd	sf	gf	N	α	Aantal vragen
Natuurkunde	open	T1	V1	5.43	3.05	1.26	3.35	147	.61	28
	open	T1	V2	6.30	3.52	.59	.78	140	.70	28
	open	T2	V1	4.02	2.17	.29	.38	138	.49	28
	open	T2	V2	2.42	1.64	.70	.11	141	.39	28
	gesloten	T1	V1	10.37	3.61	-.35	.81	135	.58	28
	gesloten	T1	V2	11.42	3.96	-.70	1.21	128	.65	28
	gesloten	T2	V1	7.02	2.99	-.72	.30	137	.53	28
	gesloten	T2	V2	4.93	2.16	.43	1.82	134	.30	28
Aardrijkskunde	open	T1	V1	6.25	3.23	.20	-.28	53	.63	28
	open	T1	V2	8.57	4.03	.34	-.01	51	.72	32
	open	T2	V1	5.63	3.49	.05	-.79	49	.71	28
	open	T2	V2	8.10	4.24	.02	-.22	52	.77	32
	gesloten	T1	V1	4.87	2.45	-.37	-.52	47	.46	28
	gesloten	T1	V2	6.39	2.48	-.70	1.31	46	.29	32
	gesloten	T2	V1	5.29	2.71	-.78	2.61	45	.51	28
	gesloten	T2	V2	5.77	3.04	-.21	.15	51	.56	32

Uit Tabel 3 komt naar voren dat de meeste leerlingen vrij veel moeite hebben gehad met de examens natuur- en aardrijkskunde. Het aantal vragen dat de leerlingen gemiddeld correct beantwoord hebben is laag. Ook komt naar voren dat de interne consistentie van de examens natuur- en aardrijkskunde in de meerkeuze-vraagvorm over het algemeen zwak genoemd moet worden. De interne consistentie van de examens in de open vraagvorm voor het vak aardrijkskunde blijkt redelijk. Dat de consistentie van examens met open vragen hoger is dan die van examens met meerkeuze-vragen is in overeenstemming met de in de inleiding aangehaalde onderzoeken.

In Tabel 4 rapporteren we de toetsingsresultaten van acht modellen betreffende de natuurkunde examens. Naast de toetsingsgrootheid - t - bij ML-analyse rapporteren we de 'goodness of fit' index (gfi) en de 'root mean square residual' (RMR). Deze laatste twee rapporteren we eveneens voor de ULS analyse. Ook worden verschilgrootheden van model tot model gerapporteerd. In Tabel 5 staan de overeenkomstige resultaten betreffende de aardrijkskunde-examens.

Tabel 4. Toetsingsresultaten Natuurkunde-examens (ML = Maximum likelihood analyses; ULS = Unweighted Least Squares analyses; t = toetsingsgrootheid; gfi = goodness of fit index; RMR = root mean square residual).

Model	ML					ULS	
	t	df	p	gfi	RMR	gfi	RMR
Ia	510.56	158	.00	.807	.85	.839	.85
Ib	498.74	154	.00	.813	.85	.840	.85
IIa	397.43	153	.00	.915	.36	.969	.31
IIb	331.26	151	.00	.916	.32	.958	.39
IIIa	265.86	150	.00	.930	.40	.917	.50
IIIb	258.33	148	.00	.928	.50	.909	.42
IVa	242.96	147	.00	.894	.57	.880	.58
IVb	242.18	146	.00	.901	.54	.877	.59
Ia-Ib	11.82	4	.025				
Ib-IIa	101.31	1	.00				
IIa-IIb	66.17	2	.00				
IIb-IIIa	65.40	1	.00				
IIa-IIIb	7.53	2	.025				
IIIb-IVa	15.37	1	.00				
IVa-IVb	.78	1	.25				

Tabel 5. Toetsingsresultaten Aardrijkskunde-examens (voor toelichting zie: Tabel 4).

Model	ML					ULS	
	t	df	p	gfi	RMR	gfi	RMR
Ia	284.42	158	.00	.740	1.00	.745	1.00
Ib	261.54	154	.00	.747	1.07	.714	1.04
IIa	261.54	153	.00	.747	1.06	.731	.99
IIb	256.48	151	.00	.772	.88	.786	.81
IIIa	240.36	150	.00	.880	.47	.883	.61
IIIb	200.74	148	.003	.878	.49	.943	.44
IVa	201.51	147	.002	.859	.47	.942	.44
IVb	185.46	146	.015	.860	.47	.939	.45
Ia-Ib	22.88	4	.00				
Ib-IIa	.00	1	.99				
IIa-IIb	5.06	2	.10				
IIb-IIIa	16.12	1	.00				
IIIa-IIIb	39.62	2	.00				
IIIb-IVa	-.77	1	.99				
IVa-IVb	16.05	1	.00				

Uit de gegevens in Tabel 4 kan geconcludeerd worden dat de toetsingsgrootheden bij ML-analyse bij geen enkel model voor de natuurkunde examens chi-kwadraat verdeeld zijn. Omdat getwijfeld kan worden aan een multivariaat normale verdeling der scores (zie Tabel 3), hoeft dit nog niet een onvoorwaardelijke verwerping van alle modellen in te houden. Zo blijkt de 'goodness of fit' index bij zowel ML- als ULS-analyse in een aantal gevallen redelijk ($> .90$).

Bij het vergelijken van de modellen dient ook de verschil-grootheid met omzichtigheid te worden gehanteerd. Vier verschillen zijn de moeite van het interpreteren waard. Bij de overgang van model Ib op IIa wordt de restrictie opgeheven dat de restscore-varianties voor twee tijdstippen gelijk aan elkaar moeten zijn. Dit leidt tot een aanzienlijke verbetering in passing. Het blijkt dat de restscore-varianties op het tweede tijdstip kleiner zijn dan op het eerste tijdstip.

Bij de overgang van model IIa op model IIb treedt ook een flinke passingsverbetering op. Het blijkt dat genoemd fenomeen (van verschillen in restscore-varianties) vooral geldt voor de toetsen met open vragen en nauwelijks voor de toetsen met gesloten vragen.

Bij de overgang van model IIb op IIIa hoeft de regressie van waargenomen op ware scores niet meer voor beide tijdstippen gelijk te zijn. Het blijkt dat deze op het tweede tijdstip kleiner is dan op het eerste tijdstip.

Bij de overgang van model IIIb op IVa wordt de restrictie opgeheven dat de ware scores voor twee tijdstippen perfect correleren. De conclusie lijkt gerechtvaardigd dat op beide tijdstippen voor een deel ongelijksoortige attributen worden gemeten.

Van belang is de zeer geringe winst bij overgang van model IVa op IVb. De ware score correlatie bij twee keer gelijke vraagvormen verschilt nagenoeg niet van die bij twee keer verschillende vraagvormen. Deze correlatie wordt in het eerste geval geschat (ML) als .305 (se = .165), in het tweede geval als .496 (se = .140). Bij de natuurkunde-examens is geen enkel effect van de vraagvorm aangetoond. Daarentegen blijkt er wel een effect te zijn van het afnametijdstip.

In Tabel 6 rapporteren we enkele parameterschattingen in model IVb, die relevant zijn voor een betrouwbaarheidsbepaling (ρ^2). Deze is gebaseerd op de betrouwbaarheidsdefinitie (zie bij voorbeeld: Jöreskog, 1971) en de Spearman-Brown formule voor homogene verdubbeling van de testlengte (Lord & Novick, 1968, p. 90). Het betreft immers parallelle testhelften.

De gegevens in Tabel 6 laten zien dat de natuurkunde-examens op het eerste tijdstip betrouwbaarder scores opleverden dan op het tweede tijdstip. Combineren we dit gegeven met de bevinding dat varianties van de restscores (bron van onbetrouwbaarheid) op het tweede tijdstip kleiner zijn dan op het eerste, terwijl tevens de ware score correlatie kleiner is dan één, dan is er grond voor het vermoeden dat onderling overleg en oefening tussen beide afnamen heeft plaats gevonden.

Bij de toetsing van acht modellen betreffende aardrijkskunde-examens (Tabel 5) vinden we resultaten vergelijkbaar met de natuurkunde-examenmodellen. Bij de vergelijking van de modellen vallen wel enige verschillen op met die uit de voorafgaande analyses. De overgang van model Ia op Ib houdt in het toestaan van covariërende restscores bij gelijke versies. Deze covarianties blijken negatief, zij het dat ze geschat worden met een erg grote standaardfout. Het idee dat geheugeneffecten hier een rol spelen, wordt door deze bevinding niet ondersteund.

De overgang van model IIb op model IIIa suggereert dat de regressie van waargenomen op ware scores tijdstipafhankelijk is. In tegenstelling tot de natuurkunde-examens blijkt deze op het tweede tijdstip groter dan op het eerste.

Een aanzienlijke verbetering blijkt bij de overgang van model IIIa op model IIIb. Men vergelijk ook de 'goodness of fit' index bij ULS-schatting. De ware score regressie blijkt voor gesloten vragen veel kleiner dan voor open vragen.

Een effect van tijdstippen op de ware score correlatie kan niet worden aangetoond. Daarentegen vinden we gezien de passingsverbetering van model IVa op model IVb bij aardrijkskunde een duidelijke steun voor de veronderstelling dat de vraagvorm van invloed is op het aangesproken attribuut. Zo blijkt de ware score correlatie bij twee keer dezelfde vraagvorm (ML schatting) gelijk aan .794 (se = .154), en bij twee keer verschillende vraagvorm -.303 (se = .217). Volgens ons eerder uitgewerkte criterium voor een effect zouden we, uitgaande van .794, bij een correlatie $< .668$ moeten spreken van een substantieel effect. Daar is ruimschoots aan voldaan.

Tabel 6. ML-betrouwbaarheidsschattingen Natuurkunde-examens (Standaardfouten tussen haakjes).

Parameter	Open-T1		Open-T2		Gesloten-T1		Gesloten-T2	
λ	1.326	(.093)	.646	(.071)	1.513	(.110)	.791	(.109)
θ	1.989	(.168)	1.032	(.088)	2.494	(.221)	2.122	(.183)
$\rho^2 = 2\lambda^2/(2\lambda^2 + \theta)$	.64		.45		.65		.37	

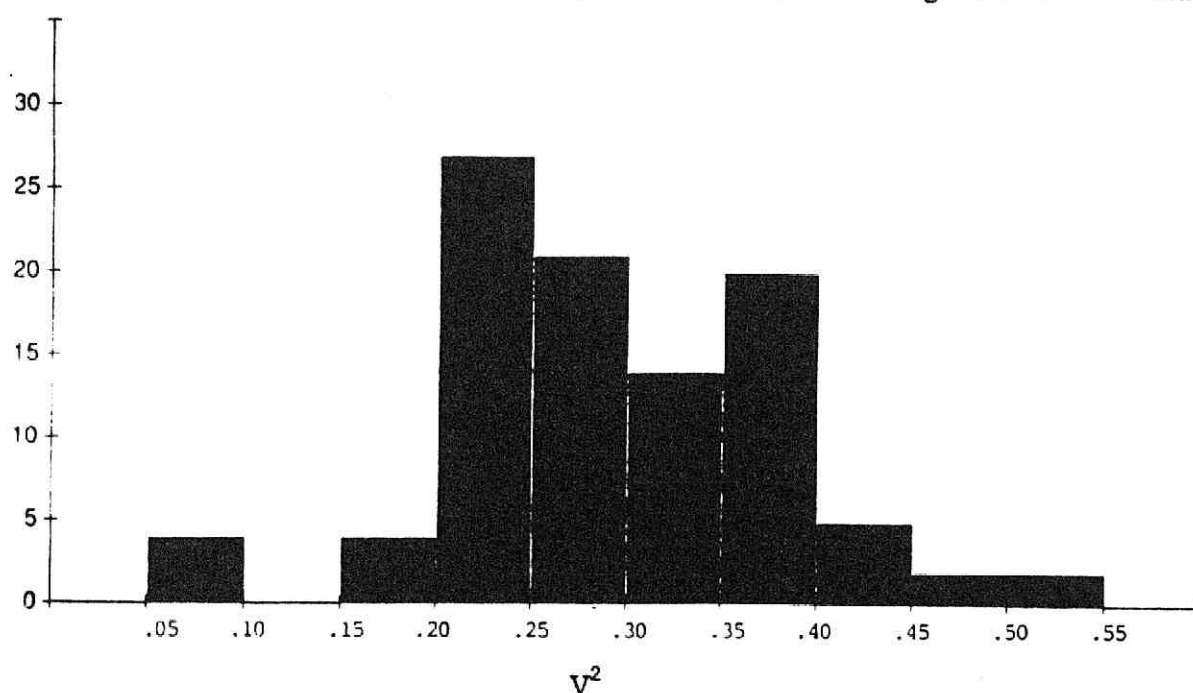
Tabel 7. ML-betrouwbaarheidsschattingen Aardrijkskunde-examens (Standaardfouten tussen haakjes).

Parameter	Open-T1		Open-T2		Gesloten-T1		Gesloten-T2	
λ	1.504	(.163)	1.617	(.165)	-.705	(.184)	1.149	(.147)
θ	2.022	(.291)	1.896	(.278)	1.956	(.289)	1.578	(.237)
$\rho^2 = 2\lambda^2/(2\lambda^2 + \theta)$	.69		.73		.34		.63	

Wel moet worden opgemerkt dat de schatting van de ware score regressie van gesloten vragen op het eerste tijdstip negatief is (zie ook Tabel 7). Absorberen we het negatieve teken in de ware scores, dan zullen deze feitelijk negatief correleren met die van de gesloten vragen op het tweede tijdstip. Hoewel de gesloten vragen op het eerste tijdstip weinig betrouwbaar blijken (zie Tabel 7), was een en ander toch aanleiding tot een exploratief model, waarin de ware score correlaties tussen alle vier de vraagvorm-combinaties vrij geschat konden worden (dus niet meer alleen voor gelijke versus ongelijke vraagvormen op beide tijdstippen). Echter, dit model blijkt nauwelijks beter te passen dan het vorige model ($t = 185.33$, $df = 144$), maar het levert beter interpreteerbare schattingen op. Laten we de onbetrouwbare gesloten vragen op het eerste tijdstip buiten beschouwing, dan vinden we voor de combinatie open (T1) - open (T2) een geschatte ware correlatie van .788 ($se = .165$) en voor de combinatie open (T1) - gesloten (T2) een correlatie van -.346 ($se = .262$). Op basis van de meer betrouwbare examens kunnen we de bovenstaande conclusie daarom handhaven.

In Tabel 7 tenslotte rapporteren we de betrouwbaarheidsschattingen voor de aardrijkskunde-examens. Deze schattingen blijken voor het tweede tijdstip groter dan voor het eerste, maar dat geldt vooral voor de gesloten vraagvormen.

In Figuur 2 wordt de frequentieverdeling gepresenteerd van de hoogte van de samenhangen



Figuur 2. Relatieve Frequentieverdeling van Cramers' V^2 voor de Samenhang van Antwoordcategorieën voor Open en Gesloten Vragen Natuurkunde.

(Cramer's V^2) tussen de gecategoriseerde antwoorden op de 56 open en gesloten vragen natuurkunde.

Uit Figuur 2 komt naar voren dat er een matige samenhang bestaat tussen het beantwoorden van een vraag in de open of gesloten vorm. De interpretatie van de hoogte van de samenhangen wordt bemoeilijkt door mogelijke rekenfouten van de leerlingen. In tegenstelling tot denkfouten zullen deze rekenfouten bij de open vragen vaak in de categorie niet-scoorbare antwoorden worden ondergebracht. Daarbij komt dat onbetrouwbaarheid van de categorisering een negatief effect heeft op de samenhangsmaten.

Bij de aardrijkskunde-examens bleek analyse van de antwoorden met contingentietabellen helaas niet mogelijk. De diversiteit van antwoorden op open vragen bleek zo groot, dat categorisering in de categorieën van de gesloten vragen niet zinvol was.

6. ALGEMENE DISCUSSIE

Onze in de inleiding vermelde hypothese is weerlegd. Wij stelden dat ongeacht het onderwerp een effect van vraagvorm zou blijken. Dit effect kon wel aangetoond worden voor de examens aardrijkskunde, doch niet voor de examens natuurkunde. Wel bleken er verschillen in attribuut afhankelijk van het tijdstip waarop de leerling het examen maakte. Verder bleken er bij de examens voor het vak natuurkunde geen verschillen in geschatte betrouwbaarheid ten gevolge van de vraagvorm. Wel werden er verschillen in betrouwbaarheid van de examens gevonden ten gevolge van afname-tijdstip: op het tweede tijdstip was de betrouwbaarheid lager dan op het eerste.

Voor de examens van het vak aardrijkskunde bleek een aanzienlijk verschil in attribuut afhankelijk van de vraagvorm. De betrouwbaarheid van de examens met open vragen bleek groter dan die van de examens met gesloten vragen. In tegenstelling tot de natuurkunde-examens bleek de betrouwbaarheid op het tweede tijdstip hoger dan op het eerste.

Een mogelijke verklaring voor deze differentiële effecten van vraagvorm kan gevonden worden in de ook door andere auteurs naar voren gebrachte veronderstelling, dat zowel herkennen als herinneren een rol spelen bij het beantwoorden van examenvragen. Men gaat er dan vanuit dat deze processen verschillend zijn. Zo zou bij het herkennen ook partiële kennis een rol spelen. Bij het beantwoorden van open vragen speelt deze partiële kennis nauwelijks een rol. Het belang van partiële kennis bij het beantwoorden van meerkeuze-vragen is afhankelijk van de mate waarin de gegeven alternatieven het oplossingsproces van de leerling sturen. Het oplossingsproces bij het beantwoorden van de natuurkunde-vragen is onafhankelijk van het aanbieden van alternatieven; de leerling moet hetzelfde oplossingsproces volgen, ongeacht het feit of het uiteindelijke antwoord gegeven is of niet. De leerling moet meestal een 'recept' uitvoeren bij een gegeven opdracht, dat uiteindelijk tot het juiste antwoord leidt. We vermoeden dat het aanbieden van alternatieven bij aardrijkskunde-vragen wel een sturende werking heeft op het oplossingsproces van de leerling. Immers, de aanwijzingen, verschaft door de antwoordalternatieven, kunnen het zoekproces in het geheugen aanzienlijk ondersteunen.

Tot nog toe hebben we een vakinhoudelijk onderscheid gemaakt (natuur- versus aardrijkskunde), ook in deze laatste redenering. Dit is misschien niet terecht. Het is waarschijnlijk beter om een onderscheid te maken tussen typen vragen; vragen waarbij een recept toegepast kan worden en vragen waarbij de aanwezige kennis in de een of andere vorm gereproduceerd moet worden. Ook bij gesloten aardrijkskunde-vragen is recept-uitvoering mogelijk, hoewel dit bij onze examens sporadisch voorkwam. Een voorbeeld van een dergelijke vraag voor het vak aardrijkskunde is: 'Bereken met behulp van de volgende tabel het geboorte-overschot in de gemeente Amsterdam in 1984'.

Bovenstaande interpretatie wordt gesteund door de betrouwbaarheidsschattingen. Wanneer de verschillende items van een toets een beroep doen op andere vaardigheden, dan is de homogeniteit van deze toets lager dan die van een toets waarin alle items in gelijke mate een beroep doen op dezelfde vaardigheden. De geschatte betrouwbaarheden van de natuurkunde-examens verschillen niet ten gevolge van vraagvorm, hetgeen bij de examens voor het vak aardrijkskunde wel het geval

bleek te zijn en wel precies in de goede richting: de examens met gesloten vragen bleken minder betrouwbaar dan die van de examens met open vragen.

Resten ons nog de lage ware score correlaties tussen de natuurkunde-examens op verschillende tijdstippen. Deze lage correlatie is misschien te wijten aan de wijze waarop de leerstof in de klas behandeld wordt. De leerstof voor het vak natuurkunde wordt sterk gefaseerd en gestructureerd behandeld, terwijl dit bij aardrijkskunde misschien minder het geval is (zie Van den Bergh & Triesscheijn (1988) voor aanwijzingen dienaangaande). Bij het vak aardrijkskunde, zo is ons vermoeden, wordt de leerstof meer verspreid behandeld.

LITERATUUR

- Anastasi, A. (1958). *Differential Psychology*. New York: McMillan.
- Benson, J. (1981). A redefinition of content validity. *Journal of Educational and Psychological Measurement*, 741, 793-802.
- Bergh, H. van den (1987). Open en meerkeuze-vragen tekstbegrip: een onderzoek naar de relaties tussen prestaties op open en meerkeuze-vragen. *Tijdschrift voor Onderwijsresearch*, 12, 304-312.
- Bergh, H. van den & Eiting, M.H. (1988 in press). The estimation of rater reliability. *Journal of Educational Measurement*.
- Bergh, H. van den & Triesscheijn, B. (1988). *Effecten van Examenveranderingen: Een Onderzoek naar de Tijdbesteding in de Lessen Nederlands en Natuurkunde in Eindexamenklassen MAVO en LBO*. Amsterdam: SCO.
- Bloom, B.S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill Book Company.
- Bonset, H. (1973). CITO-materiaal opnieuw kritisch getoetst. *Moer*, 4, 142-146.
- Bonset, H. (1974). Objectieve schrijfvaardigheidstoetsen voor basis- en voortgezet onderwijs. *Levende Talen*, 306, 204-218.
- Bonset, H. (1975). Enkele sociolinguïstische bezwaren tegen objectieve toetsen voor taalvaardigheid. *Levende Talen*, 313, 335-345.
- Bonset, H. & Griffioen, J. (1972). Met toets en band door het ganse land. *Vernieuwing van Opvoeding en Onderwijs*, 311, 140-151.
- Bracht, C.H. & Hopkins, K.D. (1970). The communality of essay and objective tests of academic achievement. *Educational and Psychological Measurement*, 30, 359-364.
- 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.
- Cheketts, K.T. & Christensen, M.G. (1974). The validity of awarding credit by examination in English composition. *Educational and Psychological Measurement*, 33, 357-361.
- Choppin, B.M. (1968). *The report of the crossover study*. Ongepubliceerd rapport: IEA/TR/6: London.
- Cook, D.C., geciteerd in Ebel, R.L. (1965). *Measuring Educational Achievement*. Englewood Cliffs, N.J.: Prentice-Hall.
- 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.
- Culpepper, M.M. & Ramsdale, R. (1983). A comparison of multiple-choice and essay test of writing skills. *Research in the Teaching of English*, 41, 295-298.
- Daniëls, J.C. (1954). Testing geography at the ordinary level of the general certificate of education. *The British Journal of Educational Psychology*, 24, 180-189.
- Davis, F.B. & Fiffer, G. (1957) The effect of test reliability and validity of scoring aptitude and achievement tests with weights for every choice. *Educational and Psychological Measurement*, 19, 159-170.
- Emmett, W.G. (1954). Secondary modern and primary school performance predicted by tests given in primary schools. *The British Journal of Educational Psychology*, 24, 91-98.
- Frary, R.B. (1985). Multiple-choice versus free-response: A simulation study. *Journal of Educational Measurement*, 22, 21-31.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Gates, A.I. (1921). The true-false test as a measurement of achievement in college courses. *Journal of Educational Psychology*, 12, 276-287.
- Hays, W.L. (1974). *Statistics for the social sciences* (2nd edition). London: Holt, Rinehart & Winston.
- 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.
- Hopkins, K.D., Dobson, J.C. & Oldridge, O.A. (1962). The concurrent and congruent validities of the wide range achievement test. *Educational and Psychological Measurement and Evaluation*, 22, 791-793.
- Hopkins, K.D. & Stanley, J.C. (1981). *Educational and Psychological Measurement (6th ed.)*. Englewood Cliffs, N.J.: Prentice-Hall.
- Houtman, C. (1986). Bevrijd ons van het meerkeuze-examen. *Levende Talen*, 412, 367-369.
- Hurd, A.W. (1932). Comparison of short-answer and multiple-choice test covering identical subject content. *Journal of Educational Research*, 26, 28-30.
- Hurlburt, D. (1954). The relative value of recall and recognition techniques for measuring precise knowledge of word meaning-nouns, verbs, adjectives. *Journal of Educational Research*, 47, 561-567.
- Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-132.
- Jöreskog, K.G. & Sörbom, D. (1986). *LISREL: Analysis of Linear Structural Relationships by the Method of Maximim Likelihood (Version VI)*. Uppsala: University of Uppsala.
- Kempa, R.F. & L'Odiaga, J. (1984). Criterion-referenced interpretation of examination grades. *Educational Research*, 26, 56-64.
- Knight, F.B. (1922). Data on true-false test as a device for college examination. *Journal of Educational Psychology*, 13, 75-80.
- Koffler, S.L. (1986). *Assessing students' writing skills: A comparison of direct and indirect methods*. Paper presented at the 68th Annual Meeting of the American Educational Research Association. New Orleans, L.A.
- Kruglak, H. (1965). Experimental study of multiple-choice and essay tests: *American Journal of Physics*, 33, 1036-1041.
- Lawley, D.N. & Maxwell, A.E. (1971). *Factor Analysis as a Statistical Method*. London: Butterworths.
- Lord, F.M. & Novick, M.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.
- Nisbett, J.D. (1965). English composition in secondary school selection. *The British Journal of Educational Psychology*, 25, 51-54.
- Nußbaum, A. (1984). Multivariate generalizability theory in educational measurement: an empirical study. *Applied Psychological Measurement*, 8, 219-230.
- Peel, E.A. & Armstrong, H.G. (1956). Symposium: The use of essays in selection of the 11+ examination. *The British Journal of Educational Psychology*, 27, 163-171.
- Peters, G.G. & Martz, H.B. (1931). A study of the validity of various types of examinations. *School and Society*, 33, 336-338.
- Peterson, D.L. (1926). Do old and new type examinations measure different mental functions? *School and Society*, 24, 246-248.
- Rowley, G.C. (1974). Which examinees are most favoured by the use of multiple-choice tests. *Journal of Educational Measurement*, 11, 15-23.
- Ruch, C.M. & Stoddard, G.D. (1927). *Test and Measurement in High School Instruction*, Yonkers: World Book.
- Shalson, V. & Crawford, C.C. (1928). Experimental comparisons of true-false and completion tests. *Journal of Educational Psychology*, 19, 580-583.
- Shohamy, E. (1984). Does the testing method makes a difference? The case of reading comprehension. *Language Testing*, 1, 147-169.
- Svinicki, M. & Koch, B. (1983). The great essay/multiple choice debate: different strokes for different folks. *Center of Teaching Effectiveness Newsletter*, 5, 2.
- Toops, H.A. (1921). *Trade Tests in Education* (Teachers college contribution 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, 335-369.
- Vernon, P.E. (1962). The determination 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 aptitude tests. *Applied Psychological Measurement*, 6, 1-11.

- Weiss, D. & Jackson, R. (1983). *The validity of descriptive tests of language skills: Relationships to direct measures of writing skill ability and grades in introductory college English courses*. College Entrance Examination Board, New York, N.Y. (report: LB-83-4; ETS-RR-83-27).
- Westrhenen, J. van (1973). *Eindrapport proefexamens 1973*. Arnhem: CITO.
- Williams, R.C., Davis, W., Anderson, P. & Favor, K. (1978). Test format as a form of bias for black students. *Journal of Non-white Concerns in personal Guidance*, 6, 141-147.
- Wiseman, S. (1956). Symposium: The use of essays in the selection at 11+, III-Reliability and Validity. *The British Journal of Educational Psychology*, 26, 172-179.

Manuscript ontvangen 18-3-1988

Definitieve versie ontvangen 1-7-1988.

Notities en Commentaren

DE WENS NAAR EEN NAUWERE SAMENHANG TUSSEN O.T.G.- ONDERZOEK EN OPDRACHTONDERZOEK.

P. Vedder

*Instituut voor Onderzoek van het Onderwijs, Den Haag**

ABSTRACT

(Adapted for persons not conversant with the organisation of educational research in the Netherlands)

In Dutch educational research, due to the funding system, two types of research may be distinguished. The first type is research initiated by researchers themselves. This research may be either more fundamental, theoretical research or more practice-oriented research. The second type is research that is set up in accordance with wishes of policy makers and representatives of schoolboards. This latter type is normally practice-oriented and seldom fundamental or theoretical in nature. In a recent discussion researchers involved in the first type of research suggested that they should be enabled to help in structuring the organisation of the second type of research. In this article this wish is evaluated as not feasible and undesirable in view of the primary task of the first type of research, viz. contributing to theory building and to the explanation of educational phenomena.

1. INLEIDING

In de VOR-publicatie 'Naar een strategische programmering van onderwijsonderzoek' (Beintema, Van den Dool, Roede & Van der Sijde, 1988) worden in de bijdragen van respectievelijk Roede en Van den Dool voorstellen gedaan om de S.V.O.- programmering te verbeteren door middel van op de O.T.G.'s gerichte of door de O.T.G.'s genomen initiatieven. Beide auteurs doen voorstellen om tot een betere afstemming van O.T.G.-en opdrachtonderzoek te komen en om ervaringen opgedaan met de programmering en beoordeling van onderzoek in O.T.G.-kaders, vruchtbaar te maken voor het opdrachtonderzoek. Daarbij wordt de O.T.G. vereenzelvigd met het streven naar cumulatie van wetenschappelijke kennis en het, in aansluiting bij theorieën, zoeken naar verklaringen voor verschijnselen in onderwijsland.

Een nauwere samenhang tussen O.T.G.- en opdrachtonderzoek veronderstelt dat er door onderzoekers toenaderingsactiviteiten worden ondernomen of dat de toenaderingsactiviteiten door hoger hand, de subsidiegevers, worden vastgelegd in de keuze en prioritering van te subsidiëren onderzoeken. Roede en Van den Dool schrijven duidelijk over de eerste optie. De voorstellen die zij doen zijn min of meer op te vatten als voorstellen voor O.T.G.- gebiedsuitbreiding. Roede stelt voor om het aantal O.T.G.'s uit te breiden, om te bezien in hoeverre veld-en beleidgerichte onderzoeksvragen uitgewerkt kunnen worden in O.T.G.-kaders en om de programmeringsstudies van S.V.O. uit te laten voeren door O.T.G.-onderzoekers. Van den Dool stelt voor dat O.T.G.'s onderzoeksvragen gaan formuleren die in serie kunnen worden uitge-

* De in dit artikel gepresenteerde evaluatie komt voor rekening van de auteur en mag niet worden opgevat als het oordeel en de mening van het Instituut voor Onderzoek van het Onderwijs. Dank aan Loes van Tilborg die commentaar gaf bij een eerdere versie van dit artikel.
SVO, Sweelinckplein 14, 2517 GK 's-Gravenhage

voerd in het kader van de opdrachtresearch. O.T.G.'s zouden verder bepaalde problemen die in het kader van het opdrachtonderzoek worden gepresenteerd kunnen overnemen en uitwerken en onderzoeken in O.T.G.-kaders. Een derde mogelijkheid die hij presenteert houdt in dat er in het domein van de opdrachtresearch een aantal steeds weer terugkerende problematieken en thematieken worden gedefinieerd die aanleiding zouden kunnen zijn tot een organisatievorm die sterk doet denken aan O.T.G.'s. Dit commentaar is geschreven met de bedoeling om duidelijk te maken dat met betrekking tot de vraag of een nauwere samenhang/afstemming wenselijk is de consequenties voor het edele streven naar de groei van kennis niet zijn geanalyseerd. In dit commentaar geef ik een voorzet in de hoop dat anderen met veel geduld en inzet aan de bal willen blijven.

In paragraaf twee ga ik in op de wijze waarop de probleemverkenner van S.V.O., in eerste instantie, en de aanvragers (O. & W. en veld) samen met de beoordelingscommissie opdrachtonderzoek, in tweede instantie, omgaan met een mogelijke samenhang met O.T.G.-onderzoek. In de laatste paragraaf kom ik terug op de vraag naar de wenselijkheid van een nauwere samenhang in relatie tot de hiervoor beschreven voorstellen. In deze paragraaf ga ik ook in op de wijze waarop in het kader van de WOV mogelijk inhoud kan worden gegeven aan de samenhang tussen dat wat tegenwoordig O.T.G.- en A.N.G.O.-onderzoek is en opdrachtonderzoek.

2. VERKLARING EN THEORIEVORMING; SLUITSTUKKEN VAN HET OPDRACHTONDERZOEK

Hiervoor werd de wenselijke aard van O.T.G.-onderzoek omschreven, waarbij het streven naar de groei van kennis en het zoeken naar verklaringen voorop staat. In mijn speurtocht om te achterhalen in hoeverre deze groei- en verklaringsdrift ook geldt voor opdrachtresearch, heb ik mij, uit efficiëntie overwegingen, gericht op de onderzoeksvoorstellen die zijn geformuleerd in de programmeringsstudies 1 tot en met 9. Het gaat om in totaal 106 voorstellen.

Bij mijn speurtocht heb ik drie criteria gehanteerd:

1. In het voorstel wordt expliciet als doel geformuleerd het toetsen of ontwikkelen van theorieën/theoretische modellen.
2. In het voorstel wordt als probleem beschreven het verklaren van een sociaal fenomeen.
3. In het voorstel wordt expliciet als taak van het onderzoek beschreven het ontwikkelen en valideren van een onderzoeksinstrument.

De eerste twee criteria hebben een evident verband met de hiervoor beschreven wenselijke aard van O.T.G.-onderzoek. Bij het derde criterium is dat wellicht minder het geval. Derhalve een toelichting. In de eerste plaats is validatie van onderzoeksinstrumenten, met name van tests en dergelijke, veelal ten nauwste verbonden met theoriegebonden conceptuele analyses. Testontwikkeling en -validatie vraagt om het aanscherpen van een concept. Het concept wordt verbijzonderd in termen van gedrag. Aan de andere kant zijn de analyse en de interpretatie van de resultaten van het gebruik van een test-in-ontwikkeling op te vatten als een toetsing van de achterliggende theorie. In de tweede plaats is het in het kader van het streven naar groei van kennis van belang dat er meer aandacht komt voor de standaardisatie van onderzoeksinstrumenten, waardoor de vergelijkbaarheid van studies toeneemt.

Het gebruik van de drie criteria leverde bij mij het volgende resultaat op. Van de 106 onderzoeksvoorstellen die zijn gedaan in de negen programmeringsstudies hebben naar mijn oordeel 12 betrekking op onderzoek dat als fundamenteel gekwalificeerd zou kunnen worden. Van deze 12 zijn daadwerkelijk vijf in uitvoering genomen, twee in O.T.G.-kaders en drie in het kader van opdrachtresearch. Wat de sector opdrachtresearch betreft lijkt het erop dat er door de aanvragers, zoals te verwachten, weinig op fundamenteel (op verklaring gericht) onderzoek wordt aangedrongen.

De conclusie lijkt gerechtvaardigd dat aanvragers, probleemverkenner (prioritering) en bestuurders (subsidieringsbeslissingen) het er binnen het kader van de opdrachtresearch niet zo op hebben om fundamenteel onderzoek te doen. De opstellers van de programmeringsstudies

die zelf veelal opdrachtonderzoeker zijn, beleiden weliswaar veelvuldig het belang van fundamenteel onderzoek in het kader van de opdrachtresearch met de pen, maar doen weinig concrete voorstellen om de aandacht voor fundamenteel onderzoek binnen het kader van het opdracht-onderzoek te versterken.

Een algemene opmerking over de programmeringsstudies tot slot. Eén van de middelen waarmee S.V.O. hoopt in het kader van deze studies zicht te krijgen op de belangrijkste handelingsproblemen waarmee potentiële S.V.O.-cliënten geconfronteerd (zullen) worden op een bepaald werkveld is de reconstructie van reeds doorgevoerde of vastgestelde ingrepen en achterliggende ideeën. Dit middel wordt mijns inziens tot nog toe door probleemverkenner op zeer oppervlakkige wijze gebruikt. Het lijkt er veelal op dat er, voorzover er iets aan een theoretische reconstructie wordt gedaan, veeleer sprake is van het presenteren van een vooropgezet model en het geven van referenties die betrekking hebben op uitgevoerd onderzoek. De relatie (spanning) tussen de onderzoeken en het model wordt niet geanalyseerd en niet aangegrepen voor het formuleren van onderzoeksvragen.

3. EEN GEWENSTE NAUWERE SAMENHANG?

De wens van Roede en Van den Dool om de invloedssfeer van O.T.G.- onderzoek te vergroten lijkt niet te worden gedeeld door personen en instanties die het voor het zeggen hebben op het gebied van de opdrachtresearch. Het heeft er een beetje de schijn van dat de wetgever wel aan de kant van Roede en Van den Dool staat. In de wet op de onderwijsverzorging (WOV) is expliciet sprake van een samenhang. In de WOVI staat te lezen:

‘Het instituut voor Onderzoek van het Onderwijs heeft tot taak:

- a) het op verzoek van onderwijsorganisaties, andere instellingen en erkende perifere instellingen doen uitvoeren van onderzoek ten dienste van het onderwijs;
- b) het op verzoek van Onze minister in het kader van het landelijk onderwijsinnovatiebeleid adviseren daarover en opstellen, onderscheidenlijk mee opstellen van uitvoeringsplannen, alsmede het doen uitvoeren, onderscheidenlijk doen mee uitvoeren van deze plannen door onderzoeksinstellingen;
- c) het -ten dienste van het onderzoek, bedoeld onder a en b- op eigen initiatief, dan wel op initiatief van een onderzoeksinstelling, doen uitvoeren van fundamenteel onderzoek en bevorderen van wetenschappelijke kennisontwikkeling en instrumentontwikkeling.”

Kortom, in termen van O.T.G.- en opdrachtonderzoek geformuleerd, het O.T.G.- onderzoek zou dienstbaar moeten zijn aan het opdrachtonderzoek. De meest gangbare betekenis van dienstbaar is volgens de dikke Van Dale ‘dienende, in een ondergeschikte betrekking geplaatst’. Het woord kan evenwel ook een meer neutrale betekenis hebben die in de Van Dale wordt omschreven als ‘werkzaam in het belang, ten nutte of ter wille van’. Zou de term dienstbaar uit de WOVI worden omschreven volgens de eerste betekenis dan zou de strekking van de betreffende passage uit de wet haaks staan op de enigszins opdringerige houding van Roede en Van den Dool. De tweede betekenis, daarentegen laat ruimte voor opdringerigheid als deze kan worden gerechtvaardigd onder verwijzing naar een of ander nut of belang van een nog te bepalen referent. Wie of wat het nut en het belang bepaalt wordt niet omschreven in de wet. Hiervan kunnen Roede en Van den Dool gebruik maken. Zijzelf kunnen het nut omschrijven, zoals dat altijd eigenlijk is gebeurd in O.T.G.- kaders, als de groei van kennis die uiteindelijk via technologie de welvaart en het welzijn van een groep of groepen mensen dient.

In deze zin hebben zij de wet aan hun kant. Evenwel, het lijkt realistisch om de wet te zien als een kader voor rechtvaardiging van een bestaande situatie en niet als een breekijzer voor veranderingen. De wet bevestigt de oorspronkelijke taakstelling van O.T.G.s.

We kunnen er niet omheen dat het fundamentele onderzoek, of beter gezegd het vrije sector onderzoek, slechts een klein bescheiden kamertje heeft toegewezen gekregen in het grote herenhuis van het onderwijsonderzoek. Bovendien dwingt het fundamentele onderzoek dat er in dat kleine kamertje plaats vindt niet echt veel respect af. In naam mag O.T.G.- onderzoek dan

gericht zijn op het zoeken naar verklaringen in samenhang met theorievorming, maar feitelijk is veel O.T.G.- onderzoek niet veel meer dan hap-snap-werk. O.T.G.- onderzoek lijkt deels het resultaat van gezond hobbyisme, deels van een koopmansgeest gevoelig voor trends (bijv. de hausse in onderzoek waarin computers een plaats hebben) en deels van het lonken naar het idee van toepassingsgerichtheid om respect af te dwingen en om naam en faam te vergaren.

Een persoonlijke evaluatie van de lopende O.T.G./A.N.G.O.- onderzoeken levert wat mij betreft de bevinding op dat 16 van de 47 projecten gericht zijn op verklaring en theorieontwikkeling of op validatie van onderzoeksinstrumenten in aansluiting bij geëxpliciteerde theoretische noties, 13 daartoe neigen en de rest weinig diepgaande, opportunistische projecten zijn; wat mij betreft een niet al te gunstig beeld, dat niet geheel strookt met de suggestie die Roede en Van den Dool wekken.

Naar de aard is opdrachtonderzoek weinig theoriebeladen en niet in de eerste plaats gericht op de ontwikkelingen van onderzoeksprogramma's met een continue stroom van onderzoeken die bijdragen aan theorievorming t.a.v. onderwijsfenomenen. Veel opdrachtonderzoek is in hoge mate tijd- en situatiegebonden en maakt daardoor een versnipperde, onsaamenhangende indruk. Het streven naar een nauwere samenhang tussen O.T.G.- onderzoek en opdrachtonderzoek heeft het gevaar in zich dat de genoemde karakteristieken van opdrachtonderzoek het O.T.G.- onderzoek besmetten. O.T.G.- kaders blijken nu al geen garantie te bieden dat dergelijke kenmerken buitengesloten zijn. Een meer intense samenhang zou allicht kunnen betekenen dat O.T.G.- onderzoekers meegaan in de stroom van toepassingsgericht/ beleidsgericht onderzoek. De dienstbaarheid is dan vervaagd tot vereenzelviging. In dit verband is het opvallend dat er wat betreft de personele bezetting van onderzoeksprojecten sprake is van een aanzienlijke overlap tussen vrije sector- en opdrachtonderzoekers. Van de 76 tot aan 1988 gesubsidieerde O.T.G.- projecten zijn of worden 27 uitgevoerd bij de S.V.O.- instituten die vooral bezig zijn met opdrachtonderzoek. Laten we de O.T.G.- onderwijsleerprocessen buiten beschouwing dan gaat het zelfs om 21 van 36 projecten.

Opdrachtonderzoekers verschillen qua ervaring niet van onderzoekers in de vrije sector. Ze werken soms in een andere setting en houden zich ten dele bezig met een ander type onderzoek en met een andere soort onderzoeksvragen. Zij zelf, maar ook de aanvragers van onderzoek, zullen vermoedelijk geen betutteling van O.T.G.- onderzoekers tolereren als het gaat om de onderzoeksprogrammering en de selectie en prioritering van onderzoek dat zij uitvoeren. Deze activiteiten kunnen derhalve niet uitdrukking zijn van de term dienstbaar in de tweede hiervoor beschreven betekenis.

Kortom, al dat streven naar samenhang op initiatief van de O.T.G.'s lijkt niet haalbaar en niet realistisch gezien de machtsverhoudingen tussen vrije sector en opdrachtonderzoek en het lijkt niet wenselijk gezien de bedreiging van het fundamentele karakter van vrije sector onderzoek.

Wellicht moet de opdrachtresearch enigszins anders worden ingericht, opdat het grote en snelle historische verval van de waarde van onderzoeken wat wordt tegengegaan. Het bevechten van dergelijke veranderingen kan vermoedelijk in eerste instantie nergens anders gebeuren dan binnen de kaders van de opdrachtresearch. Hier ligt in ieder geval geen taak voor de O.T.G.'s, omdat de toch al niet zo duidelijke aard van O.T.G.- onderzoek de eigen positie in de strijd wel eens zou kunnen verzwakken. Als O.T.G.- onderzoek inderdaad het streven naar verklaringsgerichtheid, theorievorming en de cumulatie van kennis tot uitdrukking wil brengen, dan lijkt het dienstig als de O.T.G.'s binnenskamers werken aan een duidelijke profilering in de richting van fundamenteel onderzoek. Het bevechten van het recht van bestaan van het eigen gebied lijkt daarmee gediend. De vrije sector moet voorkomen dat het nog meer de kenmerken van opdrachtonderzoek gaat vertonen dan nu al het geval is en dat er uiteindelijk een klein scheidingsmuurtje in het grote herenhuis wordt weggekapt, waarna het vrije sector onderzoek niet meer bestaat.

Wat betreft de dienstbaarheid hoeven vrije sector onderzoekers zich volgens mij niet zo druk te maken. De temporele dimensie is in het geval van dienstbaar onderzoek in het kader van de WOV niet gespecificeerd. Dienstbaarheid moet niet per se blijken van vandaag op morgen. Als

we Kerlinger (1977) nog steeds mogen geloven dan is de dienstverlenende kracht van fundamenteel onderzoek uiteindelijk duurzamer en sterker dan van veld- of beleidsgericht onderzoek.

REFERENTIES

- Beintema, K., Dool, P.C. van den, Roede, E., & Sijde, P.C. van der (red.) (1988). *Naar een strategische programmering van onderwijsonderzoek*. Almere: Versluys.
- Kerlinger, F.N. (1977). The influence of research on educational practice. *Educational researcher*, 6, September, pp 5-12 (ook gepubliceerd in *Tijdschrift voor Onderwijsresearch*, 1977, 2, 241-252).

Ontvangen 27-4-1988.

Boekbesprekingen

De Graaf, P. M.

De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen. Instituut voor Toegepaste Sociale Wetenschappen, Nijmegen, 1987, 183 pp.

Schoolloopbanen van leerlingen worden in aanzienlijke mate bepaald door hun milieu van herkomst. Meestal wordt dat laatste gedefinieerd in termen van het beroepsniveau van de ouders of, nog simpeler, van dat van de kostwinner, meestal de vader. Het beroepsniveau moet worden beschouwd als een proxy: de meest relevante determinanten gaan er achter schuil. Globaal genomen zijn dat er twee: (1) de financiële en (2) de culturele kenmerken van een gezin. In het verleden werd aan het verklarende karakter van de financiële hulpbronnen veel waarde gehecht. Geringe financiële armslag zou veel ouders uit de lagere beroepsniveaus doen afzien van relatief dure schoolloopbanen voor hun kinderen. Bovendien zouden ongunstige huisvestingsomstandigheden, te weinig speelgoed en het ontbreken van boeken, tijdschriften e.d. de intellectuele ontwikkeling van de kinderen remmen. Tegelijkertijd werd uit onderzoek duidelijk dat de verschillen tussen de gezinnen uit diverse sociale lagen ook in belangrijke mate cultureel van aard zijn. Interactiepatronen, prestatiemotiverend gedrag van de ouders, taalgebruik en het cultuurpedagogisch niveau variëren naar sociale milieus. Het feit dat beide verschijnselen, i.c. de financiële en de culturele hulpbronnen empirisch niet onafhankelijk van elkaar zijn, levert voor het bepalen van het gewicht van elk der determinanten problemen op. Databestanden waarin beide typen variabelen tegelijkertijd voorkomen, bewerkt met een multivariate analyse zijn dan een vereiste. De Graaf heeft in zijn dissertatie, gebaseerd op een ZWO-project, getracht de relatieve gewichten van beide hulpbronnen voor de schoolloopbanen van leerlingen te achterhalen. Daarbij is geen nieuw empirisch materiaal verzameld, maar zijn secundaire analyses gepleegd op bestaande gegevens. Geen enkel bestand bleek bij uitstek geschikt te zijn, of wel omdat begrippen te oppervlakkig gemeten waren dan wel de onderzochte schoolloopbaanperiode te kort was of relevante variabelen ontbraken.

De Graaf heeft daarom zijn toevlucht gezocht tot het afzonderlijk analyseren van een drietal databestanden, ondermeer met behulp van LISREL-modellen.

Halverwege zijn verslag introduceert de auteur, verrassenderwijs, nog een derde type hulpbronnen, te weten: 'ongelijke talenten'. De definiering daarvan blijft nogal in het vage. Successievelijk gaat hij na wat de invloed is van de drie typen. Hij stuit daarbij op een aantal interessante bevindingen. Zo neemt de invloed van de financiële positie van de ouders over de generaties heen, af. Voorts blijkt de correlatie tussen sociale herkomst en prestatieniveau te verdubbelen van $v = .25$ in de eerste klas van het lager onderwijs tot $v = .50$ in de zesde klas.

Tenslotte constateert hij aanzienlijke effecten van de culturele hulpbronnen in het gezin van herkomst op de onderwijsprestaties. Of deze invloed over de generaties heen is toegenomen, kan door ontbrekende gegevens niet worden aangetoond. De data staan evenmin uitspraken toe over de ontwikkeling van het gewicht van elk der hulpbronnen tijdens de laatste decennia. Daarmee is tevens het belangrijkste kritiekpunt geïntroduceerd: de discrepantie tussen de geformuleerde pretenties en de analysemogelijkheden. De verwachtingen die worden gewekt door aan te sluiten bij o.a. het gedachtengoed van de Franse socioloog Bourdieu en door te refereren aan concurrerende sociologische theorieën (functionalistische vs. conflictologische) doen de lezer met steeds meer interesse naar de resultaten uitzien. De gepresenteerde analyses werken vervolgens min of meer als een anti-climax. Veel van het beloofde kan niet worden waargemaakt.

Ondanks het niet inlossen van verwachtingen kan toch van een geslaagd proefschrift worden gesproken. De poging op zich om reeds bekende empirische bevindingen naar een hoger abstractieniveau te tillen werkt verfrissend en inspirerend. Eens te meer blijkt dat aandacht voor theorievorming een waardevolle meerwaarde introduceert op een onderzoeksterrein waar al

veel empirisch onderzoek is verricht. Relatieve buitenstaanders kunnen soms een verrassende nieuwe visie introduceren op voor insiders reeds lang bekende empirische verschijnselen. Dat geldt ook voor De Graaf die zich als auteur zeker niet als onderwijssocioloog zal afficheren.

G.W. Meijnen

Universiteit van Amsterdam

F.W. Siero, *Feedback en motivatie in de klas*.

Proefschrift R.U. Groningen, 1987, ISBN 90-71382-18-4.

De beoordeling van taakgedrag kan twee vormen aannemen: klassikale en geïndividualiseerde beoordeling (p. 6). Bij de eerste vorm werken leerlingen allen aan dezelfde taken en worden ze aan de hand van eenzelfde criterium geëvalueerd; de leerling wordt beoordeeld door vergelijking met de anderen in zijn groep. Bij de tweede vorm kiest de leerkracht de eerdere verrichtingen van de leerling als criterium; de leerling wordt beoordeeld op basis van de individuele voor- of achteruitgang in zijn prestaties. In hoofdstuk 1 wordt gesteld dat deze beoordelvormen van invloed zijn op de manier waarop leerlingen hun eigen taakuitkomsten percipiëren en attribueren en dat deze 'kausale cognities' op hun beurt weer bepalend zijn voor latere prestaties – een proces waarmee met name de attributietheorie van Weiner zich bezig houdt. Deze theorie, alsmede de verwante, enigszins algemenere attributerings-theorie van Kelley, worden summier maar duidelijk en toereikend uiteengezet. Betoogd wordt dat leerkrachtoordelen alleen dan motiverend werken en leiden tot effectieve bijsturing van taakgedrag, als leerlingen die oordelen via hun gedrag kunnen beïnvloeden: het oordeel moet betrekking hebben op gedrag dat door de leerling controleerbaar is, zoals bijvoorbeeld inspanning; er is dan sprake van kontingente beoordeling. Klassikale beoordeling kan nauwelijks kontingent zijn, individuele beoordeling daarentegen wel. De mate van kontingentie van het leerkrachtoordeel zal tot uitdrukking komen in de attributies van leerlingen en in latere taakresultaten. Deze verwachtingen worden onderzocht in enkele veldexperimenten waarover gerapporteerd wordt in de hoofdstukken 2 en 3.

De faktor 'beoordeling' werd gemanipuleerd door aan leerlingen in de vijfde klas van een aantal basisscholen een reeks van 5 lessen te geven, toetsen af te nemen en daarop het oordeel te baseren. In het eerste experiment (klassikale beoordeling) werden de leerlingen op grond van hun prestatienivo op de twee eerste toetsen ingedeeld in drie groepen (laag, midden en hoog). Bij het begin van de lessen drie tot en met vijf werd een prestatiebeoordeling gegeven die aansloot bij het vastgestelde nivo (laag nivo = ongunstige beoordeling, enz.). Onderzocht werd hoe de leerlingen hun eigen prestaties in deze situatie attribueerden (naar bekwaamheid, inzet, taakmoeilijkheid en toeval). Het tweede en derde experiment (individuele beoordeling) hadden een analoge opzet. De beoordeling was nu echter gebaseerd op veranderingen in individuele prestaties en werd gegeven in vier varianten (toenemende zichtbaarheid van de kontingentie en de controleerbaarheid). Naast de attributies waren hier ook de verrichtingen op de derde en vierde toets de afhankelijke variabelen.

De resultaten stemmen grotendeels overeen met de verwachtingen. Bij klassikale beoordeling voeren onder de maat presterende leerlingen vooral toeval en nauwelijks inzet aan ter verklaring van hun ongunstige resultaten; zij zien de beoordeling klaarblijkelijk als weinig controleerbaar. Goede presteerders noemen inzet als belangrijke faktor en vinden toeval niet zo van invloed. Uit de experimenten met individuele beoordeling bleek dat de leerlingen de beoordelingen inderdaad over de vier varianten toenemend als meer controleerbaar percipieerden en toenemend hogere prestaties leverden.

Na deze beknopte samenvatting die het werk nauwelijks recht kan doen, enkele evaluatieve opmerkingen. Het boek is heel leesbaar geschreven; de opzet van de experimenten en de

verwerking van de gegevens getuigt van inventiviteit en grote zorgvuldigheid. Het theoretisch kader is betrekkelijk arm: veel direct relevante literatuur wordt niet vermeld, bijvoorbeeld het werk van Rheinberg over referentienormen bij het beoordelen van leerlingen en het werk van Meyer over de relatie tussen beoordeling, zelfbeeld van eigen bekwaamheid, attribuering en taakmotivatie. Het onderzoek blijkt overigens ook niet over feedback in relatie tot *motivatie* te gaan, maar uitsluitend over beoordeling in relatie tot attributies en prestaties. Natuurlijk is er een relatie tussen attributies en motivatie, en tussen motivatie en prestaties, maar de betreffende literatuur komt in het proefschrift niet ter sprake en in de experimenten worden nergens motieven of motivatie gemeten, noch als persoonlijkheidskenmerk, noch in de vorm van taakgedrag. Het onderzoek levert een verdienstelijke bijdrage aan onze kennis over de manier waarop attributies voor prestaties op school tot stand komen, met name onder invloed van leerkrachtoordelen. Deze problematiek is trouwens al belangrijk genoeg, niet alleen omdat ze direct van betekenis is voor de onderwijspraktijk, maar ook omdat iemand als Weiner zich er zelf nauwelijks mee bezig houdt. Weiner richt zich vooral op de cognitieve, affectieve en gedragsmatige konsekventies van attributies. Een voor de onderwijskunde relevante reeks van onderzoeken gericht op die thematiek wordt gerapporteerd in het vrijwel gelijktijdig verschenen proefschrift van Frank van Overwalle, dat een mooie aanvulling is op dat van Siero.

REFERENTIE

Van Overwalle, F. (1987). *Causes of success and failure of freshmen at university: an attributional approach*. Brussel, Vrije Universiteit.

John van Berken
Vakgroep Orthopedagogiek
Erasmusplein 1
6800 HD 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, Faculteit der Toegepaste Onderwijskunde, Vakgroep OCTO, Universiteit Twente, Postbus 217, 7500 AE Enschede, tel. 053-893643/893554.

Jrg. 12, nr. 5

October 1988

Inhoudsopgave

VOR-Bulletin, Jrg 12. nr. 5

Congressen	296
Mededelingen	297
Publicaties	298
Agenda VOR-ledenvergadering	299
Verslag VOR-ledenvergadering 31 mei 1988	300

Output: kwaliteit of kwantiteit

P.C. van der Sijde
Universiteit Twente

Inleiding

In de eerste twee nummers van het VOR Bulletin van dit jaar is door Claessen en door Ax en Creemers de beoordeling van de wetenschappelijke productie onder de loep genomen. Ax en Creemers (1988) doen een concreet voorstel tot een kwantitatief beoordelingssysteem, maar tevens merken zij aan het eind van hun bijdrage op dat kwantitatieve beoordelingen slechts een beperkte waarde hebben. Claessen (1988) laat zich in gelijkstreckende bewoordingen uit, als hij schrijft: 'het getalscriterium dreigt een magische betekenis te gaan aannemen'.

Het probleem dat geschetst wordt, is dat van kwaliteit versus kwantiteit. Hoe reëel dit probleem is, weten we allemaal, maar realiseren we ons dit ook? Om het probleem heel concreet te maken worden hieronder twee cases geschetst.

Twee cases

Case 1: Deze case betreft een niet nader aan te duiden, buitenlands onderzoeker.

In het kader van een tijdens de AERA door een Special Interest Group georganiseerd symposium houdt prof. A. een paper, waarin hij een internationale stand- van- zaken presenteert. Hij reviewde 70 studies. Naar aanleiding van dit symposium redigeert hij het daarop volgende jaar een bundel met daarin opgenomen zijn eigen stand- van- zaken paper.

In dezelfde periode werkt hij aan een boek over dit onderwerp, het boek verschijnt weer een jaar later en hoofdstuk 2 bestaat opnieuw uit dezelfde paper.

Binnenkort zal hetzelfde artikel in een vertaling nog eens verschijnen. Een en ander betekent het volgende:

- een congrespresentatie op de AERA, 1985;
- artikel in de door hemzelf geredigeerde proceedings, 1986;
- een hoofdstuk in zijn eigen boek, 1987;
- een vertaling voor opname in een boek, 1988;

Kwantitatief gezien uitstekende score, echter geen enkel van de genoemde publicaties verschilt niet meer van elkaar dan hooguit de interpunctie en uiteraard de titel.

Case 2: Het betreft hier een gefingeerde Nederlandse onderzoeker.

Zo rond november-december van ieder jaar wordt er in onderwijskundig Nederland voor het eerst weer eens aan de komende ORD gedacht. Daar moet wel empirisch materiaal worden gepresenteerd, maar je kan ook wel eens wat bluffen en een schat aan empirisch materiaal beloven. Resultaat hiervan is een paperpresentatie en met wat geluk ook in een thema. Dat laatste is van belang omdat je anders niet in de bundel komt. Voor de bundel maak je je artikel interessant, maar je laat niet het achterste van je empirische data zien (die trouwens ook nog niet geanalyseerd zijn). De volledige dataset bewaar je voor een artikel in *Pedagogische Studiën*. Het artikel laat je vervolgens in het Engels vertalen en biedt het aan een buitenlands tijdschrift aan. Verder maak je een bewerking voor een vaktijdschrift en biedt je SVO het eindrapport aan. Het resultaat is:

- een ORD-presentatie
- een artikel in een ORD-bundel
- een artikel in *Pedagogische Studiën*
- een artikel in een Engelstalig tijdschrift
- een vakpublicatie
- een eindrapport

Kortom, zes verschillende bijdragen aan de wetenschap waar een ieder jaloers op zou zijn. Echter, hetzelfde empirische databestand voor alle publicaties.

De waardering

Volgens het door Ax en Creemers voorgestelde beoordelingssysteem kan prof. A. rekenen op 8 punten:

– congrespresentatie plus opname in bundel:	2
– hoofdstuk in een boek (omvangrijk):	3
– hoofdstuk in een boek (omvangrijk):	3
totaal:	8 punten

De gefingeerde, maar zeer reële Nederlandse onderzoeker kan zelfs rekenen op 15 punten:

– congrespresentatie plus opname in bundel:	2
– artikel in een Nederlandstalig tijdschrift:	3
– artikel in een Engelstalig tijdschrift:	3
– vakpublicatie:	1
– eindrapport:	6
totaal:	15 punten

Het bovenstaande illustreert de kwantitatieve waardering, maar hoe zou de kwalitatieve waardering kunnen worden uitgedrukt? In beide cases worden de onderzoekers rijkelijk beloond voor hun inspanningen. Echter in beide gevallen is de bron waaruit de eerste publicatie geput werd dezelfde als die reeds voor de overige publicaties werd gebruikt.

Kwaliteit?

Levert een onderzoeker die op een soort gelijke wijze werkt (en welke onderzoeker doet dit in feite niet?) nog wel een bijdrage aan de wetenschap? Misschien moeten we ons wel afvragen of we dit als onderzoeker wel willen. Is het vaak niet zo, dat we hetzelfde empirische materiaal op verschillende wijzen aanbieden, omdat het voor een verschillend lezerspubliek toegankelijk gemaakt moet worden? Maar zijn we dan nog wel met wetenschap bezig?

We kunnen ons ook afvragen of de titel van de bijdrage van Ax en Creemers ('Kwaliteitsbepaling wetenschappelijke produktie in de onderwijskunde') wel een juiste titel is? Als het gaat om de kwaliteitsbepaling van de *wetenschappelijke produktie* is hun beoordelingssysteem dan wel een terechte, het is een systeem om de output van de onderwijswetenschappen te bepalen.

Wie bepaalt de kwaliteit van een publicatie of van een onderzoek? Het wetenschappelijk forum of een visitatiecommissie?

Slotopmerking

Persoonlijk geef ik de voorkeur aan het wetenschappelijk forum en daarom is misschien de citatie-index toch niet eens zo'n gek instrument. Maar wie gaat dit voor het Nederlandstalige gebied opzetten? Zolang er door niemand nog een voorstel gedaan is om het kwaliteitsprobleem in de onderwijskunde op een bevredigende manier op te lossen, troost ik mij met de gedachte dat dit korte stukje mij wellicht toch nog één punt op gaat leveren.

REFERENTIES

- Ax, J. & Creemers, B. (1988). Kwaliteitsbepaling wetenschappelijke produktie in de onderwijskunde. *VOR Bulletin*, 12, 118-121.
- Claessen, J.F.M. (1988). Outputbepaling onderwijswetenschappen op juiste spoor. *VOR Bulletin*, 12, 60-61.

Congressen

TWEEDE EUROPESE LOGO-CONFERENTIE 1989

Datum : 30 augustus - 1 september 1989

Plaats: Rijksuniversiteit Gent (België)

Dienst Pedagogiek

EUROLOGO is een tweejaarlijkse conferentie. De eerste vond in 1987 in Dublin plaats. Er was een representatief aantal deelnemers uit geheel Europa aanwezig.

Volgende thema's komen gedurende EUROLOGO '89 aan bod:

- * Ervaringen met Logo in de klas;
- * Logo en het leerplan;
(in basis-, middelbaar- en buitengewoon onderwijs);
- * Logo-projecten;
- * Onderzoeksprojecten;
- * Bijscholing i.v.m. Logo;
- * Nieuwigheden op technisch vlak.

De conferentie richt zich zowel naar beginners als naar experts, zowel naar leerkrachten als naar vorschers.

Activiteiten tijdens de conferentie zijn o.a. het naar voor brengen van bijdragen door de deelnemers, demonstraties, workshops, postersessies.

Als u actief aan deze conferentie wil deelnemen, gelieve een résumé (één A4-bladzijde) van uw bijdrage naar het programmacomit  te sturen:

Rijksuniversiteit Gent

Dienst Pedagogiek - EDIF

Eurologo '89

H. Dunantlaan 1

B-9000 Gent

BELGIE

Mededelingen

'Minicongres'

Het VOR-bestuur wil onderzoeken of er onder de leden van de VOR belangstelling bestaat voor het organiseren van een 'minicongres'. Met 'minicongres' wordt bedoeld een korte bijeenkomst van enkele deskundigen op een beperkt vakgebied, afkomstig uit West-Europa. Mede oogmerk zal publikatie moeten zijn.

Ideeën of evt. zelfs initiatieven kan men zenden aan de secretaris van de VOR, mevr. dr G. van der Werf, p/a RION, Postbus 1286. 9701 BG Groningen.

ORD'89

De ORD'89 zal plaatsvinden op 24 en 25 mei 1989 in Leiden.

SIG 'The Study of Learning Environments'

Binnen de American Educational Research Association bestaat sinds enkele jaren een Special Interest Group 'The study of learning environments'. Deze groep stelt zich als doel verspreiding van resultaten van onderzoek naar leeromgevingen te bevorderen. Het gaat daarbij om klasomgeving, schoolomgeving en buitenschoolse omgevingen. Er is in de groep traditioneel veel aandacht voor leeromgevingsmetingen via vragenlijsten afgenomen aan participanten in de leeromgeving maar de laatste tijd worden ook observaties gebruikt.

De groep heeft ongeveer 120 leden, waaronder een vrij aanzienlijk deel niet-Amerikanen. Ze organiseert elk jaar enkele sessies op het AERA-congres, publiceert jaarlijks een bundel met geselecteerde bijdragen van het congres en kent een prijs toe voor het beste paper. Verder is er halfjaarlijks een nieuwsbrief.

Voor het lidmaatschap van de SIG kan men zich aanmelden bij ondergetekende of bij de secretaris Chad D. Ellet, 113-E Peabody Hall, Louisiana State University, Baton Rouge, LA 70803, U.S.A. De kosten bedragen \$ 3.

De tot nu toe verschenen bundels zijn alle geredigeerd door Barry Fraser. Deel 1 bevat ondermeer een overzicht van 20 jaar leeromgevingsonderzoek. Deel 2 opent met een paper van Rudolf Moos over verbanden tussen leeromgevingen op school, andere werkplekken en thuis. Deel 3 tenslotte geeft onder meer Walberg's visie op de stand van het onderzoek naar leeromgevingen. Bundel 1 kan besteld worden bij Chad D. Ellet (\$ 16) en deel 2 en 3 via Barry Fraser, Science and Mathematics Education Centre, Curtin University, GPO Box U 1987, Perth 6001, Australia (20 Australische dollars). Van deel 3 is ook een beperkt aantal via ondergetekende te verkrijgen (kosten fl. 32,=)

De volledige inhoudsopgaven zijn ook bij ondergetekende verkrijgbaar.

Dr. Th. Wubbels

Rijksuniversiteit

Vakgroep Natuurkunde Didactiek

Postbus 80.008

3508 TA Utrecht.

Publicaties

- Straetmans, G.J.M., Dijkstra, S. (1987). *Een evaluatiemissie in het kader van het project 'Modulering opleidingen metaalberoepen'*. (Toegepaste Onderwijskunde: f 8,-).
- Mulder, M. (red.) (1987). *Aspecten van bedrijfsopleidingen: gepresenteerde papers van een symposium...* (Toegepaste Onderwijskunde: f 17,50).
- Remmers, J.L.M., Visscher, A. J., Bentem, I.C.M. van, & Bergsma, M.H.M. (1987). *Computerondersteunde absentieregistratie en -beleidsvorming: een evaluatie van de implementatie en het gebruik van het Absentie Registratie Systeem*. (Toegepaste Onderwijskunde: f 18,-).
- Mulder, M., Akkerman, J. & Bentvelsen, N. (1987). *Bedrijfsopleidingen in Nederland: twee oriënterende case studies en een beknopte analyse van enkele onderwijskundige knelpunten*. (Toegepaste Onderwijskunde: f 10,-).
- Pelgrum, W.J. & Bruggencate, G.C. ten (1986). *Second International Science Study: technical report*. (Toegepaste Onderwijskunde: f 30,-).
- Heijden, M.P. van der (1987). *Tekst- en videobeeld: een literatuurstudie*. (Toegepaste Onderwijskunde: f 6,50).
- Vries, F.J. de (1987). *Interactieve tekstpresentatie versus interactieve video*. (Toegepaste Onderwijskunde: f 17,-).
- Hofstra, M.L., Hobus, P.P.M., Boshuizen, H.P.A. & Schmidt, H.G. (1988). Diagnostiek bij huisartsen: het belang van gegevens uit de context van de patiënt. Maastricht: R.L.
- Bouhuijs, P.A.J. & Wijnen, W.H.F.W. (1988). Onderwijskundige evaluatie tele-educatieproject Open Universiteit 1987. Maastricht: RL
- Vernieuwing, jrg. 47, nummer 4/5, april/mei 1988.
- RVE (1988). Advies over de harmonisatie van collegegelden in het hoger onderwijs en de invoering van één inschrijvingsduur in de eerste fase van het hoger onderwijs. Den Haag: RVE
- Vernieuwing, jrg. 47, nummer 6, juni 1988
- Min Biza (1988). Regeringsstandpunt inzake minderhedenonderzoek, -beleid en privacy. Den Haag: BiZa
- CITO (1988). Jaarverslag 1987.
- Rion (1988). Jaarverslag 1987.
- Leesgoed, jrg. 15, juni 1988.
- Nekkers-Kapitein, M.P. (1988). Samenwerking in de dienstverlening en gezondheidszorg Rijndelta. COA Rijndelta.
- F.J. de Vries. Interactieve tekstpresentatie versus interactieve video: literatuurverslag. Enschede, 1987. - 35p. (hfl. 6,50)
- Scheerens, J., Stoel, W.G.R., Vermeulen, C.J.A.J., Pelgrum, W.J. (1988). De haalbaarheid van een onderwijsindicatorenstelsel voor het basis- en voortgezet onderwijs. Enschede: TO.
- Knol D.L. (1988). Stapsgewijze itemselectieprocedures in het Raschmodel. Enschede: TO.
- Adema, J.J. (1988). LP-modellen voor maximilisatie van coefficient α . Enschede: TO.
- Van der Burg, E., De Leeuw, J., (1988). Nonlinear redundancy analysis. Enschede: TO.
- Van der Linden W.J., Adema, J.J. (1988). Algorithmic test design using classical item parameters. Enschede: TO.
- Boekkooi-Timminga, E. (1988). A cluster-based method for test construction. Enschede: TO.
- Adema, J.J. (1988). A note on solving large-scale zero-one programming problems. Enschede: TO.
- Van der Linden, W.J. (1988). Optimizing incomplete sample designs for item response parameters. Enschede: TO.
- Vos, H. J. (1988). The use of decision theory in the Minnesota. Adaptive Instructional System. Enschede: TO.
- Rikers, J.H.A.N. (1988). Towards an authority system for item construction. Enschede: TO.
- Jelsma, O., Bijlstra, J.P. (1988). PROCESS: program for research on operator control in an experimental setting. Enschede: TO.
- Scholten, J.V.A. (1988). Het ontwerpen en ontwikkelen van een computerondersteund instructieprogramma voor het aanvankelijk rekenonderwijs. Enschede: TO.
- Thijssen, J. A., Mulder, M. (1988). Onderwijsleermiddelen in het MTO, afdeling werktuigbouwkunde. Enschede: TO.
- Ten Brummelhuis, A.C.A., Brandsma, T.F. (1988). Vergelijkende evaluatie LTO- vernieuwingsprojecten (+ aparte bijlagen). Enschede: TO.
- Hol, D. (1988). Het leren van economische begrippen in de socratische dialoog. Enschede: TO.

Agenda VOR-ledenvergadering

Plaats: Stiltecentrum, Radboutzaal, Hoog Catharijne Utrecht.

Tijd: 14.00 uur.

Datum: 10 november 1988.

1. Opening/ mededelingen
2. Verslag vorige ledenvergadering (zie dit VOR-bulletin)
3. Begroting 1989 (op te vragen bij de penningmeester, drs. E. Roede, tel. 020-5550325).
4. Nieuwe Bestuurssamenstelling voor 1989.

Aftredende bestuursleden: Prof. dr. P. Span

Dr. B. van Hout Wolters

Door het bestuur worden voorgedragen: Prof. dr. G. Kanselaar

Dr. J.F.M.J. van Hout

Tegenkandidaten kunnen tijdens de vergadering worden ingebracht.

5. W.v.t.t.k.

6. Sluiting.

Verslag VOR-ledenvergadering 31 mei 1988

Aanwezig: Slavenburg, Smulders, Ten Cate, Wierstra, Wubbels, Claessen, Gerritsen van der Hoop, Van Hout, Wolters, Roede, Span en Van der Werf.

1. *Opening.*

De vijf leden worden welkom geheten.

2. *Verslag vorige vergadering (26 mei 1987).*

Het verslag wordt vastgesteld. Naar aanleiding van het verslag wordt verzocht als aanvulling op de agenda te bespreken hoe de vakdidactiek op de ORD's een plaats kan krijgen.

3. *Jaarrekening 1987.*

Financieel ziet de VOR er gezond uit. Momenteel blijft het eigen vermogen stabiel. De kascontrolecommissie spreekt haar waardering uit voor het financieel beheer. De jaarrekening wordt vastgesteld. De suggesties van de kascontrolecommissie (eerdere inning van inkomsten van SVO, werving IJER-leden, organiseren van seminars) zullen in het bestuur aandacht krijgen.

4. *Begroting 1988.*

Deze wordt vastgesteld.

5. *Verslag verenigingsjaar 1988.*

Vastgesteld.

6. *Samenstelling nieuwe kascontrolecommissie.*

Wardenaar zal uit de commissie gaan. Voor de vacante plaats heeft geen van de aanwezigen belangstelling. Na zijn afscheid uit het bestuur zal Span de vacante plaats innemen.

7. *Vakdidactiek.*

Er wordt verzocht om te stimuleren dat op de ORD's en bij de uitgave van bundels meer aandacht wordt besteed aan de vakdidactiek. Het gaat vooral om drempelverlaging: vakdidactici moeten meer gestimuleerd worden om op de ORD te presenteren, eventueel in de vorm van symposia. Opname van de vakdidactiek als apart thema is lastig omdat de inhoud van de thema's, evenals die van de bundels, reglementair vast ligt. Een organisatiecommissie kan echter wel een thema invullen met vakdidactiek, zoals dat bijv. in Utrecht is gebeurd.

Het bestuur zal zich over dit punt buigen, in samenspraak met de Commissie Thema planning en de ORD organisatie commissie Leiden.

Greetje van der Werf
secretaris VOR-bestuur.

Het Lineair Logistisch Test Model en heterogeniteit van cognitieve strategieën bij balansproblemen

L. van Maanen ¹⁾, P. H. Been ¹⁾, en K. Sijtsma ²⁾

¹⁾ *Rijksuniversiteit, Groningen.*

²⁾ *Vrije Universiteit, Amsterdam.*

ABSTRACT

Some properties of the Linear Logistic Test Model (LLTM) are discussed in this paper. To test hypotheses about problem-solving behavior with respect to test items by means of the LLTM, two conditions must be satisfied. First, the items should comply with the requirements of the Rasch model. Second, the population of interest must consist of persons all using the same problem-solving strategy when confronted with the items. It is shown that this second condition is not fulfilled for a test consisting of balance items. In this study, persons are allocated to four distinct groups, each group characterized by the use of a specific problem-solving strategy. The LLTM gives a satisfying explanation of problem-solving behavior within one of these groups. Problems concerning the application of the LLTM to empirical data are discussed. It is concluded that the LLTM should only be applied within groups of persons all using the same problem-solving strategy.

Cognitieve procesmodellen kunnen worden getoetst met het Lineair Logistisch Test Model (LLTM) (Scheiblechner, 1972; Fischer, 1973; 1974; Spada, 1976; Fischer & Formann, 1982; Been, Jorna & Sijtsma, 1984), dat een uitbreiding is van het item response model van Rasch. In het Rasch model wordt de moeilijkheid van de items in een test verdiskonteerd in een moeilijkheidsparameter per item. In het LLTM wordt deze parameter opgevat als de gewogen som van een aantal basisparameters. Iedere basisparameter staat voor de moeilijkheid van een deelvaardigheid die ten grondslag ligt aan de taak die in een item wordt opgedragen. Op deze wijze wordt de moeilijkheid van een item verklaard uit de moeilijkheden van de onderliggende deelvaardigheden. De empirische geldigheid van een cognitief procesmodel kan worden getoetst met behulp van het LLTM door per item de benodigde deelvaardigheden te expliciteren, en ze te wegen met de frequentie waarin ze zijn vereist bij de oplossing van een item.

Empirisch is gebleken dat het LLTM strenge eisen stelt aan de gegevens. In de literatuur zijn dan ook geen voorbeelden te vinden van bij de gegevens passende modellen (bijvoorbeeld, Spada, 1976; Haeussler, 1977; Kubinger, 1979; Kubinger, 1980; Nährer, 1980; Whitely & Schneider, 1981; Sijtsma, 1982; Zwarts, 1983). Alvorens een cognitief procesmodel met het LLTM kan worden getoetst, dient aan twee voorwaarden te zijn voldaan. Ten eerste moeten de items waarop het cognitief procesmodel wordt getoetst, een Rasch-homogene schaal vormen. Ten tweede dient voor elk item een cognitief procesmodel te worden gespecificeerd dat geldig is voor iedere persoon in de te onderzoeken populatie. Dit betekent dat wordt aangenomen dat alle personen in de populatie bij het oplossen van één specifiek item dezelfde strategie zullen gebruiken. Indien bij het oplossen van één specifiek item verschillende strategieën worden gebruikt, wat bij verscheidene taken empirisch is aangetoond (Broadbent, 1984), dan kan deze strategie-heterogeniteit een mogelijke oorzaak zijn van het ontbreken van een passend LLTM (zie ook Spada & McGaw, 1985). Doel van dit onderzoek is na te gaan of een passend LLTM kan worden verkregen door rekening te houden met strategie-heterogeniteit. Voor dit doel is gekozen voor balansproblemen. Reden om balansproblemen te kiezen is dat de bij het oplossen van deze taken betrokken cognitieve processen (Siegler, 1976; Siegler & Klahr, 1982) bekend zijn.

Adres: Interdisciplinair Centrum voor de ontwikkeling van Computercoaches en Expertsystemen, Postbus 800, 9700 AV Groningen.

Tenslotte refereren we aan Spada en McGaw (1985) en Van de Vijver (1988) die er op wijzen dat het LLTM kan worden gebruikt in de stadia van item- en testontwerp. Dit type toepassing is in het huidige artikel verder niet aan de orde.

Modellen voor balansproblemen

Het model dat in dit onderzoek wordt gebruikt is het balansmodel van Siegler en Klahr (1982). Het gaat om een specifiek type balansopgaven. Op beide armen van de balans zijn op onderling gelijke afstanden pennen aangebracht. Op iedere pen kunnen één of meer ringen van gelijk gewicht worden gelegd. Het model van Siegler en Klahr leent zich goed voor ons onderzoek. In de eerste plaats is het model toepasbaar op een brede range van leeftijdsklassen. Vijfjarigen hebben al een notie van gewicht terwijl veel volwassenen de formele kruisproductregel (gewicht * afstand) voor het oplossen van balansproblemen niet beheersen. In de tweede plaats zijn er voor de oplossing van een balansprobleem maar twee factoren van belang, de afstand van de gewichten tot het kantelpunt en de grootte van de gewichten. Dit betekent dat het probleemdomen begrensd is. Het aantal typen items dat kan worden gekonstrueerd is beperkt, en tevens kan er slechts een beperkt aantal oplossingsstrategieën worden geformuleerd.

Tevens kan worden vermeld dat het LLTM reeds op balansproblemen is toegepast door Spada (1976), waarbij bleek dat het model niet bij de gegevens paste. Een mogelijke verklaring voor dit resultaat is dat de onderzochte populatie van respondenten niet homogeen is wat betreft het gebruik van een oplossingsstrategie, zodat aan de tweede konditie voor toepassing van het LLTM niet is voldaan.

Siegler en Klahr (1982) vonden vier mogelijke strategieën voor het oplossen van balansproblemen (zie Figuur 1), variërend van een elementaire strategie tot een "expert"-strategie.

Personen die de elementaire strategie (Strategie 1) toepassen letten alleen op de hoeveelheid gewichten aan beide kanten van de balans. Als het gewicht aan beide kanten gelijk is, wordt evenwicht voorspeld. Is het aantal gewichten aan beide kanten ongelijk dan wordt voorspeld dat de arm met het grootste aantal gewichten naar beneden gaat.

De tweede oplossingsstrategie (Strategie 2) bevat Strategie 1, maar is met de volgende regel uitgebreid. Als de afstand van de gewichten tot het kantelpunt verschilt en de hoeveelheid gewichten aan beide kanten gelijk is, dan wordt voorspeld dat die kant waar de gewichten op de grootste afstand van het kantelpunt zijn geplaatst naar beneden zal gaan.

De derde oplossingsstrategie (Strategie 3) bevat Strategie 2 (en dus Strategie 1). De volgende regel is toegevoegd: als de hoeveelheid gewichten en de afstand tot het kantelpunt ongelijk zijn en de grootste hoeveelheid bevindt zich op de grootste afstand van het kantelpunt dan gaat die arm naar beneden.

Strategie 5 bevat Strategie 3 (en dus Strategie 1 en Strategie 2). Toegevoegd is de kruisproductregel, die in de volgende situatie wordt toegepast: op de ene arm is de grootste hoeveelheid gewichten geplaatst en op de andere arm is de afstand tot het kantelpunt het grootst.

Uit een vooronderzoek bleek dat er een aanvulling op Strategie 3 van het model van Siegler en Klahr mogelijk is. Bij items waar zowel de gewichten als de afstanden ongelijk zijn, geeft een deel van de proefpersonen systematisch aan dat er sprake is van evenwicht. Het algoritme dat wordt gebruikt, is: neem het stokje met de grootste hoeveelheid gewichten en verplaats de gewichten zolang tot de afstand gelijk is aan de afstand aan de andere zijde van de balans. Trek bij elke verplaatsing telkens één gewicht af van het stokje met de grootste hoeveelheid gewichten. Deze Strategie 4 in Figuur 1 levert bij sommige konfliktitems een goed antwoord op.

De vraag of personen deze regels bij het oplossen van balanstaken ook gebruiken kan worden onderzocht met zes typen van items (zie Figuur 1). Drie typen kunnen worden opgelost zonder gebruik te maken van de korrekte Strategie 5. Deze itemtypen zijn:

- balans: zowel de gewichten als de afstanden zijn aan beide zijden gelijk;
- gewicht: de afstanden zijn gelijk, de gewichten verschillen;
- afstand: de gewichten zijn gelijk en de afstanden verschillen.

De drie resterende typen items worden konfliktitems genoemd omdat één zijde meer gewicht en de andere zijde een grotere afstand heeft, wat eventueel een konflikt bij de persoon kan

veroorzaken. De drie soorten items van dit type zijn:

- konfliktgewicht: de zijde met het grootste gewicht gaat naar beneden;
- konfliktafstand: de zijde met de grootste afstand gaat naar beneden;
- konfliktbalans: het produkt van gewicht*afstand is aan beide zijden gelijk, de balans blijft in evenwicht.

De combinatie van regels en itemtypen maakt het mogelijk te voorspellen welk antwoordpatroon op zal treden. In Figuur 1 is per combinatie het verwachte percentage goede antwoorden gegeven. Aangezien er bij ieder item drie antwoordalternatieven zijn - de balans slaat door naar links of rechts, of de balans is in evenwicht - levert Strategie 3 bij blind gissen op de konfliktitems een percentage gelijk aan 33%. Strategie 5 is de formeel juiste oplossingsstrategie en leidt derhalve altijd tot het goede antwoord.

De vijf strategieën die worden gebruikt bij het oplossen van balansproblemen kunnen als basisparameters in het LLTM dienen. In ons onderzoek hebben we echter personen die strategie 5 gebruiken buiten beschouwing gelaten, omdat in de door ons onderzochte leeftijdsgroep deze strategie nauwelijks voorkomt. Uit het eerder genoemde vooronderzoek bleek dat alle personen

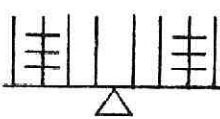
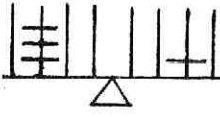
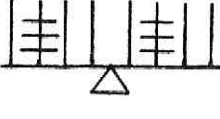
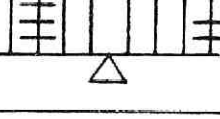
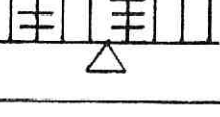
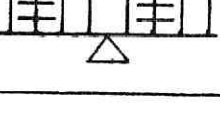
Item type	Strategie				
	1	2	3	4	5
'Balans' 	balans (100)	balans (100)	balans (100)	balans (100)	balans (100)
'Gewicht' 	links omlaag (100)	links omlaag (100)	links omlaag (100)	links omlaag (100)	links omlaag (100)
'Afstand' 	balans (0)	links omlaag (100)	links omlaag (100)	links omlaag (100)	links omlaag (100)
'Konfliktgewicht' 	links omlaag (100)	links omlaag (100)	gokken (33)	balans (0)	links omlaag (100)
'Konfliktafstand' 	rechts omlaag (0)	rechts omlaag (0)	gokken (33)	balans (0)	links omlaag (100)
'Konfliktbalans' 	rechts omlaag (0)	rechts omlaag (0)	gokken (33)	balans (100)	balans (100)

Fig. 1: Voorspellingen van het antwoord en tussen haakjes het percentage goede antwoorden bij verschillende strategieën. Strategie 1, 2, 3 en 5 zijn strategieën volgens het model van Siegler en Klahr. Strategie 4 is een aanvulling op dit model.

de balans items goed beantwoorden. Met dit type item kan derhalve geen onderscheid worden gemaakt ten aanzien van de vaardigheid die personen bezitten om dit type item op te lossen. De balansitems zijn daarom ongeschikt voor verdere analyse. De empirische geldigheid van Strategie 1 tot en met Strategie 4 zal met het LLTM worden getoetst voor de vijf resterende itemtypen.

Formalisering van oplossingsstrategieën in het LLTM

De vier te onderzoeken oplossingsstrategieën kunnen op hun geldigheid worden getoetst met behulp van het LLTM. Per oplossingsstrategie wordt een formeel model gedefinieerd, dat bestaat uit een gewogen lineaire combinatie van de basisparameters. De basisparameters (N_j) corresponderen met de moeilijkheden van de strategieën en hun gewogen som levert de moeilijkheid van een item. De gewichten (q_{ij}) van de basisparameter(s) geven aan hoe vaak de personen gebruik maken van een deelvaardigheid bij het oplossen van een item. De relevante strategieën en bijbehorende basisparameters voor het toetsen van de cognitieve strategieën zijn:

1. De persoon let op gewicht (N_1). Als voor het oplossen van item i ($i = 1, 2, \dots, k$) deze strategie nodig is dan krijgt N_1 een gewicht $q_{i1} = 1$, in het andere geval geldt $q_{i1} = 0$.
2. Er wordt met afstand rekening gehouden als de gewichten gelijk zijn (N_2). Deze strategie speelt een rol bij het oplossen van afstanditems. Het gewicht q_{i2} heeft dezelfde betekenis als in het vorige geval.
3. De persoon let op gewicht en afstand (N_3). Bij konfliktitems worden afstand en gewicht met elkaar vergeleken. Het gewicht q_{i3} neemt weer de waarden nul of één aan.
4. Afstand en gewicht worden vergeleken (N_4). Bij konfliktitems wordt de opgave herleid tot een eenvoudiger op te lossen balansitem. Het gewicht q_{i4} geeft weer aan- of afwezigheid aan van de deelvaardigheid.

In het LLTM wordt voor elk item vastgelegd welke strategieën een persoon toe moet passen om een item goed te kunnen beantwoorden. Binnen het LLTM worden de voor het oplossingsproces van een item benodigde cognitieve processen dus gespecificeerd aan de hand van itemkenmerken en niet met behulp van persoonskenmerken. Indien personen bij het oplossen van één balansitem verschillende strategieën kunnen gebruiken dan kan dit consequenties hebben voor het passen van het LLTM bij de empirische gegevens. De keuze van een persoon voor een bepaalde strategie kan niet worden verdisconteerd in het LLTM, dat alleen een model specificeert op basis van itemkenmerken.

Zoals we hebben gezien kunnen balansitems met verschillende strategieën worden opgelost. De vraagstelling van dit onderzoek is of een passend LLTM voor balansproblemen voor de hele populatie te vinden is, uitgaande van een Rasch-homogene schaal. We verwachten dat dat niet het geval zal zijn omdat proefpersonen verschillende strategieën gebruiken. Mocht een passend LLTM voor de hele populatie niet worden gevonden, dan zal vervolgens worden onderzocht of een passend LLTM per strategie-homogene subgroep te konstrueren valt.

METHODE

Steekproef

Het onderzoek werd uitgevoerd op 12 basisscholen in de provincies Friesland en Groningen. In totaal namen 484 leerlingen (235 in groep 7 (klas 5) en 249 in groep 8 (klas 6)) aan het onderzoek deel. De testgegevens van vier leerlingen bleken onbruikbaar te zijn.

Items en procedure

Het balansmodel van Siegler en Klahr diende als uitgangspunt bij het vervaardigen van de items. Er werd gebruik gemaakt van vijf typen items van dit model: gewicht, afstand, konfliktgewicht, konfliktafstand en konfliktbalans. Van elk type item werden vijf verschillende items gekonstrueerd (zie Figuur 1), zodat aan de leerlingen in totaal 25 balansitems werden gepresenteerd. Vijf items hebben 'blijft in balans' als goede oplossing. Bij de overige twintig items gaat de balans bij de helft van de items naar links en bij de andere helft naar rechts. De antwoordmogelijkheden zijn willekeurig over de verschillende itemtypen verdeeld.

Op iedere school werd in groep 7 en groep 8 klassikaal het doel van het onderzoek uitgelegd en volgde er een gestandaardiseerde instructie over het beantwoorden van de items. Vervolgens werden de leerlingen in de gelegenheid gesteld de 25 items schriftelijk te beantwoorden. De test werd als power-test aan de leerlingen voorgelegd. Om eventuele leer- en vermoeidheidseffecten te ondervangen werden de items in elk boekje in willekeurige volgorde geplaatst. Voor elke leerling werd per item de uitkomst als goed (1) of fout (0) gescoord.

Analyses op de gegevens

Teneinde na te kunnen gaan of het LLTM een passende beschrijving geeft van de vier cognitieve oplossingsstrategieën, werden twee stadia van gegevens-analyse doorlopen. Ten eerste werd nagegaan of de 25 items in de onderzochte populatie schaalbaar zijn volgens het Rasch model. Schaalbaarheid is een nodige voorwaarde voor het passen van het LLTM, dat een verbijzondering is van het Rasch model. Ten tweede werd, na gebleken schaalbaarheid van de items volgens het Rasch model in de deelgroepen van personen, voor elk van vier oplossingsstrategieën onderzocht in hoeverre het gedrag van de leerlingen kon worden verklaard met behulp van het LLTM. Aangezien in dit onderzoek de toepasbaarheid van het LLTM op de balansitems centraal staat, zullen de resultaten met betrekking tot het passen van het Rasch model slechts summier worden genoemd.

RESULTATEN

Rasch-homogeniteit

De schaalbaarheid van de 25 balansitems volgens het Rasch model werd onderzocht met behulp van de procedure die is beschreven door Sijsma (1983). Deze procedure houdt in dat eerst wordt onderzocht of het diskriminerend vermogen van alle items in de test identiek is. Indien items een afwijkend diskriminerend vermogen hebben, worden zij uit de test verwijderd. Vervolgens wordt voor de resterende items onderzocht of zij te zamen een eendimensioneel konstrukt meten. Ook in dit stadium van het onderzoek kunnen weer items afvallen, of kunnen ééndimensionele subtests worden gevormd. Eventueel begint de analyse op de resterende items of subtests weer opnieuw en wordt voortgezet totdat de gegevens aan het model voldoen.

De verzameling van items bleek niet te voldoen aan de eisen van het Rasch model. Zo bleek in de eerste fase van de zojuist beschreven procedure voor itemanalyse dat de gebruikte chi-kwadraat verdeelde toetsingsgrootte (Andersen's likelihood ratio test; zie Andersen, 1973) gelijk was aan 983 bij 24 vrijheidsgraden. De consequentie van dit resultaat is dat het zinloos is om het LLTM toe te passen op de hele groep van leerlingen in een verder onderzoek naar de cognitieve procesmodellen.

Rasch-homogeniteit binnen strategie-homogene deelgroepen

Aangezien op basis van de literatuur (Siegler, 1976; Siegler & Klahr, 1982), alsmede een verricht vooronderzoek, bekend was dat de leerlingen verschillende strategieën gebruiken, werd vervolgens getracht om strategie-homogene groepen van leerlingen te vormen. Binnen een strategie-homogene groep werd vervolgens nagegaan of het Rasch model paste.

Voor deze twee fasen van gegevensanalyse werd gebruik gemaakt van een partitie van de gegevens in twee deelmatrices: één gebaseerd op vijf items en een ander gebaseerd op de overige twintig items. Strategie-homogene groepen werden gevormd met behulp van de gegevens betreffende de vijf items. Rasch-homogeniteit werd daarna onderzocht met behulp van de gegevens betreffende de twintig items. Deze werkwijze garandeert dat de indeling in groepen onafhankelijk is van het onderzoek naar Rasch-homogeniteit, alsmede het later te verrichten onderzoek naar het passen van het LLTM binnen deze groepen.

Vorming van strategie-homogene groepen

In deze paragraaf wordt nagegaan of het mogelijk is om strategie-homogene subpopulaties van elkaar te onderscheiden. Voor dat doel werd een hiërarchische clusteranalyse met het pro-

gramma Clustan uitgevoerd op de antwoordpatronen van de leerlingen op vijf items. Van elk type item (zie Figuur 1) was één item in de gegevens vertegenwoordigd. Er werden clusters gevormd van personen die een gelijksoortig antwoordpatroon hebben.

Personen werden op basis van hun antwoordpatroon door Clustan ingedeeld in één van vier, ongeveer even grote clusters. In Tabel 1 zijn de aantallen personen per cluster opgenomen. De antwoordpatronen in elk van de vier clusters komen in het algemeen redelijk tot perfect overeen met de antwoordpatronen zoals deze door het balansmodel worden voorspeld in geval van de Strategieën 1, 2, 3, en 4.

Aangezien antwoordpatronen gebaseerd op slechts vijf items nogal onbetrouwbaar zijn ten gevolge van steekproeffluctuaties, werd de clusteranalyse nog eens herhaald op de antwoordpatronen gebaseerd op de totale verzameling van 25 items. Een belangrijke gevolgtrekking was in dit geval dat de formeel juiste Strategie 5 niet voorkwam. Vandaar dat bij de clusteranalyse gebaseerd op vijf items besloten werd om vier clusters te vormen, zodat antwoordpatronen die "toevallig" bestaan uit vijf enen niet een aparte vijfde cluster vormen.

Tenslotte vermelden we dat twee personen buiten de analyse werden gelaten, aangezien zij scorepatronen hadden bestaande uit vijf nullen. Een zinvolle indeling van deze personen leek niet mogelijk.

Schaalbaarheid van items binnen strategie-groepen

Schaalbaarheid van de items volgens het Rasch model werd vervolgens per strategie-homogene groep onderzocht. In Tabel 1 zijn de resultaten met betrekking tot het onderzoek naar identiek diskriminerend vermogen van de items weergegeven. Het is duidelijk dat op dit punt geen belangrijke afwijkingen van het Rasch model in de gegevens aanwezig zijn. Wat betreft het

Tabel 1: Resultaten van onderzoek naar identiek diskriminerend vermogen van items binnen strategie-homogene groepen. Uitgevoerd is de Andersen-toets op een indeling in hoge en lage ruwe scores.

Strategie	n	aantal items	χ^2	df	p	KR20
1	118	12	25.1	11	0.01	0.54
2	135	17	32.0	16	0.01	0.49
3	111	17	28.2	16	0.03	0.35
4	114	17	29.8	16	0.02	-0.12

Noot: De aantallen items zijn kleiner dan twintig. De reden hiervoor is dat sommige items door alle personen goed of fout werden gemaakt. Dergelijke items zijn ongeschikt voor een Rasch-analyse.

onderzoek naar eendimensionaliteit kan worden gekonkludeerd dat de Groepen 2 en 3 niet voldoen terwijl de Groepen 1 en 4 wel voldoen. Ter illustratie zijn in Tabel 2 resultaten opgenomen met betrekking tot Groep 3. Deze resultaten zijn verkregen door de splitter-item techniek volgens Van den Wollenberg (1982) toe te passen op items met niet-extreme proporties van goede antwoorden. In Tabel 2 is duidelijk dat schending van de eis van eendimensionaliteit gezien de consistentie in de resultaten, zeer plausibel is.

Als eindkonklusie geldt dat het Rasch model binnen de Groepen 1 en 4 een adequate beschrijving geeft van de testgegevens. Binnen de Groepen 2 en 3 is de assumptie van eendimensionaliteit geschonden en geldt het Rasch model niet. Strikt genomen mag het LLTM dus alleen binnen de Groepen 1 en 4 worden getoetst.

Het LLTM en cognitieve procesmodellen

Per strategie-homogene groep werd tenslotte onderzocht of het LLTM een passende beschrijving geeft van de oplossingsstrategie die per groep wordt verondersteld. Daarbij werden ook de Groepen 2 en 3 onderzocht, hoewel de items in deze groepen niet schaalbaar bleken. De resultaten van deze groepen moeten dan ook met de nodige voorzichtigheid worden geïnterpreteerd.

Tabel 2: Resultaten van onderzoek naar eendimensionaliteit binnen strategie-groep 3. Uitgevoerd is de Andersen-toets volgens de splitter-item techniek.

Itemnummer	proportie goed	χ^2	df	p
1	0.32	29.3	15	0.02
3	0.40	44.6	15	< 0.01
7	0.30	65.3	15	< 0.01
10	0.58	53.8	15	< 0.01
14	0.55	59.5	15	< 0.01
19	0.82	40.7	15	< 0.01

Ter illustratie bespreken we de resultaten van de toepassing van het LLTM voor Groep 2 uitgebreid. De resultaten van de overige groepen geven we kort weer.

In Groep 2 worden voor de oplossing van de items naar verwachting twee deelvaardigheden gebruikt:

1. Er word alleen op gewicht gelet (basisparameter N_1).
2. Met afstand word rekening gehouden indien de gewichten gelijk zijn (basisparameter N_2).

Items die door alle respondenten uit Groep 2 goed dan wel fout werden beantwoord, zijn niet geschikt voor analyse met behulp van het LLTM, en zijn derhalve weggelaten. In Tabel 3 zijn voor de resterende items de gewichten q_{i1} en q_{i2} aangegeven.

De schattingen van de basisparameters zijn te vinden onder de kolommen q_{i1} en q_{i2} . Ook zijn de geschatte moeilijkheidsparameters ($\hat{b}_i$) van de items op basis van het Rasch model gegeven, alsmede de schatting ($\hat{b}_{i+}$) hiervan op basis van het LLTM. Naarmate de schattingen $\hat{b}_i$ en $\hat{b}_{i+}$ sterker verschillen, zal het LLTM slechter bij de gegevens passen.

In Tabel 4 is achter Strategie 2 te zien dat op basis van de chi-kwadraat verdeelde toetsingsgrootte van het LLTM de nulhypothese van een passend model moet worden verworpen op een 1% significantieniveau. Desondanks verklaart het LLTM 92% (R^2) van de variantie van de itemmoeilijkheden ($\hat{b}_i$) op basis van het Rasch model.

Teneinde een beter passend LLTM te verkrijgen, lijkt het nuttig om de residuen van de

Tabel 3: Resultaten van de LLTM-toetsing voor Strategie-groep 2.

Itemtype	Itemnr	q_{i1}	q_{i2}	$\hat{b}_i$	$\hat{b}_{i+}$	$\hat{b}_i - \hat{b}_{i+}$
Konfl. bal.	1	0	0	-2.57	-1.95	0.62
" "	2	0	0	-2.03	-1.95	0.08
" "	3	0	0	-1.25	-1.95	0.70
" "	5	0	0	-2.17	-1.95	0.22
Konfl. afst.	6	0	0	-2.33	-1.95	0.38
" "	7	0	0	-2.64	-1.95	0.69
" "	9	0	0	-2.27	-1.95	0.32
" "	10	0	0	-1.44	-1.95	0.51
Konfl. gew.	14	1	0	1.17	2.23	1.06
" "	15	1	0	1.83	2.23	0.40
Afstand	16	0	1	1.11	1.12	0.01
" "	17	0	1	0.84	1.12	0.28
" "	19	0	1	1.17	1.12	0.05
" "	20	0	1	1.17	1.12	0.05
Gewicht	21	1	0	3.25	2.23	1.02
" "	23	1	0	2.92	2.23	0.69
" "	25	1	0	3.25	2.23	1.02
	$\hat{N}_j$	4.18	3.07			

Tabel 4: Resultaten van de analyses met het LLTM.

Strategie	n	aantal basispar.	aantal items	χ^2	df	p	R
1	118	1	12	52.4	10	< 0.01	0.96
2	135	2	17	68.2	14	< 0.01	0.96
3	111	3	17	60.0	13	< 0.01	0.96
4	114	3	17	29.6	13	< 0.01	0.98

Noot: R staat voor de multiële correlatie tussen de item-moeilijkheid geschat door middel van de basisparameters en de itemmoeilijkheid rechtstreeks geschat door middel van het Rasch model.

schattingen van de moeilijkheidsparameters van de items nader te beschouwen. Uit deze residuen (Tabel 3) blijkt dat relatief grote afwijkingen voorkomen bij items van diverse typen. Een inhoudelijke verklaring voor de afwijkendheid van deze items valt vanwege deze diversiteit moeilijk te geven. Verwijdering van enkele items met grote residuen zou kunnen leiden tot een passend LLTM, maar bij het ontbreken van een inhoudelijke verklaring voor hun afwijkendheid zien we van verdere analyses af.

De resultaten met betrekking tot de overige groepen geven we summier weer.

Voor de Groepen 1 en 3 geldt, evenals voor Groep 2, dat het LLTM strikt genomen wordt verworpen op een 1% significantieniveau. Inspectie van de residuen per item laat voor Groep 1 zien dat item 15 van het type konflikt gewicht extreem afwijkt (residu gelijk aan 3.08). Aangezien in Groep 1 item 15 het enige item van dit type is, werd het LLTM na verwijdering van item 15 nogmaals getoetst op de gegevens. Voor het gemodificeerde model geldt nu dat de toetsingsgrootte gelijk is aan 50.6 bij 9 vrijheidsgraden, met $p < 0.01$. Het weglaten van item 15 levert dus nauwelijks een beter passend model op. Verdere modificaties werden achterwege gelaten.

Binnen Groep 3 geldt dat, ondanks dat het LLTM niet past, geen items werden gevonden met relatief grote residuen. Derhalve werd hier van verdere modificatie van het model afgezien.

Tenslotte blijkt uit Tabel 4 dat het LLTM voor Groep 4 tamelijk goed past als verklarend model voor het testgedrag, hoewel het LLTM strikt genomen wordt verworpen. Dit resultaat zou kunnen impliceren dat de drie deelvaardigheden:

1. er wordt alleen op gewicht gelet;
2. bij gelijke gewichten wordt er met afstand rekening gehouden;
3. bij konflikt items wordt getracht het probleem te herleiden tot een eenvoudig op te lossen balansprobleem;

het oplossingsgedrag van de respondenten in Groep 4 afdoende verklaren.

KONKLUSIES

Hoewel balansitems 'op het oog' een homogene verzameling items lijken te zijn, bleek het niet mogelijk om met de complete verzameling items een Rasch-homogene schaal te vormen. De indeling in vier subpopulaties leverde binnen de Strategieën 1 en 4 Rasch-homogene schalen op. Bij de toetsing van het LLTM werd strikt genomen geen passend LLTM gevonden. Als uitzondering kan wellicht het model volgens Strategie 4 gelden. Hoewel dit model moet worden verworpen op 1% significantieniveau lijkt de afwijking van model en gegevens toch zeer gering.

Het is duidelijk dat er verschillende strategieën bestaan bij het oplossen van balansproblemen. De resultaten wijzen er op dat de drie strategieën die het balans model (Siegler & Klahr, 1982) voorspelt en de vierde strategie die in het vooronderzoek werd gevonden in de onderzochte populatie op consequente wijze worden gebruikt. Deze strategie-heterogeniteit is van invloed op de bruikbaarheid van het LLTM.

Ofschoon strategie-heterogeniteit een beletsel kan zijn voor de toepassing van het LLTM, kan ook strategie-homogeniteit tot problemen leiden. Dit is als volgt in te zien.

Een strategie-homogene groep bestaat uit personen die in principe hetzelfde patroon van itemscores genereren. Empirisch zijn er natuurlijk verschillen mogelijk, maar deze verschillen zijn vrij gering. Het gevolg is dat de ruwe score van de respondenten weinig spreiding vertoont. In het Rasch model betekent dit dat de personen allen gelokaliseerd zijn op een klein interval van de schaal, waarop gemeten wordt. Wanneer nu alle personen precies op hetzelfde punt zouden zijn gelokaliseerd, zou het Rasch model per definitie passen (Wood, 1978). Kortom, onderzoek naar het passen van het Rasch model is in deze situatie zinloos, en een serieuze toetsing van het LLTM kan niet plaatsvinden.

In de vier strategie-homogene groepen uit dit onderzoek is de variantie van de ruwe scores steeds relatief gering. In Tabel 1 is te zien dat KR20 voor de Strategieën 1, 2 en 3 een middelmatige waarde heeft, hetgeen er op wijst dat de variantie van de ruwe scores voor een belangrijk deel systematisch is. Voor Strategie 4 geldt waarschijnlijk dat, gegeven KR20, de variantie van de ruwe scores voor een belangrijk deel toevallig is. Hieruit zou men de conclusie kunnen trekken dat het zojuist genoemde artefact vooral doorwerkt in de resultaten voor Strategie 4. Men kan achteraf naar voren brengen dat de gegevens misschien met een latente klassen model kunnen worden geanalyseerd. Dat was echter nadrukkelijk niet de bedoeling van dit onderzoek. In de literatuur zijn - althans zover wij weten - geen resultaten vermeld waarbij het LLTM bij de gegevens past. In dit onderzoek hebben wij geprobeerd condities te creëren die voor toepassing van het LLTM ideaal zijn. Daarbij kunnen zich, zoals nu blijkt, nieuwe problemen voordoen die in de literatuur over het LLTM nog niet eerder zijn gesignaleerd.

Los van een eventueel artefact in de resultaten, blijft staan dat de onderzochte situatie inhoudelijk zinvol is te noemen. De uitspraak van Newell (1973): "Ken de methode die proefpersonen gebruiken bij het uitvoeren van experimentele taken" en "middel nooit over methoden", is hier van toepassing.

Samengevat luidt de voorlopige konklusie met betrekking tot de bruikbaarheid van het LLTM als volgt. Indien men het LLTM wil gebruiken om cognitieve procesmodellen op hun empirische geldigheid te toetsen, dan is het een vereiste dat de te onderzoeken populatie strategie-homogeen is. Een onderzoek naar strategie-homogeniteit dient dan ook vooraf te gaan aan een toepassing van het LLTM. Wanneer men strategie-homogene groepen heeft gevormd, kan het echter gebeuren dat de respondenten binnen een groep niet meer verschillen op de gemeten eigenschap. Een serieuze toetsing van het LLTM kan dan ook niet meer plaatsvinden. Al met al is het duidelijk dat aan de toepasbaarheid van het LLTM veel haken en ogen zitten, en dat resultaten die met dit model zijn verkregen met grote voorzichtigheid moeten worden geïnterpreteerd.

LITERATUUR

- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.
- Been, P. H., Jorna, R. J., & Sijtsma, K. (1984). Over kijken, meten en modellen bij breukrekentaken. In P. G. Vos, K. B. Koster, & J. Kingma (eds.). *Rekenen. Balans van standpunten in theorievorming en empirisch onderzoek*. Lisse: Swets & Zeitlinger.
- Broadbent, D. E. (1984). The Maltese cross, a new simplistic model for memory. *The Behavioral and Brain Sciences*, 7, 1, 55-95.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.
- Fischer, G. H., & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement*, 6, 397-416.
- Haeussler, P. (1977). Investigation of mathematical reasoning in science problems. In H. Spada & W. F. Kempf (eds.). *Structural models of thinking and learning*. Bern: Huber.
- Kubinger, K. D. (1979). Das Problemlöseverhalten bei der statistischen Auswertung psychologischer

- Experimente. Ein Beispiel hochschuldidaktischer Forschung. *Zeitschrift für experimentelle und angewandte Psychologie*, 26, 467-495.
- Kubinger, K. D. (1980). Die Anwendung von Fischers linear logistischem Test Modell zur Leistungssteigerung in Lehrveranstaltungen - Einige neuere Ergebnisse. *Archiv für Psychologie*, 133, 69-79.
- Nährer, W. (1980). Zur Analyse von Matrizenaufgaben mit dem linearen logistischen Testmodell. *Zeitschrift für experimentelle und angewandte Psychologie*, 27, 553-564.
- Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (ed.), *Visual information processing*. New York: Academic Press.
- Scheiblechner, H. (1972). Das Lernen und Lösen komplexer Denkaufgaben. *Zeitschrift für experimentelle und angewandte Psychologie*, 19, 3, 476-506.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481-520.
- Siegler, R. S. & Klahr, D. (1982). When do children learn. In R. Glaser (ed.), *Advances in instructional psychology*, vol 2. Hillsdale N. J. L.: Erlbaum Assoc.
- Sijtsma, K. (1982). Een lineair logistisch model ter verklaring van de moeilijkheidsparameters van breukrekenitemen. In J. G. L. C. Lodewijks & P. R. J. Simons (eds.), *Strategieën in leren en ontwikkeling*. Lisse: Swets & Zeitlinger.
- Sijtsma, K. (1983). Rasch-homogeniteit empirisch onderzocht. *Tijdschrift voor Onderwijsresearch*, 8, 104-121.
- Spada, H. (1976). *Modelle des Denkens und Lernens*. Bern: Huber.
- Spada, H., & McGaw, B. (1985). The assessment of learning effects with linear logistic test models. In S. E. Embretson (Ed.), *Test design. Developments in psychology and psychometrics*. Orlando: Academic Press, Inc.
- Vijver, F. J. R. van de (1988). Systematizing the item content in test design. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models*. New York: Plenum Press.
- Whitely, S. E., & Schneider, L. M. (1981). Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement*, 5, 383-397.
- Wollenberg, A. L. van den (1982). A simple and effective method to test the dimensionality axiom of the Rasch model. *Applied Psychological Measurement*, 6, 83-91.
- Wood, R. (1978). Fitting the Rasch model - A heady tale. *British Journal of Mathematical and Statistical Psychology*, 31, 27-32.
- Zwarts, M. A. (1983). *Criteriumtoetsen bij aansluiting van primair en secundair onderwijs*. (ongepubliceerde dissertatie). Rijksuniversiteit Utrecht.

Manuscript ontvangen 22-3-1988.

Definitieve versie ontvangen 28-7-1988

Schrijven en schrijven is twee: een onderzoek naar de samenhang tussen prestaties op verschillende schrijftaken*

Huub van den Bergh**

Stichting Centrum voor Onderwijsonderzoek

ABSTRACT

The convergent validity of writing assessment was investigated. Third grade pupils ($n=619$) from secondary education wrote five essays each (two long ones and three short ones). The results showed that only a small proportion of the variance in scores could be explained with a writing skill factor. Topic specific variance proved to be extremely important. It was suggested that 'topic variance' was in fact background knowledge or knowledge of the textstructure. To a lesser degree the significance of rating method specific variance was demonstrated.

1. INLEIDING

Het meten van de schrijfvaardigheid van leerlingen stelt onderwijsgevers en -onderzoekers vaak voor lastige problemen. Allereerst moeten de schrijfprodukten beoordeeld worden, en dat verschillende beoordelaars vaak een andere kwantificering aan hetzelfde schrijfprodukt toekennen mag als bekend verondersteld worden (zie bij voorbeeld: Blok, 1984a, 1984b, 1985; Hert & Rijmenans, 1987; McColly, 1970; Meuffels, 1984, 1985; Wesdorp, 1981, 1983). Het spreekt haast voor zich dat een betrouwbare beoordeling een noodzakelijke, maar nog niet voldoende voorwaarde is voor een valide meting van de schrijfvaardigheid.

Een tweede probleem dat inherent lijkt aan de meting van schrijfvaardigheid wordt zichtbaar, wanneer de samenhang van prestaties van leerlingen op verschillende opdrachten bestudeerd wordt. Verwacht zou mogen worden dat de prestatiescores van de leerlingen op verschillende opdrachten onderling hoog correleren, daar met verschillende schrijfopdrachten getracht wordt hetzelfde concept schrijfvaardigheid te meten. Dat blijkt vaak niet het geval, ook niet na een correctie voor de beoordelingsonbetrouwbaarheid van de jury die de schrijfprodukten beoordeelde (zie o.a. Farrell, 1971; Hogan & Mishler, 1980; Moss e.a., 1982). In de Voorstudie Periodieke Peiling van het Onderwijsniveau (Wesdorp e.a., 1986) bleek de gemiddelde correlatie tussen de prestaties van zesdeklassers op (in totaal tien) schrijfopdrachten, ondanks vaak hoge jurybetrouwbaarheden, niet veel hoger dan nul. Ook Blok (1987) vond geen hoge samenhang tussen prestaties van LBO- en MAVO-eindexamen kandidaten op verschillende opdrachten. Quellmalz (1977, 1982) concludeerde dat lange schrijfopdrachten een meer valide meting voor de meting van schrijfvaardigheid zijn dan korte schrijfopdrachten. De Glopper (1986b) demonstreerde dat de samenhang tussen prestaties van leerlingen uit het derde

* Dit onderzoek is tot stand gekomen dankzij een subsidie van het Instituut voor onderzoek van het onderwijs (SVO).

** Met dank aan G.J. Mellenbergh, R. Schoonen en G. Rijlaarsdam voor hun commentaar op een eerdere versie van dit artikel

Adres: Grote Bickersstraat 72, 1013 KS Amsterdam

leerjaar in het voortgezet onderwijs op langere schrijfoopdrachten hoger is dan die tussen korte schrijfoopdrachten, waarmee hij de conclusie van Quellmalz kon ondersteunen.

Natuurlijk kunnen geen extreem hoge correlaties tussen schrijfpredstaties van leerlingen verwacht worden; het lijkt plausibel te stellen dat een schrijfoopdracht een één-item toets is, wanneer we van de gangbare beoordelingsprocedures (één opstel, één oordeel) uitgaan. Het probleem is dan dat we niet eenvoudig de toets kunnen verlengen om de betrouwbaarheid van de toets te verhogen. Schrijfvaardigheidsonderzoek is duur en arbeidsintensief. Bovendien moeten leerlingen vaak lang werken aan één opdracht (één item) en moet het schrijfprodukt daarna door verscheidene beoordelaars beoordeeld worden. Het is daarom zaak om een zo goed (valide) mogelijke opdracht aan de leerlingen voor te leggen.

In het voorgaande zijn twee vormen van invaliditeit van het schrijfvaardigheidsmetingen aangestipt (beoordelaarsbetrouwbaarheid en een lage samenhang tussen prestaties op verschillende schrijfoopdrachten). Wesdorp (1974) en Coffman (1966) beargumenteren op grond van eigen en ander onderzoek dat er aan een leerling minimaal vier schrijfoopdrachten afgenomen moeten worden, en dat de schrijfprodukten door vijf beoordelaars beoordeeld moeten worden, voor een valide meting van zijn schrijfvaardigheid.

In dit artikel wordt getracht inzicht te verkrijgen in mogelijke oorzaken van de lage samenhangen tussen verschillende schrijfpredstaties van een leerling. Deze vraag is niet alleen van theoretisch, maar ook van praktisch belang, daar in 1987 het onderdeel schrijfvaardigheid in het Centraal Schriftelijk Eindexamen LBO/MAVO gewijzigd is. LBO-leerlingen moeten nu drie à vier schrijfoopdrachten maken, die gebaseerd zijn op schrijfsituaties die in de praktijk van alle dag voor kunnen komen (zie o.a. Wesdorp e.a., 1986; Blok, 1987; Van den Bergh, 1988). Voor MAVO-leerlingen mag de school een keuze maken: de leerlingen moeten in het Centraal Schriftelijk Eindexamen dezelfde functionele opdrachten maken als de LBO-kandidaten, of zij moeten een traditioneel (lang) opstel (blijven) schrijven.

Parallel aan deze verandering in type opdracht is ook een verandering in beoordelingsprocedure doorgevoerd; de prestaties van de leerlingen op de functionele opdrachten worden door de klasseleerkracht gescoord aan de hand van een analytisch scoringsvoorschrift. In dit scoringsvoorschrift zijn de relevante kenmerken waaraan een (ideale) prestatie moet voldoen gespecificeerd. De beoordeling van het opstel is niet veranderd; dit blijft het traditionele globale oordeel van de klasseleerkracht. De ratio achter deze verandering in beoordelingsprocedure is de wens om een grotere objectiviteit van de beoordelingen van schrijfprodukten uit het Centraal Schriftelijk Eindexamen te realiseren. Echter, bekend is dat de toepassing van een analytische beoordelingsprocedure niet zonder meer tot een hogere overeenstemming (grotere objectiviteit) tussen verschillende beoordelaars leidt (zie o.a. Meuffels, 1985).

2. SCHRIJFOPDRACHTEN EN BEOORDELINGSPROCEDURES

In totaal zijn vijf schrijfoopdrachten afgenomen aan 619 leerlingen uit de derde klas van het LBO en MAVO. Elke leerling maakte drie korte functionele opdrachten en twee lange opstellen van minimaal één à twee A4. Hieronder wordt een korte omschrijving van de vijf opdrachten gegeven.

1. De zin of onzin van huiswerk.

Voor deze opdracht moet de leerling in twee lesuren een betoog schrijven over de zin (nut) of de onzin ('onnut') van huiswerk. Ter ondersteuning heeft de leerling de beschikking over vijf pagina's tekst, waarin aan de hand van citaten een aantal meningen over het nut van huiswerk toegelicht wordt. De leerling moet in twee lesuren een betoog van 1 à 2 kantjes A4 schrijven. De opdracht is ontleend aan Rijlaarsdam (1986).

2. Een examenopdracht.

Voor de tweede stelopdracht is een keuze gemaakt uit de eindexamentitels waaruit de leerlingen

de afgelopen 5 jaar konden kiezen. De opdracht aan de leerlingen luidt: schrijf een opstel van $\pm$ 500 woorden over één van de volgende tien titels. Deze opdracht is afkomstig uit het eindexamen MAVO-C van 1984. De tijd die de leerlingen voor deze opdracht hebben is wederom twee lesuren van 50 minuten.

3. Een ooggetuigeverslag.

De leerling wordt geconfronteerd met een situatiebeschrijving, waarin hij (fictief) getuige is van het gehaast vertrek van een motorrijder. Enkele dagen later leest de leerling in de krant een oproep waarin getuigen van een motordiefstal gezocht worden. Het is aannemelijk dat de leerling de diefstal gezien heeft. Voorts heeft de leerling de beschikking over een kleurenfoto waar de (vermoedelijk) gestolen motor en dief op staan. De opdracht aan de leerling luidt: schrijf een brief aan de politie waarin je verslag doet van deze gebeurtenis. Deze opdracht, waarvoor leerlingen 30 minuten de tijd krijgen, is ontleend aan Blok en Wolf (1985).

4. Een sollicitatiebriefje.

Op grond van een advertentie, waarin een vakantiehulp gevraagd wordt, moet de leerling een sollicitatiebrief aan een groot warenhuis schrijven. Voor het schrijven van het (zeer korte) sollicitatiebriefje hebben de leerlingen 20 minuten de tijd. Deze opdracht is ontwikkeld in het project 'Internationale vergelijking van de stelvaardigheid' (De Glopper, 1986a).

5. Een formulier invullen.

Voor de laatste functionele opdracht moet de leerling, op grond van een fictieve correspondentie met de eigenaar van een kampeerboerderij, een formulier invullen. Voor deze opdracht, die ontleend is aan de Voorstudie Periodieke Peiling in het Voortgezet onderwijs (Van den Bergh e.a., 1986), krijgen de leerlingen 25 minuten de tijd.

Zoals uit de omschrijvingen blijkt, zijn vier van de vijf opdrachten examen of examenachtige opdrachten. Alleen de eerste opdracht over de zin en onzin van huiswerk wijkt in dit opzicht af. Een dergelijke opdracht is geen reguliere examenopdracht, maar lijkt meer op een opdracht zoals die in het project 'Gericht schrijven' gehanteerd wordt. Deze opdracht is opgenomen, omdat de leerlingen aan de hand van documentatiemateriaal een (lang) opstel (betoog) moeten schrijven. Wat dit betreft is de opdracht 'De zin en onzin van huiswerk' een mengvorm van de korte functionele opdracht, waarbij informatie verschaft wordt, en een lang opstel.

2.1. BEOORDELINGSPROCEDURES

De schrijfsprodukten van de leerlingen op de bovenstaande opdrachten zijn op verschillende manieren beoordeeld. In totaal zijn drie verschillende beoordelingsprocedures gebruikt: analytische scoringsvoorschriften, schaalbeoordelingen en docentenoordelen; voor drie van de vijf opdrachten zijn twee beoordelingsprocedures toegepast, namelijk voor de opdrachten: Ooggetuigeverslag, Sollicitatiebriefje en De zin en onzin van huiswerk. Bij de analytische scoringsprocedure werden correct genoemde inhoudselementen geteld (voor de opdrachten: Formulier, Ooggetuigeverslag en Sollicitatiebriefje) en de correctheid van door de leerling gehanteerde conventies en lay-out voorschriften (voor de opdrachten: Ooggetuigeverslag en Sollicitatiebriefje). Ter illustratie zijn in Tabel 1 twee elementen uit het scoringsvoorschrift voor het Sollicitatiebriefje weergegeven.

De leerlingprestaties zijn gescoord door één beoordelaar, honderd leerlingprestaties zijn daarna, ter controle op de scoringsprocedure, door een tweede beoordelaar gescoord. Naast de in Tabel 2 weergegeven overeenstemming tussen de beide beoordelaars is ook de homogeniteit (α) voor de scoringsvoorschriften berekend. Deze bleek voor alle drie de opdrachten bevredigend (de homogeniteitsschattingen (α) varieerden van: .80 voor het Formulier tot .75 voor het Sollicitatiebriefje).

De tweede beoordelingsprocedure, de zogenaamde schaalbeoordelingsprocedure, verdient

Tabel 1. Twee items uit het scoringsvoorschrift betreffende het scoren van de aanhef in de sollicitatiebriefjes.

Element	Aanvaardbare formulering(en)/positie(s)	Onaanvaardbare formulering(en)/ positie(s)
Aanhef: positie	de positie moet aan de volgende eisen voldoen: - op een verder ongebruikte regel; - door een witregel gescheiden van de regel eronder; - door een witregel gescheiden van hetgeen er direct boven staat (indien van toepassing); - op de linkerhelft of het midden van de regel. (score: 1)	anders (score: 0)
Aanhef: formulering:	- Geachte heer - Geachte heren - Mijne heren - Mijne dames en heren - Geachte leden van de sollicitatiecommissie - L.S. - Aan het hoofd van de personeelsafdeling (score: 1)	- Geachte directeur - Geachte Postkoop - Geacht postorderbedrijf - Geachte meneer Postkoop - Geachte - Aan de heer van de personeelsafdeling (score: 0)

wellicht enige toelichting. Bij de schaalbeoordelingen beschikten de beoordelaars over vijf in kwaliteit oplopende voorbeeldopstellen: een opstelschaal. (In de bijlage bij dit artikel is een opstelschaal weergegeven, het betreft de schaal voor het beoordelen van de Inhoud & organisatie van de sollicitatiebriefjes) Aan elk voorbeeldopstel is, op grond van de resultaten van een vooronderzoek (Van den Bergh, 1987) een kwantificering toegekend zo dat de schaal een gemiddelde van 100 en een standaardafwijking van 15 heeft. De taak van de beoordelaars bestaat uit het vergelijken van het te beoordelen opstel met de voorbeeldopstellen, en het te beoordelen opstel als het ware op de schaal te plaatsen en de bijbehorende waardering af te lezen (zie: Blok & Hoeksma, 1984; Van den Bergh & Rijlaarsdam, 1987). Met behulp van deze methode van schaalbeoordeling worden vaak zeer hoge inter- en intrabeoordelingsovereenstemmingen gerealiseerd (zie o.a. Wesdorp e.a., 1986; Rijlaarsdam, 1986).

In het onderhavige onderzoek zijn drie opdrachten beoordeeld met behulp van opstelschalen: Ooggetuigeverslag, Sollicitatiebriefje en De zin en onzin van huiswerk. Alle prestaties op deze opdrachten zijn beoordeeld op de aspecten Globale kwaliteit, Taalgebruik en Inhoud & Organisatie door twee beoordelaars per aspect. Bij de beoordeling van de Globale kwaliteit vormden de beoordelaars zich een indruk over de kwaliteit van het schrijfprodukt als geheel. In hun oordeel betrokken zij diverse relevante kenmerken: met name inhoud, organisatie, stijl, interpunctie en 'netheid'. Bij de beoordeling van het Taalgebruik werd niet enkel gelet op de grammaticaliteit maar ook op de stijl (de gepastheid van het taalgebruik gegeven de context) van het schrijfprodukt. Het oordeel over de Inhoud & Organisatie tenslotte werd gelet of een aantal relevante inhoudselementen genoemd werden en of deze op een overzichtelijke of logische wijze geordend waren.

De derde beoordelingsprocedure die onderscheiden werd is de beoordeling door docenten. Dertien docenten kregen de opdracht: beoordeel deze schrijfprodukten over 'De zin of onzin van huiswerk' al waren het de prestaties van uw leerlingen. Vijf docenten kregen eenzelfde taak voor het 'Vrije opstel'. Elke docent beoordeelde slechts een deel van de stelprodukten (tussen de 74 en de 112 voor de stelprodukten over De zin en onzin van huiswerk en tussen de 207 en 267 voor het Vrije opstel). Hierbij was de beoordelingsprocedure zo vorm gegeven dat elke schrijfprestatie steeds door twee docenten beoordeeld werd (zie: Van den Bergh & Eiting, 1988).

In Tabel 2 worden de beoordelingsprocedures per opdracht per aspect alsmede de gerealiseerde jurybetrouwbaarheden en enkele psychometrische gegevens (jurygemiddelden en

Tabel 2. De gehanteerde beoordelingsmethode, een betrouwbaarheidsschatting (betr), het gemiddelde (gesommeerde) jury-oordeel (gem) en de standaarddeviatie (sd) per opdracht (N = aantal beoordeelde schrijfprodukten).

Opdracht	Beoordeeld aspect	Beoordelingswijze	betr.	gem.	sd	N
Formulier	inhoud	scoring	.94	11.02	6.43	607
Vrij opstel	globale kwaliteit	docentenbeoord.	.67	10.90	2.79	581
Ooggetuigeverslag	inhoud & convent.	scoring	.89	8.42	4.23	614
	globale kwaliteit	schaalbeoord.	.93	170.38	37.08	614
	taalgebruik	schaalbeoord.	.92	179.81	32.09	614
	inhoud & organis.	schaalbeoord.	.93	179.11	34.25	614
	globale kwaliteit	schaalbeoord.	.93	204.07	30.78	619
Sollicitatiebriefje	taalgebruik	schaalbeoord.	.91	186.78	32.25	619
	inhoud & organis.	schaalbeoord.	.95	200.89	27.67	619
	globale kwaliteit	docentenbeoord.	.68	10.60	2.98	586
De zin en onzin van huiswerk	globale kwaliteit	schaalbeoord.	.78	195.02	105.43	586
	taalgebruik	schaalbeoord.	.77	187.66	98.93	586
	inhoud & organis.	schaalbeoord.	.80	195.99	110.27	586

standaarddeviatie van de juryscores) samengevat.

We mogen op grond van de in Tabel 2 weergegeven betrouwbaarheidscoëfficiënten stellen dat alle opdrachten redelijk tot zeer betrouwbaar beoordeeld zijn. Vooral de betrouwbaarheid van de schaaloordelen is bijzonder hoog maar ook de betrouwbaarheid van de docentenoordelen valt niet tegen, zeker als we bedenken dat de docenten een zeer vrije opdracht hadden: 'beoordeel deze schrijfprodukten alsof het de schrijfprodukten uit het eindexamen van uw (MAVO-/LBO-) leerlingen zijn'.

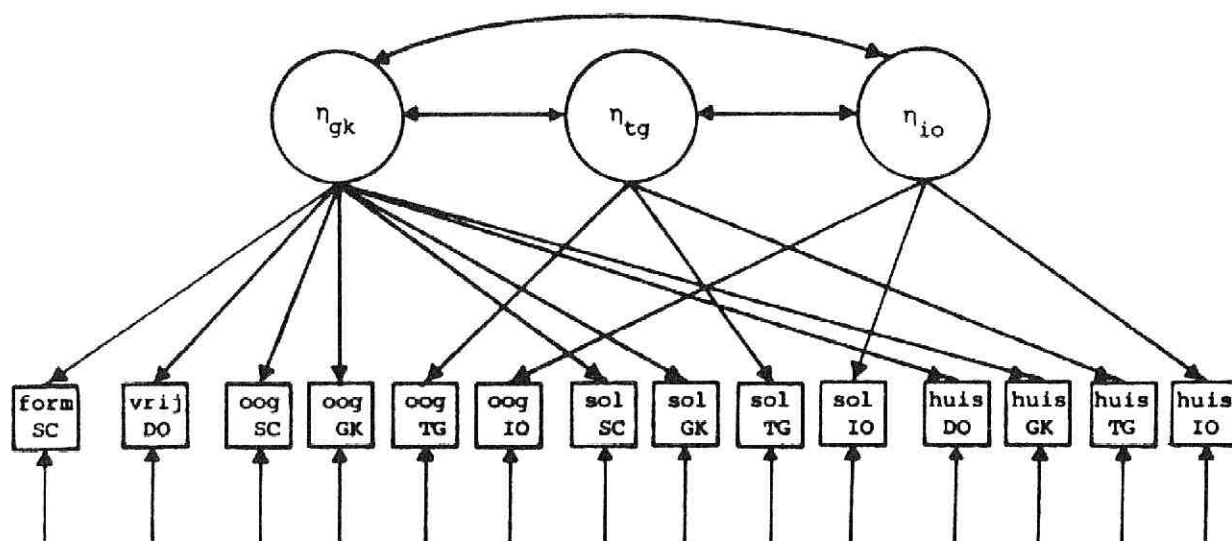
Opmerkelijk zijn de hoge standaardafwijkingen bij de opdracht De zin en onzin van huiswerk. Deze werden veroorzaakt door de minimale prestaties van enkele LBO-leerlingen en de zeer goede prestaties op deze opdracht van een gering aantal MAVO-leerlingen.

3. RESULTATEN

Allereerst zijn de correlaties tussen de juryscores per aspect berekend. Bij bestudering van de in Tabel 3 weergegeven correlatiematrix blijkt dat de correlaties tussen de verschillende aspecten voor één opdracht relatief hoog zijn, en dat de samenhang tussen hetzelfde aspect beoordeeld bij verschillende opdrachten laag is. Zo is de onderlinge samenhang tussen de beoordeelde aspecten van het ooggetuigeverslag hoger dan die met enig andere opdracht. De gemiddelde correlatie tussen aspecten binnen een opdracht bedraagt .60, de gemiddelde correlatie tussen opdrachten slechts .12 (zie Tabel 3).

De consequentie van bovenstaande is, dat een model waarin alleen drie gecorreleerde schrijfvaardigheden (voor globale kwaliteit (η_{gk}), taalgebruik (η_{tg}) en Inhoud & Organisatie (η_{io})) gespecificeerd zijn weinig plausibel lijkt (zie Figuur 1). Toetsing bevestigde dit. Onderstaand model met drie gecorreleerde schrijfvaardigheidsfactoren paste zeer slecht (zie Tabel 4 (model I) voor de toetsingsgrootheden).

Op grond van de hierboven beschreven globale en enigszins intuïtieve analyse van de correlatiematrix bleek al dat de verschillende beoordeelde aspecten binnen één opdracht hoog gecorreleerd zijn. Klaarblijkelijk heeft een aspect van een gegeven opdracht meer gemeenschap-pelijk met de andere aspecten van dezelfde opdracht, dan met enig aspect van een andere opdracht. De oordelen over bij voorbeeld het taalgebruik in het opstel over 'de zin en onzin van huiswerk' weerspiegelen voor een deel taalgebruiksvariantie en voor een deel variantie die inherent lijkt aan de opdracht. In modeltermen: er moeten naast de gecorreleerde schrijfvaardigheidsfactoren ongecorrleerde opdrachtspecifieke factoren gespecificeerd worden. Deze opdrachtspecifieke factor kunnen alleen geschat worden voor de opdrachten waarbij verscheidene aspecten beoordeeld zijn (Ooggetuigeverslag (η_{oog}), Sollicitatiebriefje (η_{sol}) en De zin



Figuur 1. Een grafische weergave van model I, een model met drie gecorreleerde schrijfvaardigheidsfactoren (voor globale klwaiteit, η_{gk} ; voor taalgebruik, η_{tg} ; en voor inhoud & organisatie, η_{io}).

Tabel 3. De correlaties tussen de scores per beoordeeld aspect per opdracht (GK = Globale kwaliteit; TG = Taalgebruik; IO = Inhoud & Organisatie; SC = Gescoord; DO = Holistisch Docenten Oordeel; NB: decimale punten zijn wegelaten).

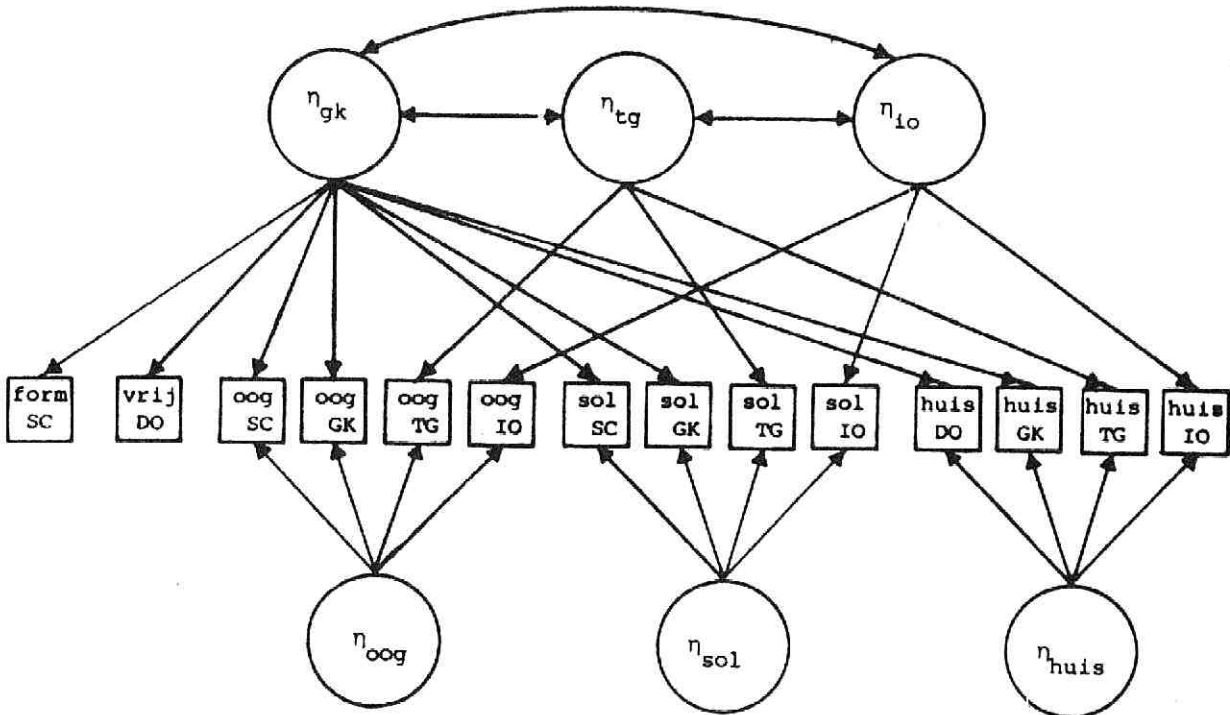
Beoordeeld aspect		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Formulier	: SC	—													
2. Vrije opstel	: DO	04	—												
3. Ooggetuige- verslag	: SC	16	12	—											
4.	: GK	13	35	62	—										
5.	: TG	14	27	48	77	—									
6.	: IO	10	34	58	91	71	—								
7. Sollic. briefje	: SC	-04	04	-06	06	-02	07	—							
8.	: GK	00	09	05	23	18	22	62	—						
9.	: TG	04	05	11	21	27	19	41	69	—					
10.	: IO	-03	08	-06	18	12	18	68*	87	65	—				
11. De zin en on-	: DO	42	03	12	11	16	12	-06	05	12	02	—			
12. zin van huis-	: GK	60	07	09	14	19	13	01	05	14	04	34	—		
13. werk	: TG	57	10	08	09	13	09	-07	-01	08	-05	35	58	—	
14.	: IO	51	08	19	21	22	19	-11	02	07	-03	64	47	45	—

en onzin van huiswerk (η_{huis}). (Zie Figuur 2).

De uitbreiding van het eerste model met drie opdrachtspecifieke factoren bleek aanzienlijk beter bij de gevonden correlatiematrix te passen (zie Tabel 4, model II).

Welke vaardigheden aan het schrijven toegedicht kunnen worden is afhankelijk van hetgeen beoordeeld is. Wanneer schrijfprodukten op een andere wijze beoordeeld zijn, dan kan dit dus van invloed zijn op de samenhang tussen de geobserveerde variabelen. Het voorgaande model kan dan ook verfijnd worden, door de toevoeging van een beoordelingsmethodespecifieke factor. Een verdere verbetering van de passing van het model werd verkregen door de toevoeging van een ongecorrleerde beoordelingsmethodespecifieke factor voor alle oordelen die middels schaalbeoordelingen (η_{sch}) verkregen waren (zie Figuur 3 en Tabel 4 (model III) voor de toetsingsgrootheden).

Tevens bleek dat de samenhang tussen de drie onderscheiden schrijfvaardigheidsfactoren



Figuur 2. Een grafische weergave van model II, een model met drie gecorreleerde schrijfvaardigheidsfactoren (voor globale kwaliteit, η_{gk} ; voor taalgebruik, η_{tg} ; en voor inhoud & organisatie, η_{io}) en drie ongecorrleerde opdracht specifieke factoren (voor het ooggetuigeverslag, η_{oog} ; voor het sollicitatiebriefje, η_{sol} ; en voor het opstel over de zin van huiswerk, η_{huis}).

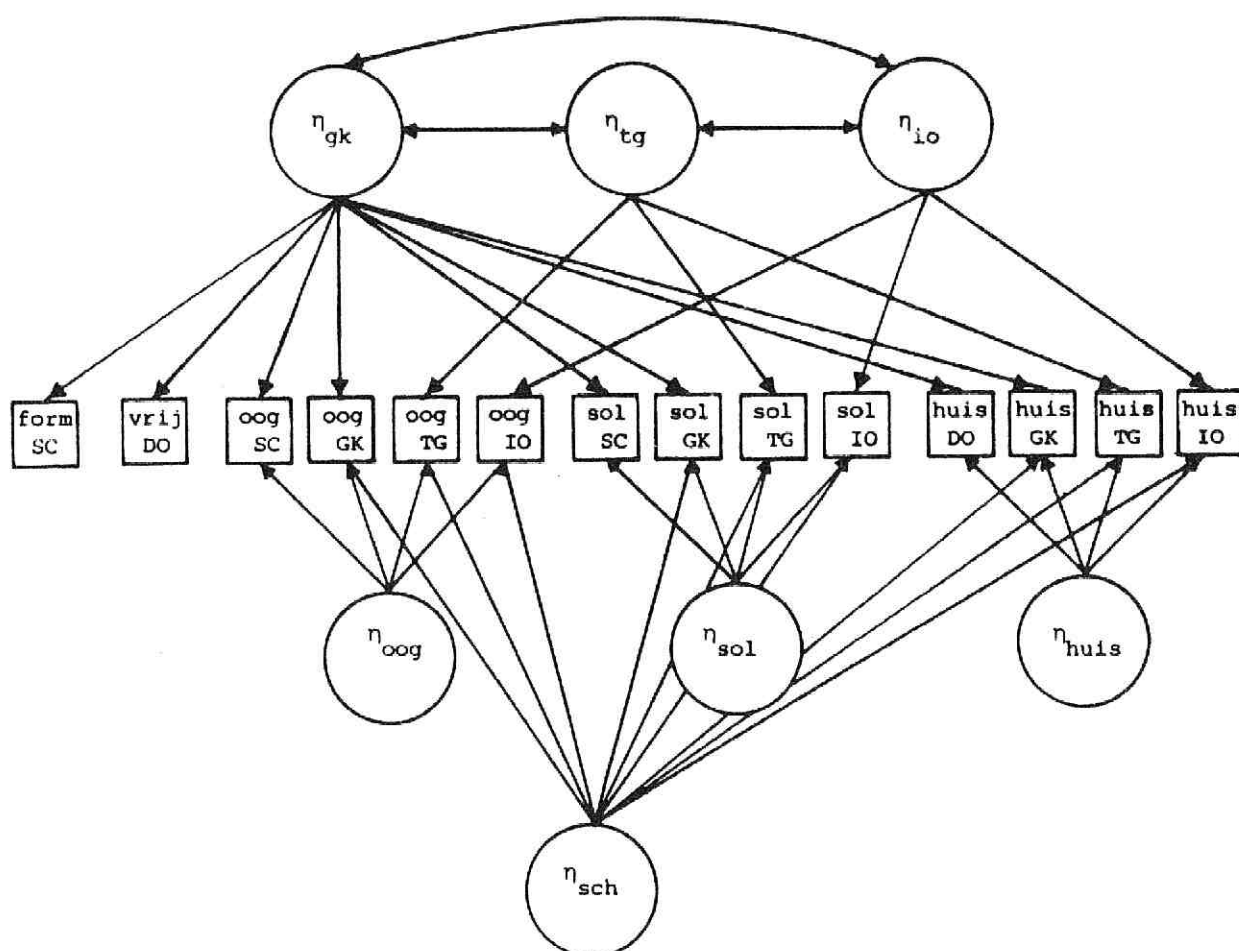
bijzonder hoog was ($\bar{r} = .99$; $\overline{se} = .07$). Het lijkt derhalve weinig zinvol om drie gecorreleerde schrijfvaardigheidsfactoren te onderscheiden; één lijkt voldoende (η_{sv}) (zie Figuur 4 en Tabel 4 (model IV) voor de toetsingsgrootheden).

Gezien de passing van de vier modellen moet de voorkeur gegeven worden aan model III of model IV. Tussen deze twee modellen kan op basis van statistische overwegingen geen onderscheid gemaakt worden (het verschil in passing ($\chi^2 = .35$; $df = 3$) is niet significant).

Op grond van het spaarzaamheidsprincipe, wat inhoudt dat de geobserveerde gegevens zo goed mogelijk beschrijven moeten worden met zo weinig mogelijk variabelen, wordt de voorkeur aan het meer spaarzame model IV gegeven.

Het lijkt misschien alsof geen van de vier modellen bijzonder goed bij de gegevens past, gezien de klassieke toetsingsgrootte chi-kwadraat (met het bijbehorende aantal vrijheidsgraden). Bovendien kan de passing van model IV eventueel nog verbeterd worden, door bij voorbeeld de introductie van een extra relatie tussen de opdrachtspecifieke factor van het Ooggetuigeverslag en het docentenoordeel over de prestaties op het Vrije opstel ($\chi^2 = 122.63$; $df = 54$). Een dergelijke 'overmodellering' door het toestaan van niet of zeer moeilijk interpreteerbare relaties lijkt mij niet zinvol. Het lijkt dan meer op het nastreven van een acceptabele p-waarde, dan op het nastreven van een interpreteerbaar model. Bovendien is de klassieke passingsgrootte chi-kwadraat gevoelig voor de steekproefgrootte; is het aantal waarnemingen groot, dan zorgen kleine verschillen tussen geobserveerde en geschatte correlaties voor een (relatief) grote toename in de chi-kwadraat. Vandaar dat in Tabel 4 ook de beide 'goodness of fit' indexen (gfi

¹ In een dergelijk geval kan de passingsprocedure ook gevalideerd worden door de steekproef aselekt op te splitsen in een aantal substeekproeven. De te trekken conclusies dienen dan te gelden voor alle substeekproeven. In feite is deze procedure ook gevolgd ten behoeve van dit hoofdstuk. De totale steekproef is aselekt opgesplitst in drie substeekproeven. Om redenen van overzichtelijkheid zijn alleen de indices voor de totale steekproef vermeld. Ter illustratie: voor zowel model III als model IV bleek de goodness of fit index bij alle drie de substeekproeven .96.



Figuur 3. Een grafische weergave van model III, een model met drie gecorreleerde schrijfvaardigheidsfactoren (voor globale kwaliteit, η_{gk} ; voor taalgebruik, η_{tg} ; en voor inhoud & organisatie, η_{io}) en drie ongecorreleerde opdrachtspecifieke factoren (voor het ooggetuigeverslag, η_{oog} ; voor het sollicitatiebriefje, η_{sol} ; en voor het opstel over de zin van huiswerk, η_{huis}) en een beoordelingsmethodespecifieke factor, η_{sch} .

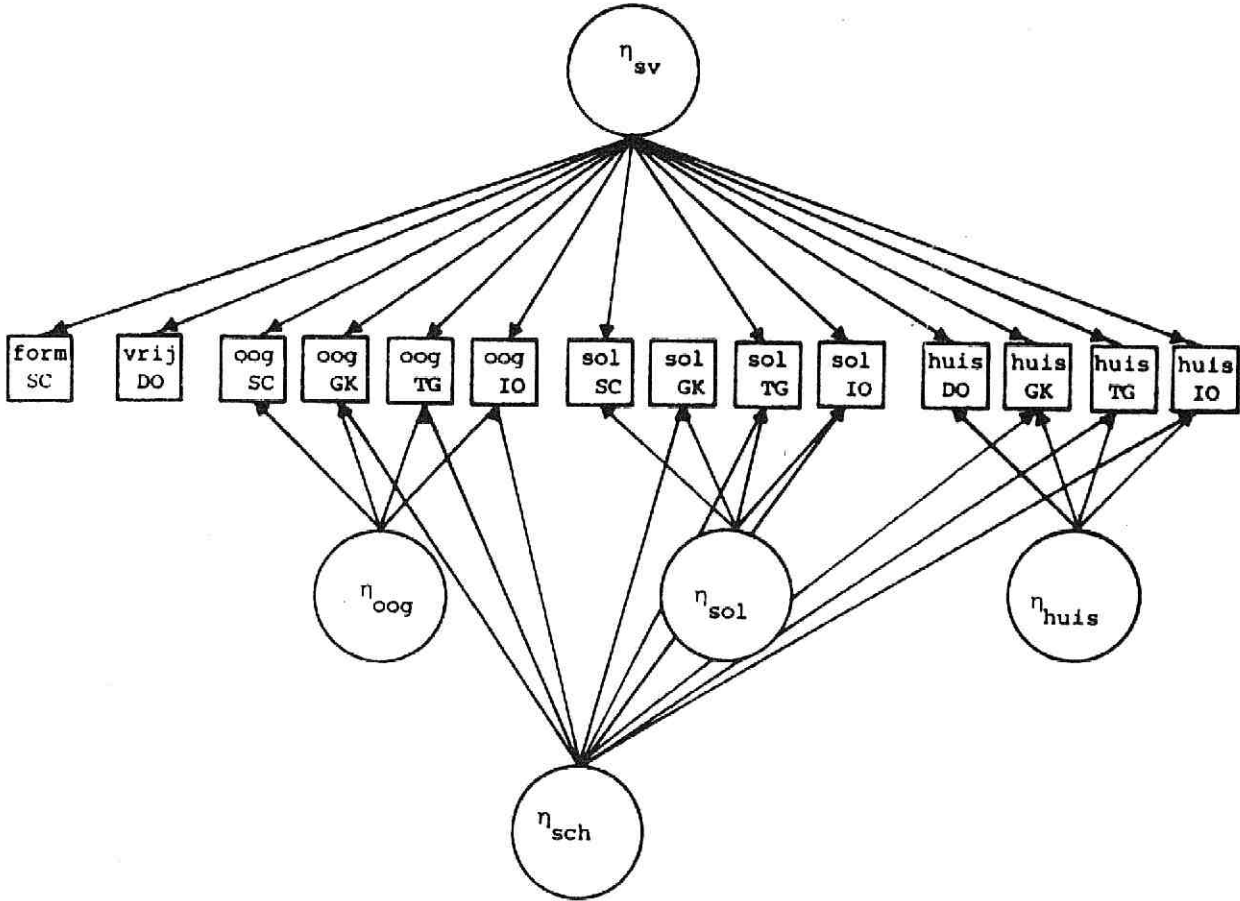
agfi) alsmede de wortel uit de gemiddelde gekwadrateerde residuen som (RMR) weergegeven zijn. Deze drie maten zijn voor model IV alleszins redelijk.

Uit Tabel 4 blijkt ook dat de toevoeging van drie ongecorreleerde opdrachtspecifieke factoren de passing van het model enorm verbetert ($\chi^2 = 2553.16$; $df = 12$). De vraag is dan ook gerechtvaardigd: wat is er eigenlijk gemeten met de opdrachten en de wijze waarop deze beoordeeld zijn? Om deze vraag te kunnen beantwoorden is Tabel 5 geconstrueerd. In deze Tabel zijn de verhoudingen van de onder model IV geschatte gestandaardiseerde varianties per aspect per opdracht weergegeven.

Uit Tabel 5 blijkt dat de schrijfvaardigheidsfactor voor drie van de vijf opdrachten slechts een miniem percentage van de variantie in leeringscores verklaart. De oordelen voor de opdrachten Formulier en De zin en onzin van huiswerk 'dragen' de schrijfvaardigheidsfactor.

Tabel 4. De passing van verschillende modellen betreffende de samenhang tussen vijf schrijfoopdrachten.

Model:	Omschrijving	Passingsmaten:	χ^2	df	p	gfi	agfi	RMR
I	drie schrijf factoren		2834.33	74	>.001	.57	.39	.22
II	I + drie opdracht factoren		281.17	62	>.001	.94	.89	.08
III	II + één beoordelingsmethode factor		203.88	53	>.001	.96	.91	.07
IV	één schrijf factor + drie opdracht factoren + één beoordelingsmethode factor		204.23	56	>.001	.96	.92	.07



Figuur 4. Een grafische weergave van model IV, een model met één schrijfvaardigheidsfactor (η_{sv}), drie ongecorreleerde opdrachtspecifieke factoren (voor het ooggetuigeverslag, η_{oog} ; voor het sollicitatiebriefje, η_{sol} ; en voor het opstel over de zin van huiswerk, η_{huis}) en een beoordelingsmethodespecifieke factor, η_{sch} .

Tabel 5. De verhouding van de varianties in percentages per opdracht per beoordeeld aspect zoals geschat in model IV (SV = Schrijfvaardigheidsvariantie; OV = Opdrachtspecifieke variantie; BV = Schaalbeoordelingsspecifieke variantie; RV = Residulele variantie).

Opdracht	Beoordeeld aspect	Beoordelingswijze	SV	OV	BV	RV
Formulier	inhoud	scoring	59	–	–	41
Vrij opstel	globale kwaliteit	docentenbeoord.	1	–	–	99
Ooggetuigeverslag	inhoud & convent.	scoring	3	38	–	59
	globale kwaliteit	schaalbeoord.	3	89	4	2
	taalgebruik	schaalbeoord.	5	45	14	34
	inhoud & organis.	schaalbeoord.	3	77	3	16
Sollicitatiebrief	inhoud & convent.	scoring	0	52	–	47
	globale kwaliteit	schaalbeoord.	0	73	13	14
	taalgebruik	schaalbeoord.	1	33	29	37
	inhoud & organis.	schaalbeoord.	0	87	4	9
De zin en onzin van huiswerk	globale kwaliteit	docentenbeoord.	31	41	–	28
	globale kwaliteit	schaalbeoord.	69	2	1	38
	taalgebruik	schaalbeoord.	53	1	0	46
	inhoud & organis.	schaalbeoord.	46	17	0	37

De opdrachtspecifieke factoren trekken voor twee van de drie opdrachten veel variantie naar zich toe. Met name bij de beide brieven (Ooggetuigeverslag en Sollicitatiebriefje) is dit frappant. Het is daarom jammer dat voor twee opdrachten (Formulier en Vrij opstel) geen ladingen op een opdrachtspecifieke factor geschat kon worden, omdat deze beide opdrachten slechts op een aspect beoordeeld waren. Toch kan door combinatie van de betrouwbaarheidsgegevens uit Tabel 2 en de residuele variantie uit Tabel 5 een indruk verkregen worden van de maximale grootte van de opdrachtspecificiteit van deze beide schrijfoopdrachten (onder de aanname dat de beoordelingsspecifieke-variantie gelijk aan nul is). Immers de proportie opdrachtspecifieke variantie is maximaal gelijk aan de residuele variantie minus de beoordelingsonbetrouwbaarheid (waarbij de beoordelingsonbetrouwbaarheid gelijk is aan één minus de beoordelingsbetrouwbaarheidsschatting). Een eenvoudige berekening leert dan dat de proportie opdrachtspecifieke variantie voor het Formulier en het Vrije opstel maximaal respectievelijk .35 en .66 bedraagt. De conclusie lijkt derhalve gewettigd dat voor drie van de vijf schrijfoopdrachten een aanzienlijk percentage variantie opdracht specifiek is.

Tot slot blijkt uit Tabel 5 dat voor de schaalbeoordelingsmethode de beoordelingsmethodespecifieke-variantie aanzienlijk kan zijn. Deze ligt in een range van .00 tot .29.

4. DISCUSSIE EN CONCLUSIES

We hebben geconstateerd dat de prestaties van leerlingen op vijf verschillende schrijfoopdrachten redelijk tot zeer betrouwbaar beoordeeld zijn door jury's van twee beoordelaars. De betrouwbaarheid van de analytisch beoordeelde produkten bleek hoger dan de traditionele docentenoordelen. Wel moeten we hierbij aantekenen dat het hier een enigszins ongelijke vergelijking betreft. Immers, de betrouwbaarheid van oordelen over lange opstellen is vergeleken met die van analytische oordelen over korte functionele opdrachten. De gemeenschappelijke variantie van de schrijfpredaties op de verschillende opdrachten bleek echter relatief gering. Voor drie van de vijf opdrachten bleek de met andere opdrachten gemeenschappelijke variantie zelfs miniem. De convergente validiteit van deze schrijfvaardigheidsmetingen is derhalve laag.

We hebben voorts gedemonstreerd dat een beoordelingsmethodespecifieke factor onmisbaar lijkt, voor alle aspecten die met behulp van opstelschalen beoordeeld zijn. De conclusie lijkt gewettigd dat voor de hoge overeenstemming tussen juryleden die met deze methode verkregen wordt, wel een tol betaald moet worden. De interpretatie van deze beoordelingsmethodespecifieke factor, die met name bij de beoordeling van korte opdrachten tot uiting komt, is moeilijk. Misschien worden de beoordelaars (te) sterk gestuurd door de voorbeeldopstellen, misschien komt in de scores de vaardigheid van de beoordelaars in het hanteren van beoordelingsschalen tot uiting. Hoe het ook zij, het lijkt voorbarig om deze factor, die we voor het gemak beoordelingsmethode specifiek genoemd hebben, te interpreteren als een invaliditeitsfactor, zolang we niet precies weten wat deze factor inhoudt.

De belangrijkste onderscheiden variantiebron is echter de opdrachtspecifieke variantie. Met elke schrijfoopdracht wordt iets unieks gemeten, wat niet gecorreleerd is met andere opdrachten. Voor drie van de vijf schrijfoopdrachten trekt deze opdrachtspecifieke factor het leeuwedeel van de variantie in leerlingprestaties naar zich toe. Het lijkt daarom zaak ons af te vragen wat deze opdrachtspecificiteit inhoudt, en hoe we de invloed van een opdrachtspecifieke factor kunnen terugdringen wanneer we de schrijfvaardigheid willen meten.

De opdrachtspecifieke factor kan misschien verklaard worden met behulp van het schrijfprocesmodel van Flower & Hayes (1980). Zij zien schrijven als een informatie verwerkingsproces, waarbij het probleem voor de schrijver is de grote hoeveelheid te verwerken informatie die verwerkt moet worden en de vele zaken waarmee (min of meer gelijktijdig) rekening gehouden moet worden. Zo moet een schrijver onder andere letten op de spelling, de grammaticaliteit, de kwaliteit van de gegenereerde ideeën, de passendheid van het taalgebruik, het beoogde publiek, de implicaties van de voorafgaande tekst etcetera. Niet voor niets wordt het schrijfproces door Flower & Hayes (1980) omschreven als 'juggling with constraints'. Het

gemak waarmee een schrijver aan al deze 'beperkingen' tegemoet kan komen bepaalt (onder meer) de kwaliteit van het schrijfprodukt. Dit kan in alle beoordeelde aspecten tot uiting komen. De opdrachtspecifieke variantie zou dan geïnterpreteerd kunnen worden als een moeilijkheidsfactor, die de variantie weerspiegelt in het omgaan met een overbelasting betreffende zowel informatie als eisen waaraan de te produceren tekst moet voldoen.

Sommigen menen dat de overbelasting van het werkgeheugen voornamelijk veroorzaakt wordt door het genereren van informatie (Bereiter & Scardamalia, 1982). Gesuggereerd wordt dat schrijfvaardigheidsoordelen naast schrijfvaardigheid ook kennis van het onderwerp of het bezitten van adequate schemata over de opdrachtsituatie of het teksttype gemeten weerspiegelen, of in elk geval het gemak waarmee deze gegenereerd kunnen worden (vergelijk: Boodoo & Garlinghouse, 1983; Van den Bergh, de Glopper & Schoonen, 1987). Het lijkt plausibel te veronderstellen dat een meer valide meting van de schrijfvaardigheid verkregen wordt wanneer de invloed van verschillen in achtergrondkennis over de opdracht c.q. de opdrachtsituatie teruggedrongen worden. Dan kunnen verschillen in schrijfvaardigheid aan bod komen in plaats van verschillen in aanwezige achtergrondkennis. Voor deze redenering vinden we aanwijzingen bij de opdrachten *Formulier* en *De zin en onzin van huiswerk*. Bij deze beide opdrachten bleek de proportie opdrachtspecifieke variantie gering. Dat alle leerlingen vertrouwd zijn met het verschijnsel huiswerk is zeer aannemelijk en bovendien kregen alle leerlingen documentatie bij de opdracht uitgereikt, waardoor eventuele verschillen in achtergrondkennis verminderd kunnen zijn. Hetzelfde gaat waarschijnlijk op voor het *Formulier*; elke leerling heeft weleens een formulier ingevuld, en de gegevens die bij deze opdracht ingevuld moesten worden konden uit de bij de opdracht behorende (fictieve) correspondentie gehaald worden. Bij deze opdrachten lijkt de variatie in achtergrondkennis tussen de leerlingen geringer dan bij de opdrachten waar de proportie opdrachtspecifieke variantie groot was. Experimenteel onderzoek naar deze hypothese is nodig.

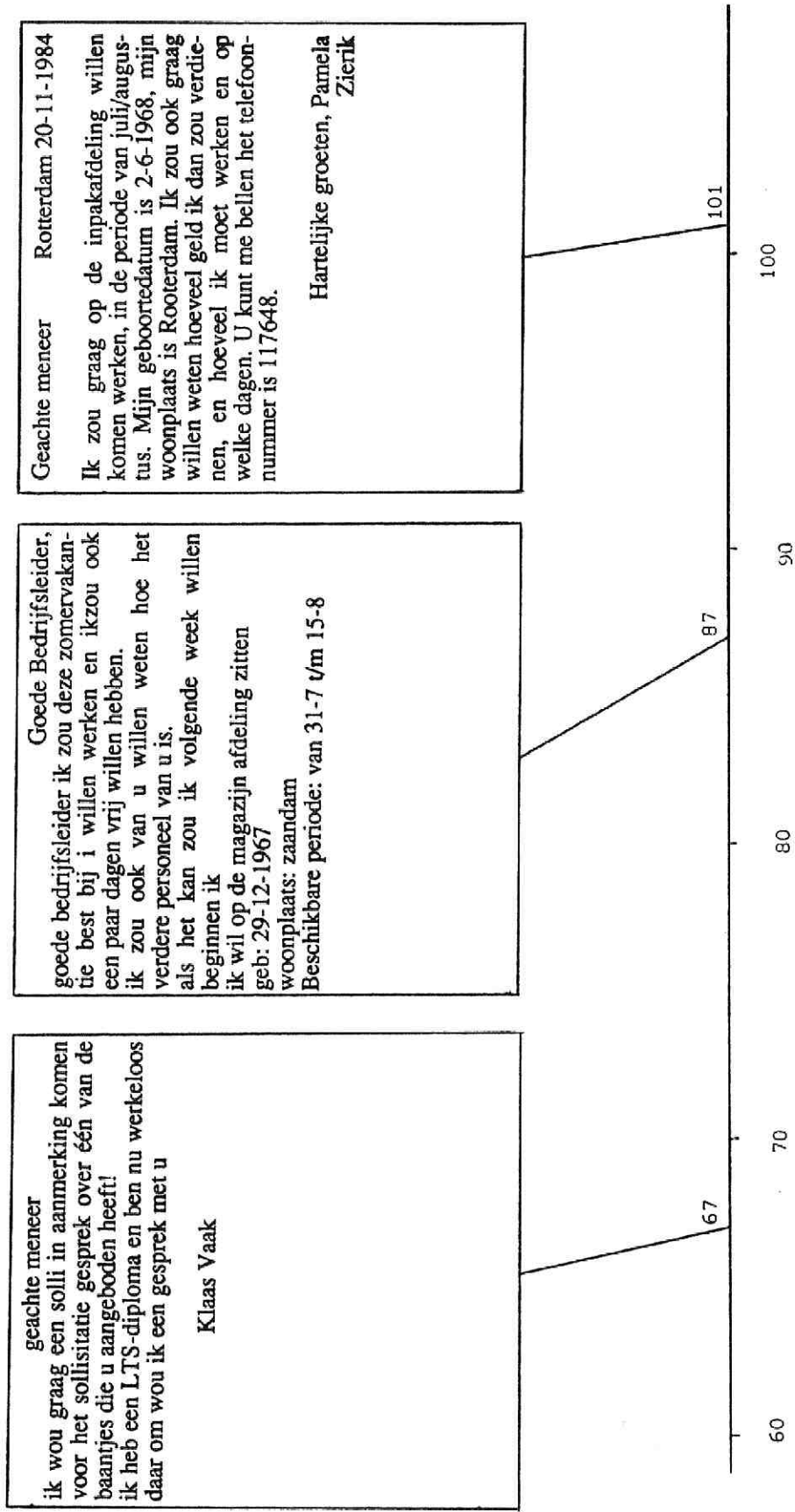
Gegeven de resultaten van dit onderzoek lijkt het gewettigd te veronderstellen dat tekstlengte niet de enige bepalende factor is, zoals sommigen gesuggereerd hebben. Ook met een korte schrijfopdracht kan een (relatief) adequate meting van de schrijfvaardigheid verkregen worden.

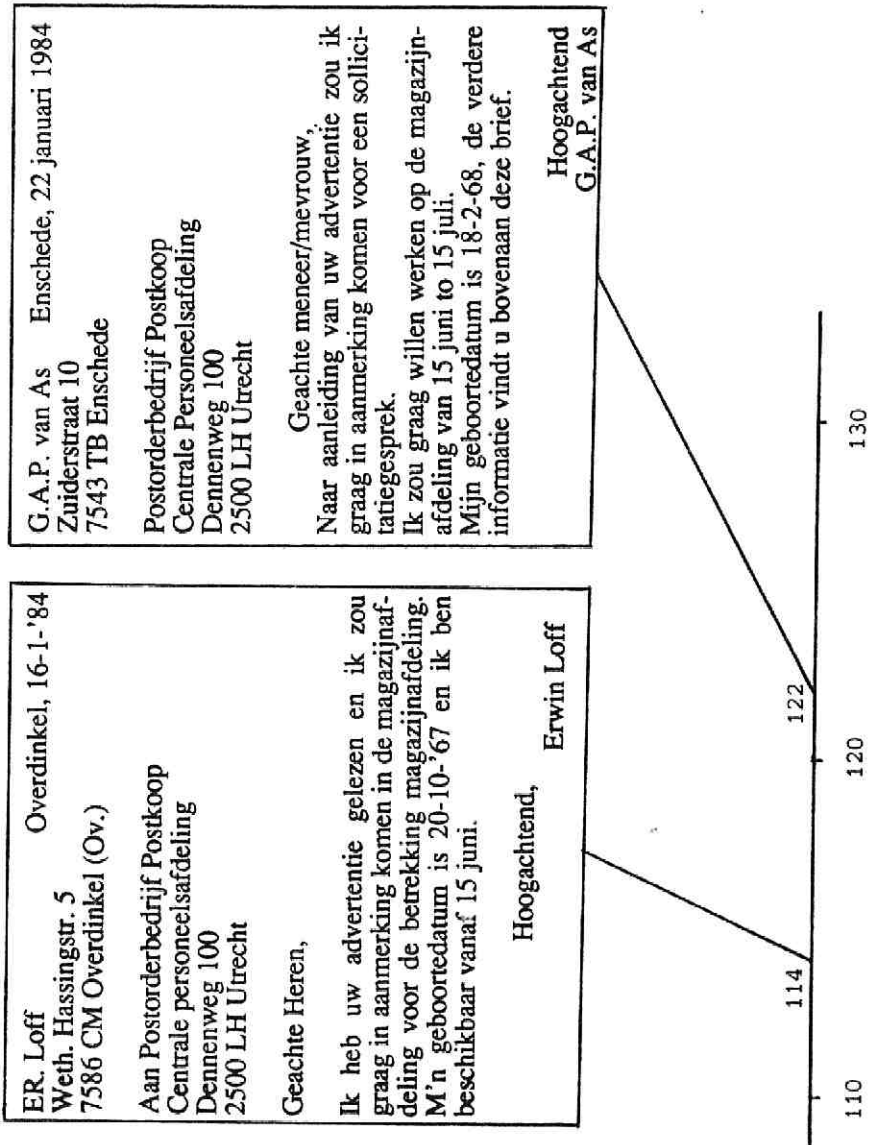
Ook moet opgemerkt worden dat de proportie residuele of niet verklaarde variantie een behoorlijke omvang kan aannemen. De reden voor de hoge proportie foutenvariantie moet in tweetal moeilijk onderscheidbare oorzaken gezocht worden. Allereerst is de proportie foutenvariantie rechtstreeks afhankelijk van de juryonbetrouwbaarheid, omdat de maximaal te verklaren proportie variantie in schrijfprestaties natuurlijk gelijk is aan één minus de juryonbetrouwbaarheid. Een tweede oorzaak is gelegen in het één item probleem; een schrijfvaardigheidsprestatie is, wanneer er een globaal oordeel over het schrijfprodukt gegeven wordt, in principe een prestatie op een 'één item toets'. We kunnen nu eenmaal geen items maken die een vaardigheid perfect meten. Laat staan dat een complexe vaardigheid als schrijfvaardigheid met één item betrouwbaar en valide gemeten kan worden. Bovendien fluctueren de leerlingprestaties van item tot item. Beide aspecten (beoordelingsonbetrouwbaarheid en het 'één item probleem') zijn rechtstreeks van invloed op de hoogte van de residuele variantie.

Nemen we de resultaten van dit onderzoek als maatgevend, dan kan onder de aanname van homogene toetsverlenging het aantal schrijfopdrachten berekend worden dat nodig is voor een betrouwbare meting van de schrijfvaardigheid. Als we hierbij uitgaan van een betrouwbaarheidscoefficiënt van .80, dan moeten er minimaal veertien schrijfopdrachten aan een leerling afgenomen worden voor een betrouwbare meting van zijn schrijfvaardigheid. De afsluitende conclusie lijkt gewettigd dat met niet alle schrijfopdrachten een goede meting van de schrijfvaardigheid verkregen wordt. Bij een aantal opdrachten zal de variantie in leerlingprestaties slechts voor een gering deel de variantie in schrijfvaardigheid weerspiegelen, en zal de opdrachtspecifieke variantie een groot deel van de variantie in leerlingprestaties verklaren. Waarschijnlijk is het verstandig om bij één meting van de schrijfvaardigheid van leerlingen de variantie in achtergrondkennis van de leerlingen zo veel mogelijk te minimaliseren. Hiervoor lijken op het eerste gezicht twee methoden in aanmerking te komen: het verschaffen van achtergrondinformatie over het onderwerp en/of het kiezen van een onderwerp dat voor alle leerlingen even bekend of vertrouwd is.

BIJLAGE: EEN VOORBEELD VAN EEN OPSTELSCHAAL

Schaal: Inhoud en Organisatie sollicitatiebriefje





GERAADPLEEGDE LITERATUUR

- Bereiter, C., & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In: R. Glaser (Ed.), *Advances in instructional Psychology* (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Assoc.
- Bergh, H. van den (1987). *Oude en nieuwe examens vergeleken: Instrumentconstructie*. (Interimrapport SVO: 4075). Amsterdam: SCO.
- Bergh, H. van den (1988). Onderzoek naar functionele doelstellingen voor het moedertaalonderwijs. *Pedagogische Studiën*, 65, 139-148.
- Bergh, H. van den, Baltzer, J.E., Marshall, F., Triesscheijn, B. & Wesdorp, H. (1986). *Instrumenten ter peiling van de taalvaardigheid van leerlingen in de derde klas van het v.o.* Amsterdam: SCO.
- Bergh, H. van den, & Eiting, M.M. (1988). A method of estimating rater reliability. *Journal of Educational Measurement* (In press).
- Bergh, H. van den, & Rijlaarsdam, G. (1987). Opstelbeoordeling? Een recept. *Levende Talen*, 413, 448-454.
- Blok, H. (1984a). Trivialiteit van jury's bij globale opstelbeoordeling. *Tijdschrift v. Taalbeheersing*, 6, 1-11.
- Blok, H. (1984b). Trivialiteit van jury's: antwoord aan Meuffels. *Tijdschrift voor Taalbeheersing*, 6, 201-205.
- Blok, H. (1985). Estimating the reliability, validity and invalidity of essay ratings. *Journal of Educational Measurement*, 22, 41-45.
- Blok, H. (1987). *Taal voor alledag: Feiten en meningen over het taalgebruik van LBO- en MAVO-leerlingen in alledaagse situaties*. 's-Gravenhage: SVO.
- Blok, H., & Hoeksma, J.B. (1984). *Opstellen geschaald: De constructie van beoordelingsschalen voor vijf schrijfopdrachten*. Amsterdam: SCO.
- Boodoo, G.M., & Garlinghouse, P. (1983). Use of essay examination to investigate the writing skills of undergraduate education majors. *Educational and Psychological Measurement*, 43, 1005-1014.
- Coffman, W.E. (1966). On the validity of essay tests of achievement. *Journal of Educ. Measurement*, 3, 151-156.
- Farell, F.J. (1971). *Implications of national writing assessment*. Speech given at the annual meeting of the American Publishers Inc., Washington, D.C.
- Flower, L.S., & Hayes, J.R. (1980). The dynamics of composing: Making plans and juggling constraints. In: L.W. Gregg & E.R. Sternberg (Eds.), *Cognitive processes in writing*. Hillsdale, N.J.: L. Erlbaum.
- Glopper, K. de (1986a). *Schrijfpredstaties in het voortgezet onderwijs: Een onderzoek naar de stelvaardigheid van leerlingen uit het derde leerjaar LTO, LHNO, MAVO, HAVO en VWO*. 's-Gravenhage: SVO.
- Glopper, K. de (1986b). *Meettechnische aspecten van het onderzoek naar de stelvaardigheid van leerlingen in het derde leerjaar LTO, LHNO, MAVO, HAVO en VWO*. Amsterdam: SCO.
- Hert, W. de & Rijmenans, R. (1987). Wie hoort er thuis in het verbeteringsgesticht? *Vonk*, 17, 258-268.
- Hogan, T.P., & Mishler, C. (1980). *Relationships among measures of writing skill*. Education commission of the states. Denver, C. National Assessment of Education Research.
- McColly, W. (1970). What does educational research say about the judging of writing ability? *Educational Research*, 64, 148-156.
- Meuffels, B. (1984). Jury's triviaal? *Tijdschrift voor Taalbeheersing*, 6, 201-205.
- Meuffels, B. (1985). Het beoordelen van opstellen: dimensionaal of typologisch? In: W.K.B. Koning (red.), *Taalbeheersing in theorie en praktijk*. Dordrecht: Floris Publications.
- Moss, P.A., Cole, N.S., & Khampalikit, L. (1982). A comparison of procedures to assess written language skills at grade 4, 7 and 10. *Journal of Educational Measurement*, 19, 37-47.
- Quellmalz, E.S. & Capell, F.J. (1979). *Defining writing effects of discourse and response mode*. California University L.A.: Center for Study of Education.
- Quellmalz, E.S., Capell, F.J., & Chou, C. (1982). Effects of discourse and response mode on the measurement of writing competence. *Journal of Educational Measurement*, 19, 241-258.
- Rijlaarsdam, G.J.W. (1986). *Effecten van leerlingrespons op aspecten van taalvaardigheid*. Amsterdam: SCO.
- Wesdorp, H. (1974). *Het meten van de produktief-schriftelijke taalvaardigheid*. Purmerend: Muusses.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. 's-Gravenhage: SVO.
- Wesdorp, H. (1983). Gelijke kansen in centrale examens. In: W.C. Weeda (red.), *Examens in discussie: Een bundel opstellen voor J.W. Solberg*. Groningen: Wolters Noordhoff.
- Wesdorp, H., Bergh, H. van den, Bos, D.J., Hoeksma, J.B., Oostdam, R.J., Scheerens, J., & Triesscheijn, B. (1986). *De haalbaarheid van periodiek peilingsonderzoek*. Lisse: Swets & Zeitlinger.
- Wolf, T., & Blok, H. (1985). *Tienertoer en uitzendwerk: Twee toetsprogramma's voor functionele taalvaardigheid in LBO en MAVO*. Amsterdam: SCO.

Manuscript ontvangen 8-4-1988

Definitieve versie ontvangen 29-7-1988.

Het spellen in het Papiaments door Curaçaose schoolkinderen

P. Vedder en H. Kook

ABSTRACT

In the schoolyear 86/87 Papiamentu was introduced as a subject in the curriculum of Curaçaoan primary schools. At the start of the schoolyear 87/88 fourth and fifth graders' spelling achievements in Papiamentu were measured in eight schools. The results were not satisfactory. A regression analyses showed that children's spelling skills in Dutch language, their home language and the school are important variables in explaining achievement differences in Papiamentu spelling.

1. PAPIAMENTS: EEN NIEUW VAK

Vanaf augustus 1986 is Papiaments, de moedertaal van 90 tot 95% van alle Curaçaose schoolkinderen, opgenomen in het curriculum van de zesjarige Curaçaose basisscholen (qua structuur te vergelijken met de aloude Nederlandse lagere school). Daarmee is een eind gekomen aan een periode van ongeveer 60 jaar waarin het Papiaments een verboden taal was op de basisschool. Het Papiaments wordt volgens voorschrift een half uur per schooldag in alle klassen onderwezen. De rest van de schooldag is de voertaal Nederlands.

In de discussie over de invoering van het vak spelen de verschillen in onderwijskansen tussen van huis uit Papiamentssprekende en van huis uit Nederlandssprekende kinderen een belangrijke rol. Het onderwijsrendement bij Papiamentssprekende kinderen is in vergelijking tot het rendement bij Nederlandssprekende kinderen laag. Papiamentssprekende kinderen doen gemiddeld bijna een jaar langer over de lagere school dan Nederlandstalige kinderen en terwijl in 1981 66% van de Nederlandstalige zesde klassers overging naar een HAVO/VWO-brugklas ging van de Papiamentssprekende kinderen slechts 6,1% over naar een HAVO/VWO-brugklas (zie De Jonge, 1986a). Prins-Winkel (1973) toonde aan dat de onderwijskansen verschillen ook als er wordt gecontroleerd voor verschillen in intelligentie en sociale achtergrond. De verwachting is dat de schoolprestaties van Papiamentstalige kinderen zullen verbeteren als het Papiaments wordt ingevoerd als voertaal in het onderwijs. De invoering van het Papiaments als vak wordt gezien als stap op weg naar de invoering van het Papiaments als voertaal.

Er bestond bij de invoering van het vak twijfel over de effecten. Sommige voorstanders verkondigden op de t.v. en in kranten grootse verwachtingen. De kinderen zouden zich meer gaan thuisvoelen op school en hun prestaties bij andere vakken zouden verbeteren. Anderen hadden twijfels en gaven uiting aan hun vrees dat indien de invoering van het vak geevalueerd zou worden de resultaten bedreigend zouden kunnen zijn voor het vak Papiaments (zie Evertsz & Faries, 1986).

In dit artikel wordt verslag gedaan van een evaluatiestudie. De probleemstelling en de omvang van de studie zijn beperkt. In de studie wordt nagegaan welk effect het programma Papiaments heeft op de spellingsvaardigheid bij leerlingen van de derde en vierde klas na één jaar spellingsonderwijs in het Papiaments

De beperking tot het derde en vierde leerjaar hangt samen met het feit dat in deze leerjaren tijdens de Papiamentse lessen vooral aandacht wordt besteed aan het leren schrijven in het

Papiaments. In klas een en twee gaat de aandacht uit naar mondeling taalgebruik en in klas vijf en zes staat het begrijpen van Papiamentstalige teksten centraal.

De leergang spelling van het Papiaments voor het basisonderwijs

De methode voor het derde en vierde leerjaar heeft als planningseenheid een week. Een leergang voor een leerjaar omvat 40 weken. Per week is voorzien dat er vijf lessen van elk 30 minuten worden gegeven. De opzet van een weekprogramma is globaal steeds dezelfde. Er is aandacht voor het luisteren naar en het spreken van het Papiaments, maar de meeste aandacht in het programma wordt besteed aan het leren spellen van het Papiaments en het lezen van Papiamentse teksten. Dit betreft volgens de tijdsplanning ongeveer 75% van de tijd. Bij het leren spellen en lezen staat het correct spellen en lezen van de Papiamentse klank-letterverbindingen centraal. Er wordt gefocused op klank-letterverbindingen en niet op bijvoorbeeld woorden, aangezien de auteurs er van uitgaan dat het Papiaments een fonologische spelling heeft. Als leerlingen alle Papiamentse klank-letterverbindingen kennen dan kunnen ze woorden zo schrijven als ze ze horen. In Nederland is deze aanpak bekend als de analytisch-synthetische methodiek.

De gehanteerde methodiek wordt desalniettemin door de auteurs van de leerboekjes omschreven als een "globaal methode". Vrijwel elk weekprogramma begint met het voorlezen van een korte tekst door de leerkracht. De tekst kan in de boekjes worden meegelezen door de leerlingen, zodat zij zelf de klank-letterverbinding van het Papiaments kunnen ontdekken. Per week wordt vervolgens systematisch aandacht besteed aan één of soms twee klank-letterverbinding(en). Het gaat hierbij om klinkers, medeklinkers, tweeklanken en combinaties van klinkers. Verdeeld over het schooljaar worden 52 klank-letterverbindingen behandeld.

De klank-letterverbinding die in een week centraal staat wordt goefend bij het maken van korte invuloefeningen, het opschrijven van namen van voorwerpen en korte leesteksten. Aan het eind van een week wordt meestal een woorddictee gegeven waarin de behandelde klank-letterverbinding centraal staat. Voor enkele voorbeelden van klank-letterverbindingen zij verwezen naar tabel 5.

2. DE METHODE

De proefpersonen

Het onderzoek is uitgevoerd in 12 vierde en in 14 vijfde klassen van acht Curaçaose basisscholen. Aangezien er beperkte middelen beschikbaar waren voor de scoring en de verdere gegevensverwerking zijn voor elke klas de gegevens van tien a-select gekozen leerlingen gebruikt. De gegevens van in totaal 120 vierde en 140 vijfde klassers werden geanalyseerd.

Het onderzoek is uitgevoerd bij vierde en vijfde klas kinderen na één jaar spellingsonderwijs. Er is gewerkt met een "post test only design". De treatment is niet random toegedeeld aan klassen. Zes van de acht deelnemende scholen zijn gekozen door Curaçao drie districten te verdelen en uit elk district willekeurig twee scholen te kiezen (hierna school 1-6). De overige twee scholen zijn beide scholen waarvan relatief veel leerlingen thuis geen Papiaments spreken (hierna school 7-8) maar Nederlands of een andere taal. Meer informatie over de deelnemende leerlingen wordt gepresenteerd in tabel 1.

In de tabel word zichtbaar dat er tussen de scholen 1-6 en de scholen 7-8 een aanzienlijk verschil bestaat wat betreft de overheersende taalachtergrond van de leerlingen. Op de scholen 1-6 is deze haast volledig Papiamentstalig en op de scholen 7-8 is de meest omvangrijke taalgroep de Nederlandstalige. De restcategorie betreft kinderen die thuis Spaans of Engels spreken of opgroeien in een meertalige thuissituatie.

Het leeftijdsverschil tussen de kinderen van school 1-6 en van school 7-8 is opvallend. Op school 1-6 is vaker sprake van zittenblijven.

Tabel 1: Steekproef van leerlingen van acht Curaçaose basisscholen

leerjaar	school	n	% Pap	% Ned	rest	leeftijd		% meisjes
						$\bar{x}$	s.d.	
4	1 - 6	80	95	—	5	9,8	1,0	45
	7 - 8	40	40	42,5	17,5	9,3	0,5	45
5	1 - 6	100	95	—	5	11,3	1,3	43
	7 - 8	40	32,5	45	22,5	10,2	0,7	47

De onderzoeksinstrumenten

Als belangrijkste instrumenten, de instrumenten voor het meten van de effectvariabelen, zijn woorddictees gebruikt. Voor het meten van controle variabelen is gebruik gemaakt van intelligentietests voor de vierde en vijfde klas. Verder is de leerlingen bij het maken van de dictees gevraagd aan te geven welke taal ze thuis spreken.

Bij het samenstellen van de dictees is uitgegaan van het Papiamentse spellingsprogramma. De beheersing van alle klank-letterverbindingen wordt gemeten. Er is gebruik gemaakt van woorddictees. De meeste van de gebruikte woorden zijn één- of tweelettergrepige woorden en het betreft meestal woorden die bij kinderen bekend mogen worden geacht. Elk woord is steeds eerst aangeboden in een zin. Waarna het betreffende woord twee keer werd gezegd. Twee voorbeelden uit het Papiamentse en uit het Nederlands dictee kunnen dit verduidelijken.

Papiaments

1. Kon a bai na skol awe? (Hoe ging het vandaag op school?)
Skirbi 'kon' ... 'kon'. (Schrijf 'hoe' ... 'hoe'.)
2. Straño, e mucha no gusta mangel. (Vreemd dit kind lust geen snoepjes.)
Skirbi 'mangel' ... 'mangel'. (Schrijf 'snoepje' ... 'snoepje')

Nederlands

1. De kinderen zingen in koor. Schrijf 'koor' ... 'koor'.
2. Die plant is giftig. Schrijf 'giftig' ... 'giftig'.

De betrouwbaarheid van de gebruikte toetsen is geschat m.b.v. Cronbach's alpha. De betreffende alpha-waarden samen met andere informatie over de instrumenten worden gepresenteerd in tabel 2.

Tabel 2: Kenmerken van de gebruikte instrumenten

instrument	bron / gebruik gemaakt van*	aantal items	alpha
Papiaments dictee	lijst gestandaardiseerde woorden (ILA, 1986)	122	.96/.96**
Nederlands dictee	frequentie woordenlijst (Van der Geest & Swüste, 1978) en spellingtoetsen leerjaar 1 en 2 (CITO, 1980)	113	.95/.96**
Intelligentietest			
onderdelen:	PMA 2-4 (Kema, 1980) (vertaling:		
woordenschat	Schooladviesdienst Curaçao)		
- Papiamentse versie	PMA 5-6 (Kema, 1976)	60	.77
- Nederlandse versie		60/30**	.66/.87**
ruimtelijk inzicht		27/26**	.78/.81**
figuur sorteren		27	.62/.73**
waarneming		50	.91/.92**

* Alle instrumenten zijn uitvoerig beschreven in Kook en Vedder 1988.

** De eerste waarde heeft betrekking op leerjaar vier en de tweede op leerjaar vijf.

Zowel het Papiamentse als het Nederlandse dictee blijken goed betrouwbaar te zijn.

In de tabel is verder te zien dat ook de meeste onderdelen van de intelligentietests redelijk tot goed betrouwbaar zijn. De PMA 2-4 is gebruikt in klas 4 en de PMA 5-6 in klas 5. Voor leerjaar 4 bestaat van het onderdeel woordenschat een Nederlandse en een Papiamentse versie. De betrouwbaarheid van de Papiamentse versie is bevredigend, maar die van de Nederlandse versie is enigszins teleurstellend. Ook de betrouwbaarheid van het onderdeel figuur sorteren voor klas 4 is aan de lage kant.

De onderzoeksprocedure

Het onderzoek is uitgevoerd in de periode vanaf de eerste week van oktober, zes weken na aanvang van het schooljaar 87/88, tot aan de tweede week van december. De dictees zijn allemaal gemaakt in de eerste drie weken van oktober. Het Papiamentse en het Nederlandse dictee zijn elk opgesplitst in vier subdictees. Leerlingen maakten per afnamedag twee subdictees. Elk subdictee nam 10 tot 15 minuten in beslag.

Vijf Papiamentstalige onderzoeksassistenten hebben geholpen bij het afnemen van de Papiamentse dictees. Op vijf scholen zijn de Nederlandse dictees eveneens afgenomen door de onderzoekssassistenten. Op de andere drie scholen hebben de schoolleiders de Nederlandse dictees afgenomen.

De PMA-tests zijn afgenomen door drie ervaren medewerkers. In overleg met hun leerkrachten zijn de leerlingen van school 1-6 geïnstrueerd in het Papiaments. De vierde klas kinderen van deze scholen maakten het onderdeel woordenschat in het Papiaments. De leerlingen van school 7-8 zijn geïnstrueerd in het Nederlands. De vierde klas leerlingen van deze scholen maakten het onderdeel woordenschat in het Nederlands.

De scoringswijze

Zowel voor wat betreft het Papiamentse als wat betreft het Nederlandse dictee worden de resultaten gepresenteerd in termen van goed of fout geschreven klank-letterkoppelingen.

Wijzelf hebben de norm geformuleerd die wordt gebruikt om te beoordelen of de spellingsprestaties van de leerlingen al of niet bevredigend zijn. Bij de beoordelingen richten we ons niet op een compleet woord, maar op één specifieke klank-letterverbinding in een woord. We stellen dat de kinderen minimaal 80% van de gepresenteerde klank-letterverbindingen goed mochten beheersen. De 80% norm is een gangbare maat bij strategieën voor beheersingsleren. Een meer strenge norm, een score van 95% correct, is ook verdedigbaar. Immers, het kennen van de regels voor de klank-letterkoppeling in een taal is zeer fundamenteel voor schriftelijke communicatie in die taal.

Bij alle onderdelen van de intelligentietest krijgen de leerlingen een punt per goed antwoord. De scores zijn verder niet omgerekend. Voor de correctiemodellen zij verwezen naar de handleidingen (Kema, 1976, 1980).

3. DE LEEREFFECTEN

De 120 vierde klassers schreven gemiddeld 70,1 klank-letterverbindingen goed (s.d. 21,3). Dit betreft 57,5% van de aangeboden woorden. De conclusie is dat de bij beheersingsleren veel gehanteerde norm van 80% beheersing van basisvaardigheden niet wordt gehaald. De beheersing van de klank-letterverbinding in het Papiaments is in de vierde klas slecht.

In klas 5 is de beheersing iets beter. De gemiddelde goed score is 77,4 (s.d. 20,6) dit betreft 63,4% van de woorden. Ook deze mate van beheersing is verre van de gewenste.

Vervolgens is gekeken of de prestaties wellicht verschillen voor de onderscheiden scholenclusters en voor de twee belangrijkste taalgroepen (Papiamentstalige en Nederlandstalige leerlingen). De betreffende informatie wordt gepresenteerd in tabel 3 en in tabel 4.

Tabel 3: Het verschil in spellingsprestatie tussen school 1-6 en school 7-8

leerjaar	school	n	$\bar{x}$	s.d	t	d.f.	p
4	1 - 6	80	70,2	22,0	,00	118	,99
	7 - 8	40	70,2	20,1			
5	1 - 6	100	77,7	19,7	,21	138	,83
	7 - 8	40	76,9	23,0			

Tabel 4: Gemiddelde spellingsprestaties van kinderen met een Papiamentse taalachtergrond (P) en kinderen met een Nederlandse taalachtergrond (N)

leerjaar	taalgroep	n	$\bar{x}$	s.d	t	d.f.	p
4	Papiaments	92	72,1	22,0	3,0	107	,00
	Nederlands	17	55,9	10,8			
5	Papiaments	108	78,8	19,9	2,2	124	,03
	Nederlands	18	67,3	23,6			

Zoals in tabel 3 en tabel 4 is te lezen is er geen sprake van prestatieverschillen tussen de scholen 1-6 en de scholen 7-8, maar wel tussen de taalgroepen.

Alle deelnemende Nederlandstalige kinderen zitten op de scholen 7-8. In tabel 4 wordt duidelijk dat deze kinderen relatief slecht presteren op het Papiamentse dictee. Een en ander houdt in dat de Papiamentstalige kinderen op deze scholen beter presteren dan de Papiamentstalige kinderen op de andere scholen. De Papiamentstalige vierde klassers van de scholen 7-8 hebben een gemiddelde score van 84,0 (s.d. 10,4) en de vijfde klassers 87,6 (s.d. 18,7).

De Nederlandstalige kinderen hebben over het algemeen iets meer problemen met de Papiamentse spelling, maar vooral met die klank-letterverbindingen die verschillen voor het Nederlands en het Papiaments. Dit betreft ondermeer het gebruik van accenten in het Papiaments waar dat niet hoort of het weglaten van accenten of trema's waar ze juist wel horen in het Papiaments. Over deze problemen wordt informatie gegeven in tabel 5. De beheersing van een bepaalde klank-letterverbinding is in de betreffende gevallen steeds gemeten met drie woorden. De getallen betreffen gemiddelde percentages fouten over de drie woorden. De informatie is gespecificeerd voor leerjaar 4 en voor leerjaar 5 en voor de groep Papiamentstalige en de groep Nederlandstalige leerlingen.

Voor de klanken die in het Papiaments worden geschreven met een accent of een trema leveren veel problemen op. In de tabel wordt verder duidelijk dat de verschillen tussen de taalgroepen vooral groot zijn in de vierde klas. Dat de verschillen in de vijfde klas kleiner zijn is vooral te danken aan de, t.o.v. de vierde klas, relatief betere prestaties van de Nederlandstalige leerlingen.

Het zoeken naar een verklaring voor de verschillen in spellingsprestaties

Met behulp van een stapsgewijze regressie-analyse is nagegaan in welke mate de volgende variabelen de verschillen tussen de scores van de kinderen op het Papiamentse dictee verklaren: de scores op het Nederlandse dictee, de woordenschat, de redkundige intelligentie (de som van de scores voor de onderdelen figuur sorteren, ruimtelijk inzicht en waarneming) de thuistaal en school. De variabele thuistaal is gedichotomiseerd door slechts de Papiamentstalige en de Nederlandstalige kinderen in de analyse te betrekken. De variabele school is in de analyse opgenomen als een zogenaamde dummy-variabele. De variabelen zijn in de analyse opgenomen in de gepresenteerde volgorde (forced entry). Dit is gedaan, omdat van belang is te weten of de scholen er iets toe doen. Van de overige variabelen is al in andere, reeds aangehaalde onderzoeken, in meerdere of mindere mate aangetoond dat ze de leerprestaties van de kinderen beïnvloeden. Door de variabele school als laatste variabele op te nemen kan worden nagegaan welke rol deze variabele nog speelt als alle andere variabelen al in ogenschouw zijn genomen. De resultaten worden gepresenteerd in tabel 6. Dit gebeurt voor het vierde en vijfde leerjaar afzonderlijk.

Tabel 5: Spellingsproblemen bij letters met accenten en trema's

letter	woorden/Nedse vertaling	Leerjaar 4		Leerjaar 5	
		Pap.	Ned.	Pap.	Ned.
a	bals, kas, fam	3,2	17,6	2,5	11,1
o	kauwgom, huis, achternaam	10,2	31,4	16,1	16,7
	bolo, bos, skol				
e	taart, stem, school	10,9	15,7	14,8	14,8
	brek, brel, slep				
u	rem, bril, slippen	16,3	31,4	17,0	24,1
	lus, pus, skur				
i	licht, etter, donker	0,4	9,8	2,8	3,7
	kil, fli, spil				
ò	kruis, vlieger, spiegel	47,0	82,4	42,6	68,5
	nò, shòrt, blòf				
è	nee, short, bluffen	36,6	72,6	38,6	55,6
	krèm, skèr, kèts				
ù	creme-kleur, schaar, gympies	44,5	78,5	31,5	57,4
	krùl, drùm, brùg				
ü	krullen, vat, brug	31,5	72,6	21,0	48,1
	stür, zür, hür				
e (shua)	stuur, zuur, huren	9,1	23,5	7,1	11,1
	feter, mangel, hamber				
	veter, snoepje, honger				

Tabel 6: Samenvatting van de stapsgewijze regressie analyse over de scores op het Papiamentse dictee

leerjaar 4	df	R ²	R ² -change	F-change	P
Nederlands	1	.23	.23	32.1	.000
woordenschat	2	.23	.00	.2	.68
red. intell.	3	.29	.06	8.2	.005
thuis taal	4	.43	.14	26.1	.000
school 1-8	11	.63	.20	7.4	.000
rest	97	.37			
leerjaar 5					
Nederlands	1	.26	.26	44.0	.000
woordenschat	2	.27	.01	.4	.512
red. intell.	3	.27	.01	1.2	.275
thuis taal	4	.45	.18	38.5	.000
school 1-8	11	.66	.21	10.4	.000
rest	114	.34			

N.B. De F-change heeft betrekking op een toetsing van gevonden effecten afgezet tegen de rest R².

De resultaten vertonen een patroon dat grotendeels gelijk is voor het vierde en vijfde leerjaar. De score op het Nederlandse dictee verklaart een groot deel van de variantie in de scores op het Papiamentse dictee (23 en 26%). De woordenschat voegt daaraan eigenlijk niets toe. De redekundige intelligentie doet er in klas vier nog een beetje toe, maar in klas vijf eigenlijk helemaal niet. De thuistaal draagt sterk bij aan de verklaring van de verschillen tussen de kinderen wat betreft hun prestaties op het Papiamentse dictee. Nog interessanter is het om vervolgens vast te stellen dat ook als de thuistaal reeds in de beschouwing is betrokken de variabele school een groot deel van de verklaring voor zijn rekening neemt (resp. 20 en 21% meer verklaarde variantie). Kortom, niet alleen het verschil tussen de scholen met overwegend

Papiamentstalige kinderen en de scholen met overwegend Nederlandstalige kinderen is bepalend. De variabele school is in de analyse een verzameling van verschillende subvariabelen. Deze betreffen bijvoorbeeld klassekenmerken van leerlingen, de kwaliteit van de leerkrachten, specifieke materiaal voorzieningen, het organisatievermogen van de leerkracht, effectieve leertijd en dergelijke.

Bij de uitgevoerde analyse dient te worden aangetekend dat niet expliciet is gecontroleerd voor de milieu-achtergrond van de leerlingen anders dan de thuistaal. Op grond van eerdere studies wordt aangenomen dat de invloed van het milieu grotendeels is verdisconteerd in de variabelen Nederlandse spellingsprestatie, woordenschat, redkundige intelligentie en thuistaal (zie De Jonge, 1986b; Prins-Winkel, 1973; Vedder, 1987c).

4. BESCHRIJVING VAN HET ONDERWIJS IN HET PAPIAMENTS

Aanvankelijk is geprobeerd na te gaan of er een relatie bestaat tussen leerkrachtkenmerken en instructievariabelen aan de ene kant en leerlingprestatie aan de andere kant. Tijdens de dataverzameling bleek dit heel moeilijk, aangezien aan het begin van het nieuwe schooljaar klassen niet meer te koppelen waren aan een bepaalde leerkracht. Hetzij de leerkracht was vertrokken naar een andere school, of de kinderen die in het schooljaar 86/87 een klas vormden waren aan het begin van het schooljaar 87/88 verspreid over verschillende klassen en voor een deel over verschillende scholen. Vervolgens is besloten om toch bij de leerkrachten van klas drie, vier en vijf van de deelnemende scholen informatie te verzamelen over hun houding t.o.v. het Papiaments, over de wijze waarop zij vorm geven aan het Papiaments en over de door de leerkrachten ervaren interesse van de ouders voor het vak Papiaments. Deze informatie geeft enig inzicht in de wijze waarop het vak Papiaments vorm krijgt op Curaçaose lagere scholen. Deze informatie geeft een kader waarbinnen de resultaten op leerlingniveau kunnen worden geïnterpreteerd.

De leerkrachten is gevraagd een vragenlijst in te vullen. Op één school wilden de leerkrachten niet meedoen aan dit onderdeel. In totaal 29 leerkrachten vulden de vragenlijst wel in.

Van deze 29 leerkrachten gaven 22 te kennen dat ze in het schooljaar 86/87 Papiamentse les hebben gegeven. Zeven hebben geen Papiaments gegeven. Van de 22 leerkrachten die wel Papiaments gaven hebben negen volgens eigen zeggen alle stof voor dat cursusjaar behandeld. Acht hebben 75 - 99% van de stof behandeld en vijf 50 - 74% van de stof. Achttien van de 22 leerkrachten gaven dictees voor het vak Papiaments. Van die 18 gaven elf cijfers voor de dictees en van deze elf gaven acht leerkrachten ook de een of andere beoordeling op het rapport.

Elf van de 22 leerkrachten gaven geen huiswerk voor het vak Papiaments (huiswerk is een normaal verschijnsel op Curaçaose lagere scholen). Eén leerkracht gaf één keer in de maand, vijf gaven één keer per twee weken, drie gaven één keer per week en twee leerkrachten gaven twee keer per week huiswerk mee naar huis. Tenslotte is nagegaan of ouders interesse toonden voor het vak Papiaments en de prestaties van hun kinderen bij dit vak. Twintig leerkrachten gaven aan dat geen der ouders enige interesse toonde. Eén leerkracht bespeurde bij 10% van de ouders of minder enige interesse en één leerkracht gaf aan dat elf tot veertig procent van de ouders interesse toonden.

Al met al kan worden geconcludeerd dat er noch door het gros van de leerkrachten noch door de ouders met het vak Papiaments wordt omgegaan alsof het een serieus vak is.

Wat de leerkrachten betreft valt het resultaat wellicht nog mee. Officieel is het vak Papiaments ingevoerd onder het beding dat de prestaties voor het vak niet mogen worden betrokken bij de selectie op school. In het licht van deze maatregel is het misschien verbazingwekkend dat relatief veel leerkrachten cijfers geven, een beoordeling op het rapport geven en de leerlingen soms huiswerk laten maken. De leerkrachten lijken voor een deel verder te willen gaan dan de beleidmakers. Deze houding vinden we terug in de resultaten van een attitudetest. De meeste leerkrachten blijken een relatief positieve houding te hebben t.o.v. het vak Papiaments. De verschillen tussen leerkrachten zijn echter groot. Er zijn ook leerkrachten met een minder positieve houding t.o.v. het Papiaments.

5. AANBEVELINGEN EN DISCUSSIE

Op Curaçao zijn al vele spellingscursussen georganiseerd voor leerkrachten en andere volwassenen. De effecten van deze cursussen laten slechts de conclusie toe dat de Papiamentse spelling in een betrekkelijk korte tijd goed leerbaar is. In dit licht is het bevreemdend dat de prestaties in de onderzochte groepen leerlingen zo matig zijn. Dat de kinderen na één jaar spellingsonderwijs niet een meer acceptabel prestatie leveren kan verklaard worden door verschillende factoren.

- Een weinig serieuze houding t.o.v. het vak Papiaments. Dit is voor een deel het gevolg van het feit dat het vak duidelijk geen selectierol heeft en dus in eerste instantie geen rapportvak is. Hierdoor kunnen de leerkrachten het vak als een restvak zien dat niet al te veel inzet en moeite mag kosten. Ouders menen dat ze de kinderen niet hoeven te stimuleren om hun best te doen.
- Er is buiten school weinig stimulans tot het oefenen van de geleerde Papiamentse spelling. Niet alle artikelen in de kranten worden volgens de heersende spelling geschreven. De meeste ouders, grote broers en zussen schrijven het Papiaments volgens eigen goed denken.
- Het lijkt er op dat in de methode te weinig gericht aandacht wordt besteed aan het herhaaldelijk oefenen van de grote problemen van de Papiamentse spelling (accenten, trema's, klank-letterverbindingen waarbij verwarring met het Nederlands kan optreden).

Het eerste punt, de weinig serieuze houding t.o.v. het vak Papiaments verdient boven alles aandacht als het gaat om de verbetering van het Papiamentse spellingsonderwijs. Er kan worden overwogen om van Papiaments een gewoon vak te maken met duidelijk omschreven doelen en te onderscheiden niveaus van beheersing, dat ook op het rapport meetelt. Het is te verwachten dat, indien Papiaments geen serieus vak wordt met goede leereffecten, terechte efficiëntie-overwegingen (veel geld en tijd investeren in iets met weinig effect) er toe kunnen leiden dat de positie van het vak Papiaments verder wordt gemarginaliseerd.

Meer in het algemeen lijkt het aanbevelenswaardig om verder te gaan met het evalueren van de gehanteerde spellingsmethode en de opbrengsten en dit te combineren met het doorvoeren van veranderingen in de spellingsmethode, om aldus op meer onderzoeksmatige wijze te zoeken naar een methode die leidt tot gewenste effecten. Het zou dienstig zijn om een grondige implementatie-analyse op te zetten, bijvoorbeeld in de vorm van een begeleidingsproject.

In aansluiting bij deze opmerking kunnen tenslotte nog enkele methodologische kenmerken van het onderzoek worden gememoreerd die de zeggingskracht van de uitkomsten beperken. Er was sprake van een ontwerp met een nameting bij een a-selecte steekproef met een bestaande, niet gecontroleerde treatment. Het was niet mogelijk om informatie te verzamelen over de beheersing van de Papiamentse spellingsvaardigheid bij de onderzochte leerlingen voor de invoering van het vak Papiaments. De wijze waarop de lessen Papiaments vorm kregen kon slechts via een kleine survey worden onderzocht. Meer directe observaties van het doen en laten van leerkrachten en leerlingen tijdens de lessen konden niet worden gedaan.

REFERENTIES

- Cito: Spellingstoetsen voor klas een en klas twee. Arhem: Cito, 1980.
- Evertsz, F. & Faries, T. (1986). Introdukshon di papiamentu komo materia na skil. Unpublished paper, Willemstad.
- Geest, A. N., Van der & Swüste, W. (1978). Spellingwijzer. Praxisreeks nr. 14, Den Bosch: Malmberg.
- I.L.A./Komishon Standardisashon di Papiamentu (1986). Lista di palabranan standardias, no 1. Curaçao: I.L.A.
- Jonge, J., de (1986). De overgang van basis- naar voortzetzet onderwijs op curacao (cohort 1981): DE voorspellende waarde van de selectie-instrumenten. Curaçao: Departement van Onderwijs (a).
- Jonge, J., de (1986). Onderwijs en werk. Curaçao: Departement van Onderwijs (b).
- Kema, G. N.: P.M.A. 2-4 (1980). Handleiding en verantwoording. Amsterdam: Swets & Zeitlinger.
- Kema, G. N.: P.M.A. 5-6 (1976). Handleiding en verantwoording. Amsterdam: Swets & Zeitlinger.
- Kook, K. B. & Vedder, P. (1988). Het spellen van het papiaments door curaçaose schoolkinderen. Curaçao: Komapa.
- Prins-Winkel, N. (1973). KABES DURU? Assen: Van Gorcum.

- Vedder, P. (1978). Learning in a second language: A study in curacaon primary schools. Curacao: Kondesa (a).
- Vedder, P. (1987). De Nederlandse taal in het Curaçaose lager onderwijs. Comenius, 7, nr. 3, pp. 290-304(b).
- Vedder, P. (1987). De relatie tussen schoolse en buitenschoolse socialisatie bij curacaose schoolkinderen. Curaçao: Sedukal (c).

Manuscript ontvangen 11-4-1988

Definitieve versie ontvangen 21-9-1988

BEKROOND ORD PAPER 1988

Dit artikel is een bewerking van het door de VOR bekroonde ORD-paper.

De Rijksinspectie en het Voortgezet Onderwijs.

J.F.A. Braster

RISBO - Afdeling Onderwijsonderzoek, Erasmus Universiteit Rotterdam

ABSTRACT

This article deals with the functioning of the Dutch School Inspectorate. Data are provided by 329 headmasters of schools for secondary education. The central topic of this paper is the relationship between the statutory tasks the Inspectorate has to fulfil and the use of specific means by which those goals can be achieved. In this respect attention will be paid to the contacts between school and Inspectorate, the influence of the Inspectorate on the educational process in general and to the headmaster's appreciation of the actual work done by the Inspectorate. It will be shown that differential types of contact between school and Inspectorate are just partially responsible for the 'effectiveness' of the Inspectorate.

1 INLEIDING

'Geen leraar kent de inspecteur. Soms zien de leraren bij het eindexamen een man buiten adem binnen komen om snel de afstand tussen de tafeltjes in de examenzaal te meten en nog sneller weer te verdwijnen. Dat is de inspecteur. Aan de inspecteur heb je als school niets. Onderwijskundig stelt hij ook niets voor'. Op deze wijze opent een artikel dat in 1981 verschijnt in het onderwijstijdschrift *School* (Ahlers, 1981). Het rijtje negatieve uitspraken kan gemakkelijk worden uitgebreid. Negatieve verhalen over het functioneren van de inspectie van het onderwijs doen immers al jaren de ronde. Van een empirische weerlegging of bevestiging van dergelijke verhalen is tot voor kort echter nauwelijks sprake geweest.

Pas in 1987 verschijnen er, praktisch tegelijkertijd, twee empirische studies die betrekking hebben op het functioneren van de inspectie van het onderwijs of het Rijksschooltoezicht. Zo wordt op 26 mei van dat jaar door de Algemene Rekenkamer een rapport inzake de inspectie van het onderwijs aangeboden aan de Tweede Kamer der Staten-Generaal (Algemene Rekenkamer, 1987). En in juni 1987 verschijnt er een rapport van de Vakgroep Onderwijssociologie en Onderwijsbeleid van de Erasmus Universiteit te Rotterdam getiteld 'De inspecteur en de school' (Braster, 1987). Dit laatste onderzoek is een deelstudie in het kader van een meeromvattend promotie-onderzoek inzake het functioneren van de Rijksinspectie dat in 1988 is afgerond (Elte, 1988). In dit artikel zal met name aandacht worden besteed aan de resultaten van deze deelstudie.

2 PROBLEEMSTELLING

In artikel 23 van de Nederlandse grondwet is vastgelegd dat het (doen) geven van onderwijs vrij is. Daarmee wordt in essentie bedoeld dat particulieren vrij zijn om scholen op te richten van welke levensbeschouwelijke of maatschappelijke aard dan ook. Er is echter geen absolute vrijheid om die scholen naar eigen goeddunken te modelleren. 'Het geven van onderwijs is vrij', zo is te lezen in de grondwet, maar deze passage wordt direct gevolgd door de zinsnede 'behoudens het toezicht van de overheid...'. De rijksoverheid behoudt zich daarmee het recht voor een vinger aan de pols van het onderwijs te houden (vgl. Mentink & Akkermans, 1986, p.74-81).

Adres: Postbus 1738, 3000 DR Rotterdam

In diverse onderwijswetten wordt dit toezicht op het onderwijs opgedragen aan de minister van onderwijs. Onder zijn gezag wordt het uitgeoefend door een aantal hoofdinspecties dat een onderdeel vormt van het ministerie van Onderwijs en Wetenschappen (Leune, 1987, p.319 e.v.).

In dit artikel komt alleen de Hoofdinspectie Voortgezet Onderwijs aan de orde. Deze hoofdinspectie is belast met het toezicht op scholen voor algemeen voortgezet onderwijs (AVO) en lager beroepsonderwijs (LBO). In de sector van het AVO werken ongeveer 35 veldinspecteurs. Gemiddeld is een AVO-inspecteur belast met het toezicht op 39 scholen. De inspectie belast met het LBO telt 29 veldinspecteurs die elk te maken hebben met ongeveer 45 scholen.

De taken van de AVO- en LBO-inspectie worden nader genoemd in de Wet op het voortgezet onderwijs van 1963. Blijkens een ministeriële nota van 1984 zou het met name gaan om de volgende vier hoofdtaken (Ministerie van Onderwijs en Wetenschappen, 1984, p.4):

- het toezien op de naleving van de wettelijke voorschriften (controleren);
- het bekend blijven met de toestand van het onderwijs (evalueren);
- het bevorderen van de ontwikkeling van het onderwijs (stimuleren);
- het gevraagd en ongevraagd rapporteren en adviseren aan de minister (rapporteren).

Om deze genoemde taken naar behoren te kunnen uitvoeren heeft de inspectie de beschikking over een aantal beleidsinstrumenten. In de Wet op het voortgezet onderwijs worden de volgende middelen genoemd:

- de vrije toegang tot zowel openbare als bijzondere scholen;
- de verplichting van personeel en bestuur van een school om alle vereiste inlichtingen te verstrekken;
- het schoolbezoek;
- het voeren van overleg met betrokkenen op schoolniveau en lokaal niveau.

De probleemstelling van dit artikel heeft betrekking op het functioneren van het Rijksschooltoezicht in het voortgezet onderwijs. Daarmee wordt de relatie tussen de taken van de inspectie (de doelen) en de inzet van beleidsinstrumenten (de middelen) centraal gesteld. Het nader analyseren van deze relatie betekent dat antwoord moet worden gegeven op een drietal vragen:

- Op welke wijze maakt de inspectie gebruik van de beschikbare beleidsmiddelen?
- In hoeverre boekt de inspectie resultaten als het gaat om het uitoefenen van haar taken?
- In welke mate zijn de resultaten van het inspecteren afhankelijk van het gebruik van beleidsmiddelen en van kenmerken van de te inspecteren scholen?

3 DATA

Gelet op het feit dat er nagenoeg geen empirische kennis aanwezig is aangaande het functioneren van de inspectie in de alledaagse schoolpraktijk, is besloten om een globale vragenlijst op te stellen die is voorgelegd aan een groot aantal schoolleiders van scholen voor voortgezet onderwijs (Bijdevaate, 1986). Deze lijst bestaat uit enkele tientallen gestructureerde vragen die betrekking hebben op vier groepen van variabelen:

- kenmerken van de school;
- de aard van het contact tussen school en inspectie;
- de invloed van de inspectie op de gang van zaken op school;
- de waardering voor het feitelijk functioneren van de inspectie.

De vragenlijst is eind april 1986 toegezonden aan 500 schoolleiders. Daarmee is ongeveer één op de vijf scholen voor AVO en LBO verzocht om medewerking te verlenen. Bij de aselechte trekking van de steekproef is rekening gehouden met twee stratificatiecriteria: de richting van de school en de schoolstructuur. Uitgangspunt was dat elke combinatie van deze twee criteria in ongeveer gelijke mate vertegenwoordigd moest zijn. Op deze regel werd één uitzondering gemaakt: aan alle rijksscholen in Nederland (54 in getal) is een vragenlijst toegestuurd. Op deze wijze werd het mogelijk om vergelijkingen te maken tussen de door het Rijk, de gemeente en privaatrechtelijke organen bestuurd scholen voor voortgezet onderwijs. Rijksscholen zijn met name interessant in verband met de extra bevoegdheden die het Rijksschooltoezicht heeft op deze scholen. Zo worden in het Reglement Rijksscholen onder meer de volgende taken

onderscheiden: het plegen van overleg met de rector over de toetsingsmiddelen, de verwijdering van leerlingen, het schoolwerkplan en de benoeming van leerkrachten en het onderhouden van contacten met het ministerie over de benoeming en installatie van de rector.

De vraag rijst of het met de gekozen steekproefopzet nog wel mogelijk is om conclusies te trekken die representatief zijn voor de totale groep van scholen voor AVO en LBO. Een vergelijking van het oorspronkelijke databestand met een bestand dat is verkregen na het uitvoeren van een correctie voor de disproportionaliteit van de steekproef wijst echter uit dat er slechts geringe verschillen bestaan tussen deze twee databestanden (Braster, 1987, p.163 e.v.). De te trekken conclusies zijn daarmee in zekere zin generaliseerbaar naar de totale populatie.

Van de 500 verzonden enquêtes zijn er uiteindelijk 358 geretourneerd (72%). Het aantal bruikbare vragenlijsten was iets lager tengevolge van 20 weigeringen en 9 onbekende adressen. Het totale analysebestand bestond dus uit 329 scholen, waarmee het responspercentage uitkwam op 66. Deze response is, mede gelet op de huidige ervaringen van andere onderwijsonderzoekers, bevredigend te noemen. Er is daarom afgezien van het verzenden van een (tweede) rappel aan de eerder geselecteerde scholen.

4 CONTACT

Het schoolbezoek is het voornaamste beleidsinstrument dat de inspectie ter beschikking staat om de na te streven inspectiedoelen te verwezenlijken. Andere in de wet genoemde middelen als de vrije toegang tot scholen en de verplichting van scholen om informatie te verstrekken, zijn immers op te vatten als voorwaarden voor een adequaat gebruik van het schoolbezoek als beleidsinstrument. Op basis van de gegevens die zijn verzameld tijdens het schoolbezoek kan de inspectie dan vervolgens in overleg treden met het bevoegd gezag, de leiding en/of het personeel van de school.

Het contact dat de inspecteur heeft met de school kan op verschillende manieren worden getypeerd. We dienen voor een nadere typering in ieder geval aandacht te besteden aan een zestal contactvariabelen:

- het aantal malen dat de inspecteur de school onaangekondigd heeft bezocht (vgl. Dekker & Gijtenbeek, 1975);
- het aantal malen dat de school aangekondigd is bezocht;
- het al of niet bijwonen van de landelijke afname van het centraal schriftelijk eindexamen door de inspecteur;
- de personen met wie de inspecteur tijdens het laatste schoolbezoek gesprekken heeft gevoerd;
- het aantal malen dat de inspecteur lessen heeft bijgewoond;
- het aantal malen dat de schoolleider de inspecteur of één van zijn medewerkers telefonisch heeft geraadpleegd.

Deze variabelen worden geacht onderling met elkaar samen te hangen. Het vaker voorkomen van de ene vorm van contact zal met andere woorden samengaan met het vaker voorkomen van een andere vorm van contact. Opmerkelijk is echter dat er slechts voor een beperkt aantal combinaties van steeds twee contactvariabelen dergelijke samenhangen worden gevonden. Anders gezegd: het contact tussen school en inspectie laat zich niet ééndimensionaal beschrijven. Het terugbrengen van de genoemde zes contactvariabelen tot één variabele behoort niet tot de mogelijkheden.

Het is daarentegen wel mogelijk om op basis van deze variabelen een vijftal clusters te onderscheiden waarbinnen een bepaald type contact overheerst. Deze conclusie werd getrokken na een analyse van de contactvariabelen met behulp van HOMALS. Deze techniek is in het bijzonder geschikt voor de analyse van meerdere categorische variabelen. Met HOMALS is het mogelijk om groepen scholen op te sporen die qua contactpatroon met elkaar overeenkomen (zie verder: Van den Berg, 1986; Gifi, 1980). De met behulp van HOMALS onderscheiden clusters of contactprofielen laten zich als volgt beschrijven.

Cluster I -Minimaal contact- Op de scholen die binnen deze groep vallen, is er nauwelijks contact tussen school en inspecteur. Het bezoek is beperkt tot maximaal één maal in een periode

van zo'n tien maanden. Dan wordt er meestal alleen gesproken met de rector of directeur. Lessen worden bijna nooit bijgewoond. Telefonisch contact komt ook sporadisch voor. Het centraal schriftelijk eindexamen wordt ook meestal niet bezocht. In totaal gaat het om 29% van de scholen.

Cluster II -Modaal contact- In vergelijking met de eerste groep is hier sprake van een hogere contactfrequentie. Er wordt meer dan in cluster I telefonisch overleg gevoerd (zo'n vijf tot tien keer). Het aantal malen dat de inspecteur de school (aangekondigd) bezoekt ligt rond de twee keer per tien maanden. Daarbij wordt ook vrij vaak met leerkrachten gesproken. Het centraal schriftelijk eindexamen wordt echter bijna nooit bezocht. Soms wordt één van de lessen bijgewoond. Het gaat om zo'n 35% van de scholen.

Cluster III -Intensief contact- Een kleine maar bijzondere groep is cluster III. De scholen die binnen dit contactprofiel vallen worden gemiddeld drie maal aangekondigd bezocht. Ook is er sprake van een relatief hoge mate van onaangekondigd bezoek. Het eindexamen wordt soms wel, soms niet bijgewoond. Het telefonisch contact is in vergelijking met de andere groepen het hoogst. Het kan oplopen van 16 tot 20 keer in een periode van tien maanden. Verder is het opvallend dat de inspecteur op de scholen binnen deze groep veel aandacht besteedt aan het leggen van contacten met leerkrachten, leerlingen, ouders en leden van het niet-onderwijzend personeel. In een meerderheid van de gevallen worden ook meerdere lessen bijgewoond. In totaal gaat het om zo'n 5% van de scholen.

Cluster IV -Frequent contact- Ook binnen deze groep die zo'n 12% van de scholen omvat, is er een relatief frequent contact tussen schoolleider en inspecteur. Dat komt echter niet tot uiting als we het telefonisch contact bekijken (dat blijkt in dit cluster weinig voor te komen), maar wel als we letten op het (on)aangekondigd bezoeken van de school. Onaangekondigd bezoek komt gemiddeld één maal voor, terwijl het aangekondigd bezoek op kan lopen tot vier keer of meer in de onderzochte periode. Het bijwonen van lessen komt ook geregeld voor, evenals een relatief hoge mate van contact met leerkrachten. Op de scholen die binnen dit cluster vallen wordt het centraal schriftelijk relatief vaak bezocht.

Cluster V -Examen contact- Voor de scholen binnen cluster V geldt dat het centraal schriftelijk eindexamen altijd wordt bezocht. Het lijkt er echter op dat het eindexamen de enige gebeurtenis is waarbij de schoolleider in contact komt met de inspecteur. Lessen worden immers niet bijgewoond, het telefonisch contact is tot een minimum beperkt en met leerkrachten en/of leerlingen wordt niet gesproken. Voor drie kwart van de scholen geldt dat de inspecteur zich slechts één maal in de onderzochte periode op school laat zien. In totaal gaat het dan om 20% van de scholen.

De aard van het contact tussen school en inspectie (het contactprofiel) wordt verondersteld te variëren naar schooltype. De inspectie heeft namelijk een aantal neventaken in het openbaar onderwijs (benoemingskwesies), terwijl scholen voor beroepsonderwijs, gelet op de materiële toerusting, eveneens een andere vorm van contact nodig hebben. In totaal worden in dit artikel op basis van de bestuursvorm en de structuur van de school zeven verschillende schooltypen onderscheiden. Het aantal geselecteerde scholen per type is vermeld kolom 2 van tabel 1.

Uit tabel 1 is af te lezen dat sommige combinaties van bestuursvorm en schoolstructuur niet voorkomen. Zo bestaan er in Nederland geen categoriale rijksscholen of rijksscholen voor LBO. Gecombineerde scholengemeenschappen voor AVO en LBO komen daarentegen wel voor, maar scholen van dit type waren in onze steekproef slechts in geringe mate aanwezig. Dat gold eveneens voor gemeentelijke scholengemeenschappen AVO-LBO en voor gemeentelijke scholen voor LBO. Gezien de kleine aantallen in onze steekproef worden voor deze schooltypen geen conclusies getrokken.

In tabel 1 wordt het aantal scholen per schooltype vermeld en tevens het gemiddeld aantal leerlingen per school. Bovendien zijn twee kenmerken die een indicatie geven voor de 'stabiliteit' van de school vermeld:

- het bestaan van concrete fusieplannen;
- het voordoen van belangrijke wijzigingen in de samenstelling van de school (bijvoorbeeld door fusie, uitbreiding van het aantal schooltypen of splitsing) in de laatste vijf jaar.

Tabel 1. Schoolkenmerken, contactprofiel, invloedsgebieden en waardering naar bestuursvorm en schoolstructuur

	SCHOOL- KENMERKEN	CONTACT ¹	INVLOED ²	WAARDERING ³
RIJK				
s.g.AVO	32 scholen 1088 leerlingen 9% fusieplannen 3% wijzigingen	38% frequent 25% examen 19% modaal	68% benoeming SL 59% benoeming LK 22% lessentabel 22% eindexamen 19% s-werkplan	63% toezicht 56% kennis 31% sturing 25% conflict. 19% bemiddeling
GEMEENTE				
cat.AVO	21 scholen 383 leerlingen 14% fusieplannen 0% wijzigingen	38% examen 33% minimaal 19% frequent	55% benoeming SL 38% eindexamen 29% lessentabel 20% benoeming LK	57% kennis 48% toezicht 19% sturing 19% conflict. 19% bemiddeling
GEMEENTE				
s.g.AVO	17 scholen 1276 leerlingen 6% fusieplannen 12% wijzigingen	41% examen 29% modaal 24% minimaal	56% benoeming SL 41% eindexamen 18% lessentabel	38% toezicht 31% kennis 25% sturing 20% bemiddeling 19% conflict.
BIJZONDER				
cat.AVO	62 scholen 356 leerlingen 21% fusieplannen 6% wijzigingen	40% minimaal 24% examen 24% modaal 11% frequent	32% eindexamen 31% lessentabel	77% toezicht 61% kennis 24% sturing 23% conflict. 21% bemiddeling
BIJZONDER				
s.g.AVO	67 scholen 1141 leerlingen 15% fusieplannen 16% wijzigingen	36% examen 36% minimaal 18% modaal	33% eindexamen	67% toezicht 64% kennis 23% bemiddeling 21% sturing
BIJZONDER				
AVO-LBO	23 scholen 719 leerlingen 17% fusieplannen 48% wijzigingen	43% modaal 22% minimaal 13% intensief 13% frequent	50% inventaris 27% huisvesting 27% lessentabel 19% leermiddelen 18% vernieuwing	83% toezicht 78% kennis 39% bemiddeling 22% sturing 17% inspiratie 17% kwaliteit
BIJZONDER				
LBO	82 scholen 460 leerlingen 52% fusieplannen 21% wijzigingen	70% modaal 17% minimaal	67% inventaris 43% huisvesting 41% lessentabel 34% leermiddelen 32% vernieuwing 30% eindexamen 23% s-werkplan	84% toezicht 59% kennis 44% sturing 40% bemiddeling 27% vertaling 23% conflict. 23% inspiratie 23% kwaliteit

Uit kolom 2 van tabel 1 blijkt dat met name scholen voor LBO in de toekomst te maken zullen krijgen met fuseringen. Er zijn in Nederland veel kleine scholen voor LBO die voortdurend in hun bestaan worden bedreigd (Elte, 1988, p.82). Enerzijds hebben zij te maken met dalende leerlingaantallen waardoor zij onder de opheffingsnormen van de wet dreigen te komen en anderzijds worden zij geconfronteerd met landelijke schaalvergrotingsoperaties. Uit tabel 1 is tevens af te lezen dat het met name de gecombineerde scholengemeenschappen voor AVO-LBO zijn geweest die in het verleden te maken hebben gehad met ingrijpende wijzigingen van de schoolstructuur.

Er kan worden verondersteld dat beide schoolkenmerken van invloed zijn op de relatie tussen school en inspectie. Een minder stabiele situatie zou voor de inspecteur bijvoorbeeld aanleiding kunnen zijn om de betreffende school vaker te bezoeken. We geven echter eerst een antwoord op de vraag in hoeverre de aard van het contact tussen school en inspectie varieert als we het schooltype in ogenschouw nemen.

In kolom 3 van tabel 1 is het voorkomen van een bepaald contactprofiel uitgesplitst naar bestuursvorm en schoolstructuur. (Alleen percentages hoger dan 10 zijn vermeld). Het modale contactprofiel blijkt te overheersen op de scholen voor LBO, terwijl dit type contact ook relatief vaak voorkomt op scholen voor AVO-LBO. Van examencontact is nauwelijks sprake op scholen met een LBO-afdeling. Deze vorm treffen we daarentegen relatief vaak aan op scholen voor AVO. Binnen deze sector geldt bovendien dat de inspecteur frequent in contact treedt met de scholen waar het rijk optreedt als bevoegd gezag. In globale zin kan er, gelet op het contactprofiel, een onderscheid worden gemaakt tussen drie clusters van schooltypen: de rijksscholen, de scholen met een LBO-afdeling en de niet door het rijk bestuurde scholen voor AVO. De relatie die de inspectie heeft met de laatstgenoemde groep van scholen laat zich daarbij het best omschrijven in termen van minimaal en beperkt tot het bezoeken van het centraal schriftelijk eindexamen.

De meer intensieve vorm van contact tussen inspectie en school komt vrij weinig voor. Alleen op de gecombineerde scholengemeenschappen voor AVO-LBO is het contact in 13% van de gevallen te typeren als intensief. In het algemeen kan echter worden gesteld dat de betrekkingen tussen inspecteurs en de scholen gebreken vertonen. Grote groepen scholen hebben slechts in beperkte mate contact met de inspecteur. Minimaal contact en examen contact komen vaak voor. Het meer frequent bezoeken van de school komt relatief weinig voor. Binnen de inspectie wordt er van uitgegaan dat elke inspecteur twee à drie dagen per week bezig hoort te zijn met het bezoeken van scholen (inclusief voorbereiding en uitwerking). Minimaal twee dagen per week aandacht voor het schoolbezoek blijkt in de praktijk een ijdel streven te zijn. Inspecteurs geven dit in gesprekken zelf ook aan. Veel tijd wordt besteed aan andere zaken zoals het overleg binnen de inspectie, administratieve werkzaamheden en het bijhouden van beleidsontwikkelingen.

Het schoolbezoek is het belangrijkste instrument om bekend te blijven met de toestand van het onderwijs en om toe te zien op de naleving van de wettelijke voorschriften. Op basis van de bij het schoolbezoek verkregen informatie kan de inspecteur een bijdrage leveren aan de kwaliteit van het onderwijs, onder andere via het rapporteren en adviseren aan de minister. In de praktijk dreigt het schoolbezoek door een gebrek aan tijd echter het kind van de rekening te worden (Algemene Rekenkamer, 1987, p.26-27).

5 RESULTATEN VAN HET INSPECTEREN

In deze paragraaf komen de resultaten van het inspecteren aan de orde. Het gaat daarbij achtereenvolgens om de invloed van de inspectie op de dagelijkse gang van zaken op een school en om de waardering van de schoolleider voor het feitelijk functioneren van de inspectie. In beide gevallen gaat het om indirecte metingen van het functioneren van de inspectie. Deze gegevens kunnen een indicatie geven van de resultaten van het inspectiewerk op schoolniveau. Met nadruk moet worden vermeld dat het hier gaat om gepercipieerde en niet om feitelijke

resultaten. Uitspraken over de effectiviteit van het inspecteren worden niet gedaan. De meting van die effectiviteit, opgevat als de mate waarin het verwezenlijken van de inspectiedoelen exclusief aan het opereren van de inspectie is toe te schrijven, is slechts mogelijk als gekozen wordt voor een experimentele onderzoeksofzet. Het is in de praktijk echter onmogelijk om met de onderwijsinspectie te experimenteren (vgl. Elte, 1988, p.161).

De invloed van het Rijksschooltoezicht op openbare scholen heeft met name betrekking op de benoeming van onderwijsgevend personeel, terwijl op de scholen voor LBO de inspectie een belangrijke plaats inneemt in de middelenvoorziening en op basis hiervan een zekere mate van allocatiemacht heeft over de school. Vaklokalen, machines voor houtbewerking, grootkeuken, fijnmetaal, motorvoertuigentechniek en dergelijke, alsmede een constante toestroom van grondstoffen zijn voor het LBO van vitaal belang. De inspecteur is het eerste station op weg naar bekostiging, vervanging of uitbreiding van de inventaris binnen de LBO-sector (Elte, 1988, p.83-84). Deze verschillen in bemoeienis komen duidelijk naar voren in kolom 4 van tabel 1.

Ook voor alle andere in tabel 1 genoemde invloedsgebieden geldt dat er (statistisch) significante verschillen bestaan tussen de schooltypen. Op het terrein van de huisvesting en de vaststelling van de lessentabel wordt op de scholen voor LBO relatief veel invloed toegeschreven aan de inspecteur, terwijl er op deze scholen eveneens sprake is van een zekere mate van bemoeienis op het gebied van de leer- en hulpmiddelen, onderwijskundige vernieuwingen en het centraal schriftelijk eindexamen. Op de bijzondere scholen voor AVO heeft de inspectie alleen op dit laatste terrein enige invloed. De schoolleiders van gemeentelijke scholen ervaren eveneens een zekere invloed indien het eindexamen in geding is. De inspecteur blijkt zich op deze scholen echter meer te bemoeien met de benoeming van nieuwe schoolleiders. Deze uitspraak is eveneens van toepassing op de rijksscholen. Op dit schooltype blijkt de inspecteur echter ook een belangrijke rol te spelen indien het gaat om de benoeming van leerkrachten. Van een dergelijk invloed is op de gemeentelijke scholen in veel mindere mate sprake. Op de rijksscholen lijkt de inspectie met name een rol te vervullen als verlengstuk van het bevoegd gezag, i.c. de minister van Onderwijs en Wetenschappen.

In het algemeen wordt de invloed van het Rijksschooltoezicht het sterkst ervaren op terreinen die de uitvoering van landelijke regelingen betreffen. Zo geldt voor de totale groep scholen bijvoorbeeld dat de bemoeienis met het centraal schriftelijk eindexamen het hoogst wordt ingeschat. Ongeveer 29% van de schoolleiders percipieert (zeer) veel invloed van de inspectie op dit terrein. Van een totale lijst met veertien potentiële invloedsgebieden zijn er negen opgenomen in kolom 4 van tabel 1. Op de resterende vijf terreinen werd door minder dan 15% van de schoolleiders veel invloed aan de inspecteur toegeschreven. Het gaat dan om zaken als het vaststellen van algemene doelstellingen of doelstellingen per vak, de leerlingadministratie, didactiek en methodiek en toelatingsvoorwaarden.

Een tweede manier om in te gaan op de resultaten van de inspectorale arbeid betreft het vaststellen van de mate waarin het feitelijk functioneren van de inspecteur door de schoolleider wordt gewaardeerd. Daartoe is aan de respondenten een aantal uitspraken voorgelegd die elk betrekking hebben op een bepaalde functie die het Rijksschooltoezicht kan vervullen. Deze functies, die voortvloeien uit de specifieke taakstelling van het inspectie-apparaat, zijn opgenomen in tabel 2.

In het algemeen kunnen de volgende conclusies worden getrokken. Een duidelijke meerderheid van de schoolleiders is van mening dat de inspecteur op een effectieve wijze toeziet op de naleving van de wettelijke voorschriften en goed op de hoogte is van de gang van zaken op een school. De waardering voor de uitoefening van andere taken is in het algemeen niet zo hoog. Als het gaat om de sturings- of bemiddelingsfunctie beoordeelt ongeveer 30% van de schoolleiders de inspectie positief. Indien het gaat om de oplossing van onderwijskundige meningsverschillen, om het 'vertalen' van circulaire, om het verhogen van de kwaliteit van het onderwijs of om het leveren van inspirerende ideeën, dan spreekt hooguit één op de vijf schoolleiders zijn waardering voor de inspectie uit. Het controleren blijkt de inspecteur niet slecht af te gaan, maar van een meer sturende of stimulerende rol is volgens de schoolleiders in mindere mate sprake.

Tabel 2. Waardering voor het functioneren van de inspectie (%)

FUNCTIE -item	(zeer) eens	neutraal	(zeer) oneens
TOEZICHT -De inspecteur ziet op effectieve wijze toe op naleving van de voorschriften	71	19	10
KENNIS -De inspecteur is goed op de hoogte van de gang van zaken op mijn school	60	22	18
STURING -De inspecteur stimuleert de onderwijskundige ontwikkeling van de school nauwelijks	46	25	30
BEMIDDELING -Dankzij bemiddeling van de inspecteur worden veel verzoeken van de school door het ministerie van O & W gehonoreerd	29	39	32
CONFLICTOPLOSSING -Bij onderwijskundige meningsverschillen hechten leraren veel waarde aan de mening van de inspecteur	19	30	51
VERTALING -De inspecteur levert een grote bijdrage aan de 'vertaling' van (moeilijke) circulaire's	15	30	55
KWALITEITSBIJDRAGE -De inspecteur levert een grote bijdrage aan de kwaliteit van het onderwijs	13	40	47
INSPIRATIE -Voor mij vormen de ideeën van de inspecteur betreffende het onderwijs een grote bron van inspiratie	13	33	55

De vraag rijst of de waardering voor het functioneren van de inspectie verschilt per schooltype. Het feit dat de inspecteur op verschillende schooltypen een andere invloedspositie inneemt tengevolge van een aantal neventaken, hoeft nog niet te betekenen dat de perceptie van het inspectorale functioneren op een gelijke wijze zal variëren. Deze stelling dient in ieder geval op te gaan indien het gaat om het houden van toezicht op de naleving van voorschriften of om het kennis hebben van de gang van zaken op een school. In het laatste geval blijken er in de praktijk inderdaad geen (statistisch) significante verschillen te bestaan tussen de schooltypen (kolom 5 van tabel 1). De waardering voor het uitoefenen van toezicht is daarentegen bijzonder hoog binnen de LBO-sector en relatief laag op de gemeentelijke scholen voor AVO. Gemeentelijke scholen hebben echter niet zelden ook te maken met de gemeentelijke inspectie, een onderdeel van de gemeentelijke overheid. Het is mogelijk dat deze inspectie de controlefunctie van de rijksinspectie voor een deel heeft overgenomen.

De waardering voor de uitoefening van alle andere functies, met uitzondering van het oplossen van conflicten, verschilt eveneens per schooltype. Daarbij blijken met name de scholen voor LBO er positief uit te springen. Het LBO is het enige schooltype waar alle acht onderscheiden functies door meer dan minstens 15% van de schoolleiders positief wordt beoordeeld. Bovendien valt op dat een relatief hoge waardering voor de onderwijskundige beïnvloeding van deze scholen samengaat met een relatief hoge waardering voor het uitoefenen van toezicht. Deze twee op het oog tegenstrijdige hoofdtaken van de inspectie (controleren en stimuleren) hoeven in de praktijk dus niet met elkaar op gespannen voet te verkeren.

6 RELATIES

In de voorafgaande paragraaf is naar voren gekomen dat er op de onderzochte scholen verschillend wordt geoordeeld over de resultaten van het inspectiewerk. Er is echter nog geen antwoord gegeven op de vraag in hoeverre deze verschillen afhankelijk zijn van een anderssoortig contactprofiel tussen inspectie en school of van bepaalde kenmerken van de scholen. De beantwoording van deze vraag komt in deze paragraaf aan de orde. Daartoe is om

te beginnen de inspectorale invloed op de eerder genoemde veertien facetten van het schoolleven teruggebracht tot één concept: de algehele invloed van de inspecteur op de gang van zaken op een school. Statistische analyses wezen uit dat er tegen een dergelijke samenvoeging van invloedssitems geen bezwaar bestond. De Cronbach alpha voor deze items bedraagt 0,87, hetgeen zonder meer acceptabel is. De aldus geconstrueerde invloedsschaal heeft een bereik van 0 tot en met 10. Eenzelfde werkwijze is toegepast op de acht waarderingsitems. Ook hier wezen schaalanalyses uit dat de items zonder enig bezwaar konden worden samengevoegd. De Cronbach alpha bedroeg in dit geval 0,83. De waarderingsschaal heeft eveneens schaalwaarden die variëren van 0 tot 10.

Met behulp van de aan elke school toegekende invloed- en waarderingsscore is het nu mogelijk meer eenduidige vergelijkingen te maken tussen de contactprofielen en de schooltypen. Deze scores staan vermeld in de tweede en vierde kolom van tabel 3.

De invloedsscore die de inspectie krijgt toebedeeld blijkt vrij laag te zijn. Een maximale waarde van 3,6 is immers betrekkelijk gering als we in ogenschouw nemen dat de invloedsscore in theorie kan oplopen tot 10. Voor de waardering voor het functioneren van de inspecteur komt een iets rooskleurig beeld naar voren. De maximale waarde is nu 6,0 wat in ieder geval boven het theoretisch midden van 5 ligt.

Uit tabel 3 blijkt verder dat, indien de inspecteur een intensief contact onderhoudt met de school, hij kan rekenen op een relatief hoge waardering voor zijn werk. Bovendien wordt hem dan relatief veel invloed toegeschreven. Minimaal en examen contact gaan samen met betrekkelijk weinig invloed en waardering. Deze uitkomsten baren weinig opzien. Wat wel opvallend is zijn de verschillen tussen intensief en frequent contact. Er bestaat een zekere afstand tussen de scores voor deze twee contactprofielen. Op grond hiervan zou kunnen worden gesteld dat indien het Rijksschooltoezicht haar invloedsspositie op en haar relatie met de scholen wil verbeteren, zij de scholen eerder intensiever dan alleen maar vaker zou moeten bezoeken.

Lettend op het schooltype vinden we de hoogste invloed- en waarderingsscores voor de scholen voor LBO. Ordenen we de schooltypen op basis van de toegekende score, dan valt op dat we voor de gepercipieerde invloed een andere volgorde van schooltypen vinden dan voor de uitgesproken waardering voor het inspectiewezen. Zo wordt op de rijksschool relatief veel invloed aan de inspecteur toegekend, terwijl de waardering voor zijn functioneren naar

Tabel 3. Resultaten van het inspecteren naar contactprofiel en schooltype

	INVLOED		WAARDERING	
	a	b	a	b
CONTACTPROFIEL				
intensief contact	3.4	3.0	6.0	5.9
frequent contact	2.6	2.6	4.9	5.0
modaal contact	2.9	2.4	5.0	4.8
minimaal contact	1.9	2.1	4.1	4.2
examen contact	1.7	2.1	4.2	4.5
F-ratio (p-waarde)	.00	n.s.	.00	.00
SCHOOLTYPE				
bijzondere school LBO	3.6	3.5	5.3	5.2
rijks scholengemeenschap AVO	2.8	2.6	4.5	4.3
gemeentelijke categoriale AVO	2.4	2.5	4.0	4.1
bijzondere scholengem. AVO-LBO	2.4	2.3	5.0	4.9
gemeentelijke scholengem. AVO	1.9	2.0	4.1	4.2
bijzondere categoriale AVO	1.6	1.6	4.7	4.8
bijzondere scholengem. AVO	1.3	1.4	4.1	4.2
F-ratio (p-waarde)	.00	.00	.00	.01

(a = ongecorrigeerde score; b = gecorrigeerde score)

verhouding wat lager wordt ingeschat. Het omgekeerde geldt voor de scholengemeenschappen AVO-LBO: een relatief hoge waardering gaat hier samen met een verhoudingsgewijs wat slechtere invloedsspositie.

Een vraag die restert is de volgende: is het resultaat dat de inspectie boekt (uitgedrukt in een invloeds- en waarderingsscore) nu een gevolg van een differentiële inzet van middelen of juist van specifieke kenmerken van de te inspecteren eenheden? Het antwoord wordt gegeven in de derde en vijfde kolom van tabel 3. In deze kolommen staan de gecorrigeerde scores vermeld die verkregen zijn na het uitvoeren van een twee-wegs variantieanalyse. Na controle voor de verschillen in schooltype blijkt er geen (statistisch) significant verschil meer te bestaan tussen de verschillende invloedsscores per contactprofiel. Anders gezegd, het is niet het verschil in de aard van het contact tussen school en inspectie dat verantwoordelijk is voor het bestaan van verschillen in de aan de inspecteur toegekende invloed, maar het is het verschil tussen de verschillende schooltypen dat hiervoor verantwoordelijk is. Weer anders gezegd: de inspectie heeft niet op alle schooltypen een gelijke invloed. Deze verschillen in invloed blijven bovendien bestaan als we de verschillende contactprofielen in het verhaal betrekken. De inzet van beleidsmiddelen is niet van dien aard dat de verschillen in invloedsspositie wegvallen. Samenvattend: de invloed die aan de inspecteur wordt toegekend is gerelateerd aan kenmerken van de school en niet aan kenmerken van het contact tussen school en inspectie.

In een voorgaande paragraaf is opgemerkt dat het bestaan van fusieplannen of in het verleden voorgekomen wijzigingen in de schoolstructuur kunnen worden gerekend tot die schoolkenmerken die een rol zouden kunnen spelen bij de nadere vormgeving van de relatie tussen inspectie en school. Uit het nogmaals uitvoeren van de eerder genoemde twee-wegs variantieanalyse, nu echter met opname van deze twee variabelen als covariaat, bleek dat er geen noemenswaardige verschillen optreden. Beide schoolkenmerken leveren geen significante bijdrage in de verklaring van de verschillen in inspectorale invloed. Het is nog steeds het schooltype dat moet worden gezien als determinant voor deze invloed, niet de onstabiliteit van de te inspecteren scholen.

De tweewegs variantie-analyse met het contactprofiel en het schooltype als onafhankelijke variabelen en met de waarderingsscores als afhankelijke variabele levert een ander beeld op dan bij de analyse van de invloed naar voren komt. We verwijzen opnieuw naar de vijfde kolom van tabel 3. Zowel het schooltype als de aard van het contact is nu van invloed op de verschillen in waardering. Beide effecten zijn significant en van een interactie-effect is (evenals bij de analyse van de invloedsscores overigens) geen sprake. Opnieuw heeft het opnemen van de twee covariaten (fusieplannen en wijzigingen) geen invloed op de reeds vermelde conclusies. Aan de inspectie wordt de hoogste waardering toegekend door de directeuren van scholen voor LBO en door de schoolleiders van scholen die door de inspecteur intensief worden bezocht. Een differentiële inzet van beleidsinstrumenten is ten dele verantwoordelijk voor een verschil in de resultaten van de inspectie-arbeid.

NOTEN

1. Leesvoorbeeld: op 38% van de rijksscholengemeenschappen AVO is er sprake van frequent contact tussen inspectie en school.
2. Leesvoorbeeld: 68% van de schoolleiders van rijksscholengemeenschappen AVO ervaart (zeer) veel invloed van de rijksinspectie inzake de benoeming van schoolleiders.
3. Leesvoorbeeld: 63% van de schoolleiders van rijksscholengemeenschappen voor AVO is het (zeer) eens met de uitspraak dat de inspecteur op effectieve wijze toeziet op de naleving van de voorschriften.

LITERATUUR

- Ahlers, J. (1981). Het geheim van de inspectie. *School*, 9(5), 13-28.
 Algemene rekenkamer (1987). Inspectie van het onderwijs. In: *Tweede Kamer der Staten-Generaal*, Vergaderjaar 1986-1987, Bijlage 19978, Nr.1-2.

- Berg, G.M. van den (1986). *Homals voor beginners*. Leiden: Vakgroep Datatheorie.
- Bijdevaate, D. (m.m.v. J.F.A. Braster) (1986). *Het functioneren van de inspecteur van het voortgezet onderwijs*. Rotterdam: Vakgroep Onderwijssociologie en Onderwijsbeleid (Doctoraalscriptie).
- Braster, J.F.A. (1987). *De inspecteur en de school*. Rotterdam: Vakgroep Onderwijssociologie en Onderwijsbeleid.
- Dekker, A. & C. Gijtenbeek (1975). *De evaluerende functie van het rijksschooltoezicht op microniveau*. Utrecht.
- Elte, R. (1988). *'Van het toezicht op het onderwijs'*. De Lier: Academisch Boeken Centrum.
- Gifi, A. (1980). *Niet lineaire multivariate analyse*. Leiden: Vakgroep Datatheorie.
- Leune, J.M.G. (1987). Besluitvorming en machtsverhoudingen in het Nederlandse onderwijsbestel. In: J.A. van Kemenade e.a. (red.), *Onderwijs: bestel en beleid. Deel 2. Onderwijs en samenleving*. Groningen: Wolters-Noordhoff.
- Mentink, D. & P.W.C. Akkermans (1986). *Deugdelijk onderwijsbeleid*. Den Haag: SVO.
- Ministerie van Onderwijs en Wetenschappen (1984). *De inspectie van het onderwijs in Nederland*. (Docinform nr. 328). Zoetermeer.

Manuscript ontvangen 4-7-1988

Definitieve versie ontvangen 18-10-1988

Evaluating an Item and an Option Statistic Using the Bootstrap Method

Dato N.M. de Gruijter

University of Leiden

ABSTRACT

Item statistics like p and r_{it} are useful for the detection of flaws in items or in test administration. In the interpretation of these indices the magnitude of their sampling fluctuations should not be ignored. Little is known, however, about sampling properties of r_{it} and similar coefficients. The bootstrap might, therefore, be a useful tool. The bootstrap method is demonstrated in connection with a corrected r_{it} and an option index.

INTRODUCTION

Item statistics like p , the item proportion correct, and r_{it} , the item-total correlation, or similar statistics are used frequently, together with information on the incorrect item options, to detect errors in multiple-choice items. These errors might be errors in the formulation of the item, in the relation between the item and previous teaching or in the administration and correction of the test. The latter is the case when there is an error in the scoring key. This error shows in a relatively low p in combination with a r_{it} close to zero or negative and a high proportion of relatively good scoring examinees on a distractor or incorrect option.

The interpretation of the item indices should take sampling fluctuations into account. With a small group of examinees p might be relatively low by chance.

The sampling distribution of p presents no problems as will be argued in a subsequent section. Coefficients like r_{it} and corrected r_{it} do. Perry and Michael (1958) derived the sampling variance of the point biserial correlation using an assumption against which - as they admit - valid objections can be raised. Other authors have based their suggestions on studies in which the reliability of r_{it} is computed (see, for example, De Groot & Van Naerssen, 1969, p. 256). As recently as 1987 reliability of item indices like r_{it} has been used in the literature (Van Berkel, 1987). Two objections can be raised against the use of reliabilities. First, reliability depends on true variation between indices and error variance. The true variation can differ from application to application and, consequently, it is difficult to imagine how general guide-lines can be formulated. Secondly, the relevance of reliability as a criterion is doubtful. Reliability is relevant when the items in the test are rank ordered on the basis of r_{it} -values. Frequently, however, the researcher has other purposes in mind. A more obvious concern is the question how small the population value of r_{it} is. In such a context the variance of sampling error is of prime interest.

In case the sampling distribution of a statistic is unknown, the researcher has to resort to a sampling strategy. The bootstrap is a tool for the construction of samples from the available data. With the bootstrap, statistics like r_{it} can be recomputed a number of times, resulting in some indication of the variability of this index. The bootstrap based on sampling of examinees from the total group which took the test, is restricted to the items included in the test. Generalization of results to tests composed of other items is not warranted.

This study is concerned with the use of the bootstrap in connection with a corrected item-total correlation and an option index. First these indices are introduced. Next, the bootstrap is discussed. Finally, an illustrative example is presented.

A CORRECTED r_{it}

The r_{it} has a spuriously high value: the item is included in the total test. Therefore, alternatives have been suggested. One of these alternatives is the item-rest correlation, r_{ir} . Index r_{ir} has another disadvantage, however, as each item is correlated with a slightly different test. We decided to use a corrected item-total coefficient, r'_{it} , instead. We opted for a coefficient proposed by Henrysson (1963), except for the correction for attenuation which sets this coefficient apart from other coefficients (Berk, 1978) and adds a factor to its variability. Another possibility would have been to use Cureton's concept of a rationally equivalent item (Cureton, 1966).

First, notice that for dichotomous items the item-total correlation can be written as

$$r_{it} = \frac{s_i^2 + (n-1)\bar{c}_i}{s_i s_x} = \frac{(\bar{x}_{i+} - \bar{x})}{s_x} p_i^{1/2} / (1-p_i)^{1/2}, \quad (1)$$

with

- n = the number of test items
- s_i = $p_i^{1/2} (1-p_i)^{1/2}$ = the standard deviation of scores on item i
- s_x = the standard deviation of total test scores
- $\bar{c}_i$ = the average covariance between item i and the other items
- $\bar{x}$ = the average total score
- $\bar{x}_{i+}$ = the total score averaged over the examinees with a correct response to the item.

The spuriousness in r_{it} is due to the presence of s_i^2 in the numerator of Equation 1. Henrysson replaced s_i^2 by $\bar{c}_i$, which is an estimate of the covariance between the item and the universe score in generalizability theory (Cronbach, Rajaratnam & Gleser, 1963).

The corrected item-total correlation, r'_{it} , can be written as

$$\begin{aligned} r'_{it} &= \frac{n\bar{c}_i}{s_i s_x} \\ &= \frac{n}{n-1} \left(\frac{s_i s_x r_{it} - s_i^2}{s_i s_x} \right) \\ &= \frac{n}{n-1} \left[\frac{(\bar{x}_{i+} - 1) - (\bar{x} - p_i)}{s_x} \right] p_i^{1/2} / (1-p_i)^{1/2}. \end{aligned} \quad (2)$$

Equation 2 differs from Equation 1 in that $\bar{x}_{i+} - 1$ is used instead of $\bar{x}_{i+}$ and $\bar{x} - p_i$ instead of $\bar{x}$, and in the multiplicative factor $n/(n-1)$. When the average score on the remaining items in the group of correctly responding examinees exceeds the average rest score in the total group, r'_{it} is positive.

Equation 2 differs from the corrected item-total correlation coefficient proposed by Henrysson in that Henrysson multiplies s_x in the denominator of the equation by the square-root of the KR-20. In other words, Henrysson's coefficient corrects for attenuation in the total score as well as for spuriousness. Ideally, his coefficient does not depend on test length.

AN OPTION INDEX

From the introduction it is clear that information on distractors can be useful in addition to item indices like p , and r_{it} or similar coefficients. One of the option indices is the a -value of an option, the proportion of persons endorsing that option. The a -value of a distractor is comparable to p for the correct option.

Other relevant information concerns the extent to which high scoring examinees endorse a

particular incorrect option. Some authors have suggested an alternative-total correlation, r_{at} , analogous to r_{it} for the correct option (Van Hees, 1985). Other authors propose to compute the average score for those examinees who endorsed a particular option a on item i , $\bar{x}_{ia}$, and to compare the resulting values (Dousma & Horsten, 1980, p. 159-160). These two approaches recur in the literature on option weighting (Bejar & Weiss, 1977). In maxalpha weighting, for example, weights are chosen which maximize coefficient alpha (Serlin & Kaiser, 1978). The weights in this approach, also known as homogeneity analysis or the method of reciprocal averages, can be determined iteratively from (weighted) averaged scores of examinees who endorse the various options.

Standardization of the option index can be helpful in the interpretation of the size of the averages and their differences. We propose to correct for spuriousness at the same time, comparable to the correction of r_{it} . The uncorrected standardized option mean can be written as:

$$\frac{(\bar{x}'_{ia} - \bar{x} + p_i + d_{ia})}{s_x}$$

with

$$d_{ia} = l - p_i \text{ if the option is correct} \\ = -p_i \text{ otherwise}$$

and

$$\bar{x}'_{ia} = \text{the average restscore of the examinees endorsing the option.}$$

Now, the contribution of the item to the difference between the option mean and the overall mean, d_{ia} , is replaced by the average contribution of the remaining $n-1$ items. We obtain the option index

$$t_{ia} = \frac{n}{n-1} \left(\frac{\bar{x}'_{ia} - \bar{x} + p_i}{s_x} \right). \quad (3)$$

There is a relation between t_{ia} for the correct option, t_{i+} , and the corrected item-total correlation from Equation 2:

$$r'_{it} = t_{i+} p_i^{1/2} / (1-p_i)^{1/2}. \quad (4)$$

A similar relation is obtained between the corrected option-total correlation and t_{ia} for incorrect options, using the a -values of the options. When t_{ia} for an incorrect option exceeds the item's value t_{i+} , one knows that that particular option has been chosen by examinees who are better on the average than the examinees who endorsed the correct option. When r'_{at} exceeds r'_{it} , the a -value of the option and p_i have also to be taken into account. Further, when two distractors have identical values t_{ia} , the most extreme r'_{at} is obtained for the distractor with the highest a -value, as can be seen from Equation 4.

Clearly, corrected option-total correlations r'_{at} and indices t_{ia} can rank options differently. The ordering of the options on the basis of option locations t_{ia} seems more natural than the ordering on the basis of option correlations r'_{at} .

THE BOOTSTRAP

Assume that we have N observations, which can be considered as a random sample of size N from an unknown probability distribution. The bootstrap assumes that the distribution obtained in the sample does not differ too much from this probability distribution. The empirically obtained distribution is then chosen to play the role of the probability distribution. A number of, say, k new samples of size N is obtained by random sampling with replacement

from the sample. The samples are used for inference with respect to sampling variability (Efron, 1979). A reasonably large value of k should be chosen when an accurate estimate of the sampling variance is in order.

One of the examples given by Efron (1979) to illustrate the bootstrap method concerns the binomial distribution with parameter π . A random sample of size N gives a point estimate p of π . When the bootstrap is used, samples of size N are generated from the binomial distribution with parameter p . The bootstrap sampling distribution has mean equal to p and variance equal to $p(1-p)/n$. Standard sampling theory suggests the same variance as an estimate of sampling variability.

In the binomial case more accurate results can be obtained. It is possible to construct exact confidence intervals (Pearson & Hartley, 1970). Application of these intervals in connection with p in multiple choice tests has been suggested by De Gruijter (1982). Doing so, no attention is paid to the underlying item response model. This is admissible, for the examinees are not fixed over replications but sampled; replications with both examinees and items fixed are not feasible in practice and not possible on a theoretical basis only (cf. the distinction between the weak and strong assumption of local independence; De Gruijter & Van der Kamp, 1984, p. 19). To be sure, covariances between p -values depend on the underlying response processes.

In many other cases exact sampling distributions are unknown and then one has to resort to techniques like the bootstrap in order to estimate sampling variability. Such is the case with indices r'_{it} and t_{ia} in the present study. However, the bootstrap is no panacea. The bootstrap is, for example, not a remedy against small samples: it is based on approximations which increase in accuracy with increasing sample size. Unfortunately, information on variability is needed most when sample sizes are small. In the present study, where the variability of r'_{it} and t_{ia} is studied, this is especially the case with t_{ia} . Some options might be chosen by only a few examinees and information on the corresponding values t_{ia} is inherently weak. In case of an extreme value p also information on r'_{it} is weak. In the demonstration which follows, we will use the bootstrap only in a informal way, with a special eye on the sample size problem.

AN EXAMPLE

The data for the illustration of the bootstrap with r'_{it} and t_{ia} are from a test of thirty four-choice items, administered to a group of 151 examinees. Resampling with replacement was based on this group of examinees. One hundred samples of 151 were drawn, and for all samples r'_{it} and t_{ia} were recomputed. The standard errors for r'_{it} varied between .06 and .09. Approximate confidence intervals $\bar{x}_r \pm 1.96s_r$, where $\bar{x}_r$ is the average bootstrap value of r'_{it} and s_r its standard deviation, are given in Figure 1. The intervals are represented by vertical line segments. The bootstrap results indicate that items 8 and 18, both of which had negative values r'_{it} in the original sample, might have a (small) positive true value.

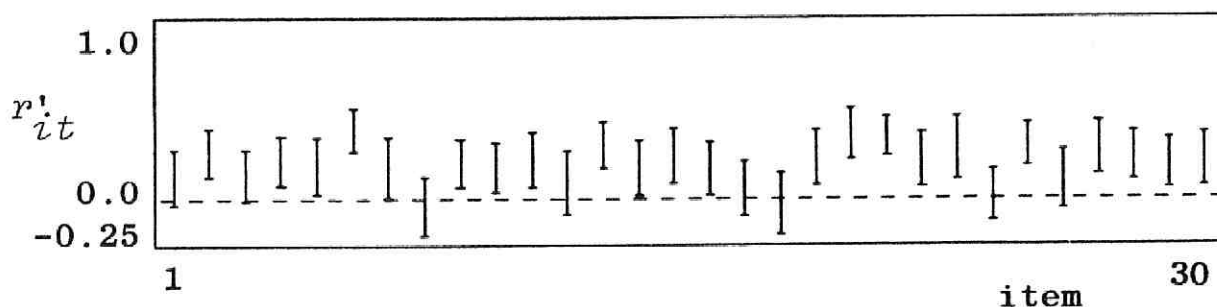


Figure 1. Approximate confidence intervals for r'_{it} .

With the given sample size no strong conclusions can be drawn: the confidence intervals are too large in general. A similar conclusion can be obtained on the basis of the results given by Perry and Michael (1958): a true point-biserial equal to .3 in combination with a proportion correct equal to .6 (or .4) gives a standard error equal to .07, with sample size 151.

The bootstrap standard error s_r is subject to sampling fluctuations, and the size of these fluctuations depends on the number of replications k . The entries in the table presented by Perry and Michael can be useful in choosing k . One might start with a first estimate of the standard error of r'_{it} for a particular item on the basis of the table. Next, assuming a normal distribution for resample values r'_{it} , a value k can be obtained which is likely to result in a suitable standard error.

Figure 2 gives the bootstrap results for the values t_{2a} of the four options of item 2 based on 10 bootstrap values. Even such a small number of replications can give us a feel of the variation to be expected. The presentation is easily interpreted, also by persons who are not accustomed to statistics. It is similar to the presentation of bootstrap results of a homogeneity analysis presented in Gifi (1981; Gifi also presents some interesting asymptotic results on error variances).

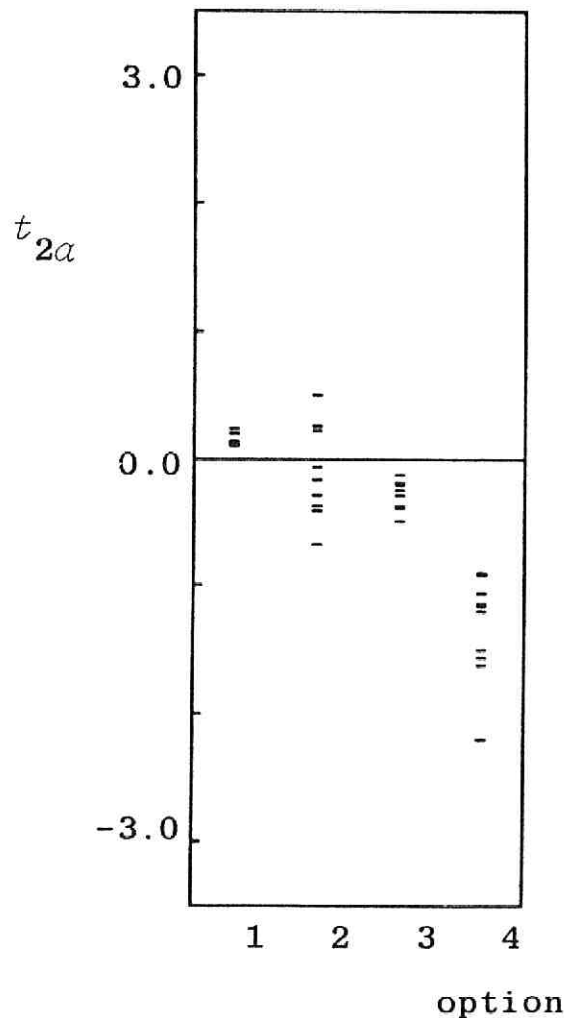


Figure 2. Bootstrap values t_{ia} for the options of item 2.

The variation of t_{2a} in Figure 2 obviously differs from option to option. The correct option, option 1, attracted most examinees. It gives the most stable results. The fourth option shows the largest variation. This option attracted only a few examinees. In this case the dependency of

variation on group size is clear. However, one should be careful with the interpretation of the bootstrap result for an option when the number of examinees endorsing that option in the original sample, N_{ia} , is small. Consider the extreme case with an option chosen by a single examinee. All bootstrap samples with a number of examinees for this option larger than zero, are based on one single total score. The true variation in t_{ia} is seriously underestimated in this particular case.

Figure 3 gives the bootstrap results for item 18, one of the two items with a negative r'_{it} . Option one, the correct option, is chosen by the largest group of examinees. The variation in t_{18a} is smallest for this option. The average bootstrap value of $t_{18,1}$ is close to zero.

Incorrect option 2 has been chosen by a relatively large number of examinees. The value of $t_{18,2}$ was positive, indicating that this option was chosen by examinees who were better than average in terms of total test score. The bootstrap shows that the result is stable: $t_{18,2}$ is positive for all bootstrap samples. This result supports the interpretation that this option is attractive to relatively high scoring examinees.

Figure 3

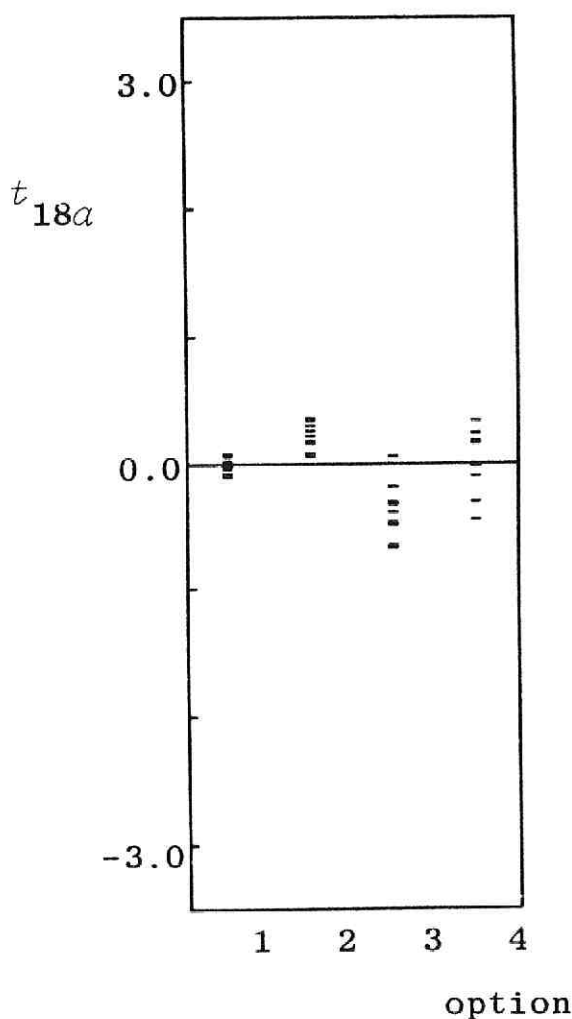
Figure 3. Bootstrap values t_{ia} for the options of item 18.

Figure 4 gives an overview of bootstrap standard deviations or standard errors associated with options for which N_{ia} exceeds 4. The standard errors are based on $k=100$. There is a strong dependency of the size of the standard error on N_{ia} . According to the expectation, large values of N_{ia} correspond to small bootstrap standard deviations. For relatively small values of N_{ia} a considerable amount of variation remains after the value of N_{ia} has been taken into account.

This variation has two possible sources. First, notice that the standard error depends on the variance of scores associated with examinees who endorse a particular option, and on N_{ia} . The score variance can vary from option to option, and so can the standard error. The resulting variation is largest in absolute terms for small N_{ia} . Secondly, the standard errors are sensitive to bootstrap sampling fluctuation. This variation is larger too, when standard errors are large. In principle, the problem of inaccurately estimated standard errors could be alleviated by enlarging k . It is doubtful, however, whether this is a fruitful approach in connection with very low values of N_{ia} . In this case reliance on the bootstrap approximation can be unwarranted.

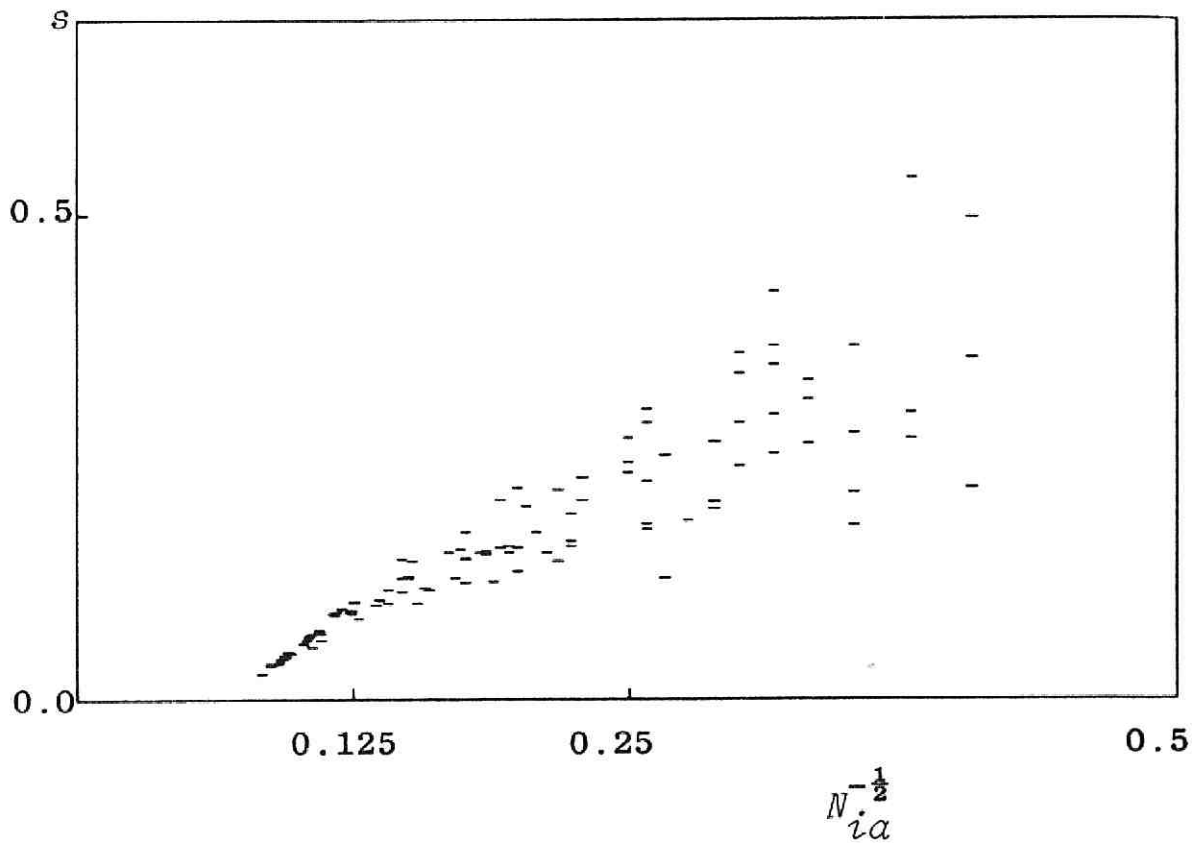


Figure 4. Bootstrap standard deviations for t_{ia} as a function of $N_{ia}^{-1/2}$.

DISCUSSION

This study has demonstrated the possibilities for the application of the bootstrap with item and option indices. It has been pointed out that bootstrap results can be unreliable with small examinee groups. This is true especially for item options chosen by only a few examinees. In the illustrative example the bootstrap has been used more as a descriptive tool than as a technique to construct exact confidence intervals.

REFERENCES

- Bejar, I.I., & Weiss, D.J. (1977). A comparison of empirical differential option weighting scoring procedures as a function of inter-item correlation. *Educational and Psychological Measurement*, 37, 335-340.
- Berk, R.A. (1978). Empirical evaluation of formulae for correction of item-total point-biserial correlations. *Educational and Psychological Measurement*, 38, 647-652.

- Cronbach, L.J., Rajaratnam, N., & Gleser, G.C. (1963). Theory of generalizability: a liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137-163.
- Cureton, E.E. (1966). Corrected item-test correlations. *Psychometrika*, 31, 93-96.
- De Groot, A.D., & Van Naerssen, R.F. (1969). *Studietoetsen*. Den Haag: Mouton.
- De Gruijter, D.N.M. (1982). *Tentamineren en beslissen*. Den Haag: SVO-reeks, nr. 63.
- De Gruijter, D.N.M. de, & Van der Kamp, L.J.Th. (1984). *Statistical models in psychological and educational testing*. Lisse: Swets & Zeitlinger.
- Dousma, T., & Horsten, A. (1980). *Tentamineren*. Utrecht: Het Spectrum.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1-26.
- Gifi, A. (1981). *Nonlinear multivariate analysis*. Leiden: Department of Data Theory, Faculty of Social Sciences.
- Henrysson, S. (1963). Correction of item-total correlations in item analysis. *Psychometrika*, 28, 211-218.
- Pearson, E.S., & Hartley, H.O. (1970). *Biometrika tables for statisticians Vol. I*. Cambridge: University Press (3rd ed.).
- Perry, N.C., & Michael, W.B. (1958). A note concerning the reliability of a point biserial coefficient for large samples. *Educational and Psychological Measurement*, 18, 139-143.
- Serlin, R.C., & Kaiser, H.F. (1978). A method for increasing the reliability of a short multiple-choice test. *Educational and Psychological Measurement*, 38, 337-340.
- Van Berkel, H.J.M. (1987). De betrouwbaarheid van drie discriminatie-indices. *Tijdschrift voor Onderwijs-research*, 12, 365-368.
- Van Hees, E.J.W.M. (1985). Toetsverwerking en - analyse. In E.J.W.M. van Hees, G.A. Vonk & J.M. Schoenmaker. (Red.), *Handboek voor Onderwijs en Computer*. Alphen a/d Rijn: Samsom.

Manuscript ontvangen 19-5-1988

Definitieve versie ontvangen 17-11-1988

Notities en Commentaren

A FURTHER NOTE ON THE NECESSITY OF INCLUDING A RETROSPECTIVE PRE-TEST IN SELF-REPORT PRETEST-POSTTEST DESIGNS TO DETECT TRAINING EFFECTIVENESS

Mirjam Sprangers
University of Amsterdam

ABSTRACT

The present study compared the sensitivity of the posttest minus retrospective pretest (thentest) difference scores versus the posttest minus pretest difference scores in detecting a training effect for a text studying program. Psychology freshmen ($N = 131$), fulfilling a course obligation, were randomly assigned to either the experimental text studying training or the control training. All subjects filled out a performance task at pre- and posttraining sessions. In addition, subjects filled out a self-report instrument as a pretest, posttest, and thentest. Results showed that the text studying training was successful: experimental subjects improved their note taking ability significantly. With respect to the self-report measures, a response-shift was found. In addition, posttest minus thentest difference scores proved to legitimately document the training effectiveness that the posttest minus pretest difference scores failed to find.

Attempts to eliminate the detrimental effects of response-shifts in studies employing pretest-posttest self-report indices of improvement, included retrospective reports of pre-training states (thentests) (Howard, Ralph, Gulanick, Maxwell, Nance, & Gerber, 1979). At the posttesting subjects are asked how they then perceive themselves to have been prior to the training. This is done in order to assure that both ratings - post- and thenratings - are made with respect to the same evaluation standard, thus attenuating response-shift bias effects. A considerable amount of studies favored the posttest minus thentest difference scores in providing a more accurate estimate of a training effect, while the posttest minus pretest difference scores often mask the training effect (Howard, Schmeck, & Bray, 1979; Bray & Howard, 1980; Pohl, 1982; Hoogstraten, 1982, 1985; Perry & Apostol, 1986; Porras & Singh, 1986).

The present study compared the sensitivity of the posttest minus thentest difference scores versus the posttest minus pretest difference scores in detecting a training effect for a text studying program.

METHOD

Subjects were psychology freshmen at the University of Amsterdam who were fulfilling a course requirement. They were randomly assigned to either an experimental condition ($n = 65$) or a control condition ($n = 66$).

The *experimental* training consisted of a 15-page booklet on a method to study texts, called

This study was supported by grants from the Foundation for Educational Research (SVO; project-1107). I am indebted to Pieter Emeis, Marian Koeman, Ed Lebert, and Irene Guerand for their contribution in carrying out the experiment presented in this paper. Requests for reprints should be sent to Mirjam Sprangers,
Address: Department of Psychology, Weesperplein 8, 1018 XA Amsterdam, The Netherlands.

the 'topic-comment technique' (adapted from chapter three of Elshout-Mohr, 1984). This information processing technique distinguishes between topics and comments by following a stepwise procedure. Subjects practiced the technique on several texts. This training took one hour. The *control* treatment consisted of 13 summaries of research procedures for each of which subjects had to give their opinion on the ethical permissibility. The control treatment also took one hour.

The *Self-report* instrument consisted of 27 items on subjects' perception of their actual knowledge with regard to the experimental training topics and their perception of how well they were able to take notes from an article. Subjects responded to 14 items on a 7-point scale on which 1 referred to 'not at all applicable to me' and 7 to 'completely applicable to me'. For 13 items these labels were reversed. Two examples are: 'I am able to give a clear and concise summary of a study text' and 'I know how to decide which topics are important in an intricate study text'. After deletion of 4 items with a moderate item-rest correlation, Cronbach's alpha for the remaining 23 items was .88 at the pretest (minimumscore 23, maximumscore 161).

The *performance* instrument consisted of a note taking task that invited subjects to take notes from a two page study text on social psychology. The notes were rated according to a rating system (Elshout-Mohr, Van Daalen-Kapteijns, & Sprangers, 1988) that distinguished three criteria, for each of which subjects could receive maximally 4 points (minimumscore 0, maximumscore 12). All notes were scored independently in a fixed random order by two senior undergraduates. Each of four raters rated half of all notes. The interrater reliability of pre- and posttraining notes was calculated for experimental and control subjects separately. The interrater reliability ranged from .79 to .90.

All subjects first completed the self-report and the note taking task as a pretest. After training, subjects filled out the note taking task as a posttest. Finally, the self-report instrument was administered as a posttest and as a thentest. The resulting design is displayed in Table 1.

Table 1: Experimental design

Condition	<i>n</i>	Self-report pre	Performance pre	Treatment	Performance post	Self-report post	Self-report then
1	65	yes	yes	experimental	yes	yes	yes
2	66	yes	yes	control	yes	yes	yes

RESULTS AND DISCUSSION

The means and standard deviations of the self-report pretest, posttest, and thentest, and of the pre- and posttraining performance task (note taking task) are displayed in Table 2.

With respect to the performance task, ANCOVA using prescores as covariate indicated a significant treatment effect ($F(1,128) = 77.86, p < .001$). Since experimental subjects did and control subjects did not improve their note taking ability significantly, it must be concluded that the training produced the desired result.

With respect to the self-report measures, the pretest minus thentest difference scores reached significance in the experimental condition ($t(63) = 5.25, p < .001$)¹ and not in the control condition ($t(65) = 1.47, p = .08$). It can be concluded that a response-shift occurred.

It is apparent that the posttest minus pretest difference scores of experimental subjects indicated a decline, while the posttest minus thentest difference scores indicated a positive and

¹ One experimental subject failed to fill out the self-report pretest due to late arrival at the training session.

Table 2: Means and Standard Deviations for Self-report Measures and Performance Task

Condition	n		Self-report			Performance	
			Pre	Post	Then	Pre	Post
1	65	M	110.63	108.00	103.52	4.52	8.33
		SD	14.81	17.87	17.95	2.04	2.68
2	66	M	111.82	110.77	110.47	5.27	5.90
		SD	15.51	17.31	17.98	1.73	3.29

significant training result ($t(64) = 3.48, p = .001$). As expected, posttest minus thentest difference scores of the control subjects did not differ significantly ($t(65) = .39$). It can be concluded that posttest minus thentest difference scores provided a similar result as objective indices of change, while the posttest minus pretest difference scores did not.

Again, it has been demonstrated that when self-report measures are used in a conventional pre-post manner to evaluate a training intervention, the results may be confounded by response-shift bias. In addition, the posttest minus thentest difference scores proved to legitimately document the training effectiveness that the posttest minus pretest difference scores failed to find. Formal evaluations suggest that study skills programs are of questionable worth, despite their prevalence and face validity (Conway & Ross, 1984). However, the existent research on the study skills training effectiveness which relies upon conventional pre-post self-report measures, may contain response-shift bias and consequently, finds results that are erroneously conservative.

REFERENCES

- Bray, J. H., & Howard, G. S. (1980). Methodological considerations in the evaluation of a teacher-training program. *Journal of Educational Psychology, 72*, 62-70.
- Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology, 47*, 738-748.
- Elshout-Mohr, M. (1984). *Dikke Boeken Weinig Tijd*, Coutinho, Muiderberg.
- Elshout-Mohr, M., Van Daalen-Kapteijns, M., & Sprangers, M. (1988). The topic-comment technique to study expository text. *Journal of Experimental Education, 56*, 83-90.
- Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education, 50*, 200-204.
- Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education, 53*, 207-210.
- Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement 3*, 1-23.
- Howard, G. S., Schmeck, R. R., & Bray, J. H. (1979). Internal invalidity in studies employing self-report instruments: A suggested remedy. *Journal of Educational Measurement, 16*, 129-135.
- Perry, D. C., & Apostol, R. A. (1986). Modifying attitudes of business leaders toward disabled persons. *Journal of Rehabilitation, 52*, 35-38.
- Pohl, M. F. (1982). Using retrospective pre-ratings to counteract response-shift confounding. *Journal of Experimental Education, 50*, 211-214.
- Porrás, J. I., & Singh, J. V. (1986). Alpha, beta, and gamma change in modelling-based organization development. *Journal of Occupational Behaviour, 7*, 9-24.

Boekbesprekingen

J. Neuvel, M.E. Otter en D.J. Bos

Functioneel aanvankelijk leesonderwijs. Een beschrijving en een evaluatie
Amsterdam/Lisse: Swets & Zeitlinger, 1988. SVO-reeks 86. VIII + 230 pp, f 47.50

In vergelijking met bijvoorbeeld Engeland en Amerika is in Nederland de discussie over de beste manier om kinderen te leren lezen nauwelijks van de grond gekomen. Al jaren beheerst één methode, Veilig leren lezen, de markt. De meest recente concurrent, Letterstad, lijkt qua methodiek vrij veel op Veilig leren lezen. Toch hoort men regelmatig geluiden waarin om een vernieuwing van het leesonderwijs wordt gevraagd. Eén van de vernieuwingsstromingen betreft het zogenaamde functioneel aanvankelijk leesonderwijs. De Stichting Centrum voor Onderwijsonderzoek (SCO) te Amsterdam heeft op verdienstelijke wijze in het hier besproken boek een aantal van deze zgn. 'FAL'-methodes beschreven en geëvalueerd.

Het rapport gaat helemaal over FAL. Wat echter FAL precies is wordt niet duidelijk. Het is een verzamelnaam van alternatieven voor de gangbare didactiek van het aanvankelijk leesonderwijs. Op blz. 7 staat: (Het is) 'Aanvankelijk leesonderwijs waarin meer aandacht wordt besteed aan de betekenisverlening, waarin het leren decoderen ligt ingebed in communicatieve situaties'.

Na een inleidend hoofdstuk volgen drie hoofdstukken waarin een beschrijving van FAL wordt gegeven. In hoofdstuk 2 staan de uitgangspunten en de keuzes van FAL. Bij FAL is het doel van het (kunnen) lezen iets met de verkregen informatie te doen: ontspanning, bevrediging van nieuwsgierigheid, het kunnen installeren van een apparaat enz. (blz. 18). In de methodiek dient dit doel centraal te staan. Opmerkelijk is evenwel, dat het er sterk op lijkt dat dit doel niet het wezenlijke van FAL is, maar veeleer dat de gehanteerde methodiek dat is. Uitgangspunt van die methodiek is de taal en het denken (ervaringen) van de kinderen (blz. 26). De neerslag hiervan, in de vorm van een tekst, vormt het startpunt van de ontdekkingsstocht van het kind in de wereld van het lezen. Deze ontdekkingsstocht doet een groot beroep op het leervermogen van kinderen middels het zelfontdekkend leren en ook op de didactische kwaliteit van de leerkracht. Men zou zich in dit verband af kunnen vragen of in plaats van functioneel leren lezen (FAL) niet beter van ontdekkend leren lezen (OLL) gesproken kan worden.

Dit idee wordt versterkt door de beschrijving van een viertal buitenlandse varianten van FAL in hoofdstuk 3: twee Engelse (het programma van Doris Nash en 'Breakthrough to literacy'), één Franse (Freinet) en één Amerikaanse ('language experience approach' van o.a. Stauffer). Deze volgen in hoofdlijnen dezelfde methodische lijn, startend bij teksten van het kind zelf van waaruit in hoge mate middels het ontdekkend lezen de structuur van ons schriftsysteem wordt verkend. Ook in de beschrijving van de Nederlandse varianten van FAL in hoofdstuk 4 staat de notie van het ontdekkend leren lezen centraal. In dat hoofdstuk worden een drietal stromingen besproken: het Jenaplanonderwijs, het Freinetonderwijs en de Geïntegreerde lees-taalaanpak van het Regionaal Pedagogisch Centrum Zeeland (waarin overigens het werken met teksten niet het startpunt vormt). Degene die een beeld wil verkrijgen van de inhoud en de werkwijze van een aantal FAL-aanpakken, vindt in deze drie hoofdstukken, die zo'n 130 bladzijden beslaan, een systematische en heldere beschrijving.

Het tweede deel van het boek bestaat uit twee hoofdstukken waarin onderzoeksgegevens over de effectiviteit van FAL staan; hoofdstuk 5 vat het buitenlandse onderzoek samen en in hoofdstuk 6 wordt een onderzoek beschreven dat de SCO in Nederland heeft uitgevoerd.

Wat betreft de buitenlandse onderzoeken hangt het er een beetje vanaf hoe streng men is in methodologisch opzicht. Als men zeer strikt is, moet geconcludeerd worden dat er geen dergelijk onderzoek is verricht naar FAL. Als men iets minder strenge eisen stelt dan blijkt volgens de onderzoekers het volgende (blz. 151):

- Als er voldoende aandacht wordt besteed aan het decoderen, dan is FAL zeker niet minder effectief dan andere methoden.

- Als er te weinig aandacht wordt besteed aan het decoderen, dan lopen met name zwakke en kansarme leerlingen risico's.
- Op de langere termijn worden met FAL geen betere resultaten geboekt dan met andere methoden.
- Deze punten hebben betrekking op technisch en begrijpend lezen. Over enkele belangrijke doelstellingen van FAL, het stimuleren van de leesmotivatie en een goede leeshouding blijkt in de buitenlandse literatuur nauwelijks goed onderzoek gedaan te zijn.

Men kan hieruit konkluderen dat noch vóór, noch tegen FAL overtuigende onderzoeksargumenten zijn aan te dragen. De enige uitzondering vormt het belang van de aandacht voor het decoderen in de methode.

In het zesde hoofdstuk wordt het eigen onderzoek beschreven, dat de SCO in Nederland heeft verricht. Een pikant detail is dat een onderzoek van Bus & Kruizinga naar de leesdidactiek van een Vrije School géén onderzoek naar FAL genoemd wordt, omdat men daar begint met het aanleren van losse letters en pas later betekenisvolle taaleenheden (nl. bestaande teksten) aan de orde stelt. (Gaat het nu toch om 'ontdekkend leren' en minder om 'functioneel leren'?).

Dan nu het onderzoek. Uit een bestand van bijna 40 FAL-scholen worden de scholen geselecteerd die al een aantal jaren volgens FAL werken. Er blijven 'maar net tien scholen' over (blz. 175): vijf Jenaplan-scholen, die Freinet-scholen en twee scholen met Geïntegreerde leestaalpak van het RPC-Zeeland. Op deze scholen wordt de technische leesvaardigheid en het begrijpende lezen gemeten in groep 3 en 5 en vervolgens worden deze resultaten vergeleken met de resultaten van het project Preventie van leesmoeilijkheden (verder genoemd: PPL). Hoewel enigszins is nagegaan of de tien FAL-scholen op bepaalde punten afwijken van de PPL-scholen speelt hier natuurlijk levensgroot het bekende probleem: moeten de verschillen in leerprestaties tussen FAL- en PPL-scholen en -kinderen toegeschreven worden aan de methode óf aan andere zaken, zoals allerlei selectie-effecten. Een tweede beperking van het onderzoek is dat ook hier (om overigens goede redenen) alleen wordt gekeken naar het technisch en begrijpend lezen en niet naar leeshouding en leesmotivatie. De conclusies die de onderzoekers trekken zijn, mijns inziens, correct (blz. 195-196):

'Geconcludeerd kan worden, dat de leesontwikkeling van kinderen in de FAL-scholen uit dit onderzoek na één jaar leesonderwijs een achterstand vertoonde met de leesontwikkeling van kinderen die hebben leren lezen met Veilig Leren Lezen of Letterstad (de PPL-groep). Die achterstand deed zich echter hoofdzakelijk voor in FAL-aanpakken waarin relatief weinig cursorisch lesonderwijs werd gegeven en werd met name veroorzaakt door een grotere groep zeer trage lezers in die scholen.'

Er kan ook geconcludeerd worden dat de leesontwikkeling van de leerlingen in de 10 FAL-scholen aan het eind van groep vijf niet verschilde van die van leerlingen uit het PPL-onderzoek. Dit geldt zowel voor het technisch als het begrijpend lezen; het geldt in het algemeen, maar ook als alleen naar het percentage zwakke lezers in beide groepen wordt gekeken.

Kan nu ook geconcludeerd worden, dat de aanvankelijke achterstand in de leesontwikkeling ingelopen wordt in de loop van groep vier en vijf?

De resultaten wijzen in die richting. Echter de cross-sectionele opzet van het FAL-onderzoek laat zo'n conclusie niet volledig toe. In de eerste plaats is niet helemaal duidelijk in hoeverre de FAL-kinderen in groep drie en vijf vergelijkbaar waren, hoewel de resultaten op intelligentie-tests ook niet doen vermoeden dat er al te grote verschillen bestonden. In de tweede plaats is er het probleem van uitval van kinderen. In het FAL-onderzoek kon niet systematisch nagegaan worden hoeveel en welke kinderen in de onderzochte FAL-scholen, anders dan om redenen van verhuizing, naar een andere school zijn gegaan, met name naar het Speciaal Onderwijs.

Afsluitend kan men zeggen dat het boek een goed beeld verschaft van wat FAL is. Het doet ook verslag van buitenlandse en een eigen onderzoek naar FAL. Daar kleven nog al wat methodologische (methode-vergelijking) en inhoudelijke (de leeshouding en -motivatie worden niet meegenomen) beperkingen aan. Het pleidooi om bij het werken met FAL veel aandacht te hebben voor de pedagogisch-didactische bekwaamheid van de leerkracht (b.v. in de vorm van

het 'diagnostiserend onderwijzen' kan ik alleen maar van harte onderschrijven.

Dré van Dongen,
IVA, Postbus 90153, 5000 LE Tilburg

Rik Min

Computersimulatie als leermiddel: een inleiding in methoden en technieken
Schoonhoven: Academic Service, 1987, 221 pagina's.

In het boek 'Computersimulatie als leermiddel: een inleiding in methoden en technieken' geschreven door Rik Min, wordt geprobeerd de lezer een kijk te geven op 'enkele mogelijkheden voor het ontwerpen van attractieve instructieve educatieve simulatie-software'. Het is samengesteld uit gedeelten van collegedictaten zoals die bij de Faculteit der Toegepaste Onderwijskunde van de Universiteit Twente in gebruik zijn. Het boek bestaat uit vier delen voorafgegaan door een inleidend hoofdstuk waarin het vakgebied Instrumentatietechnologie duidelijk wordt gemaakt.

De instrumentatietechnologie is een discipline binnen de onderwijskunde. Ze heeft betrekking op, op de computer gebaseerde interactieve individuele leermiddelen. Een van deze leermiddelen is computersimulatie, te gebruiken in situaties die vergelijkbaar zijn met practica. Een computersimulatie moet allereerst als leeromgeving worden bekeken. Daarna zijn pas vormgeving en bedieningsgemak van de computer en programmatuur aan de orde.

Instrumentatietechnologisch onderzoek is sterk interdisciplinair van aard. Voor het doen van goed onderzoek is kennis vereist uit de psychologie, informatica, onderwijskunde, ergonomie en communicatieleer. Wanneer het gaat om computersimulaties is bovendien kennis nodig van wiskundige modellen.

In deel I van het boek wordt computersimulatie als methode van onderwijs uit de doeken gedaan. Computersimulatie is een vorm van Computer Ondersteund Onderwijs (COO), waarbij met een nagebootste situatie wordt geëxperimenteerd. Besproken worden computersimulaties waarbij een wiskundig model de werkelijkheid nabootst. Deze modellen worden in een computersimulatieprogramma geformaliseerd waarmee de lerende op eenvoudige wijze bepaalde variabelen in het model kan aanpassen. Het stelt leerlingen in staat om beslissingen te nemen waarna de computer reageert. Het leerdoel dat aan dergelijke programma's verbonden is, is het leren herkennen van de invloed die modelparameters hebben op de verschillende variabelen. Met computersimulatieprogramma's wordt dus niet nagestreefd om te leren modelleren.

In deel II worden allerlei technieken uiteengezet waarmee een computersimulatieprogramma vorm kan worden gegeven. *Presentatietechnieken* worden gebruikt om de opbouw van het beeldscherm vorm te geven. Behandeld worden het presenteren van tekst, graphics, geluid en bewegende beelden. Onderwerpen die aan de orde komen zijn o.a. hoeveelheid tekst, layout van tekst, muziek, spraaksynthese, het gebruik van kleur, windowtechnieken, opbouw van grafieken etc. Al deze technieken ondersteunen de gebruiker bij het volgen van het verloop van een simulatie en geven aldus feedback op de door de gebruiker genomen beslissingen. Met behulp van *acceptatietechnieken* kan een gebruiker zijn beslissingen aan het programma kenbaar maken. Aan de orde komen het gebruik van o.a. toetsenbord, muis, lichtpen en aanraakscherm.

Tenslotte worden *communicatietechnieken* behandeld waarmee verbindingen tussen verschillende gegevensbestanden kunnen worden gelegd.

Nadat zowel methoden als ook technieken voor het ontwerpen van simulaties zijn behandeld, worden deze in deel III geïntegreerd in een ontwikkelomgeving. Er wordt uitdrukkelijk aangegeven dat het gaat om het ontwikkelen van een leeromgeving en dat voor het ontwikkelen hiervan veel meer nodig is dan een computersimulatieprogramma. In dit deel wordt ook veel aandacht gegeven aan het opzetten en vastleggen van een model.

Het boek wordt afgesloten met een deel waarin een aantal voorbeelden wordt gegeven.

Het zal duidelijk zijn dat over een zeer breed terrein informatie wordt gegeven. Voor een boek van ruim 200 pagina's heeft dit wel tot consequentie dat sommige onderwerpen te kort wordt gedaan. Zo wordt aan bijvoorbeeld een complex veld als 'gaming' slechts 5 zinnen besteed. Ook het ontwikkelen van leeromgevingen had diepgaander besproken kunnen worden.

Voor wie is dit boek nu bedoeld? De achterflap van het boek vermeldt dat het een brug probeert te slaan tussen onderwijskundigen en informatici wier vak het is (of wordt) goede attractieve, educatieve simulatieprogrammatuur te ontwikkelen. Hierbij zijn twee kanttekeningen te plaatsen.

Reeds in de begintijd van de computer werd deze gebruikt om grote modellen door te rekenen (simulaties). Informatici bedachten slimme trucs om zoveel en zo snel mogelijk te kunnen rekenen. Ook binnen het onderwijs en de onderwijskunde zijn tal van gebieden aan te wijzen waarin modellen en simulaties een rol spelen. De brug die getracht wordt te slaan is dus al aanwezig, hij wordt alleen fraaier en sterker gemaakt. De huidige personal computers bieden mogelijkheden voor betere gebruikersvriendelijkheid (fraaier) en hoge rekensnelheden (sterker). Dit boek geeft een overzicht van deze mogelijkheden.

De andere kanttekening heeft te maken met het feit dat de belangrijkste groep wordt vergeten: de docenten en leraren die uiteindelijk moeten werken met het ontwikkelde COO. Het zijn uiteindelijk deze mensen die als gebruikers zullen optreden. Elke ontwikkelingsmethodiek onderstreept het belang van het betrekken van de gebruikers bij het ontwikkelingsproces. Dit boek is juist voor leraren en docenten een zeer nuttig boek omdat ze als gebruikers van COO-programma's betere gesprekspartners kunnen zijn voor ontwikkelaars van COO. Simulatie als leermiddel kan zeer krachtig zijn, maar zal eerder ingang vinden als de vraag om dit soort programma's gestimuleerd wordt vanuit de gebruikers.

Aan de uitvoering van het boek had meer aandacht kunnen worden besteed. Dit geldt zowel voor de tekst als voor de illustraties. Dit blijkt direct al bij het voorwoord. In de eerste alinea wordt verteld dat het boek uit 3 delen bestaat. In de volgende alinea's (op dezelfde bladzijde) wordt uitgelegd wat men zoal kan vinden in het eerste, tweede, derde en *vierde* deel).

Ook de illustraties kunnen beter (m.u.v. de omslagfiguur, deze is prachtig). Twee voorbeelden: figuur 2.12 heeft als ondertekening: 'Een voorbeeld van een complex computersimulatie-systeem voor het medisch onderwijs (het RLCS-systeem, Min, 1982)'. Uit de illustratie is noch op te maken dat het systeem complex is, noch dat het om een computersimulatie gaat, noch dat het bedoeld is voor het medisch onderwijs.

Bij figuur 6.3. luidt het ondertekening: 'Dit beeldscherm met aanraakscherm wordt gebruikt in een bibliotheek waar de klanten (dus ook kinderen) het boek van hun keuze zelf kunnen uitzoeken'. De illustratie geeft niet een bibliotheek en klanten die een boek opzoeken weer, maar we zien een jongetje dat kijkt en wijst naar een beeldscherm met daarop afgebeeld een boom, een huis en een kat.

Concluderend kunnen we het volgende stellen. Er wordt inleidende informatie gegeven over een zeer divers aantal onderwerpen. Soms is de hoeveelheid informatie over een onderwerp te klein. Het boek kan niet alleen voor informatici en onderwijskundigen nut hebben, maar kan ook zeer goed bruikbaar zijn voor docenten die onderwijs geven in vakgebieden waarin modellen een rol spelen. De uitvoering van het boek is mager.

H. de Jong

Faculteit der Sociale Wetenschappen, R.U. Utrecht, Postbus 80140, 3508 TC Utrecht

Peschar, J.L.

Zo vader - zo zoon, zo moeder - zo dochter? Vergelijkende analyses naar de processen van statusverwerving en onderwijsmobiliteit in Nederland, Hongarije en Polen.

Lisse: Swets & Zeitlinger, 1987. 183 pagina's. f 42.00.

In dit boek wordt verslag gedaan van een vergelijkend onderzoek naar de effecten van sociale achtergronden die met de ouders gegeven zijn, vooral hun opleiding en beroep, op de hoogten van de opleiding en het beroep van hun kinderen. Deze effecten worden vergeleken 1. tussen een viertal leeftijdsgroeperingen die cohorten worden genoemd 2. tussen mannen en vrouwen en 3. tussen Nederland, Hongarije en Polen. De data waarop de studie is gebaseerd, zijn cross-sectioneel (1981/82). Het onderzoek is daarmee te typeren als een sociaal mobiliteitsonderzoek én als een internationaal vergelijkend onderzoek. En als zodanig is het van meet af aan geconfronteerd met alle problemen van beide typen te meer omdat het gaat om landen met een totaal verschillend sociaal politiek stelsel. Niet voor niets staat dan ook in het voorwoord: 'Zo op het oog lijkt het een nogal ambitieuze en hachelijke onderneming, waarvan het resultaat niet op voorhand duidelijk zou zijn.'

In het voorliggende boek heeft Peschar op zeer verdienstelijke wijze het onderzoek verslagen. Het boek is overzichtelijk opgebouwd, geschreven in een heldere en beknopte stijl (in totaal 183 pagina's) en wat erg belangrijk is op een terrein waar uitwisseling van ervaring met probleemoplossingen zo nodig is, het bevat naast de inhoudelijke resultaten het methodische verslag; verantwoording van methodieken, keuzen van modellen, tussentijdse resultaten krijgen uitvoerig aandacht. Ook de data voorzover het de mobiliteitstabellen betreft, ontbreken niet. Het zou aardig geweest zijn als ook voor de LISREL analyses een correlatie matrix gegeven was, maar het probleem zal wel geweest zijn welke dan precies.

We zullen nu hoofdstuksgewijs het boek doorlopen en bij elk hoofdstuk kanttekeningen plaatsen daar waar we die althans hebben. Voorop zij gesteld dat daarmee geenszins onze indruk van zorgvuldigheid en gedegenheid wordt weggenomen.

Hoofdstuk 1 beschrijft de achtergronden van het onderzoek; eerst globaal de inhoudelijke (later gedetailleerder), uitvoeriger de methodische. Met name wordt ingegaan op het probleem van het creëren van meetinstrumenten waarmee vergelijkingen tussen landen kunnen worden gemaakt. Gekozen wordt (aanvankelijk) voor functioneel equivalente meetinstrumenten; dat zijn instrumenten die voor de betreffende landen naar dezelfde begrippen refereren, maar verder 'uiterlijk' (indicatoren, factorstructuur) per land verschillend mogen zijn. Voor ogen staat voor elk land LISREL modellen te hanteren waarin dezelfde latente variabelen voorkomen maar waaraan per land verschillende meetmodellen gekoppeld zijn. De centrale vraag is of de causale structuur tussen de latente variabelen voor de onderzochte landen identiek is. Omdat LISREL een nogal strenge techniek is met veel assumpties, kondigt de auteur ook aan log-lineaire analyses toe te passen die minder assumpties kennen.

Hoofdstuk 2 beschrijft eerst van Hongarije, Nederland en Polen de onderwijs- en beroepsstructuur en het kwalificatieniveau van een aantal globale beroepsgroepen. Zeer beknopt worden ook de relevante ontwikkelingen op onderwijskundig en sociaal-politiek gebied geschetst. Verschillende theoretische aspecten komen daarna aan de orde. Het status attainment model (waarin een relatie gelegd wordt tussen de sociale achtergrond en de sociale positie van een persoon) komt dicht bij wat hierboven is gezegd en sluit goed aan bij een LISREL benadering; Turner's notie van 'sponsored' (meer in Nederland) en 'contest' mobiliteit (meer in de Oost Europese landen); de reproductie-theorie(en) en tenslotte het vaak gesignaleerde theoretische advies om in deze rekening te houden met de economische ontwikkeling van een samenleving waarin de rol van de industrialisering sterk wordt benadrukt. Met de drie laatste theoretische aspecten wordt de keuze van de landen gemotiveerd en wordt de gezichtshoek bepaald waaronder de landen worden vergeleken. Een vijftal basis-hypothesen worden opgezet, te weten:

- In socialistische landen is de reproductie keten minder sterk dan in niet-socialistische landen.
- Binnen de socialistische landen is er een duidelijk verschil in dit opzicht in vergelijking met de voorgaande niet socialistische periode.

- De verschillen tussen de seksen zijn in socialistische landen geringer dan in niet-socialistische landen.
- Verschillen in industrialisatieniveau en -tempo spelen een ondergeschikte rol in vergelijking met socio-politieke systeemverschillen.
- Verschillen in waarnemings- en analysemethoden spelen geen rol in de vaststelling van de genoemde verschillen.

In hoofdstuk 3 wordt met behulp van LISREL gestreefd naar een structureel model voor statusverwerving. Omdat voor Polen de gegevens ontbreken worden alleen Nederland en Hongarije met elkaar vergeleken. De basishypothesen worden geconcretiseerd en na de variabelen met HOMALS optimaal geschaald te hebben zodat de relaties meer lineair zijn, wordt een viertal modellen binnen elk land getoetst. Ondanks de inspanningen (het gaat om een 80 LISREL analyses, goed leesbaar middels grafieken verslagen) is het resultaat zoals Peschar zelf zegt, nogal teleurstellend. Dat zit hem nog niet eens zozeer in de inhoud van de conclusies. Ze gaan enigszins in de veronderstelde richting, maar ze missen overtuigingskracht omdat ze op slecht passende modellen zijn gebaseerd. De indruk bestaat dat de moeilijkheid vooral in het meetgedeelte van het model zit. Vandaar dat in model 4 het principe van functionele equivalentie losgelaten is en een padanalyse op factorscores is uitgevoerd. De tekst ter plekke (p. 71-72) is echter wel erg beknopt en roept nogal wat vraagtekens op. Waarom impliceert een fasegewijze werkwijze (eerst het meetgedeelte afwerken en daarna een padanalyse uitvoeren) het loslaten van het principe van functionele equivalentie? Binnen elk land zijn toch vooraf op basis van verschillende indicatoren factorscores te berekenen? Bovendien door alleen de indicatoren die gemeenschappelijk zijn te beschouwen gaan de aantallen in tabel 3.6 wel erg sterk afwijken van die in voorafgaande tabellen. Eerst gaat het om 1471 Nederlanders en 4357 Hongaren; nu zijn er dat respectievelijk 2208 en 10967. En waarom lopen de pijlen in figuur 3.13 nu anders dan in de voorafgaande figuren, van indicator naar latent begrip? In het beperkte model zijn verder de vier pijlen betreffende de directe effecten van sociale achtergrond naar eerste en laatste beroep vervallen. Dat zijn vier vrijheidsgraden. Het volledige model blijkt echter twee vrijheidsgraden te tellen. Wat zijn dan de restricties van het volledige model? Is het niet echt volledig? In de tekst staat verder dat in tabel 3.6 'in géén van de cohorten in de beide landen, noch voor mannen, noch voor vrouwen, de directe invloed van de sociale achtergrond op de bereikte positie kan worden gemist'. Maar zelfs bij het strenge criterium dat Peschar aanhoudt (een model past wanneer de chi-kwadraat ongeveer gelijk is aan het aantal vrijheidsgraden), is de chi-kwadraat voor 3 van de 16 beperkte modellen kleiner dan 4. In de tekst staat verder dat er zowel in Nederland als in Hongarije 'een duidelijke invloed van sociale achtergrond aanwezig is, met name bij beta 32 (opleiding respondent-eerste beroep respondent)'. Maar de opleiding van de respondent hoort toch niet tot de sociale achtergrond? Om het een en ander te controleren moet je bovendien naar de bijlage toe (p. 170) en niet afgaan op de grafieken in Figuur 3.14. Niet beta 21, beta 31 en beta 32 worden daar achtereenvolgens gegeven, maar vermoedelijk beta 32, beta 42 en beta 43. Er staan overigens ook nog enige typfouten òf in de bijlage òf in tabel 3.6 en figuur 3.14. Kortom in deze passage komt het een en ander niet goed over en dat is jammer omdat het ten onrechte een schaduw zou kunnen werpen op de rest van het boek.

Hoofdstuk 4 behandelt log-lineaire analyse modellen met betrekking tot onderwijsmobiliteit. De beperking tot alleen het onderwijs wordt bediscussieerd en de basishypothesen worden uitgewerkt. Mobiliteit wordt onderscheiden in structurele mobiliteit (de verandering van het gemiddelde onderwijsniveau van de kinderen ten opzichte van dat van de ouders) en circulatie mobiliteit (het aantal gewisselde posities na controle van de structurele mobiliteit). Wat de structurele mobiliteit betreft wordt verwacht dat deze positief samenhangt met een toename van de industrialisatie. Verwacht wordt dat er tussen Hongarije en Polen enerzijds en Nederland anderzijds duidelijke verschillen geconstateerd zullen worden. Over de richting van deze verschillen wordt evenwel niets medegedeeld (p. 83) en dat laat een wat onduidelijk beeld achter van de betekenis van de industrialisatie. Het is moeilijk vat te krijgen op welke

richting van de verschillen bedoeld wordt, getuige dat volgens de Engelse samenvatting (p. 181) het meer geïndustrialiseerde Nederland een grotere structurele mobiliteit moet kennen, terwijl volgens de samenvatting op p. 140 de industrialisatie theorie aangeeft dat juist in socialistische landen de structurele mobiliteit groter moet zijn omdat er vanuit een grotere achterstand moet worden gewerkt. Maar wat zeggen de data? In de eindconclusies (p. 110) lezen we dat de verschillen in structurele, met name niet-uniforme, mobiliteit tussen Nederland versus Hongarije en Polen duidelijk significant zijn (tabel 4.5 en 4.10). Eraan toegevoegd wordt: 'Met name is opvallend dat de structurele component over de gehele linie in Hongarije niet groter is dan in Nederland, maar dat dit pas voor de geboortecohorten van 1945 en later geldt.' Vermoedelijk steunt deze toevoeging op de figuren 4.3 en 4.5, maar deze figuren hebben alleen betrekking op de uniforme structurele mobiliteit. Gelet op het percentage verklaarde variantie in de tabellen 4.3, 4.4, 4.7 en 4.8, verdient dit beeld als eindconclusie m.i. nader genuanceerd te worden. Ten aanzien van de circulatie mobiliteit is er volgens Peschar een tegenstrijdige verwachting: vanuit de industrialisatiethese gedacht, zal de circulatie mobiliteit in Nederland groter zijn dan in Hongarije en Polen en moeten de verschillen in circulatie mobiliteit geleidelijk aan van de oudere naar de jongere cohorten tussen de landen geringer worden; vanuit de reproductiethese gezien moet de circulatie mobiliteit in socialistische landen groter zijn en in deze landen voor jongere cohorten groter dan voor oudere. Geconstateerd wordt (p. 111) dat de circulatie mobiliteit voor Nederland en Hongarije voor alle cohorten nagenoeg constant blijft en tussen de landen niet significant verschilt. Peschar leidt eruit af dat voor beide landen de onderwijskansen voor kinderen uit de lagere sociale klassen niet zijn toegenomen; er is ondanks hervormingspogingen een betrekkelijk constante reproductie. Een opmerking hier betreft verder het kwadratisch model van de circulatie mobiliteit. De reden waarom dit ingevoerd is, wordt duidelijk gemaakt op p. 90. Er wordt van gezegd dat in het kwadratisch model de mogelijkheden om hoger of lager uit te komen dan de vaders – kwadratisch – afhankelijk zijn van het aantal niveaus dat men stijgt of daalt (hoe verder – hoe moeilijker). Mijn vraag is waarom de kwadratische circulatie mobiliteit zo centraal staat in dit hoofdstuk en de niet-kwadratische niet aan bod komt (meestal weliswaar gering maar soms niet bv. Nederland cohort 55 23% p. 91). Waarom niet de totale circulatie mobiliteit ineens beschouwd? En tenslotte op p. 100 (voor de vergelijking Nederland-Polen) wordt naast de kwadratische, de verticale circulatie mobiliteit onderscheiden om reden dat met name in Nederland de kwadratische component nog onvoldoende verklaart. Exact dezelfde Nederlandse mobiliteitstabel wordt echter eerder gebruikt in de vergelijking Nederland-Hongarije op p. 91. Wanneer daar de kwadratische component wordt ingevoerd, komt het model 'erg goed' met de gegevens overeen en is geen verticale component nodig.

Hoofdstuk 5 combineert in zekere zin de beide voorafgaande hoofdstukken. Het statusverwervingsproces uit hoofdstuk 3 staat weer centraal. Nu wordt het evenwel niet met LISREL dat slechts gering succes opleverde, onderzocht maar met log-lineaire analyses met formeel equivalente variabelen. Het gaat om de causale keten 'opleiding (beroep) vader → opleiding kind → beroep kind', tegelijkertijd voor de vier cohorten gemodelleerd zodat een tijdseffect kan worden nagegaan. Zowel voor vrouwen en mannen als voor Hongarije en Nederland worden analyses verslagen en verschillen getoetst. De reproductie keten blijkt goed te passen en tijdseffecten werden niet geconstateerd; een conclusie in overeenstemming met het voorafgaande. Een kleine opmerking is waarom niet een zg. gemodificeerde pad-analyse uitgevoerd is. Een causale volgorde is immers vooraf duidelijk gegeven. De analyse zoals ze nu is uitgevoerd, leidt tot een relatie opleiding (beroep) vader met opleiding kind gecontroleerd voor het beroep van het kind hetgeen gezien de causale volgorde niet wenselijk is. De reden waarom niet voor een dergelijke analyse gekozen is zal wel zijn dat veel relaties tussen de condities (land, cohort, sekse, opleiding vader) die dan niet aan het licht zouden komen, ook interessant zijn. Een gescheiden aanpak zou de voorkeur verdienen hoewel dat moet gezegd worden, in de praktijk de verschillen in resultaten wel gering zullen zijn.

Hoofdstuk 6 geeft een zeer lezenswaardige samenvatting van de bevindingen zowel in methodologisch als inhoudelijk opzicht.

Terug naar de basishypothesen zijn de bevindingen:

- De reproductie keten bleek in de socialistische landen niet minder sterk dan in niet-socialistische landen.
- Er is geen duidelijk onderscheid binnen de socialistische landen met de voorafgaande niet-socialistische periode.
- Verschillen tussen de seksen liggen in het gemiddeld opleidingsniveau: het sterkst zijn deze in Polen, dan Hongarije, dan Nederland en de verschillen worden in latere cohorten geringer. Er zijn geen sekse verschillen in de kansenstructuur.
- Mobiliteitsverschillen zijn terug te voeren tot niveau en tempo van industrialisering. Socio-politieke systeemverschillen spelen geen rol.
- Verschillende technieken lieten hetzelfde beeld zien. De log-lineaire analyse is echter strategisch (sneller, minder assumpties) beter dan LISREL.

De conclusies worden tenslotte in het licht van ander recent vergelijkend onderzoek geplaatst en met enkele aanbevelingen voor het landenvergelijkend onderzoek op dit terrein wordt het boek afgesloten.

J.G.M. Lammers

K.U. Nijmegen, Sociologisch Instituut, Th. van Aquinostraat 4, Nijmegen.

W. van Os

Evaluatie in het hoger onderwijs

Groningen: Wolters-Noordhoff, 1987. f 39,50, ISBN 90 01 67477, 218 pagina's.

De laatste jaren valt in Nederland een toenemende belangstelling te constateren voor de kwaliteit van het hoger onderwijs. Veranderingen in de wet- en regelgeving ten aanzien van het hoger onderwijs en bezuinigingen op de overheidsuitgaven ten behoeve van dat onderwijs hebben ertoe geleid, dat het WO en HBO meer aandacht moeten besteden aan de onderwijskwaliteit. De meest recente plannen met betrekking tot kwaliteitsbewaking kunnen gelezen worden in de HOOP-nota (1987). Daarin komt onder andere naar voren dat de overheid streeft naar een globalere en meer afstandelijke regelgeving die gepaard moet gaan met een vergroting van de autonomie van de betrokken instellingen (bijvoorbeeld wat betreft profielkeuze in onderzoek en onderwijs). In de HOOP-nota (pp. 103, 1987) valt daarover het volgende te lezen.

'Bij de huidige besturingswijze hebben procedurele voorschriften over de inrichting van het onderwijs mede de functie een garantie vooraf voor de kwaliteit te bieden. Bij de toekomstige wijze van 'sturen op outputs' zal de verantwoordelijkheid van de overheid een andere zijn. De overheid zal het functioneren van het stelsel van hoger onderwijs en wetenschappelijk onderzoek volgen en toetsen aan de functie die het stelsel moet vervullen. Dit betekent niet dat de instellingen hierin geen taak hebben. Integendeel, de instellingen dragen een grote verantwoordelijkheid bij het bewaken van de kwaliteit van hun produkten'.

Een dergelijke verandering in de regulering van het hoger onderwijs zal ingrijpende gevolgen hebben voor de opzet en uitvoering van het onderwijs. Bovenstaande citaat geeft namelijk al aan dat instellingen voor het WO en HBO de komende jaren ernst moeten maken met de systematische bewaking van de onderwijskwaliteit. Naast de confrontatie met visitatiecommissies en inspectie, die onder de noemer van externe kwaliteitsbewaking vallen, zullen de betrokken instellingen tevens een stelsel van interne kwaliteitsbewaking moeten opbouwen om zodoende aan de overheids-eisen ten aanzien van controle op onderwijskwaliteit te voldoen. De vraag die zich daarbij uiteraard voordoet is hoe in de toekomst interne en externe kwaliteitsbewaking nader ingevuld zal worden.

In het boek van Van Os wordt met name de vraag met betrekking tot interne kwaliteitsbewaking behandeld. Het boek is opgezet als gebruikersgids voor studenten, individuele docenten en besturen die in aanraking komen met de hierboven geschetste problematiek. Dit boek maakt deel uit van de vernieuwde 'Hoger Onderwijs Reeks'. Deze reeks is opgezet om onderwijskun-

dige informatie binnen het WO en HBO te verspreiden, en verschijnt onder auspiciën van de Contactgroep Research Wetenschappelijk Onderwijs (CRWO): het landelijk samenwerkingsverband van de universitaire centra voor Onderzoek en Ontwikkeling van het hoger onderwijs. Dit laatste is met name van belang om te weten, omdat het boek verschijnt uit een hoek waarin men dagelijks geconfronteerd wordt met de, soms harde, praktijk van evaluatie in het hoger onderwijs. De auteur, bijvoorbeeld, is zelf nauw betrokken bij de ontwikkeling van een evaluatiesysteem voor de Vrije Universiteit te Amsterdam.

Dit gegeven leidt tot de interessante vraag wat de onderwijskunde te bieden heeft voor de beoogde doelgroep. Met andere woorden, welke strategieën, methoden en technieken worden in dit boek behandeld en welke standpunten worden daarover ingenomen. Deze vraag is des te meer van belang als men bedenkt dat evaluatie van het hoger onderwijs binnen de Nederlandse onderwijskunde lange tijd een tamelijk verwaarloosd en onderontwikkeld terrein is gebleven. Het boek van Van Os is wat dat betreft een unicum: het eerste handboek dat over deze problematiek verschijnt.

Van Os behandelt in het eerste hoofdstuk de doelen, achtergronden, problemen en uitgangspunten van evaluatie. De auteur beschrijft achtereenvolgens de toenemende invloed van de overheid op de kwaliteit van het hoger onderwijs, enkele theoretische overwegingen met betrekking tot onderwijsevaluatie, problemen die kunnen ontstaan bij de uitvoering van evaluatie, en een mogelijke aanpak voor de evaluatie van onderwijsprogramma's. In dit hoofdstuk wordt een aantal facetten van onderwijsevaluatie geadstrueerd aan de hand van het bekende evaluatiemodel van Stufflebeam (1971). De sterke nadruk die binnen dit model gelegd wordt op informatieverzameling ten behoeve van het beoordelen van beslissingsalternatieven, biedt Van Os de mogelijkheid twee zaken in de daaropvolgende 19 hoofdstukken uit te werken, namelijk: 1) methoden en technieken van dataverzameling, en 2) het nemen van beslissingen op basis van evaluatiegegevens.

Deze hoofdstukken zijn ondergebracht in drie delen: curriculumevaluatie, cursusevaluatie en instrumenten.

In het eerste deel (curriculumevaluatie) bepleit de auteur een evaluatie-aanpak waarin niet alleen aandacht bestaat voor het aspect van dataverzameling maar ook voor de relaties tussen evaluator en bestuursorganen. Daartoe is het noodzakelijk dat met de opdrachtgevers consensus bereikt wordt over de taakstelling van de evaluator, de te kiezen onderzoeksmethode en de minimumeisen die aan de kwaliteit van het onderwijsprogramma gesteld kunnen worden. Van Os tracht dit mede te bereiken door een evaluatiecommissie in te stellen waarin de evaluator, enkele docenten, studenten en de onderwijscoördinator zitting hebben. De commissie moet een vrijheid van handelen krijgen betreffende zaken die verband houden met de gestelde taak. Tevens moeten middelen ter beschikking gesteld worden om data te verzamelen. In het evaluatieplan dat de commissie opstelt moeten deze zaken aan bod komen.

Centraal, in de door Van Os beschreven evaluatie-aanpak, staat de gedachte dat de evaluatie van het onderwijsprogramma gefaseerd ten uitvoering wordt gebracht. Daartoe vindt eerst een inventarisatie plaats van mogelijke problemen in het onderwijsprogramma. Daarbij wordt op vier zaken gelet. In de eerste plaats is de vraag aan de orde of het curriculum voldoet aan de wettelijk gestelde minimumvoorwaarden ten aanzien van het aantal studenten dat binnen de cursus- en inschrijvingsduur de studie voltooit. In de tweede plaats wordt de vraag gesteld of een tijdsbesteding van 6800 uur, over een periode van vier jaar, voor de gemiddelde student voldoende is om de studie te voltooien. In de derde plaats is het didactisch functioneren van docenten object van onderzoek. Tenslotte kijkt men of er praktische of inhoudelijke problemen bestaan in het onderwijsprogramma.

In het tweede deel (cursusevaluatie) wordt de tweede stap van de evaluatie-aanpak beschreven. Naar aanleiding van de inventarisatie in de eerste stap wordt getracht algemene problemen in het onderwijsprogramma, te herleiden tot knelpunten in specifieke cursussen. In deze stap staat de vraag centraal of men uit het evaluatie-onderzoek informatie kan verkrijgen die uiteindelijk leidt tot verbetering van de onderwijskwaliteit in afzonderlijke cursussen. In dit deel besteedt de auteur voornamelijk aandacht aan de zogenaamde ISEK-procedure (een

procedure ontwikkeld door het COWOG van de Rijksuniversiteit Groningen). Binnen deze procedure ontwikkelt men voor iedere cursus een evaluatie-instrument dat is toegesneden op cursusspecifieke problemen (Bartelds, Joostens & Kluijter, 1983).

In het derde en laatste deel van het boek wordt nader ingegaan op een aantal methoden en instrumenten die in de voorgaande delen aan bod kwamen (studietoetsen, tijdschrijfmethoden, interviews, etc.).

Het boek besluit met drie aanhangsels. De eerste appendix bevat de voornaamste wetsteksten met betrekking tot de evaluatie van hoger onderwijs. De tweede appendix gaat nader in op enkele onderzoeksgegevens ten aanzien van verschillende methoden om studielast te meten. Tenslotte geeft de derde appendix een overzicht met betrekking tot de stand van zaken rondom onderzoek naar de betrouwbaarheid en validiteit van studentoordelen.

In het begin van deze bespreking werd de vraag gesteld hoe vanuit de onderwijskunde vorm gegeven kan worden aan interne kwaliteitsbewaking. Van Os heeft deze vraag met name beantwoord door een uitgewerkte evaluatieprocedure te beschrijven en door gedetailleerd aandacht te besteden aan verschillende onderzoeksinstrumenten (niet alleen vragenlijsten en studietoetsen, maar ook tijdschrijfonderzoek, interviews met docenten en studenten, inhoudsanalyse en observatie van het learner report). Daarin is ook de sterkte van het boek gelegen. Aan de relatie evalueren en beslissen is duidelijk minder aandacht besteed. De vraag hoe evaluatieresultaten een rol kunnen spelen bij het bekwamen en verbeteren van de onderwijskwaliteit blijft min of meer liggen. Dat is jammer. Immers, de onderwijspraktijk wijst vaak uit dat onderwijsevaluatie niet automatisch leidt tot kwaliteitsverbetering. Uit onderzoek is bijvoorbeeld gebleken dat beslissingen over veranderingen in een programma eerder tot stand komen als de evaluatie-uitkomsten stroken met de verwachtingen of percepties van bestuurders (Cousins & Leithwood, 1986). Als de uitkomsten daarmee strijdig zijn, neemt de kans af dat het programma veranderd wordt. Slagen of falen van evaluatie is derhalve niet alleen afhankelijk van de kwaliteit van de geboden informatie, maar ook van factoren die te maken hebben met de status van de evaluator in de organisatie, de organisatiecultuur (het belang dat men hecht aan onderwijskwaliteit), en de personele consequenties die aan evaluatie-uitkomsten verbonden worden voor docenten. Een of meerdere hoofdstukken over deze problematiek zouden mijns inziens dan ook zeker thuishoren in een boek als dit. Deze kritiek wordt min of meer versterkt door het ontbreken van praktijkvoorbeelden in het boek. In een handboek dat uitdrukkelijk voor de onderwijspraktijk bedoeld is zouden dergelijke voorbeelden niet misstaan.

Een ander punt van kritiek betreft het feit dat niet of nauwelijks verwezen wordt naar onderzoek dat in de Verenigde Staten op dit gebied is verricht. De ontwikkelingen die zich nu in Nederland voordoen (inspectie, visitatiecommissies, interne kwaliteitsbewaking) waren in de Verenigde Staten reeds dertig jaar geleden aan de orde. Onvrede over het indertijd heersende onderwijsniveau had tot gevolg dat men steeds meer aandacht ging schenken aan de kwaliteit van het onderwijs. Het resultaat van deze ontwikkeling is ondermeer zichtbaar in het feit dat thans bijna iedere universiteit in de V.S. een systeem voor interne kwaliteitsbewaking heeft. De ervaringen met deze systemen zijn uitermate bruikbaar voor het ontwikkelen van evaluatiesystemen voor Nederlandse universiteiten.

Als laatste punt van kritiek wil ik het ontbreken van een index noemen. Voor gebruikersgidsen of handboeken zijn deze onontbeerlijk.

Deze kritiekpunten nemen echter niet weg dat het boek van Van Os de beoogde lezer een uitgewerkte evaluatieprocedure aanbiedt plus een breed scala aan instrumenten. Terugkomend op de vraag wat de onderwijskunde te bieden heeft op het gebied van evaluatie in het hoger onderwijs, kan gesteld worden dat dit boek een eerste aanzet vormt om de komende jaren de achterstand op dit gebied weg te werken.

Referenties

Bartelds, J.F., Joostens, T.H., & Kluijter, H. (1983). ISEK: Instrumentarium voor Systematische Evaluatie

- van Kursussen. Een beschouwing over drie varianten. *Tijdschrift voor Onderwijsresearch*, 8, 49-65.
- Cousins, J.B. & Leithwood, K.A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56, 331-364.
- HOOP. Hoger Onderwijs en Onderzoek Plan. Kerndocument. (1987). 's-Gravenhage: Staatsuitgeverij.

W.H. Gijselaers

Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Rijksuniversiteit Limburg
Postbus 616, 6500 MD Maastricht.

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

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

Jrg. 12, nr. 6

December 1988

Inhoudsopgave

VOR-Bulletin, Jrg 12. nr. 6

Verslag VOR-ledenvergadering, dd 10-11-1988	368
Oproep voor nieuwe bestuursleden VOR-bestuur	369
VOR begroting 1989	370
Toelichting op de VOR-begroting 1989	371
OTC- en ANGO onderzoek, de sluitpost op de SVO-begroting?	372
Congressen	373
Ontvangen publicaties	374

Verslag VOR-ledenvergadering, d.d. 10-11-1988

Aanwezig : de leden : van Hout
de bestuursleden: Claessen, Beintema, Gerritsen v.d. Hoop, van Hout Wolters,
Roede en v.d. Werf

Afwezig m.k.: de leden: Slavenburg, Kanselaar
de bestuursleden: Scheerens, Span

1. Opening

Het enige aanwezige VOR-lid wordt hartelijk welkom geheten. Geconstateerd wordt dat de opkomst van de leden teleurstellend laag is. De volgende ledenvergadering zal op meer opvallende wijze worden aangekondigd.

2. Mededelingen

Er zijn geen mededelingen

3. Verslag ledenvergadering 31 mei 1988

N.a.v.: het voorstel over een thema vakdidactiek is besproken in het bestuur. De Cie. themapanning heeft gemeend de vaste thema's niet te veranderen. Er zijn echter op lokaal niveau altijd mogelijkheden iets te doen (bijv. symposium) aan vakdidactiek. Aan de organisatiecommissie ORD Leiden is de suggestie meegegeven. Daarmee is niets gebeurd.

De notulen worden vastgesteld.

4. Begroting 1989 (zie dit VOR-bulletin)

De volgende toelichting wordt gegeven:

- het aantal niet-betalende leden neemt steeds meer toe. Inschakeling van een incasso-bureau is overwogen, maar deze maatregel is wel erg rigoreus. Er bestaat slechts één keer per jaar de mogelijkheid om niet-betalende leden van de lijst te schappen. Dit gebeurt ook wel. Het blijkt echter ook dat leden die in een jaar niet betaald hebben dit later alsnog wel doen;
- International Journal: de VOR heeft een collectief abonnement voor 150 abonnees. Er zijn inkomsten van 100 abonnementen. SVO geeft voor 1990 een garantiesubsidie voor 50 abonnementen en wil dit waarschijnlijk ook na 1990 wel continueren;
- de vereniging heeft ca. f 20.000,- reserves in de SOR. Dit is de financiële buffer om verliezen op de ORD's te kunnen opvangen. Vergeliken met een aantal jaren geleden is de financiële situatie goed. Dit is bereikt door een goed financieel beleid, door aanpassing van de contributie en door de integratie van TOR en VOR-bulletin.

De begroting wordt vastgesteld.

5. Nieuwe bestuurssamenstelling

De bestuursleden van Hout Wolters en Span treden m.i.v. 1-1-89 af na een jarenlange trouwe staat van dienst. Als hun opvolgers zijn door het bestuur voorgedragen Kanselaar en van Hout. Door de leden zijn geen tegenkandidaten ingebracht. De voorgedragen leden zijn hierbij verkozen.

W.v.t.t.k.

Gevraagd wordt hoe de AERA reviewing door Nederlandse reviewers verloopt. Dit zal worden nagegaan.

Rondvraag

Wordt geen gebruik van gemaakt.

Sluiting

De volgende ledenvergadering is tijdens de ORD 89 op 24 en 25 mei te Leiden. Plaats en tijd worden nader bekend gemaakt.

Greetje van der Werf
secretaris VOR-bestuur

Oproep voor nieuwe bestuursleden VOR-bestuur

Het VOR-bestuur zoekt belangstellenden die zitting willen nemen in het bestuur. In aanmerking komen zij die werkzaam zijn bij onderzoeksinstituten, vakgroepen, RWO-centra, bedrijfsopleidingen of landelijke verzorgingsinstellingen en betrokken zijn bij onderwijsresearch. De voorkeur gaat uit naar kandidaten die bereid zijn het penningmeesterschap van de vereniging op zich te nemen.

VOR BEGROTING 1989

Inkomsten

100 Contributite

110 leden

120 studenten

ingeschr.

450

betalend

430 x 105,-

20 x 25,-

f 45.150,-

f 500,-

f 45.650,-

300 Intern. Journ. Ed. Res.

310 abonnementen

320 garantiesubsidie SVO

100 x f 40,-

max.

f 4.000,-

f 4.500,-

500 SOR bijdrage VOR prijs

700 Inkomsten TOR (Royalties)

800 SVO subsidie TOR

900 Diversen

f 8.500,-

f 1.000,-

f 1.800,-

f 5.000,-

f 600,-

f 62.550,-

Uitgaven

400 Activiteiten

410 ledenwerving

f 1.000,-

420 ORD-programmaboek

f 1.500,-

440 colloquia

f 1.500,-

f 4.000,-

500 Tijdschriften

510 TOR/VOR bulletin

511 Collect. abonnem.

leden 470 x 80,-

stud. 25 x 53,-

f 37.600,-

f 1.325,-

f 38.925

f 4.000,-

512 Redactiekosten TOR

513 Redactiekosten

VOR-bulletin

f 1.500,-

f 44.425,-

520 I.J.E.R.

150 abonnementen

à f 40,-

Linguistic editing

f 6.000,-

f 2.500,-

600 Bestuur

700 VOR Prijs

800 Ledenadministratie

900 Diversen

f 8.500,-

f 2.000,-

f 1.000,-

f 2.000,-

f 625,-

f 62.550,-

Toelichting op de VOR-begroting 1989

De begroting voor 1989 is vrijwel identiek aan die voor 1988. Wijzigingen betreffen de inkomsten uit de abonnementen voor het IJER, de ledenwerving, de uitgaven voor het toezenden van het ORD-programmaboek aan alle leden en de ledenadministratie.

Toelichting per post:

Inkomsten

- 100 Voor 1989 wordt gerekend op $\pm$ 475 leden, waarvan 25 student. Gezien de blijkbaar niet geheel te vermijden betalingsproblemen en enkele 'gratis' abonnementen, wordt gerekend op 430 betalende gewone en 20 betalende student-leden. De contributie is hetzelfde gebleven als voor 1988, n.l. f 105,- voor gewone leden en f 25,- voor student-leden.
- 300 Met uitgeverij Pegamon en de SVO is een overeenkomst gesloten, waarbij de VOR 150 collectieve abonnementen op het International Journal of Educational Research (IJER) neemt. Om aanloopverliezen te kunnen opvangen heeft het SVO een garantiesubsidie beschikbaar gesteld. Voor 1989 bedraagt deze maximaal f 550,-. Daarnaast subsidieert de SVO maximaal f 2.500,- voor linguistic editing ten behoeven van het IJER.
- 500 De VOR-prijs wordt gefinancierd door de Stichting Werkgroep Onderwijsresearch (SOR).
- 700 Van Swets & Zeitlinger ontvangt de VOR 10% van de TOR abonnementsinkomsten van niet-VOR-leden. Voor 1989 wordt dit geschat op f 1.800,-.
- 800 Ten behoeve van het TOR ontvangt de VOR jaarlijks f 5.000,- van de SVO.

Uitgaven

- 410 Na de intensievere ledenwerving in 1987 door middel van een op ruime schaal verspreide brochure zijn de wervingsactiviteiten voor 1989 geschat op f 1000,-.
- 420 Net als in 1988 zal het programmaboek voor de ORD 1989 vóór de ORD aan alle VOR leden worden toegezonden. De hieraan verbonden kosten worden geschat op maximaal f 1.500,-.
- 440 Colloquia zijn weer opgevoerd voor f 1.500,-. Hieruit kunnen tevens andere kosten t.b.v. inhoudelijke activiteiten worden betaald.
- 500 De uitgaven voor de tijdschriften (TOR/VOR-bulletin en IJER) zijn naar verwachting hetzelfde als voor 1988.
- 600 De uitgaven ten behoeve van bestuur zijn hetzelfde begroot als voor 1988. Een belangrijk deel hiervan is bestemd voor reiskosten van enkele bestuursleden.
- 700 De jaarlijks tijdens de ORD toe te kennen VOR-prijs bedraagt maximaal f 1.000,- (te betalen door de SOR).
- 800 De uitgaven voor de ledenadministratie zijn ten opzichte van 1988 iets verlaagd.

E. Roede, oktober 1988
(penningmeester)

OTG- en ANGO-onderzoek, de sluitpost op de SVO-begroting?

Met het van kracht worden van de Wet op de Onderwijsverzorging is een wettelijk kader gegeven voor de verdeling van de onderzoeksgelden van de SVO. Een deel van de middelen wordt besteed aan aanvragen van het 'veld', een deel is bestemd voor aanvragen van het Ministerie en een deel komt ten goede aan onderzoeksvoorstellen die door onderzoekers zelf worden ingediend. Als verdeelsleutel fungeert daarbij de verhouding 40.40.20 procent.

Deze verdeling is destijds door de wetgevers niet zonder reden gemaakt. Men heeft beoogd de verschillende belangen in balans te brengen en te voorkomen dat een bepaalde groepering een te zwaar stempel zou drukken op de allocatie van de middelen. Nadrukkelijk is daarbij ook ruimte gereserveerd voor aanvragen van onderwijsonderzoekers zelf.

Deze aanvragen vanuit het forum van de onderzoekers zelf worden gekanaliseerd via de ANGO-procedure en de verschillende Onderzoeks Thema Groepen. Voor de ANGO-aanvragen kunnen individuele onderzoekers onderzoeksvoorstellen opstellen, welke dan beoordeeld worden door de ANGO-Commissie. Voor de OTG's verloopt de procedure via de onderzoekscommissie van de Onderzoeks Thema Groep, die na een zorgvuldige procedure in meerdere ronden, de uitverkoren voorstellen doorleiden naar het bestuur van de SVO.

In de afgelopen jaren hebben de ANGO-Commissie en de OTG's zich met veel inzet en zorg gekweten van hun taak. Voor de onderzoekers was het weliswaar een langdurig en arbeidsintensief traject, maar tegelijkertijd een konkrete mogelijkheid om zelfstandig onderzoeksvoorstellen in te dienen. Het jaar 1988 lijkt evenwel een breuklijn te markeren. Van de ruim 30 ANGO-aanvragen zijn maar vier voorstellen gehonoreerd door het SVO-bestuur. Voor de OTG-aanvragen is de situatie nog veel rampzaliger. Door de onderzoekscommissies van de OTG's zijn enkele tientallen voorstellen met een positief advies voorgelegd aan het SVO-bestuur. Goedkeuring door het SVO-bestuur bleek slechts te zijn weggelegd voor het minimale aantal van twee!

Een dergelijk resultaat doet uiteraard de wenkbrauwen fronsen. Een aanzienlijk aantal onderwijsonderzoekers heeft de nodige energie gestoken in het opstellen van onderzoeksaanvragen. Hun inspanningen bleken grotendeels weggegooid geld te zijn. Ook voor de OTG's is het resultaat een absoluut dieptepunt. Zorgvuldige beoordeling en selectie door mede-onderzoekers lijken voor het SVO-bestuur geen enkele betekenis te hebben, luttelend op het uiteindelijke resultaat. Een dergelijke handelwijze leidt onherroepelijk tot een zachte dood van de OTG's. Het teleurstellende resultaat lijkt mij ook onverenigbaar met de uitgangspunten van de Wet op de Onderwijs Verzorging. Immers deze wet beoogt nadrukkelijk ook ruimte te scheppen voor onderzoeksvoorstellen van onderop als een min of meer natuurlijk tegenwicht tegen aanvragen van de zijde van veld en beleid. Een SVO-beleid waarbij ANGO en OTG als sluitpost op de begroting fungeren, kan niet in overeenstemming zijn met de W.O.V.

Prof. dr. J. Claessen
Voorzitter VOR-bestuur

Congressen ed.

Economie van het studeren

Op verzoek van het **Landelijk Overleg Studievardigheden (L.O.S.)** organiseert de Universiteit Twente in samenwerking met de Hogeschool Enschede (i.o.) de **Landelijke Dag Studievardigheden 1989** met het thema "Economie van het studeren".

Het wordt voor studenten steeds belangrijker om bij het studeren economisch te werk te gaan, o.a. door de wettelijke regelingen over de studieduur, de gewijzigde studiefinanciering met bankleningen en de grotere kans op werkloosheid bij bepaalde studierichtingen. Studenten dienen zich beter te bezinnen op de vraag hoeveel tijd, energie en geld ze aan welke activiteiten willen besteden en wat hen deze investering oplevert. Deze kosten-effecten analyses zullen op allerlei niveaus plaatsvinden: bij de studiekeuze, de keuze van een vakkenpakket, maar ook op het niveau van het studeren voor een bepaalde cursus. De universiteiten en hogescholen hebben als taak de studenten te helpen om economischer te studeren.

Op deze studiedag zal ingegaan worden op de vraag op welke wijze studenten, docenten en onderwijsbeleidsmakers (faculteiten en afdelingen) de economie van het studeren kunnen verbeteren, zowel op cursusniveau als op studierichtingsniveau.

Nadere informatie: Mevr. drs. J. A. Oosterhuis-Geere Universiteit Twente, Onderwijskundig Centrum, postbus 217, 7500 AE Enschede.

Ontvangen publicaties

- ARBO (1988). Een boog van woorden tot woorden. Advies over het onderwijs aan leerlingen uit etnische minderheden. Zeist: Adviesraad voor het basisonderwijs, het speciaal onderwijs en het voortgezet speciaal onderwijs.
- Been, P. H. & Koster, K. B. (Red.) (1988). *Onderwijstechnologie: modellen en empirie*. Amsterdam: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, no. 20).
- Collot d'Escury-Koenings, A. (1988). *Het blauwe boek: rollenspelletjes voor kleuters*. Amsterdam/Lisse: Swets & Zeitlinger.
- Gijselaers, W. H. (1988). *Kwaliteit van het onderwijs gemeten. Studies naar de betrouwbaarheid, validiteit en bruikbaarheid van studentoordelen*. Proefschrift RU Limburg.
- Hetebrij, M. & Amersfoort, J. P. M. van (1988). *Kwantiwijzer. Een researchprogramma in haar context*. Utrecht: ISOR.
- Hoogeveen, K. e.a. (1988). *Almanak onderzoek van onderwijs in Nederland. Deel I en II*. Leiden: LICOR.
- Imbos, Tj. e.a. (1988). *The development and evaluation of a measurement instrument for essay-writing*. Maastricht: RU Limburg (Onderzoek van Onderwijs, no. 35).
- Luykx, M. (1988). *Schoolsucces van Turkse en Marokkaanse meisjes. Een onderzoek naar Turkse en Marokkaanse meisjes in het mavo, havo en vwo*. Utrecht: ISOR.
- Mirande, M. J. A. & Wardenaar, E. (1988, 2e druk). *Scriptieproblemen*. Groningen: Wolters-Noordhoff (Hoger onderwijs reeks, no. 3), f 25.-.
- Mottier, I. (1988). *Emancipatie aspecten in schoolboeken: het ontwerpen van richtlijnen voor natuurkunde, techniek, geschiedenis en taal*. Proefschrift TU Eindhoven.
- Muijtjens, A. M. M. e.a. (1988). *Exploratory analysis of progress test data*. Maastricht: RU Limburg (Onderzoek van Onderwijs, no. 38).
- Mulder, M. & Lent, J. van (1988). *Kantoorautomatisering: een onderzoek voor leerplanontwikkeling*. Amsterdam/Lisse: Swets & Zeitlinger.
- Reints, A. J. C. e.a. (1988). *Het gebruik van leerboeken door leerlingen uit het voortgezet onderwijs. Eindverslag van SVO-project 6068*. Utrecht: ISOR.
- Rossum, W. van (1988). *De plaats van de ivoren toren. Een inleiding in de wetenschapssociologie*. Groningen: Wolters-Noordhoff, f 27.50.
- Rupp, J. & Jungbluth, P. (Red.) (1988). *Algemene basisvorming; schept de W. R. R. betere onderwijskansen?* Nijmegen: ITS (OOMO-reeks).
- SWIDOC/Adviescommissie Onderzoek Minderheden (1988). *Bibliografie etnische groepen*. Amsterdam: Noord-Hollandsche.
- Teunissen, J. (1988). *Etnische relaties in het basisonderwijs: "witte" en "zwarte" scholen in de grote steden*. Proefschrift RU Utrecht.
- Verhulst-Schlichting, L. & Koning, T. de (1988). *FIT, Functionele Imitatie van Taalstructuren: taalprogramma voor kinderen met een grammaticale achterstand*. Amsterdam/Lisse: Swets & Zeitlinger.
- Werf, M. P. C. van der (1988). *Het schoolwerkplan in het basisonderwijs: ontwikkeling, implementatie en opbrengst*. Amsterdam/Lisse: Swets & Zeitlinger (Proefschrift RU Groningen).
- ITS. Jaarverslag 1986.
- ITS. Jaarverslag 1987.
- Van Gennip, J. & Vierke, H. (1988). *Evaluatie-Onderzoek Schoolbegeleiding regio Delft*. Nijmegen: ITS.
- ITS. Publicaties.
- CITO. *Meerjarenplan 1989-1992 en jaarprogramma 1989*. Arnhem: CIO.
- Apple (1988). *Schoolkrant*, September.
- SVO. *Meerjarenbeleids- en onderzoeksprogramma van het Instituut voor onderzoek in het onderwijs, SVO voor de periode 1989-1992*.

- Aldenkamp, A.P., Alpherts, W.C.J., Meinardi, H. & Stores G. (Eds.) (1987). Education and Epilepsy: proceedings of an international workshop on Education and Epilepsy. Amsterdam/Lisse: Swets & Zeitlinger (Modern approaches to the diagnosis and instruction of multi-handicapped children, no. 21).
- ARBO (1988). Voorrang aan achterstand. Advies over een integraal beleid ter voorkoming en bestrijding van onderwijsachterstand. Zeist: Adviesraad voor het basisonderwijs, speciaal onderwijs en voortgezet speciaal onderwijs.
- Beem, A.L. (Red.) (1988). Weten redeneren raden: een analytische evaluatie van de basisvorming. Amsterdam/Lisse: Swets & Zeitlinger.
- Bouhuijs, P.A.J. & Wijnen, W.H.F.W. (1988). Onderwijskundige evaluatie tele-evaluatieproject open universiteit 1987. Eindrapport SVO-project 6104. Maastricht: Rijksuniversiteit Limburg (Onderzoek van Onderwijs nr. 36).
- Bus, A.G. & Pijl, S.J. (Red.) (1987). Diagnostiek en leerlingbegeleiding. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, no. 14).
- Eijl, P.J. van, Cappetti, C.P.L.F., Merks, J.J. & Muyden, J.N. van (1988). Onderwijs in modulen. Over het hoe en waarom van modulair onderwijs. Groningen: Wolters-Noordhoff (Hoger Onderwijs Reeks).
- Franssen, H.A.M., Jager, A. & Rinkema-Kohlman, V.W. (1987). Denken over doelen. Een literatuurstudie naar het inventariseren, formuleren en classificeren van onderwijsdoelstellingen als bijdragen aan de eindtermendiscussie. Utrecht: ISOR.
- Harskamp, E. & Nijhof, W.J. (Red.) (1987). Curriculum: ontwerp, implementatie en evaluatie. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, no. 15).
- Harskamp, E.G. & Hoebe, W.Th.J.G. (Red.) (1988). Leermethoden in het onderwijs. Amsterdam/Lisse: Swets & Zeitlinger (Onderwijsresearch no. 7).
- Hofstee, W.K.B. (Red.) (1987). Evaluatiemethodologie. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, no. 16).
- Hofstera, M.L., Hobus, P.P.M., Boshuizen, H.P.A. & Schmidt, H.G. (1988). Diagnostiek bij huisartsen: het belang van gegevens uit de context van de patiënt. Maastricht: Rijksuniversiteit Limburg (Onderzoek van Onderwijs nr. 37).
- Jansweijer, W.N.H. (1988). PDP: een benadering vanuit de kunstmatige intelligentie van probleemoplossen en leren door doen in een semantisch rijk domein. (Proefschrift Universiteit van Amsterdam). Enschede: Sneldruk.
- Jong, J.A. de (1988). 'Je wordt aan het denken gezet'. Ontwikkeling en evaluatie van een model voor de schoolstage van de PABO. (Proefschrift RU Utrecht). Utrecht: ISOR.
- Kanselaar, G., Linden, J.L. van der & Pennings, A. (Red.) (1988). Begaafdheid: onderkenning en beïnvloeding. Amersfoort/Leuven: ACCO, f 39,00.
- Kelderman, H. (1987). Quasi-loglinear models for test and item analysis. Proefschrift Universiteit Twente.
- Meijnen, G.W., Nieuwenhuis, A.F.M. & Peschar, J.L. (Red.) (1987). Selectie en kwalificatie in het onderwijs. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, no. 17).
- Neuvel, J., Otter, M.E. & Bos, D.J. (1988). Functioneel aanvankelijk leesonderwijs: een beschrijving en een evaluatie. (SVO reeks nr. 86). Amsterdam/Lisse: Swets & Zeitlinger.
- Overwalle, F. van (1987). Causes of success and failure of freshmen at university: an attributional approach. Proefschrift Vrije Universiteit Brussel.
- Peters, J.J. & Tillema, H.H. (Red.) (1987). Scholing in onderwijs en bedrijf. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijs, no. 18).
- Scheerens, J. & Stoel, W.G.R. (Red.) (1987). Effectiviteit van onderwijsorganisaties. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, no. 19).
- Starren, J., Bakker, S.J. & Wissel, A. van der (1988). Inleiding in de onderwijspsychologie: instructie, beoordeling en behandeling. Muiderberg: Coutinho.
- Tielemans, J. (1988). Onderwijs en persoonlijkheidsontwikkeling. Leuven/Amersfoort: ACCO, f 61,75.

- Borsboom, A.M.R., Bosch, W.G., Verhagen, P.W., (1988). Voorstudie proefschole nieuwe media. Enschede: TO.
- Rupp, J. & Jungbluth, P. (red) (1988). Algemene basisvorming; Schept de WRR betere onderwijskansen? Nijmegen: ITS.
- Buurhe, W. & Coopmans, J. (1988). Wat leren kinderen op school? Nijmegen: ITS.
- Dekkers, H. & Don, R. (1988). Allochtonen en randgroepjongeren in het volle-tijds KMBO. Nijmegen: ITS.
- SURF Bulletin, juli 1988.
- Luykx, M. (1988). Schoolsucces in Turkse en Marokkaanse meisjes. Utrecht: ISDR.
- COA (1988). Verslag van activiteiten 1987.

BULLETIN

**ORGAAN VAN DE VERENIGING
VOOR ONDERWIJSRESEARCH**

Redactie-adres

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

Jaargang 12

1988

VOR:

Convocatie ledenvergadering VOR	110
Beknopt verslag verenigingsjaar 1987	111
VOR-bijdrage aan inhoudelijke bijeenkomsten	112
VOR begroting 1988	114
Toelichting op de VOR-begroting 1988	115
Agenda VOR-ledenvergadering	299
Verslag VOR-ledenvergadering 31 mei 1988	300
Verslag VOR-ledenvergadering, dd 10-11-1988	368
Oproep voor nieuwe bestuursleden VOR-bestuur	369
VOR begroting 1989	370
Toelichting op de VOR-begroting 1989	371

Artikelen:

Kwaliteitsbepaling wetenschappelijke productie in de Onderwijskunde <i>J. Ax i.s.m. B. Creemers</i>	118
Outputbepaling onderwijswetenschappen op het juiste spoor? <i>J.F.M. Claessen</i>	60
OTC- en ANGO onderzoek, de sluitpost op de SVO-begroting? <i>J. Claessen</i>	372
Output: kwaliteit of kwantiteit <i>P.C. van der Sijde</i>	294

Congressen, Symposia, Cursussen, Trainingen	122, 177, 241, 296, 373
Mededelingen	126, 297

ORD:

De gastsprekers tijdens de ORD-Leuven	116
Oproep Prof. Dr. H.C.J. Duijkerfonds	240
Proefschriften	62, 127, 176
Ontvangen publicaties	128, 179, 244, 298, 374