

Tijdschrift voor onderwijsresearch 1989

<https://hdl.handle.net/1874/208867>

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591 X

Jaargang 14
1989

Swets & Zeitlinger B.V.
Lisse

Redactie:

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

BIBLIOTHEEK DER
RIJKSUNIVERSITEIT
UTRECHT



VERENIGING VOOR ONDERWIJSRESEARCH

VWB 620

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591 X

Jaargang 14
1989

Swets & Zeitlinger B.V.
Lisse

Redactie:

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

BIBLIOTHEEK DER
RIJKSUNIVERSITEIT
UTRECHT



VERENIGING VOOR ONDERWIJSRESEARCH

Artikelen

Leren optellen en aftrekken: een longitudinaal onderzoek. <i>J.J. Beishuizen, M. Beishuizen en E.J.H.M. Felix</i>	91
Functionele schrijfp opdrachten en stel opdrachten: een onderzoek naar de construct-validiteit van schrijfvaardigheidsmetingen. <i>Huub van de Bergh</i>	151
Bekroond ORD paper 1989: Effecten van een metacognitief trainingsprogramma. <i>Harm J.A. Biemans</i>	286
Vakinhoudelijke kennis en strategische vaardigheden bij het oplossen van rekenopgaven. <i>M. Bloemkolk</i>	111
Schrijfvaardigheid en het schrijf-denkproces. <i>Iris Breetvelt</i>	265
Complexiteit van rekenopgaven en kinderen met leermoeilijkheden. <i>C. Hamaker, J. Hoeks, M. Ekkebus en I. van Kampen</i>	77
Veranderingsdiagnostiek bij handig rekenen. <i>M.K. van der Heijden</i>	103
Gebruiksmogelijkheden van het Learner-Report. <i>Baukje J. van Kesteren</i>	13
Schoolkenmerken en leerlingfunctioneren: een replicatie-onderzoek. <i>Anja W.M. Knuver</i>	329
De effecten van een opleiding die tot doel heeft a.s. leraren te leren reflecteren. <i>F.A.J. Korthagen en Th. Wubbels</i>	214
Vergelijking van scoringsprocedures bij schaalconstructie geïllustreerd met de vragenlijst van Slade en Jenner. <i>P.M. Kroonenberg, B.C.S. Pangemanann en D. Bakker</i>	297
Computergeleid Enquêteren: een overzicht van technologische vernieuwingen. <i>E.D. de Leeuw</i>	201
Prestaties van jonge en oudere volwassenen op geheugentaken: Uitkomsten van een meta-analyse. <i>Jules Pieters en Ellen van Bommel</i>	3
Onderwijs en Maatschappelijke Ongelijkheid, een Overzicht van 25 jaar theorievorming en onderzoek in Nederland. <i>Jan C.C. Rupp en Anton A. Wesselingh</i>	237
Psychologisch onderzoek van het aanvankelijk rekenen. <i>J.A.C. Sandberg, C. Hamaker en J.J. Beishuizen</i>	65
De oordelen van kinderen over zelfverzekerd gedrag van volwassenen. <i>H.J.G. Soppe en M. van Heiningen-Scali</i>	181
Een onderzoek naar de effecten van auditieve feedback op leesfouten. <i>G.W.G. Spaai, P. Reitsma en H.H. Ellermann</i>	338
Schools over the Gangway: An Experiment on Response Improving-Procedures. <i>Cor Suhre</i>	172
Vergrijzing in het onderwijs? Een verklaring voor de ontwikkeling van de leeftijd van leerkrachten. <i>J. de Voogd en G. den Broeder</i>	30
Effecten van tweetalig onderwijs aan Marokkaanse kleuters. <i>E. Wagenaar en E.M. Scholte</i>	227
Niet-lineaire monotone transformaties van de afhankelijke variabele: Wanneer geoorloofd en wanneer niet? <i>Charlotte J. de Wolff</i>	137

Notities en Commentaren

Test-toetsgebruik in programma-evaluatie: herziene richtlijnen. <i>H.P.G.M. Caminada</i>	348
Normaal Gronings peil? Een reactie op Pijl & Harskamp: Peil van PPON-onderzoek kan hoger. <i>Theo J.H.M. Eggen, Norman D. Verhelst en Johan M. Wijnstra</i>	356
A Nonlinear Correction on Linear Equating. <i>Dato N.M. de Gruijter</i>	259
Programmering van onderwijsonderzoek: een kritisch commentaar op het beleid van SVO. <i>W.Th.J.G. Hoebe</i>	312
Evaluatie van de basisvorming. Commentaar op het kader voor het uitvoeringsplan. <i>Wijnand Th.J.G. Hoebe</i>	44

Vorrang aan achterstand? Een commentaar op het gelijknamige ARBO-advies. G.W. Meijnen	188
Over de evaluatie van de basisvorming. Een reactie op Hoeben. Jules L. Peschar	47
Peil van PPON-onderzoek kan hoger. Y.J. Pijl en E.G. Harskamp	352
De vorming van de controlegroep bij een intact veldexperiment. E.M. Scholte	306

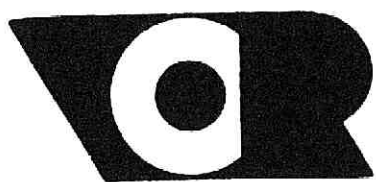
Boekbesprekingen

D. Brugman: <i>Personaliseren van de leerstof. Een interventie-onderzoek naar het ontwikkelen van waardegebieden door leerlingen tijdens lessen sociale-wereldoriëntatie.</i> (Ton Mooij)	317
W. Gijselaers: <i>Kwaliteit van het onderwijs gemeten.</i> (J.W. Willems)	364
E.G. Harskamp: <i>Rekenmethoden op de proef gesteld.</i> (L. Verschaffel)	319
W.H.N. Jansweijer: <i>PDP; een benadering vanuit de kunstmatige intelligentie van probleemoplossen en leren door doen in een semantisch rijk domein.</i> (H.P.A. Boshuizen)	193
A.J.M. de Jong: <i>Kennis en het oplossen van vakinhoudelijke problemen (een voorbeeld uit een natuurkundig domein).</i> (A. Pilot)	360
J.A. de Jong: <i>'Je wordt aan het denken gezet'. Ontwikkeling en evaluatie van een model voor de schoolstage van de PABO.</i> (F.K. Kieviet)	368
J.B.J. Koelman: <i>Kosten van de verzuiling; een studie over het lager onderwijs.</i> (J. Dronkers)	191
Pieter de Koning: <i>Programmadiifferentiatie in het voortgezet onderwijs.</i> (Anton Wesselingh)	323
C.J.W. Meijer: <i>Verwijzingen gewogen. Een studie naar determinanten van verwijzing naar speciaal onderwijs.</i> (M.J.C. Mommers)	195
A. Schouten: <i>Denken en doen in het technisch onderwijs: Leerling-strategieën bij het uitvoeren van praktijkopdrachten.</i> (J.M. Pieters)	324
J.F.M.J. Van Hout: <i>Onderzoekers in opleiding, een verklaringsmodel voor problemen van promotie-assistenten en assistenten in opleiding.</i> (J.J.A. Tacq)	366
M.P.C. van der Werf: <i>Het schoolwerkplan in het basisonderwijs: ontwikkeling, implementatie en opbrengst.</i> (Jan J.H. van den Akker)	52
Johan Zuidema: <i>Efficiënt spellingonderwijs: een leer- en expertmodel voor het spellen.</i> (Paul Eling)	362

Redactionele medewerkers

In 1989 werd redactionele medewerking verleend door:

C.A.J. Aarnoutse	KU Nijmegen
J. Bakker	KU Nijmegen
P.H. Been	RU Groningen
J.M.F. ten Berge	RU Groningen
R.J. Bosker	UT Enschede
J.F.A. Braster	EU Rotterdam
F.B. Brokken	RU Groningen
A.G. Bus	RU Groningen
H. Doddema Winsemius	RU Groningen
J. Dronkers	KUB Tilburg
P. van den Eeden	VU Amsterdam
J. Hartog	Uv Amsterdam
J. Hauer	RU Utrecht
P. van der Heijden	RU Leiden
H.J.M. Hüttner	KU Nijmegen
M.H. van IJzendoorn	RU Leiden
G.G.H. Jansen	RU Groningen
M.J. de Jong	EU Rotterdam
R. de Jong	RION Groningen
P. Jungbluth	ITS Nijmegen
W.A.M. Kok	RU Utrecht
G. Lang	RU Groningen
G.J. Mellenbergh	Uv Amsterdam
M.J.C. Mommers	KU Nijmegen
T. Mooij	ITS Nijmegen
J.H.L. Oud	KU Nijmegen
S.J. Pijl	RION Groningen
Y.J. Pijl	RION Groningen
J. Rispens	RU Leiden
J.M.M. Ritzen	EU Rotterdam
E. Roede	SCO Amsterdam
J. Scheerens	UT Enschede
J.H. Slavenburg	SAD Rotterdam
T.A.B. Snijders	RU Groningen
N.H. Veldhuijzen	CITO Arnhem
J.C. Verhoeven	KU Leuven
H. Verstralen	CITO Arnhem
K.J. Westerhof	RU Groningen



BULLETIN

**ORGAAN VAN DE VERENIGING
VOOR ONDERWIJSRESEARCH**

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

Jaargang 13

1989

VOR:

Convocatie ledenvergadering VOR	120, 262
Beknopt verslag verenigingsjaar 1988	121
De financiën van de VOR in 1988	122
Verslag VOR-ledenvergadering d.d. 25 mei 1989	198
Verslag VOR-ledenvergadering d.d. 30 november 1989	372
VOR-begroting 1990	375
Toelichting op de VOR-begroting 1990	376

Artikelen:

Evaluatie van kwaliteit of beoordeling van onderzoekers? <i>Wijnand Th.J.G. Hoebe</i>	131
De meting van wetenschappelijke productiviteit <i>W.J. van der Linden</i>	56
Congressen, Symposia, Cursussen, Trainingen	60, 128, 263, 377
Mededelingen	126
Ontvangen publikaties	63, 135, 199, 264, 382

Van de redactie

De redactie kijkt met tevredenheid terug op 1988. Het Tijdschrift voor Onderwijsresearch krijgt hoe langer hoe meer de vorm die de redactie en de Vereniging voor Onderwijsresearch voorstond. Het hoofdaccent ligt op de verslaggeving van onderwijsonderzoek, dat theoretisch, methodisch-technisch en inhoudelijk interessant is voor een brede groep lezers.

In de eerste plaats zijn dat onderwijsonderzoekers die lid zijn van de Vereniging voor Onderwijsresearch, daarnaast andere belangstellenden in het onderwijsonderzoek.

Behalve verslagen van onderzoek wil de redactie bijdragen opnemen waarin methodisch-technische problemen behandeld worden, waarin een overzicht van onderzoek op een bepaald terrein wordt gegeven of tutorials waarin auteurs lezers introduceren in nieuwe methodologisch-technische ontwikkelingen.

Ondanks de vergroting van het aanbod aan kopij slaagt de redactie erin goede bijdragen vrij kort na de eerste indiening te kunnen plaatsen. Dat is vooral een gevolg van het feit dat auteurs, beoordelaars en redactie er naar streven herschrijfronden op korte termijn af te wikkelen. Dit is bovendien de beste garantie dat een herschreven manuscript in tweede instantie geaccepteerd kan worden. Immers, wanneer de tijdspanne tussen de eerste en tweede versie te groot wordt, komen snel nieuwe kritiekpunten naar voren. Het is dan ook een ervaringsgegeven van de redactie dat een manuscript ofwel snel wordt bijgesteld, opnieuw wordt ingediend en geplaatst, ofwel lang op zich laat wachten, zodat een beoordeling in een tweede ronde meestal weer nieuwe problemen met zich mee brengt. Vandaar dat de redactie auteurs oproept om manuscripten in te zenden en om de tweede versie van het manuscript, na ontvangst van de kritiek, opnieuw in te sturen. Hierbij kan herhaald worden wat vorig jaar is opgemerkt: het Tijdschrift voor Onderwijsresearch moet niet alleen een tijdschrift *voor* onderwijsonderzoekers worden, maar ook een tijdschrift *van* onderwijsonderzoekers.

Eveneens moet worden opgemerkt dat er nog nooit een manuscript is geaccepteerd zonder dat opmerkingen zijn gemaakt waarmee bij het gereedmaken van de uiteindelijke versie rekening moest worden gehouden. In het merendeel van de gevallen wordt het manuscript in eerste instantie afgewezen, hetgeen ertoe leidt dat uiteindelijk herschreven versies van manuscripten kunnen worden geplaatst, ook indien op de eerste versie zware kritiek werd uitgeoefend.

Is de redactie tevreden over de balans van verschillende soorten bijdragen, die in de afgelopen jaren is ontstaan, minder tevreden is zij over het bereiken van het aantal doelstellingen, dat zij zich voor 1988 had gesteld. Het is nog niet mogelijk gebleken een overzichtsartikel te plaatsen. Goede overzichtsartikelen zijn volgens de redactie van uitzonderlijk belang voor onderwijsonderzoek, omdat ze richting kunnen geven aan toekomstige programmering en onderzoek op het betreffende terrein en onderzoekers op dergelijke overzichten terug kunnen vallen. Willen ze deze functie kunnen vervullen, dan dienen ze van grote kwaliteit te zijn, zowel in het overzicht van het onderzoek dat ze bieden, als in de analyse en de conclusies op basis van deze onderzoeken. In elk geval kost het de auteurs die momenteel bezig zijn met dergelijke reviews voor te bereiden meer tijd dan de redactie en wellicht ook zichzelf hadden gedacht.

Ook het samenstellen van themanummers blijkt meer voeten in de aarde te hebben. In 1988 is nog geen themanummer verschenen. Het ziet er naar uit dat nummer twee van deze jaargang een themanummer op het terrein van het rekenonderwijs zal zijn. Het gereedmaken en het beoordelen van de manuscripten ten behoeve van publikatie kost bij een themanummer minstens evenveel tijd als bij een normale aflevering van het tijdschrift. Dit, omdat behalve de afzonderlijke bijdragen, ook het geheel door de redactie beoordeeld wordt, terwijl bovendien het 'zwakste' manuscript en de 'traagste' auteur de snelheid van het verschijnen van de overige bijdragen bepalen. Bij themanummers kan gewerkt worden met gastredacteuren. De uiteindelijke verantwoordelijkheid voor een dergelijk nummer blijft bij de redactie van het tijdschrift. Om dit te realiseren vindt behalve beoordeling van de bijdragen door de gastredacteuren ook beoordeling door de redactie plaats en eventueel door haar ingeschakelde beoordelaars. Wanneer onderzoekers een themanummer willen voorbereiden, verzoekt de redactie hen zich in een vroeg stadium met de redactiesecretaris in verbinding te stellen om de uitgave voor te bereiden. Het is realistisch daarbij uit te gaan van ongeveer een periode van een jaar tussen idee en werkelijkheid.

Tegenover niet gerealiseerde plannen staat dat, volgens de aanduiding op de rug van de omslag, het Tijdschrift voor Onderwijsresearch een deel van het jaar nog in 1987 heeft verkeerd. Mocht dit ook voor de komende jaren gelden dan worden de plannen van de redactie met terugwerkende kracht alsnog gerealiseerd.

De samenwerking met het VOR-bulletin verloopt optimaal. Het VOR-bulletin stelt regelmatig pagina's ter beschikking voor het Tijdschrift voor Onderwijsresearch. Overigens, gezien het mededelend karakter van het merendeel van de bijdragen in het VOR-bulletin (behalve de onlangs gestarte discussie naar aanleiding van de voorstellen van Jan Ax en Bert Creemers over de beoordeling van output) is van vermenging met, om niet te spreken van verlaging van het niveau van, het tijdschrift, zoals door sommige lezers werd gevreesd, naar de mening van de redactie geen sprake.

De rubriek boekbesprekingen is het afgelopen jaar duidelijk gegroeid. De redactie streeft ernaar in elk geval de bespreking van (empirische) proefschriften op het terrein van het onderwijs binnen redelijke termijn na verschijning te doen plaatsvinden. Willen (a.s.) promovendi daartoe hun proefschrift opsturen aan:
Dr. M. J. M. Voeten, Instituut voor Onderwijskunde, Postbus 9103, 6500 HD Nijmegen.

In de samenstelling van de redactie zijn met ingang van 1 januari 1989 een aantal wijzigingen opgetreden. Peter Weeda heeft de redactie verlaten. Zijn plaats wordt ingenomen door Bernadette van Hout Wolters. Namens het bestuur van de VOR heeft Gellof Kanselaar in de redactie zitting genomen.

Prestaties van jonge en oudere volwassenen op geheugentaken: Uitkomsten van een meta-analyse

Jules Pieters en Ellen van Bommel
Universiteit Twente

ABSTRACT

This article reviews studies on memory performance in young and older adults. Many studies have been reported in cognitive psychological literature on perception, memory and problem solving. These studies express clear differences between young and older adults. We performed a meta-analysis in which studies on free recall, cued recall and serial recall are compared and effect sizes are achieved. Other variables taken into study were: immediate or delayed recall, intentional or incidental learning, and learning materials. Young adults showed superior memory performance, except for learning texts. In the final section relevance of the outcome of this study and of application of meta-analysis are discussed.

INLEIDING

Sinds 1970 wordt veel onderzoek verricht naar het leren en het verwerken van informatie door volwassenen, met name oudere volwassenen. In de cognitief-psychologische literatuur over perceptie, geheugen en probleemoplossen wordt over veel onderzoek gerapporteerd waarin duidelijke verschillen in informatieverwerking tussen jongeren en ouderen worden aangetoond.

Uit deze onderzoeken blijkt dat veroudering niet zozeer een effect heeft op de capaciteit van het korte-termijn geheugen, zoals vaak wordt beweerd, maar veeleer op de werking van het lange-termijn geheugen (Craik, 1977; Craik & Rabinowitz, 1983). Relatief weinig aanwijzingen zijn in de literatuur te vinden omtrent een vermindering in capaciteit van het korte-termijn geheugen. Craik (1977) constateert op grond van een literatuur-overzicht dat in die gevallen waar een verschil in capaciteit van het korte-termijn geheugen tussen jonge en oudere volwassenen wordt aangetroffen, dit meestal kan worden toegeschreven aan voor ouderen zwaardere taakomstandigheden, bijvoorbeeld: manipulatie of reorganisatie van de aangeboden informatie is vereist of er is sprake van een verdeling van de aandacht.

Een effect van veroudering op het lange-termijn geheugen kan het gevolg zijn van problemen bij de opname van informatie, bij de opslag van informatie of bij het ophalen van informatie uit het geheugen. Interessant in dit verband is het onderzoek van Schonfield & Robertson (1966). Zij vergeleken herinneringsprestaties en herkenningsprestaties bij proefpersonen van verschillende leeftijdsgroepen. Herinneringsprestaties namen af met de leeftijd, terwijl de herkenningsprestaties nagenoeg gelijk bleven over de jaren heen. Op grond hiervan kan worden geconcludeerd dat de problemen zich met name voordoen bij het ophalen van informatie uit het geheugen (zoals bij herinneringstaken) en niet zozeer bij de opname van informatie, die immers voor beide taken dezelfde is. Hartley, Harker & Walsh (1982) stellen in een overzicht van onderzoek op dit terrein dat een conclusie als deze moet worden genuanceerd. De organisatiegraad van het te leren

Adres:

Faculteit Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede

materiaal speelt bijvoorbeeld een rol, evenals het gebruik van 'retrieval cues' en de aanwezigheid van 'category cues'. Bovendien dragen Tulving & Thomson (1973) evidentie aan voor het feit dat de wijze waarop informatie wordt opgenomen in het geheugen nauw gerelateerd is aan de wijze waarop informatie uit het geheugen moet worden opgehaald.

Ten einde inzicht te verkrijgen in de complexe materie van de verandering van geheugenprestaties over leeftijd is een meta-analyse uitgevoerd op onderzoek naar de vergelijking tussen geheugenprestaties van jongeren en ouderen. Deze meta-analyse is uitgevoerd in het kader van een onderzoek naar het leren van volwassenen met de computer. Centraal in het onderzoek staat de vraag welke verschillen in informatieverwerking er bestaan tussen jongeren en ouderen, die belangrijk genoeg zijn om instructieregels voor volwassenen op te baseren. Het onderzoek waarover hier wordt gerapporteerd heeft betrekking op de verschillen tussen jongeren in de leeftijdsklasse 20-30 jaar en ouderen in de leeftijdsklasse 60-80 jaar. De taken waarop jongeren en ouderen worden vergeleken zijn: vrije herinneringstaak ('free recall'), gestuurde herinneringstaak ('cued recall') en seriële herinneringstaak ('serial recall').

De term meta-analyse werd voor het eerst gebruikt door Glass (1976) om een statistische analyse aan te duiden waarmee een groot aantal onderzoeksresultaten op integratieve wijze kan worden beschreven. Bekende studies waarin meta-analyse-technieken zijn gebruikt, zijn van Glass (Glass & Smith, 1978) naar de relatie tussen klassegrootte en schoolprestaties, van Kulik en Kulik over de effectiviteit van Computer Based Education binnen het primaire, secundaire en volwassenenonderwijs (Kulik, Kulik & Bangert-Drowns, 1985; Bangert-Drowns, Kulik & Kulik, 1985; Kulik, Kulik & Shwalb, 1986) en van Klauer (1984) over incidenteel versus intentioneel leren. Aan de uitvoering van een meta-analyse is een aantal fasen te onderscheiden. Allereerst wordt onderzoeksliteratuur gezocht aan de hand van gespecificeerde procedures. Daarna worden kenmerken van de gevonden onderzoeken gekarakteriseerd op een kwantitatieve wijze en wordt het resultaat van het onderzoek uitgedrukt in een effectgrootte. Tot slot kunnen multivariate technieken worden gebruikt om uitkomsten aan de kenmerken te relateren.

De meta-analyse-techniek is een hulpmiddel om resultaten van een groot aantal onderzoeken te integreren tot een samenvattend resultaat. De techniek dient echter bij voorkeur gebruikt te worden binnen een duidelijk theoretisch kader. Dit kader kan dan aangeven welke kenmerken van de onderzoeken relevant zijn om te beschouwen in relatie tot de gevonden effecten. Het onderzoek naar de vergelijking tussen herinneringsprestaties van jongeren en ouderen wordt uitgevoerd vanuit verschillende theoretische gezichtspunten. In de door ons uitgevoerde meta-analyse zijn deze verschillende theoretische uitgangspunten niet meer terug te vinden, aangezien wij gezocht hebben naar overeenkomst in taken en niet in theoretisch kader. Dit doet naar onze mening niets af aan de bruikbaarheid van de techniek.

In de meta-analyse die door ons is uitgevoerd is onderscheid gemaakt in: soort taak (vrije herinneringstaak ('free recall'), gestuurde herinneringstaak ('cued recall') en seriële herinneringstaak ('serial recall'), aard van het te onthouden materiaal (teksten, woorden, letters, cijfers, gebeurtenissen, en tekeningen en foto's), onmiddellijke of uitgestelde herinnering, en intentioneel of incidenteel leren. De onderzochte en vergeleken leeftijdsgroepen waren: 20 - 30 jaar, en 60 jaar en ouder.

METHODE

Literatuurverzameling

Voor de literatuurverzameling is gebruik gemaakt van het literatuurbestand ERIC

(Educational Research Information Centre), relevante tijdschriften op het terrein van 'psychology and aging' en een aantal handboeken over van leren en ontwikkeling van volwassenen.

Het zoekproces binnen ERIC leverde slechts 17 bruikbare verwijzingen op en werd verstoord door de grote hoeveelheid literatuur waar verschillen tussen kinderen en volwassenen werden onderzocht.

De literatuuroverzichten in de tijdschriften en handboeken gaven een beter overzicht. Ruim 200 titels konden op deze wijze verzameld worden. De onderzoeken bestrijken een te breed scala aan onderwerpen en deze zijn te divers om in één meta-analyse geïntegreerd te worden. In dit artikel wordt gerapporteerd over de meta-analyse uitgevoerd naar het herinneren van volwassenen. De selectie uit het grote aantal onderzoeken vond plaats aan de hand van de volgende criteria:

- In de onderzoeken dient een groep oudere volwassenen (60 jaar en ouder) vergeleken te worden met een groep jonge volwassenen (20 - 30 jaar);
- De proefpersonen worden vergeleken op een herinneringstaak;
- De proefpersonen dienen lichamelijk en geestelijk gezonde volwassenen te zijn;
- Het onderzoek dient na 1975 te zijn uitgevoerd.

Aan deze voorwaarden bleken 80 studies te voldoen.

Procedure

Van de verzamelde literatuur bleken 27 publicaties informatie te geven met betrekking tot leeftijdsverschillen bij het herinneren van volwassenen, die gebruikt konden worden in een meta-analyse. In deze publicaties worden 36 onderzoeken gerapporteerd. Met behulp van gegevens uit deze onderzoeken konden 88 effectgroottes berekend worden. De effectgroottes zijn berekend met behulp van de door Glass gegeven formule:

$$\delta = \frac{M_e - M_c}{S_c}$$

waarbij:

δ : effectgrootte

M_e : de gemiddelde score van de experimentele groep

M_c : de gemiddelde score van de controle groep

S_c : de standaarddeviatie van de controle groep

Er is gebruik gemaakt van de standaarddeviatie van de controlegroep (S_c , de jongeren) omdat vrijwel nooit een gecombineerde standaarddeviatie gerapporteerd wordt. Indien dat wel het geval is, kan in plaats van S_c , S_d , de gecombineerde standaarddeviatie, gebruikt worden. De standaarddeviatie van de experimentele groep (in dit geval de ouderen) is beïnvloed door het ouder worden en de spreiding onder ouderen is in het algemeen veel groter dan bij de jongeren. Vanwege vertekening die hier mogelijk het gevolg kan zijn, is gekozen voor de standaarddeviatie van de controlegroep.

Wanneer in de artikelen geen gemiddelde scores en standaarddeviaties gerapporteerd worden, wordt δ benaderd met behulp van $2\sqrt{df}$, indien een t-toets uitgevoerd werd, of door $2\sqrt{F/df_{\text{error}}}$, indien een F-toets gerapporteerd werd (Hedges & Olkin, 1986). Laatste formules geven een goede benadering van δ wanneer het aantal proefpersonen in beide groepen gelijk is en de variantie in de controle groep gelijk is aan die van de experimentele groep.

Tabel 1 geeft een overzicht van de in de meta-analyse betrokken experimenten. Van ieder experiment is naast de effectgrootte, waarover in de resultaten nader wordt gerapporteerd, de soort herinneringstaak, het aantal proefpersonen en de gemiddelde leeftijd van de

Tabel 1

Effectgroottes en onderzoekenmerken

Gordon & Clark	-0,84 1 1 1 1 22 71 25	Sanders, Murphy et al.	-2,48 2 1 1 1 20 74 24
Light, Zelinski & Moore	I -0,61 1 2 1 1 52 68 19	Schneider, Gritz & Jarvik	I -0,83 1 1 1 1 13 70 20
	II -1,54 1 2 1 1 72 69 26		-2,29 5 1 1 1 13 70 20
	III -1,03 1 2 1 1 24 68 25		-2,14 5 2 1 1 13 70 20
Wright	-1,99 4 3 9 1 12 68 19		-0,40 2 2 1 1 13 70 20
Taub & Kline	0,0 1 1 1 1 18 67 28		-1,96 2 2 1 1 13 70 20
	-0,32 1 1 1 1 18 67 28	II	-1,83 1 1 2 1 10 69 21
	-0,47 1 1 1 1 18 67 28		-1,40 5 1 2 1 10 69 21
	-0,63 1 1 1 1 18 67 28		-2,88 5 2 2 1 10 69 21
Erber, Herman & Botwinick	-0,36 2 1 1 2 12 69 24		-0,91 2 2 2 1 10 69 21
	-1,71 2 1 1 1 12 69 24	Shaps & Nilsson I	-1,07 2 2 2 1 10 69 21
	-1,30 2 1 1 1 12 69 24		-1,29 2 2 1 1 20 76 19
	-1,00 2 1 1 2 12 69 24	Till & Walsh	II -0,71 2 2 2 1 20 75 18
	-1,28 2 1 1 1 12 69 24	I	-0,19 1 1 1 1 9 68 20
Harwood & Naylor	-0,94 5 1 2 1 58 99 99		-1,22 1 2 1 1 8 68 20
Howard	-0,56 2 1 1 2 20 70 31		-0,35 1 1 1 1 9 68 20
Howard, McAndrews & Lasaga	-1,01 2 1 1 1 24 70 28		-0,73 1 2 1 1 8 68 20
	-1,34 2 1 1 1 24 70 28	II	-0,35 1 1 1 1 8 72 19
Kausler & Hakami	-1,87 5 1 1 2 12 66 19		-1,24 1 2 1 1 8 72 19
	-1,99 5 1 1 1 12 66 19		-0,32 1 1 1 1 8 72 19
Kausler & Puckett	-1,76 2 3 1 1 20 70 19		-0,74 1 2 1 1 8 72 19
	-0,77 2 3 1 1 20 70 19		0,44 4 1 1 1 8 72 19
	-1,02 2 3 1 1 20 70 19		0,52 4 2 1 1 8 72 19
Macht & Buschke	-0,97 2 2 1 1 12 66 32	III	0,04 1 1 1 1 7 68 21
Mueller, Rankin & Carlomusto	-1,49 2 1 1 1 12 69 19		-0,97 1 1 1 1 9 68 21
	-1,46 2 2 2 1 12 69 19		-0,98 1 2 1 1 7 68 21
	-1,03 2 2 2 1 12 69 19		0,09 1 2 1 1 10 68 21
	-0,75 4 1 1 1 12 69 19		0,22 1 2 1 1 9 68 21
Murphy, Sanders et al.	-1,76 5 3 1 1 23 69 20	Inman & Parkinson	-0,69 2 2 3 1 27 72 21
	-1,36 5 3 1 1 23 69 20		-0,76 3 1 2 1 27 72 21
Parkinson, Lindholm & Inman	-1,22 4 1 1 1 26 71 20	Barrett & Wright	-0,70 3 3 1 1 27 72 21
	-1,81 2 1 1 1 23 71 20		-1,47 2 1 2 2 12 70 21
	-1,52 2 1 1 1 26 71 20		-1,80 2 1 2 2 12 70 21
Pavur, Comeaux & Zeringue	-0,41 2 1 1 1 18 67 20	Schonfield, Davidson & Jones	0,30 2 1 2 2 12 70 21
	-0,80 2 1 1 1 18 67 20		0,74 2 1 2 2 12 70 21
Perlmutter	I -0,80 2 1 1 1 24 63 20	Rabinowitz & Craik	I -1,42 2 2 2 1 12 67 21
	II -1,32 2 1 1 1 24 62 20		-1,43 2 1 2 2 12 67 21
	-1,08 2 1 1 2 24 62 20		-1,29 2 2 2 2 12 67 21
Rabinowitz	-1,23 2 2 1 1 24 70 20	II	-0,35 2 1 2 2 16 71 20
	-1,25 2 2 2 1 24 70 20	Craik & Rabinowitz	-1,70 2 1 2 1 24 68 20
	-1,47 2 2 2 1 24 70 20	Light & Anderson	-1,25 1 2 1 1 25 69 26
			-0,83 2 3 1 1 25 69 26
			-0,46 4 3 1 1 25 69 26
			-0,55 4 3 1 1 25 69 26

Noot.

kolom 1 effectgrootte	
kolom 2 materiaal	1 tekst
	2 woorden
	3 letters
	4 cijfers
	5 anders (tekeningen, gebeurtenissen)
kolom 3 soort recall	1 free
	2 cued
	3 serial
kolom 4 tijd	1 onmiddellijke recall (geen tijd tussen aanbiedings- en herinneringsfase)
	2 uitgestelde recall (tijd of tussentaak voor herinneringstaak)
kolom 5 instructie	1 intentioneel
	2 incidenteel
kolom 6 aantal proefpersonen	
kolom 7 gemiddelde leeftijd ouderen	
kolom 8 gemiddelde leeftijd jongeren	

proefpersonen in de oudere en in de jongere groep, aangegeven.

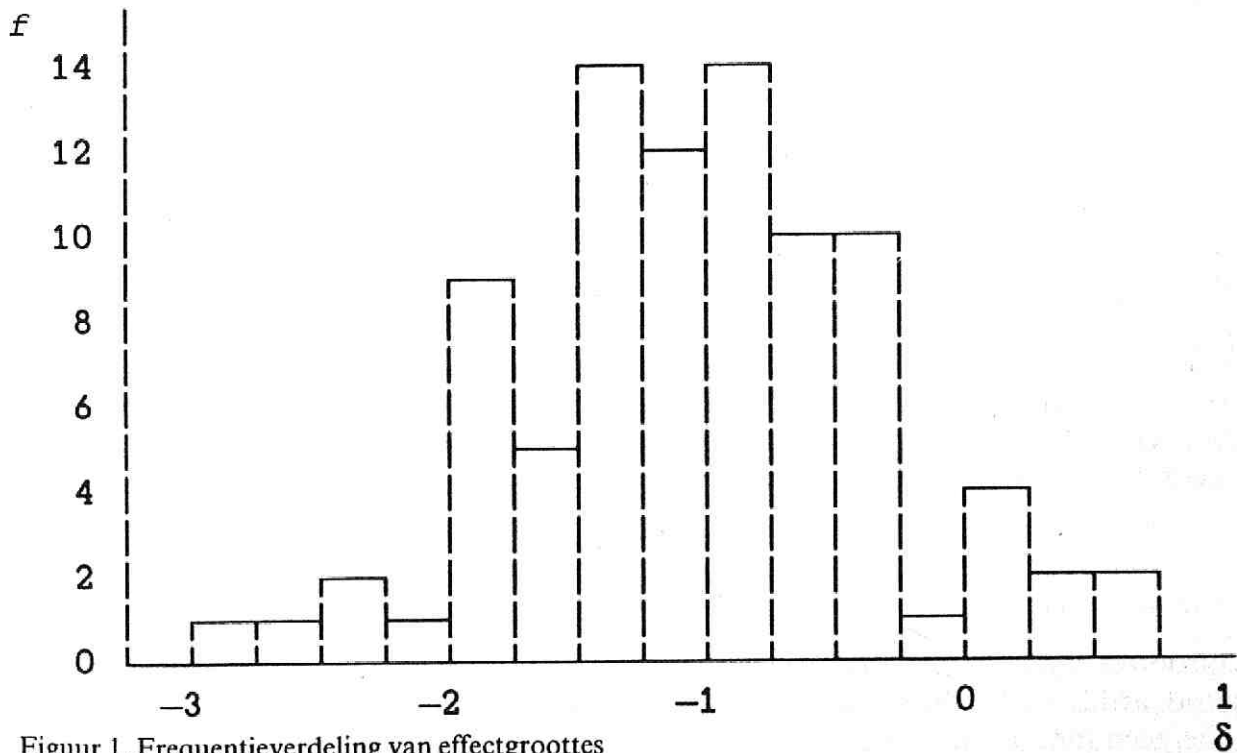
Van een experiment zijn alleen dan meerdere effectgroottes berekend wanneer het duidelijk was dat het om verschillende taken ging of dat het om verschillende proefpersonen ging.

Van de experimenten is een aantal kenmerken nader beschouwd.

De gemiddelde steekproefomvang bedraagt 17 proefpersonen (17 proefpersonen in zowel experimentele als controle groep).

De gemiddelde leeftijd van de ouderen die in deze meta-analyse zijn betrokken, is 69 jaar; de jongeren zijn gemiddeld 22 jaar.

Zoals hierboven reeds is aangegeven is een aantal kenmerken van de herinneringstaak gecodeerd. De aard van het te herinneren materiaal is in kolom 2 van tabel 1 gecodeerd. 25 effectgroottes zijn berekend voor taken waarbij teksten herinnerd moesten worden, 44 voor het herinneren van woorden, 3 voor letters, 7 voor cijfers en 9 voor diverse andere



Figuur 1. Frequentieverdeling van effectgroottes

taken (het herinneren van gebeurtenissen of tekeningen). Een onderscheid aangebracht naar de soort herinneringstaak die uitgevoerd werd, leverde de volgende verdeling op: 49 effectgroottes zijn gebaseerd op een vrije herinneringstaak, 28 op een gestuurde herinneringstaak en 11 op een seriële herinneringstaak. Tevens is bepaald of het om een onmiddellijke of uitgestelde herinneringstaak ging (respectievelijk 62 en 25 effectgroottes) en of het om een intentionele herinneringstaak ging (75 effectgroottes) of om een incidentele taak (13 effectgroottes).

RESULTATEN

In tabel 1 staan bij elk experiment de effectgroottes vermeld. Een frequentieverdeling van alle effectgroottes is weergegeven in figuur 1. De δ 's variëren tussen -2,88 en +0,74. Het gemiddelde en de mediaan van deze verdeling zijn vrijwel gelijk, respectievelijk -1,04 en -1,02. De oudere volwassenen uit deze meta-analyse scoren gemiddeld één standaarddeviatie lager dan de jonge volwassenen.

Hiermee is tevens aangegeven dat 84% van de oudere volwassenen lager scoort op de verschillende herinneringstaken dan het gemiddelde van de groep jonge volwassenen.

Vervolgens is nagegaan in hoeverre de kenmerken (materiaal, type recall-taak, tijd tussen aanbieding en herinneren, incidenteel versus intentioneel, aantal proefpersonen, leeftijd van de experimentele groep respectievelijk controle-groep) die per experiment gecodeerd zijn correleren met de gevonden effectgroottes. Tabel 2 geeft de gevonden correlaties.

Tabel 2

Correlaties tussen effectgrootte en onderzoekskenmerken

	materia	recall	tijd	instru	n_ppn	lftd_exp	lftd_c
<i>effect</i>	-0,29*	-0,05	-0,17	0,03	-0,11	0,002	0,07

* $p < 0,01$

Uit tabel 2 blijkt dat alleen een significante correlatie bestaat tussen de aard van het te herinneren materiaal en de effectgrootte. In tabel 3 zijn de gemiddelde effectgroottes aangegeven per kenmerk van de studie.

Zoals eerder vermeld, correleert alleen het type 'materiaal' met de gevonden effectgroottes. De effecten per niveau van 'materiaal' blijken volgens een uitgevoerde variantie-analyse significant verschillend ($p < 0,001$). Gezien het beperkte aantal experimenten voor de niveaus letters, cijfers en divers en de onbetrouwbaarheid van de categorie 'divers' kan slechts het verschil bij 'tekst' en 'woorden' interessant en betekenisvol zijn. Dit verschil is significant ($p < 0,02$, t toets): ouderen presteren op beide taken minder dan jongeren maar de effectgrootte bij tekstherinnering is kleiner dan bij het herinneren van woorden.

DISCUSSIE

Effectgroottes zijn voor alle onderzochte kenmerken ongeveer gelijk en schommelen rond één standaarddeviatie. Dit betekent een aanmerkelijke vermindering van het prestatieniveau, hetgeen inhoudt dat oudere volwassenen op alle taken minder goed presteren dan

Tabel 3
Gemiddelde effectgroottes per kenmerk

effectgrootte bij materialen					
	tekst	woorden	letters	cijfers	divers
<i>effect</i>	-0,75	-1,12	-0,72	-0,57	-1,85
<i>n</i>	25	44	3	7	9
effectgrootte bij type recall					
	vrije	cued	serieel		
<i>effect</i>	-1,0	-1,07	-1,08		
<i>n</i>	49	28	11		
effectgrootte bij tijd tussen aanbiedings- en herinneringsfase					
	direct	delayed			
<i>effect</i>	-0,95	-1,2			
<i>n</i>	62	25			
effectgrootte bij type leertaak					
	intentioneel		incidenteel		
<i>effect</i>	-1,04		-0,99		
<i>n</i>	75		13		

jonge volwassenen.

Alleen het herinneren van teksten lijkt minder beïnvloed te worden door het ouder worden dan het herinneren van woorden.

Geen verschillen treden op of de taak intentioneel dan wel incidenteel geleerd wordt, of het om onmiddellijk of uitgesteld herinneren gaat en geen verschillen treden op tussen de verschillende soorten van herinneringstaken (vrije, gestuurde of seriële herinnering). Drie verklaringen zijn mogelijk voor de gevonden effecten bij de vergelijking tussen het bestuderen van teksten en van woorden. Ten eerste wordt in de literatuur over het leren van volwassenen herhaaldelijk geconstateerd dat ouderen beter presteren in betekenisvolle leersituaties. De onderzoekssituatie van de in onze meta-analyse opgenomen studies is bij het herinneren van teksten betekenisvoller dan bij het herinneren van woorden.

Ten tweede blijkt bij een kwalitatieve bestudering van de in onze meta-analyse opgenomen studies dat de taak bij het herinneren van woorden relatief zwaarder is. Grote aantallen niet-samenhangende woorden moeten daarbij onthouden worden.

Dit is tevens een onderscheid met de andere soorten materiaal die over het algemeen in minder omvangrijke taken onderzocht werden dan het herinneren van woorden. Het

herinneren van letters en cijfers vond in de studies binnen onze meta-analyse over het algemeen plaats in letter- of cijferspanne taken. De omvang van de taak blijft in deze gevallen beperkt tot maximaal 6 à 9 items.

Oudere volwassenen behalen op 'verbal ability tests' eenzelfde of een hogere score dan jonge volwassenen (Knox, 1977; Sternberg & Berg, 1987). In dit licht kan als derde verklaring voor het verschil in effectgrootte bij het bestuderen van teksten en van woorden worden genoemd dat de verbale vermogens van ouderen bij taken waar tekst herinnerd moet worden kunnen compenseren voor de achteruitgang in geheugenprestaties.

Wanneer de steekproefomvang gecodeerd wordt (in groep 1 alle onderzoeken met ≤ 10 proefpersonen, groep 2 tussen 11 en 20, groep 3 tussen 21 en 30 en groep 4 meer dan 30 proefpersonen) blijkt dat het effect significant lager is in de onderzoeken met minder dan 10 proefpersonen dan in de onderzoeken met meer dan 10 proefpersonen (gemiddeld -0,7 of -1,1). Hier gaat de geringe statistische betrouwbaarheid van experimenten met slechts een gering aantal proefpersonen waarschijnlijk een rol spelen. Kwalitatieve analyse van de positieve effectgroottes geeft een aanduiding van taken waar de oudere volwassenen in onze analyse niet minder presteren of zelfs beter presteren dan de jonge volwassenen. Het aantal positieve effectgroottes is echter zo gering in aantal dat hier eerder van een statistische kans op positieve waarde gesproken moet worden dan een werkelijk relevant aantal (8 van 88, 9%). Wel zijn de positieve effecten in een aantal gevallen eenvoudig te verklaren. Zo worden de prestaties van ouderen en jongeren in het onderzoek van Barrett en Wright (1981) vergeleken op een taak waarbij woorden herinnerd moesten worden die in het verleden gangbaar waren maar nu niet meer gebruikt worden. Zij constateerden betere geheugenprestatie van de oudere volwassenen. In andere gevallen gaat het om taken die weinig mentale belasting vergen, zoals cijfer- en letterspanne-taken.

CONCLUSIE

Het is aan te bevelen, op grond van de resultaten van deze meta-analyse, dat onderzoek uitgevoerd wordt om te bepalen op welke soort leertaken oudere volwassenen meer in het nadeel zijn dan jonge volwassenen. Het hier gepresenteerde onderzoek geeft daar slechts indirecte aanwijzingen voor.

Betekenisvolheid van het te leren materiaal en wellicht omvang van de leertaak zijn mogelijke verklaringen die uit dit onderzoek af te leiden zijn, maar dienen onderzocht te worden in onderzoek waar niet ouderen en jongeren vergeleken worden maar de prestaties van ouderen op verschillende taken. Over een dergelijke experimentele vergelijking is in de literatuur nog niet gerapporteerd.

Ten aanzien van de relevantie van de studies die in deze meta-analyse zijn gebruikt, voor het ontwerpen van instructiesystemen voor volwassenen, kan het volgende worden opgemerkt. Allereerst de leeftijd van de groep ouderen: gemiddeld 69 jaar. Hoewel de scholing van ouderen toeneemt, beperkt deze scholing zich tot voornamelijk volwassenen in de leeftijd tussen 25 en 50 jaar. De groep ouderen in deze onderzoeken is niet relevant voor vakgerichte om- en bijscholing maar meer voor vrije-tijdseducatie en voor recente ontwikkelingen zoals de senioren-universiteit (Utrecht).

Daarnaast is het opleidingsniveau van de proefpersonen meestal erg hoog. Het gaat om studenten en om oud-studenten van universiteiten. In veel onderzoeken worden scores op een aantal subtesten van de WAIS gegeven. De proefpersonen scoren in het algemeen hoog op deze testen; opvallend daarbij is dat in veel gevallen de ouderen zelfs hoger scoren dan de jongeren. Dit gebeurt met name op de 'verbal ability' test. Deze groep van hoog

geschoolden hebben reeds studiestrategieën ontwikkeld. Het is juist de groep van volwassenen met weinig (formele) leerervaringen die momenteel een beroep doet op allerlei vormen van volwassenenonderwijs en over wier leren weinig of niets bekend is (zie voor enige empirische aanwijzingen voor de wijze waarop deze groep van volwassenen leert: Pieters & Van Bommel, 1986, in druk).

LITERATUUR

- Bangert-Drowns, R.L., Kulik, J.A., & Kulik, C-L.C., (1985). Effectiveness of computer based education in secondary schools. *Journal of Computer Based Instruction*, 12, 59-68.
- Barrett, T.R., & Wright, M., (1981). Age-related facilitation in recall following semantic processing. *Journal of Gerontology*, 36(2), 194-199.
- Craik, F.I.M., (1977). Age differences in human memory. In: J.E. Birren & K.W. Schaie, *Handbook of the Psychology of Aging*. New York: Van Nostrand.
- Craik, F.I.M., & Rabinowitz, J.C., (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology*, 40(3), 309-315.
- Erber, J., Herman, T.G., & Botwinick, J., (1980). Age differences in memory as a function of depth of processing. *Experimental Aging Research*, 6(4), 341-348.
- Glass, G.V., (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glass, G.V., & Smith, M.L., (1978). *Meta-analysis of research on the relationship of class-size and achievement*. San-Francisco: Far West Laboratory for Educational Research and Development.
- Gordon, S.K., & Clark, W.C., (1974). Application of signal detection theory to prose recall and recognition in elderly and young adults. *Journal of Gerontology*, 29(1), 64-72.
- Hartley, J., Harwood, E., & Naylor, G.F.K., (1969). Recall and recognition in elderly and young subjects. *Australian Journal of Psychology*, 21(3), 251-257.
- Hartley, J.T., Harker, J.O., & Walsh, D.A., (1977). Contemporary issues and new directions in adult developments of learning and memory. In: L.W. Poon (Ed.), *Aging in the 1980's*. Washington: American Psychological Association.
- Hedges, L.V., & Olkin, I., (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Howard, D.V., (1983). The effects of aging and degree of association on the semantic priming of lexical decisions. *Experimental Aging Research*, 9(3), 145-151.
- Howard, D.V., McAndrews, M.P., & Lasaga, M.I., (1981). Semantic priming of lexical decisions in young and old adults. *Journal of Gerontology*, 36(6), 707-714.
- Inman, V.W., & Parkinson, S.R., (1983). Differences in Brown-Peterson recall as a function of age and retention interval. *Journal of Gerontology*, 38(1), 58-64.
- Kausler, D.H., & Hakami, M.K., (1983). Memory for activities: Adult age differences and intentionality. *Developmental Psychology*, 19(6), 889-894.
- Kausler, D.H., & Puckett, J.M., (1979). Effects of word frequency on adult age differences in word memory span. *Experimental Aging Research*, 5(2), 161-169.
- Klauer, K.J., (1984). Intentional and incidental learning with instructional text: A meta-analysis 1970-1980. *American Educational Research Journal*, 21, 323-339.
- Knox, A., (1977). *Adult learning and developments*. San Francisco: Jossey-Bass.
- Kulik, J.A., Kulik, C-L.C., & Bangert-Drowns, R.L., (1985). Effects of computer based education on elementary school pupils. *Computers and Human Behavior*, 1, 59-74.
- Kulik, C-L.C., Kulik, J.A., Shwalb, B.J., (1986). The effectiveness of computer based adult education: a meta analysis. *Journal of Educational Computing Research*, 2, 235-252.
- Light, L.L., & Anderson, P.A., (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology*, 40(6), 737-747.
- Light, L.L., Zelinski, E.M., & Moore, M., (1982). Adult age differences in reasoning from new information. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 8(5), 435-447.

- Macht, M.L., & Buschke, H., (1984). Speed of recall in aging. *Journal of Gerontology*, 39(4), 439-443.
- Mueller, J.H., Rankin, J.L., & Carlomust, M., (1979). Adult age differences in free recall as a function of basis of organisation and method of presentation. *Journal of Gerontology*, 34(3), 375-380.
- Murphy, M.D., Sanders, R.E., Gabriesheski, A.S., & Schmitt, F.A., (1981). Metamemory in the aged. *Journal of Gerontology* 36(2), 185-193.
- Parkinson, S.R., Lindholm, J.M., & Inman, V.W., (1982). An analysis of age differences in immediate recall. *Journal of Gerontology*, 37(4), 425-431.
- Pavur, E.J., Comeaux, J.M. & Zeringue, J.A., (1984). Younger and older adults' attention to relevant stimuli in free recall. *Experimental Aging Research*, 10(1), 59-60.
- Perlmutter, M., (1979). Age differences in adults' free recall, cued recall, and recognition. *Journal of Gerontology*, 34(4), 533-539.
- Pieters, J.M. & Bommel, P.M.A. van, (1986). Leerstijlen van volwassenen bij het gebruik van de computer. In: J.S. ten Brincke, H.P. Hooymayers en G. Kanselaar (Red.) *Vakdidactiek en informatietechnologie in curriculumontwikkeling*. Lisse: Swets & Zeitlinger.
- Pieters, J.M. & Bommel, P.M.A. van, (in druk). Adult learning strategies and computer skill training. In: F. Goffree & H. Stroomberg (Eds.) *Creating Adult Learning*. London: Wiley.
- Rabinowitz, J.C., (1984). Aging and recognition failure. *Journal of Gerontology*, 39(1), 65-71.
- Rabinowitz, J.C., & Craik, F.I.M., (1986). Prior retrieval effects in young and old adults. *Journal of Gerontology*, 41(3), 368-375.
- Sanders, R.E., Murphy, M.D., Schmitt, F.A., & Walsh, K.K., (1980). Age differences in free recall rehearsal strategies. *Journal of Gerontology*, 35(4), 550-558.
- Schneider, N.G., Gritz, E.R., & Jarvik, M.E., (1975). Age differences in learning, immediate and one week delayed recall. *Gerontologia*, 21, 10-20.
- Schonfield, D., & Robertson, B.A., (1966). Memory storage and aging. *Canadian Journal of Psychology*, 20, 228-236.
- Schonfield, A.E.D., Davidson, H., & Jones, H., (1983). An example of age-associated interference in memorizing. *Journal of Gerontology*, 38(2), 204-210.
- Shaps, L.P., & Nilsson, L.G., (1980). Encoding and retrieval operations in relation to age. *Developmental Psychology*, 16(6), 636-643.
- Sternberg, R.J., & Berg, C., (1987). What are theories of adult intellectual development theories of? In: C. Schooler & K.W. Schaie (Eds.) *Cognitive functioning and social structure of the life course*. Norwood: Ablex.
- Taub, H.A., & Kline, G.E., (1978). Recall of prose as a function of age and input modality. *Journal of Gerontology*, 33(5), 725-730.
- Till, R.E., & Walsh, D.A., (1980). Encoding and retrieval factors in adult memory for implicational sentences. *Journal of Verbal learning and verbal behavior*, 19, 1-16.
- Tulving, E., & Thomson, D.M., (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- Wright, R.E., (1981). Aging, divided attention, and processing capacity. *Journal of Gerontology*, 36(5), 605-614.

Manuscript ontvangen 5-2-1988

Definitieve versie ontvangen 23-11-1988

Gebruiksmogelijkheden van het Learner Report^{1,2}

Baukje J. van Kesteren

Universiteit Twente, Onderzoekscentrum Toegepaste Onderwijskunde

ABSTRACT

The Learner Report (LR) was introduced by De Groot (1974) as an instrument to identify educational objectives. Since then it has been used in different forms, not only for its original purpose but mainly as an instrument for program evaluation, and as a didactic instrument. This article contains a survey of the various educational situations in which the LR has been used, together with the forms in which it was offered, the problems with which its users were confronted and the solutions which they found for these problems. This has resulted in a review of possibilities for the use of the LR and recommendations for its use. The present survey has been plotted against a fundamental discussion on the value of self-reports and self-evaluation in social research.

Het Learner Report werd door De Groot (1974) geïntroduceerd in het kader van het doelstellingenonderzoek in het onderwijs. Veel van het tot dan toe verrichte doelstellingenonderzoek was uitgevoerd volgens de klassieke of systematische benadering. Binnen deze stroming ziet men duidelijke en precieze doelstellingen, geformuleerd in termen van meetbaar gedrag, als startpunt van onderwijs. De Groots theorie over doelstellingen sluit aan bij een andere benaderingswijze, namelijk de cognitieve, waarin de lerende beschouwd wordt als een actief, informatieverwerkend systeem. De Groot stelt zich op het standpunt dat het onderwijs tot taak heeft leerlingen bepaalde gedragsdisposities of programma's te helpen verkrijgen en hen te leren hoe deze programma's naar eigen inzicht te gebruiken.

Uitgangspunt van De Groots theorie is het dekkingsprobleem. Het geheel van geformuleerde onderwijsdoelstellingen behoort wat men werkelijk wil bereiken te dekken. Op dit punt schieten de objectivistische meet-benaderingen tekort, omdat daarbij datgene wordt verwaarloosd wat niet gemeten kan worden. Een open benadering, waarbij de leerling (in termen van De Groot de 'beste deskundige' over wat hij/zij werkelijk geleerd heeft) wordt ingeschakeld komt aan dit probleem tegemoet. Op deze gedachte stoelt het Learner Report (LR), waarin leerlingen de leerervaringen onder woorden brengen, die zij in een onderwijs-situatie hebben opgedaan.

Twee aspecten die volgens De Groot in de gangbare denkwijze over onderwijsdoelen over het hoofd worden gezien, zijn het verwerven van zelfkennis (als domein naast kennis

Adres: Postbus 217, 7500 AE Enschede.

1. De auteur is dank verschuldigd aan Prof.Dr. A.D. de Groot en Prof.Dr. W.K.B.Hofstee voor het begeleiden van het onderzoek en voor het geven van kritisch commentaar op de achtereenvolgende versies van het manuscript. Een aantal bijlagen waarin de door de gebruikers van het Learner Report gerapporteerde problemen, oplossingen, adviezen en eigen leerervaringen zijn opgenomen, is verkrijgbaar bij B.J. van Kesteren, Universiteit Twente, OCTO, Postbus 217, 7500 AE Enschede.
2. De Groot gebruikt de term 'Learner Report' in plaats van een Nederlandse vertaling daarvan, omdat het woord 'learner' ('lerende') naar zijn smaak een ruimere betekenis heeft dan ons woord 'leerling'.

omtrent de wereld) en het ontdekken van uitzonderingen (als vorm van leren in alle leerprocessen, naast het leren van regels). Om deze lacunes op te vullen ontwierp De Groot zijn inmiddels bekende 2x2 schema (Tabel 1).

Tabel 1. Vier categorieën leerervaringszinnen: Universele versus existentiële, wereld- versus zelfervaringen. Overgenomen uit: De Groot, 1980.

Ik heb geleerd (gemerkt, begrepen, ervaren) dat:	
<i>(iets altijd zo is, of: zo moet)</i> REGELS m.b.t. de WERELD 'Zo zit dat': In Situatie S is Gedrag G nodig (juist, passend, verplicht, gewenst)	<i>(iets niet zo is, als je dacht)</i> UITZONDERINGEN m.b.t. de WERELD $S \rightarrow G$ klopt niet (hoeft niet): nieuwe feiten, tegenvoorbeelden, ontdekte vooroordelen, nieuwe mogelijkheden
<i>(ik iets altijd het beste zo kan doen, of: moet doen)</i> REGELS m.b.t. MIJZELF 'Zo zit dat bij mij': in Situaties S moet ik G kiezen	<i>(het niet waar is dat ik... – terwijl ik dat eerst wel dacht)</i> UITZONDERINGEN m.b.t. MIJZELF $S \rightarrow G$ klopt niet (hoeft niet) ik kan ook anders, er zijn ook alternatieven voor mij

Deze indeling in vier cellen valt niet samen met de verschillende manieren waarop verworven kennis geconstateerd kan worden. Daarom maakt De Groot (1982b, pp. 60-61) binnen cel A nog onderscheid tussen afvraagbare kennis en demonstreerbare vaardigheden. Cel C en cel D worden in dit verband samengetrokken tot één cel: (rapporteerbaar) zelfinzicht. Naar verworven zelfkennis en geleerde uitzonderingen kan in het LR expliciet worden gevraagd.

Sinds De Groot het LR in 1974 lanceerde, is er op verschillende manieren vorm aan gegeven en voor uiteenlopende doeleinden gebruik van gemaakt. Hieronder volgt een inventarisatie van de onderzoeken en onderwijssituaties waarin het LR een rol heeft gespeeld. Daarbij wordt getracht een antwoord te geven op de volgende vragen:

1. Hoe wordt een betrouwbaar en valide LR geconstrueerd?
2. Voor welke doeleinden is een LR bruikbaar?

De aanleiding tot deze inventarisatie was de constatering van de auteur, dat verschillende gebruikers van het Learner Report tegen dezelfde problemen aanliepen en deze ieder voor zich probeerden op te lossen. Het hierna volgende overzicht dient dan ook vooral als een praktische handreiking aan toekomstige gebruikers, die hen wellicht een deel van het werk kan besparen (zie de aanbevolen literatuur aan het eind van het artikel).

INVENTARISATIE

Materiaal

Een deel van het materiaal bestaat uit verslagen van onderzoeken en onderwijssituaties waarin van het LR gebruik is gemaakt. Deze verslagen zijn verzameld door het natrekken van literatuurreferenties en door gebruikers van het LR te vragen om documentatie en om namen van andere gebruikers van het LR. Op deze manier hebben wij, voor zover wij kunnen nagaan, een volledig overzicht samengesteld van het gebruik van het LR in de jaren 1976 tot en met 1985.

Daarnaast werd aan alle gebruikers van het LR die ons in 1985 bekend waren, een vragenlijst toegezonden. Door middel van open vragen werd daarin gevraagd voor welke doeleinden men het LR had gebruikt, welke problemen zich daarbij hadden voorgedaan,

welke oplossingen men daar eventueel voor had gevonden, voor welke gebruiksdoelen men het LR al dan niet geschikt acht, wat men toekomstige gebruikers zou willen aanbevelen, wat men had geleerd van het gebruik van het LR en of men nog andere gebruikers van het LR kende. Van de 26 verzonden vragenlijsten werden er 12 schriftelijk beantwoord en 4 in een interview. Zes gebruikers beantwoordden de vragen niet zelf, maar verwezen voor de antwoorden naar hun onderzoeksverslagen. Eén vragenlijst kwam onbestelbaar retour en op de drie overige werd geen reactie ontvangen.

In Tabel 2 treft men een overzicht aan van de geanalyseerde (series) onderzoeken. In 14 van de 22 gevallen werd alleen een open LR gebruikt, in 6 gevallen alleen een gesloten vorm (waarvan 2 keer alleen in de *try-out*) en in 2 gevallen werd zowel van een open als van een gesloten vorm gebruik gemaakt.

DE CONSTRUCTIE VAN EEN BETROUWBAAR EN VALIDE LEARNER REPORT

Vormen van het Learner Report

De centrale vraag in het Learner Report is: 'Wat heb je (in deze cursus, van dit schoolvak, enzovoorts) geleerd?' De respondent wordt verzocht antwoorden op deze vraag te geven in de vorm van leereffectzinnen. Dit zijn zinnen die beginnen met de woorden 'Ik heb geleerd ...'.

Om de respondenten een indruk te geven van wat er van hen verlangd wordt, kan men een aantal voorbeelden geven. Over het algemeen zal men de respondenten verschillende aanvulzinnen voorleggen, zoals: Ik heb geleerd dat ..., Ik heb geleerd hoe ..., Ik heb ontdekt dat ... (alle betrekking hebbend op cel A in het schema); Ik heb geleerd dat het niet altijd zo is dat ... (cel B); Ik heb geleerd dat ik ..., Ik heb gemerkt dat ik ... (cel C); Ik heb geleerd dat ik niet altijd ... maar ook ... (cel D).

Het LR kan in verschillende vormen worden aangeboden, zowel schriftelijk als mondeling en zowel individueel als groepsgewijs. Ook kan de centrale vraag dienen als uitgangspunt voor een gesprek.

Bij een schriftelijke afname kan men een open of een gesloten vorm gebruiken. Open vormen kunnen bestaan uit één of meer open vragen naar wat de respondent heeft geleerd, ontdekt, enzovoorts of uit één of meer aangeboden aanvulzinnen.³ Men kan de respondent vragen zoveel mogelijk aanvullingen per aanvulzin te geven of een beperkt aantal (bijvoorbeeld de drie leerervaringen die de meeste indruk hebben gemaakt). Tot slot is het mogelijk aanvulzinnen aan te bieden per onderwerp (bijvoorbeeld: In de les over onderwerp X heb ik geleerd dat ...) en al dan niet expliciet ingedeeld volgens het schema van De Groot.

Als men het LR in gesloten vorm aanbiedt, legt men de respondenten een aantal leereffectzinnen voor, waarvan zij aangeven of en in hoeverre deze op hen van toepassing zijn.

3. Van der Kamp spreekt bij het aanbieden van aanvulzinnen van een halfopen vorm. Omdat het uitsluitend stellen van open vragen hoogst zelden voorkomt en het aanbieden van aanvulzinnen des te meer, nemen wij deze vormen voor het gemak samen.

Tabel 2. Overzicht van onderzoeken met behulp van het Learner Report in de periode 1976-1985.

Referentie	Onderzochte activiteit	Redenen van keuze voor het LR	Respondenten
Haanstra, 1976 Van der Kamp, 1980 Van der Kamp & Oostwoud Wijdenes, 1982 Van der Kamp, 1984	expressievakken in het MAVO en HAVO	1. De doelen en effecten van kunstzinnige vakken zijn (mede) cognitief van aard. Dit cognitieve aspect kan met behulp van het LR worden geëxpliciteerd. 2. Om greep te krijgen op de moeilijk te meten leereffecten in de kunstzinnige vorming. 3. De Groots classificatieschema kan een hulpmiddel zijn om een 'dekkende' verzameling van leereffecten in de vingers te krijgen.	599 (oud-)leerlingen MAVO en HAVO, met een kunstzinnig vak als eindexamenvak
De Kat & Riemersma, 1978	het project 'School-samenleving' aan het Haags Montessori-Lyceum, in het kader van het landelijk experiment middenschool	1. Door op deze wijze de leerervaringen van leerlingen te inventariseren kan de discussie over de concretisering van de globale doelstellingen op gang gebracht worden. 2. Deze methode laat ook de moeilijk meetbare doeleinden aan bod komen. 3. De leerling zelf staat bij deze methode centraal.	70 leerlingen van 2 tweede en 2 derde klas sen.
Dresen-van der Linden, 1979	de Open School te Gouda in het jaar waarin deze startte (1977-'78)	De achtergrond van het LR (leereffecten formuleren in termen van gedragsrepertoires of disposities) sluit goed aan bij de ideeën van de Open School.	15 vrouwen, 25-55 jr, vooropl. variërend van 4 jaar LO tot moderschool
Roefs, 1980	een cursus leerlingbegeleiding, gegeven door de lerarenopleiding te Groningen aan docententeams binnen het VO	Onderzoek naar de bruikbaarheid van het LR voor het evalueren van deze cursus.	18 docenten (12 m., 6 vr.) van het Praedinius Gymnasium te Groningen
Gerritsen van der Hoop, 1981	5 keuzecolleges binnen de cursus Maatschappijwetenschappen in het 2e en 3e studiejaar van de TH Eindhoven	1. Het laten beschrijven van (ervaren) leereffecten die zich moeilijk laten meten. 2. Een eerste indruk krijgen van de vruchtbaarheid van het LR.	studenten die eind 1980 de betr. colleges volgden

Referentie	Onderzochte activiteit	Redenen van keuze voor het LR	Respondenten
Gerritsen van der Hoop, 1982a	3 uiteenlopende cursussen voor studenten, respectievelijk docenten aan de TH Eindhoven en een introductieweek voor vrouwelijke aankomende studenten N.B. Het artikel is bedoeld om informatie te geven over het LR en om aanwijzingen te geven met betrekking tot het gebruik ervan en met betrekking tot gebruiksdoelen.	Het laten beschrijven van (ervaren) leereffecten die zich moeilijk laten meten.	157, resp. 98 en 152 studenten, 15 docenten, 29 vrouwelijke aankomende studenten
Gerritsen van der Hoop, 1982b	introductieweek voor vrouwelijke studenten aan de TH Eindhoven	Zie boven	Zie boven
Van Raalte, 1981 Van Raalte, 1983	educatieve activiteiten van de balletgroep Scapino	Niet gespecificeerd	32 leerlingen (20 vr., 12 m.) van klas 5 en 21 leerlingen (11 vr., 10 m.) van de Merula-(basis)-school te Heenvliet
Meijer, 1982	stages in het MBO en HBO	Het grote meetbereik van dit type instrument, vooral met betrekking tot kennis die alleen communiceerbaar is: leereffecten van stages kunnen zeer divers zijn.	874 studenten van 19 opleidingen in het MBO en HBO
Min & Nooter, 1982 N.B. uiteindelijk zag men af van gebruik van het LR	een combinatie van lessen en een theaterproductie over fascisme, gegeven door het samenwerkingsproject Anne Frank Stichting/Werk in uitvoering en bedoeld voor de lagere stroom van het VO	Men wenste een evaluatie-instrument, dat expliciet ruimte gaf aan het meten van de affectieve doelstellingen van de theaterproductie en dat tevens aan een aantal al dan niet praktische voorwaarden voldeed.	120 leerlingen (58 m., 62 vr.) uit 11 klassen van 5 scholen in het LBO, BBO en VWJ
Oostwoud Wijdenes, 1982	het vak 'drama' in het VO	1. De diversiteit aan leereffecten waartoe vorming bij verschillende individuen leidt. 2. De noodzaak om uit te gaan van de lerende zelf, als men greep wil krijgen op leereffecten van (kunstzinnige) vorming.	120 leerlingen van 3 scholengemeenschappen
Dwarshuis, 1983	een serie lessen over onderwijsstimulering aan de Rijks PA te Meppel in '82-'83	De doelstellingen van de lessen lagen voornamelijk op het terrein vanhoudingen en (zelf)inzichten.	38 tweedejaars studenten

Referentie	Onderzochte activiteit	Redenen van keuze voor het LR	Respondenten
Godefrooij, 1983	voorbeeldlessen in creatief taalgebruik aan Rotterdamse basisscholen, georganiseerd door de Stichting Musicische Vorming Rotterdam en de School Advies Dienst Rotterdam	Het opsporen van leereffecten die noch afvraagbaar, noch duidelijk demonstreerbaar zijn.	Klas 5/6 van de Vliedbergschool te Rotterdam
Haanstra, 1983	een educatieve tentoonstelling in het Centraal Museum te Utrecht	Men wilde weten wat leerlingen zelf vonden geleerd te hebben.	755 leerlingen (270 m., 485 vr.) uit het VO
Van Kesteren, 1983 Van der Molen, 1985	twee soorten cursussen voor verlegen mensen: 1. op <i>cognitive behavioral</i> basis; 2. op <i>dynamic, humanistic</i> basis	1. Om na te gaan of er verschillen bestaan tussen de door de deelnemers van elk van beide cursussen gerapporteerde effecten. 2. Voor het opsporen van onbedoelde effecten.	33 resp. 49 deelnemers aan cursus type 1; 27 resp. 38 deelnemers aan cursus type 2
Meerum Terwogt-Kouwenhoven, 1983	propedeuse psychologie aan de Universiteit van Amsterdam, 1981-'82	Niet alle doelstellingen van de propedeuse lieten zich vertalen in objectieve maten.	51 studenten van de 174 aan wie een vragenlijst was verzonden; 5 docenten van de 13 aan wie een vragenlijst was verzonden
Rijlaarsdam, 1984	effecten van beoordeling door medeleerlingen op aspecten van stelselmatigheid	Niet gespecificeerd	81 leerlingen uit 4 derde klassen HAVO-VWO van de scholengemeenschap Noorddijk te Dordrecht
Nieuwenhuis, 1984a Nieuwenhuis, 1984b Maandag, 1984	stages in het MEAO en MDGO, waarbij een vergelijking plaatsvond tussen reguliere en experimentele scholen	Het beschrijven van onderwijseffecten (leerervaringen in het sociaal-normatieve domein) die niet direct meetbaar, maar wel communiceerbaar zijn.	(a) 43 2e-klas MEAO-leerlingen regulier; 29 2e-klas MEAO-leerl. experimenteel (b) 54 2e-klas MDGO-leerlingen regulier; 62 2e-klas MDGO-leerl. experimenteel

Referentie	Onderzochte activiteit	Redenen van keuze voor het LR	Respondenten
Vedder, 1984	het '1-1-project' waarin wiskundestudenten aan de universitaire lerarenopleiding te Utrecht, als voorbereiding op hun schoolpracticum, les geven aan individuele leerlingen	1. Het onderzoek was in belangrijke mate te typeren als <i>goal-free evaluation</i>, waarin de vraag naar de leereffecten die volgens de studenten zelf optraden, centraal stond. 2. Om verschillende redenen wordt het leren reflecteren voor de betreffende studenten zeer belangrijk geacht; het gebruik van het LR sloot goed aan bij deze doelstelling van het 1-1-project.	24 wiskundestudenten, verdeeld over twee groepen.
Van Erp & Koppen, 1984 Van Erp, 1985a Van Erp, 1985b	milieuprojecten in het voortgezet onderwijs	Niet gespecificeerd	70 2e-klas MAVO-leerlingen, 110 brugklasleerlingen
Edelenbos, De Jong en Den Boer, 1985	veldstudiecursussen gegeven door het Veldstudiecentrum te Orvelte	1. Om uiteenlopende redenen was het onmogelijk een toets te ontwikkelen voor het meten van (leer-)effecten. 2. De te evalueren cursussen waren al achter de rug, zodat alleen rapportage achteraf kon plaatsvinden.	Deelnemers aan uiteenlopende cursussen: VO-leerlingen, studenten en buitenschoolse groepen.
(a) Schouteten, 1985 (b) Gradisen, 1985	scheikundepractica in het VO	Het vergelijken van de leerervaringen die practica over het algemeen geacht worden op te leveren met de door de leerlingen gerapporteerde leerervaringen.	(a) n=? VO-leerlingen; (b) 19 VWO-leerlingen van het Niels Stenscollege, Utrecht
Bonset <i>et al.</i> , 1985	het vak Nederlands in heterogene brugklassen	Als aanvulling op toetsen en ander schriftelijk materiaal, teneinde de leereffecten op het affectieve vlak op te sporen.	27 leerlingen (9 m., 18 vr.) uit een eerste klas VO, met een prognosespreiding van LBO tot VWO

Validiteit

Verscheidene onderzoekers vermelden problemen die betrekking hebben op de validiteit van het Learner Report. Andere auteurs besteden aan validiteitsproblemen, naar onze mening ten onrechte, in het geheel geen aandacht. De problemen met betrekking tot de validiteit van het LR zijn als volgt samen te vatten:

1. In hoeverre kan men geloof hechten aan de beweringen van leerlingen over wat zij geleerd hebben?
2. In hoeverre is er verband tussen aantallen gerapporteerde leerervaringen en hoeveelheid opgedane kennis, inzicht en vaardigheden?
Vraag 2 geldt vooral het gebruik van een open LR.
3. Welke concrete aanwijzingen zijn er met het oog op de validiteit te geven voor de constructie en het gebruik van een LR?

ad 1: De juistheid van self-reports

De eerste vraag is verwant aan een oude maar actuele discussie over de juistheid van *self-reports*. Op basis van analyses van onderzoek naar onder meer cognitieve dissonantie, attributies en *learning-without-awareness* en van eigen experimenten komen Nisbett & Wilson (1977) tot de conclusie dat mensen geen toegang hebben tot hun cognitieve processen. Door Ericsson & Simon (1980) en White (1980) wordt dit bestreden. Ericsson & Simon presenteren een model voor het interpreteren van verbale data en van de relatie van verbaal gedrag met het overige gedrag van proefpersonen. Zij analyseren op hun beurt een groot aantal onderzoeken en concluderen dat de voorbeelden die Nisbett & Wilson geven van inconsistentie tussen *verbal reports* en (andere) data allemaal betrekking hebben op experimentele situaties en procedures waarbij volgens Ericsson & Simon waarheidsgetrouwe rapportage nauwelijks te verwachten valt. Beter afgestemde procedures kunnen volgens hen wel degelijk informatieve verbale verslagen van cognitieve processen opleveren. Zo moet de respondenten expliciet worden gevraagd hun herinneringen aan het cognitieve proces te raadplegen en moet de tijd tussen het uitvoeren van een taak en het beantwoorden van een vraag daarover zo klein zijn dat de informatie in het korte-termijn geheugen kan zijn bewaard.

White geeft methodologische aanbevelingen voor experimenten die de bewering moeten toetsen dat een verbaal verslag van een cognitief proces inderdaad op dit proces kan berusten en geen rationalisatie achteraf behoeft te zijn. De resultaten van een experiment dat hij volgens deze richtlijnen heeft opgezet ondersteunen zijn hypothese dat het cognitieve proces van invloed is op de nauwkeurigheid van het verbale verslag daarvan. Eén van de oorzaken van het verschil tussen White's bevindingen en die van Nisbett & Wilson is gelegen in de gevoeligheid van de analysemethoden waarmee zij hun gegevens bewerken. Wanneer White de – minder gevoelige – methoden van Nisbett & Wilson gebruikt, vindt hij niet het verschil tussen experimentele en controlegroep dat hij via zijn eigen methoden vindt. White concludeert: 'It seems likely that the null results obtained by Nisbett & Wilson (1977) are at least in part a consequence of the intrinsic insensitivity of the measures they employed' (White, 1980, p. 111).

Alle hierboven genoemde onderzoeken betreffen de toegankelijkheid van het cognitieve proces als zodanig. Slechts in enkele gevallen (Nisbett & Wilson, 1977; Shrauger & Osberg, 1981) is de validiteit van zelfrapportage over het *product* van een leerproces onderzocht. Nisbett & Wilson (1977) vermelden het volgende: uit niet gepubliceerde resultaten van onderzoek naar cognitieve dissonantie en *self-perception attribution processes* zou blijken dat proefpersonen zich vaak niet bewust zijn van de belangrijkste respons die door de manipulatie veroorzaakt wordt. Smith & Miller merken over dit type onderzoek (waaronder ook de experimenten van Nisbett & Wilson zelf vallen) op: 'The main purpose

of between-subjects methodology in social psychology is to create manipulations in which subjects will be unaware of variables that affect them and hence more willing to respond to them (Greenwald, 1976). To take the fact that this methodology succeeds in hoaxing subjects and then to generate a general perceptual rule from it carries the point too far.' (Smith & Miller, 1978, p. 356).⁴ Shrauger & Osberg (1981) analyseerden onderzoeken naar voorspellingen van respondenten ten aanzien van eigen intellectuele en werkprestaties, beroepskeuze, therapieresultaten en dergelijke. De validiteit van de voorspellingen van de respondenten evenaarde of overtrof in de meeste gevallen die van de voorspellingen op grond van feitelijke gegevens zoals schoolcijfers. De voorspellingen van de respondenten zelf zijn uiteraard gebaseerd op zelfoordelen. De bevindingen van Shrauger & Osberg bevestigden dan ook op indirecte wijze de waarde van zelfoordelen.

ad 2: Het verband tussen aantallen leerervaringen en opgedane kennis, inzicht en vaardigheden

Veel gebruikers van een open LR verwerken de geformuleerde leerervaringen door in elke leereffectzin één of meer leerervaringen te onderscheiden, deze leerervaringen in categorieën onder te brengen en de aantallen leerervaringen per categorie en/of de totale aantallen te tellen. Het bij elkaar optellen van leerervaringen is echter per definitie een onnauwkeurige werkwijze. De oorzaken hiervan zijn:

1. Leerervaringen kunnen hetzij rechtstreeks betrekking hebben op een geleerde vaardigheid, hetzij een inzicht weergeven over de mate waarin de leerling die vaardigheid beheerst. Vergelijk bijv.: 'Ik heb geleerd met machines om te gaan' en 'Ik heb ontdekt dat ik goed met machines om kan gaan.' In het eerste geval heeft de leerling zijn/haar vaardigheden uitgebreid, in het tweede geval is hij/zij aan de weet gekomen hoe vaardig hij/zij is. Volgens de boven omschreven werkwijze zal elk van deze zinnen als één leerervaring worden geteld. Het is echter niet duidelijk of men deze leerervaringen terecht eenzelfde gewicht toekent.
2. Leerervaringen kunnen een positieve, een negatieve en/of een neutrale betekenis hebben. Vergelijk: 'Ik heb ontdekt dat je veel dingen die je op school leert kunt toepassen in de praktijk' met 'Ik heb ontdekt dat het niet altijd zo is dat de theorie goed aansluit bij de praktijk.' Inhoudelijk zijn deze zinnen tot op zekere hoogte aan elkaar tegengesteld. Dit verschil in waarde komt niet tot uiting als men elk van de zinnen als één leerervaring telt.
3. Verschillende leerervaringen kunnen meer of minder omvattend zijn. Vergelijk: 'Ik heb geleerd de kassa te bedienen' met 'Ik heb meer kennis gekregen van bedrijfseconomische zaken.' Hoewel de tweede uitspraak betrekking heeft op een grotere kennisvermeerdering dan de eerste, worden ze ieder als één leerervaring geteld.

Al met al lijkt het verband tussen het aantal gerapporteerde leerervaringen en de opgedane kennis, inzichten en vaardigheden niet sterk te zijn en men moet erop bedacht zijn dat de genoemde factoren vooral bij het vergelijken van kleine aantallen leerervaringen, afkomstig van weinig respondenten, een vertekening kunnen veroorzaken. Dit zal minder het geval zijn naarmate het aantal respondenten en het aantal leerervaringen groter is, omdat de verschillen in waarde van leerervaringen elkaar dan tot op zekere hoogte zullen opheffen. Bij kleine hoeveelheden materiaal kan men beter een zo goed mogelijke samenvatting van de inhoud ervan maken dan de *aantallen* leerervaringen te analyseren.

4. Het is overigens frappant hoezeer ervaren psychologisch onderzoekers ervan overtuigd zijn, geen gebruik te maken van misleiding van hun proefpersonen, terwijl zij in vrijwel al hun experimenten hun proefpersonen op het verkeerde been zetten!

ad 3: Concrete aanwijzingen voor constructie en gebruik van het LR met het oog op de validiteit ervan

In het algemeen kan men er alleen dan op rekenen dat respondenten zo waarheidsgetrouw mogelijk rapporteren als zij geen belang hebben bij het geven van een vertekend antwoord. De Groot (1982a, pp. 113-114) wijst in dit verband op de noodzaak van het creëren van een onderzoekssituatie die dit waarborgt. Dat het LR niet als selectiemiddel dient te worden gebruikt vloeit hier rechtstreeks uit voort. Daarnaast verdient het aanbeveling dat evaluatie-onderzoek met behulp van het LR wordt uitgevoerd door anderen dan de docenten of begeleiders van de te evalueren activiteit, iets wat in veel van de hier besproken onderzoeken niet het geval is geweest.

Onnauwkeurigheden en vergissingen die zich desondanks kunnen voordoen bij het invullen van een Learner Report zijn de volgende. Respondenten kunnen beweren iets in een bepaalde onderwijssituatie te hebben opgestoken, terwijl ze het in werkelijkheid elders hebben geleerd. Vooral aan zelfkennis wordt vaak door ervaringen in meer dan één situatie bijdragen. Respondenten kunnen aangeven iets te hebben geleerd (bijv.: '...wat het verschil is tussen een concert en een symfonie'; '...een menu samen te stellen') terwijl zij het betreffende onderwerp of de genoemde vaardigheid slechts gedeeltelijk beheersen. Over het algemeen leveren fouten als deze niet zozeer onware, maar wel onnauwkeurige uitspraken op. De (praktische) mogelijkheden om uitspraken op dit soort fouten te controleren zijn vaak gering.

Door aan het LR geen kennisvragen vooraf te laten gaan kan men vermijden dat men zelf in de hand werkt, dat respondenten meer rapporteren dan zij geleerd hebben.

Het is ook mogelijk dat iets wat wel geleerd is niet wordt gerapporteerd. Voorbeelden hiervan zijn: Men kan opgedane kennis, als de situatie daarom vraagt, in praktijk brengen, terwijl men, achter een bureau gezeten, niet rapporteert hoe men in de situatie *zou* handelen. Het is mogelijk dat men iets leert door een leermeester te imiteren, zonder dat men zich daarvan bewust is (*modeling*). Men zal zich op het moment van rapporteren niet al het geleerde te binnen kunnen brengen. Ook kan men over iets wat men intuïtief heeft leren beheersen vaak slechts globaal rapporteren (De Groot, 1986).

Om deze problemen het hoofd te bieden kan men leerlingen leren te reflecteren op leerervaringen. In weinig situaties gebeurt dit expliciet (zie Bonset *et al.*, 1985 en Damsma *et al.*, 1985). Een andere manier is het instrument zo doorzichtig en uitnodigend mogelijk te formuleren. Wat respondenten in principe kunnen rapporteren, moet men ook in hun learner reports zien te krijgen (zie ook punt 1 hieronder).

Betrouwbaarheid

De concrete problemen die zich bij het gebruik van een open vorm van het LR voordoen houden vrijwel allemaal verband met de betrouwbaarheid van het instrument. Het gaat hierbij om de formulering en de invulling van het LR en de verwerking van de verzamelde gegevens. Daarnaast levert de presentatie van de gegevens vaak problemen op.

De door de verschillende auteurs gerapporteerde problemen en de oplossingen die zij gevonden hebben en/of aanbevelen, zijn:

1. De instructie bleek nogal eens te moeilijk te zijn voor de respondenten. Men kan dit zoveel mogelijk voorkomen door er voldoende tijd voor te nemen zich in de betreffende doelgroep te verdiepen en een *pilot study* of *try-out* uit te voeren, waarna de formulering kan worden aangepast. (In 8 van de hierboven genoemde 22 onderzoeken was van een dergelijke *try-out* sprake.)

Oostwoud Wijdenes (1982) nam in de instructie een verhaaltje op over een les drama, waarin gewezen werd op allerlei leermomenten, zodat de leerlingen beter zouden begrijpen wat er bij het invullen van hen werd verlangd.

2. Respondenten hadden soms moeite zich voor de geest te halen wat er in de loop van bijvoorbeeld een heel schooljaar was behandeld. Daarom werd er in enkele gevallen een geheugensteuntje gegeven.
3. Respondenten vulden hun leerervaringen niet altijd in de juiste cellen van De Groot in. Als dit vaak het geval was, probeerden de onderzoekers meestal tot een herindeling te komen. Een enkeling deelde alleen die leereffectzinnen opnieuw in waarvan het duidelijk was dat ze in de verkeerde cel stonden en in welke cel ze wel thuis hoorden. Anderen gaven na enkele pogingen het herindelen op en besloten zich aan de indeling van de respondenten te houden.

Enkele onderzoekers trokken de cellen met regels en uitzonderingen samen. In sommige gevallen (bijvoorbeeld bij jonge kinderen of bij respondenten met een geringe opleiding) is het beter om de ontkennende aanvulzinnen weg te laten, al kan dit ten koste gaan van een hoeveelheid informatie.

Overigens is de indeling in cellen volgens De Groot vooral bedoeld om zoveel mogelijk leereffectzinnen te genereren en niet om achteraf te tellen hoeveel regels, respectievelijk uitzonderingen de respondenten over de wereld dan wel over zichzelf hebben geleerd. Het is dus niet zo'n groot bezwaar als respondenten hun leerervaringen niet in de juiste cellen plaatsen.

4. Het samenvatten van het verzamelde materiaal tot een overzichtelijk geheel bleek vaak moeilijk te zijn. De meeste auteurs gingen over tot het indelen van de leerervaringen in categorieën, die bijvoorbeeld ontleend waren aan de doeleinden die men in een onderwijssituatie nastreefde of aan het materiaal zelf. Dit leidde echter niet altijd tot bevredigende resultaten.

Van der Molen (1985) ging er daarom in tweede instantie toe over om de Learner Reports 'minder atomistisch' te analyseren door ze in hun geheel op een beperkt aantal relevante kenmerken te beoordelen.

Bonset *et al.* (1985) vatten een groot deel van de leereffectzinnen samen in een aantal standaardzinnen (bijvoorbeeld: 'Ik heb geleerd dat ik activiteit ... vind', waarbij op de puntjes kan zijn ingevuld: leuk, vervelend, makkelijk, enz.).

5. Als men leerervaringen in categorieën had ingedeeld gaf dat de lezer nog niet altijd een goed beeld van deze leerervaringen. Daarom gingen veel auteurs over tot het samenstellen van een bloemlezing, al dan niet verwerkt tot 'illustraties' bij de kwantitatieve gegevens.

Samengevat blijkt een aantal problemen, dat zich voordoet bij de constructie en de invulling van het Learner Report goed op te lossen, mits men er voldoende tijd en aandacht aan besteedt. De problemen met betrekking tot het indelen in categorieën lijken niet altijd even bevredigend op te lossen. Dit is echter vooral problematisch als men verschillende groepen respondenten met elkaar wil vergelijken. Een open LR is voor dat doel dan ook minder geschikt dan een gesloten LR.

Overwegingen bij het gebruik van een gesloten LR

Bij het gebruik van een gesloten LR doen de hierboven onder 1 tot en met 5 genoemde problemen zich niet of nauwelijks voor. Een gesloten LR heeft ten opzichte van een open vorm echter enkele nadelen. In een gesloten LR legt men respondenten leerervaringen voor, waarvan zij moeten aangeven of zij deze hebben opgedaan. Dit kan het verschijnsel dat respondenten meer rapporteren dan zij geleerd hebben, bevorderen. Desondanks kan een gesloten LR discrimineren tussen verschillende (sub)groepen van respondenten en in die zin geldige resultaten opleveren. Zie bijvoorbeeld Van der Kamp (1980) over Dijkhuis-Potgieser (1977); Van Erp & Koppen (1984).

Bovendien kan een andere manier van vragen grotendeels aan het bovengenoemde

bezwaar tegemoet komen. Een voorbeeld hiervan is de uitsprakenlijst van Vedder (1984). Daarin wordt eerst gevraagd in hoeverre de respondent het met een bepaald inzicht eens is en vervolgens of de onderwijssituatie waarop het onderzoek betrekking heeft aan deze mening heeft bijgedragen. Deze manier leent zich tevens voor gebruik bij een voor- en een nameting (de tweede vraag wordt dan alleen bij de nameting gesteld) en voor gebruik bij een controlegroep of een tweede experimentele groep (bij de controlegroep wordt de tweede vraag weggelaten, bij een tweede experimentele groep wordt deze anders geformuleerd).

Voor het overige doen zich bij een gesloten LR dezelfde problemen voor als bij gesloten vragenlijsten in het algemeen en gelden voor een gesloten LR dezelfde richtlijnen als voor gesloten vragenlijsten in het algemeen. Een gesloten LR heeft als voordeel boven een open LR dat de verwerking ervan minder arbeidsintensief is. De constructie van een gesloten LR is daarentegen veel arbeidsintensiever dan die van een open LR.

De belangrijkste overweging bij de keuze tussen een open en een gesloten LR, dan wel een ander instrument, wordt gevormd door het doel dat men ermee beoogt. In tegenstelling tot een open LR leent een gesloten LR zich niet voor het opsporen van persoonlijke leerervaringen en van onderwijsdoelstellingen. Een gesloten LR is daarentegen, veel meer dan een open vorm, geschikt voor het vergelijken van verschillende groepen respondenten op de door hen gerapporteerde leerervaringen. Op de mogelijke gebruiksdoelen van een open en een gesloten LR wordt hieronder nader ingegaan.

GEBRUIKSDOELEN

- In de literatuur wordt melding gemaakt van vier gebruiksdoelen van het Learner Report:
1. het doel waarvoor het in eerste instantie is bedacht: het opsporen van onderwijsdoelstellingen;
 2. evaluatie-onderzoek;
 3. zelfevaluatie en daarmee gebruik van het LR als hulpmiddel bij het nemen van individuele didactische beslissingen;
 4. het LR als didactisch instrument.

ad 1 en 2: Doelstellingenonderzoek en evaluatie-onderzoek

Van de hier behandelde onderzoeken heeft alleen dat van Van der Kamp (1980) en Haanstra (1976) het opsporen van onderwijsdoelstellingen tot hoofddoel. Vedder (1984) en De Kat & Riemersma (1979) gebruiken het LR als evaluatie-instrument, maar vermelden het opsporen van doelstellingen als neven-doel. Alle overige in Tabel 2 genoemde publicaties gaan uitsluitend over evaluatie-onderzoeken. Een groot deel hiervan betreft kleinschalige onderzoeken waarin sprake is van formatieve evaluatie van relatief kortdurende onderwijssituaties. De methoden van gegevensverzameling en -verwerking zijn hierboven besproken en zijn voor deze beide doeleinden dezelfde.

Voor beide doeleinden achten wij het LR geschikt. Bij gebruik voor evaluatie-onderzoek zijn daarbij de volgende kanttekeningen te maken.

- Zoals hierboven besproken is het opdoen van zelfkennis niet meetbaar, maar wel rapporteerbaar. Met De Groot en gesteund door het onderzoek van Shrauger & Osberg zijn wij van mening, dat respondenten in staat zijn te beoordelen, wat zij wel en niet in een onderwijssituatie hebben opgestoken. Wel moet hiervoor worden voldaan aan de randvoorwaarden zoals genoemd in de paragraaf 'Validiteit', ad 3.
- Opgedane kennis over de wereld en verworven vaardigheden daarentegen zijn in de regel zeer goed meetbaar. Het meten ervan heeft bij evaluatie-onderzoek de voorkeur, tenzij er omstandigheden zijn die dit onmogelijk maken.

- Het LR kan een docent veel inzicht geven in de manier waarop leerlingen of cursisten de aangeboden leerstof interpreteren en verwerken (Gerritsen van der Hoop, 1982b). In die zin kan men zeggen dat het LR zeer geschikt is voor formatieve evaluatie van onderwijssituaties.

Wat het dekkingsprobleem betreft: uit de genoemde publicaties blijkt dat het LR ruimschoots mogelijkheden biedt om informatie te verzamelen over 'leren omtrent zichzelf'. Over het 'leren van uitzonderingen' vermelden de auteurs echter hooguit de aantallen geleerde regels en uitzonderingen die respondenten rapporteren. Blijkbaar wordt dit aspect relatief onbelangrijk geacht of weet men niet goed hoe men informatie over het geleerd hebben van uitzonderingen nuttig kan gebruiken.

ad 3: Gebruik van het LR met het oog op het nemen van individuele didactische beslissingen
Kreeft en Rijlaarsdam (1980) breken een lans voor het gebruik van zelfevaluatie (en daarmee van het LR) als hulpmiddel voor het nemen van individuele en gemeenschappelijke didactische beslissingen in het moedertaalonderwijs.

Het verschil tussen het gebruik van het LR voor individuele en voor gemeenschappelijke didactische beslissingen is gelegen in de verwerkingsfase. Gaat het om beslissingen ten aanzien van onderwijs zoals dat aan een groep of klas wordt gegeven, dan is men geïnteresseerd in de totale hoeveelheid leerervaringen die een groep leerlingen heeft geformuleerd. In het individuele geval bestudeert men de learner reports per leerling.

De voornaamste argumenten van Kreeft & Rijlaarsdam voor het gebruik van zelfevaluatie in het moedertaalonderwijs zijn de volgende: (a) Moedertaalonderwijs heeft een aantal sterk vormingsachtige trekken; het precies formuleren van doelen op dit gebied is een vrijwel onmogelijke zaak en derhalve het navragen of laten demonstreren van kennis eveneens. Effecten van vorming zijn echter wel communiceerbaar. (b) De evaluatie van ons onderwijs is uitsluitend gericht op positieve leereffecten. Wanneer zelfevaluatie in een sfeer van vertrouwen kan plaatsvinden, kunnen we ook informatie verzamelen over de negatieve leereffecten. (c) 'Zelfverantwoordelijkheid moet gepaard gaan met zelfevaluatie' (De Vos & Meijerink, aangehaald door Kreeft & Rijlaarsdam, 1980).

Een bijkomend voordeel is dat een combinatie van zelfevaluatie met toetsing of met de eigen waarneming van de docent een discrepantie aan het licht kan brengen tussen zelfevaluatie en de toetsgegevens of het oordeel van de docent, wat dan aanleiding kan zijn tot een gesprek tussen docent en leerling.

Op grond van de bevindingen van Kreeft & Rijlaarsdam en van wat hierboven werd opgemerkt over het inzicht dat het LR een docent kan geven over de manier waarop leerlingen met de leerstof omgaan achten wij het LR ook voor dit doel geschikt.

ad 4: Het Learner Report als didactisch instrument

Hoewel door verscheidene auteurs wordt opgemerkt dat het invullen van het LR didactische waarde kan hebben, zijn ons slechts twee gevallen bekend waarbij het LR als didactisch instrument wordt (werd) gehanteerd. Damsma *et al.* (1985) gebruikten het LR in het kader van het onderwijs- en tevens onderzoeksproject 'Taalgebruik als reflectiviteit'. Dit project had tot doel te onderzoeken (a) of leerlingen van 12-15 jaar kunnen leren reflecteren op eigen en andermans (productief en receptief) taalgebruik; (b) of leraren het reflecteren op eigen en andermans (productief en receptief) taalgebruik kunnen onderwijzen aan 12- tot 15-jarige leerlingen. Het werd gegeven in twee tweede atheneumklassen en één derde atheneumklas.

'Taalbeschouwing als reflectiviteit' kwam daarbij in de plaats van het traditionele grammatica-onderwijs. Het LR bestond in dit geval uit een schrift waarin de bij het project betrokken leerlingen na (later ook tijdens) elke les opschreven, wat zij hadden geleerd.

Het gebruik van het LR leverde in dit verband veel problemen op, die grotendeels samenhangen met een gering vermogen van de leerlingen om hun eigen leerervaringen te onderkennen. Damsma *et al.* concluderen dat leerlingen de hulp van de docent nodig hebben om zich deze vaardigheid eigen te maken.

Aan de TU Delft maakt men in de eerstejaars cursus Ontwerpen al enige jaren gebruik van het LR als didactisch instrument. Dit gebruik van het LR is in de plaats gekomen van persoonlijke begeleiding van studenten. Na iedere cursusbijeenkomst wordt aan de studenten een LR meegegeven, waarin zij naar aanleiding van deze bijeenkomst enkele belangrijke leerervaringen moeten invullen. Vanaf de tweede keer moeten de studenten tevens 'aktiepunten' formuleren. Dit houdt in dat zij bij één of meer leerervaringen opschrijven wat zij denken er in de toekomst mee te kunnen doen. Aan het eind van de cursus wordt een dergelijk LR nog eens over het gehele cursusjaar geschreven. De leerervaringen en aktiepunten worden zowel schriftelijk als mondeling bekomentarieerd. Ze worden beoordeeld op de vraag in hoeverre de studenten blijk geven, moeite te hebben gedaan en erin te zijn geslaagd hun werkelijke leerervaringen op te sporen en deze te benutten.

De specifieke leerdoelen van het 'werkstuk leerervaringen' dat op deze manier ontstaat, zijn: (a) De student krijgt een verhoogde vaardigheid in het onder woorden brengen van het eigen leerproces. (b) De student geeft de begeleiders individueel gerichte aanknopingspunten voor de begeleiding. (c) De student krijgt een grotere bewustheid van het eigen leerproces en leert bewuster besluiten te nemen in verband met het effectiever maken van het eigen (ontwerp)gedrag. (d) De student leert het formuleren van leerervaringen kennen als een belangrijk onderdeel van het leerproces waarmee hij/zij (geleidelijk) steeds meer verantwoording kan nemen voor het eigen leerproces. (e) Vooral door de vorm van het werkstuk is de student in staat om over een langere periode terug te kijken en hij/zij kan daardoor leereffecten onderkennen die anders niet op zouden vallen. (f) De student kan laten zien dat hij/zij in staat is een goed verzorgd rapport in de huisstijl te maken (De Haan *et al.*, 1985). Behalve in de eerstejaars cursus Ontwerpen wordt het LR in deze vorm ook gebruikt voor de vakken Kleinschalige Marketing en Marketing-Case Studies.

De meningen van de betrokken docenten over dit gebruik van het LR lopen zeer uiteen. Voorstanders menen dat de bovengenoemde leerdoelen met het gebruik van het LR zijn gediend. Tegenstanders zijn van mening dat studenten teveel in staat zijn zich er vanaf te maken, bijvoorbeeld door plagiaat te plegen.

Bij deze leerdoelen merken wij op dat het gevaar bestaat, dat productdoelstellingen hier gesubstitueerd worden door procesdoelstellingen, met andere woorden dat het LR een doel in zichzelf wordt.

Hoewel het idee van het LR als didactisch middel ons aanspreekt, beschikken wij op dit moment niet over wetenschappelijk onderbouwde informatie over de effectiviteit daarvan.

GEBRUIKSMOGELIJKHEDEN EN AANBEVELINGEN

Op grond van het voorgaande komen wij tot de volgende aanbevelingen voor het gebruik van het Learner Report.

Een open vorm van het LR is in onze ogen geschikt: (a) voor het opsporen van onderwijsdoelstellingen; (b) als één van de instrumenten voor het evalueren van onderwijs, bijvoorbeeld als men leerervaringen op het gebied van zelfkennis wil opsporen en in het kader van formatieve evaluatie. Naast het LR zullen voor het verkrijgen van informatie over afvraagbare (meetbare) kennis en demonstreerbare vaardigheden meestal objectievere instrumenten gebruikt moeten worden; (c) voor ieder onderzoek waarbij men

informatie wenst over wat de lerende zelf zegt, te hebben geleerd; (d) als uitgangspunt voor individuele didactische beslissingen.

Zoals in de paragraaf 'Overwegingen bij het gebruik van een gesloten LR' al is opgemerkt, is een gesloten vorm van het LR vooral geschikt voor het vergelijken van de leerervaringen van verschillende groepen respondenten en is het in die zin vooral bruikbaar voor evaluatie-onderzoek.

Het LR (in open of gesloten vorm) is niet geschikt voor het toetsen van kennis en vaardigheden, als selectiemiddel en als (enige) basis voor het nemen van beleidsbeslissingen.

Over het LR als didactisch instrument kan op dit moment geen wetenschappelijk onderbouwde uitspraak worden gedaan.

Uit de besproken literatuur is gebleken dat veel gebruikers van het LR niet goed geweten hebben 'waar ze aan begonnen'. Het is dan ook raadzaam dat toekomstige gebruikers van het LR zich op de hoogte stellen van ervaringen, die reeds zijn opgedaan bij gebruik van het LR voor soortgelijke doeleinden als die waarvoor zij het zelf willen gaan gebruiken. Als hulpmiddel hierbij vindt men aan het eind van deze paragraaf een overzicht van aanbevolen literatuur.

AANBEVOLEN LITERATUUR

Aan toekomstige gebruikers van het LR zouden wij de volgende literatuur willen aanbevelen:

1. Over de theorie achter het Learner Report: De Groot (1974, 1978).
2. Voor een uitgebreid verslag van instrumentenontwikkeling: Haanstra (1976), Van der Kamp (1980).
3. Over een originele uitwerking van het LR als evaluatie-instrument: Vedder (1984).
4. Met het oog op het waarborgen van de kwaliteit van onderzoek met behulp van het Learner Report (betrouwbaarheid, interne en externe validiteit, objectiviteit): Vedder (1984), pp. 66-68.
5. Over de problemen die zich kunnen voordoen bij het formuleren van items in een gesloten vorm van het LR: Edelenbos *et al.* (1985), Bijlage VII.
6. Over het gebruik van het LR bij personen met een geringe (voor)opleiding: Dresen-van der Linde (1979).
7. Met het oog op het nemen van didactische beslissingen: Kreeft & Rijlaarsdam (1980).
8. Over het LR als didactisch middel: Van Eijk (1984).

LITERATUUR

- Bonset, H. (Red.). (1985). *Nederlands in de heterogene brugklas. Een gevalsbeschrijving van leraarsactiviteiten, curriculum en leereffecten bij leerlingen* (Deel 1 en 2). Enschede: Stichting voor Leerplanontwikkeling.
- Damsma, H., Griffioen, J., Prins, B., Veldkamp, B., & van Willigen, J. (1985). *Taalbeschouwing als reflectiviteit*. Enschede: Stichting voor Leerplanontwikkeling.
- Dresen-van der Linde, L. (1979). *Leren in de Open School*. Groningen: Rijksuniversiteit, Werkgroep Onderwijspsychologie.
- Dwarshuis, M. (1983). *Onderwijsstimulering door en voor tweede jaars P.A. studenten*. Groningen: Rijksuniversiteit, Werkgroep Onderwijspsychologie.
- Edelenbos, P., de Jong, R., & den Boer, P.R. (1985). *Evaluatie van veldstudiecurssussen in Orvelte*. Groningen: Rijksuniversiteit, RION.
- Ericsson, K.A., & Simon, H.A. (1980). Verbal Reports as Data. *Psychological Review*, 87, 215-251.
- Erp, M. van (1985a). 'Toen we monsters aan het nemen waren, kwam er een oude man'. *Verslag van een*

- milieuproject op een scholengemeenschap te Amsterdam*. Amsterdam: Stichting Kohnstamm Fonds voor Onderwijsresearch.
- Erp, M. van (1985b). *Milieuprojecten in het voortgezet onderwijs. Samenvattend onderzoeksverslag*. Amsterdam: Stichting Centrum voor Onderwijsresearch.
- Erp, M. van, & Koppen, J.K. (1984). *'Je mocht wel niet alles, maar je was wel vrij.'* Verslag van een milieuproject op een MAVO te Zaanstad. Amsterdam: Stichting Kohnstamm Fonds voor Onderwijsresearch.
- Eijk, G. H. A. van (1984). The writing of learning experiences as a teaching tool. In R. Langdon, K. Baynes, & P. Roberts, *Design Education: Vol. 5. Proceedings of the Design Policy Conference, London, 1982* (pp. 42-45). London: Design Council.
- Gerritsen van der Hoop, J.W. (1981). *Een toepassing van het 'Learner Report' voor het vaststellen van de effecten van een cursus*. Eindhoven: Technische Hogeschool, Groep Onderwijsresearch.
- Gerritsen van der Hoop, J.W. (1982a). *Evaluatie van de Pre-introductieweek 1982; een programma voor vrouwelijke eerstejaars TH-studenten*. Eindhoven: Technische Hogeschool.
- Gerritsen van der Hoop, J.W. (1982b). 'Learner Reports' als evaluatiemethode. In C. van Dorp en A. Pilot (Red.), *Aspecten van onderwijsresearch* (pp. 63-71). Lisse: Swets & Zeitlinger.
- Godefrooij, P. (1983). *Hoe spelen leerlingen op basisscholen in Rotterdam met woorden?* Amsterdam: Stichting Centrum voor Onderwijsresearch.
- Gradisen, C.J. (1985). *Praktikum zeker niet afschaffen. Een bijvakonderzoek naar het scheikundepraktikum in het Voortgezet Onderwijs*. Utrecht: Rijksuniversiteit, Vakgroep Chemiedidaktiek.
- Greenwald, A.G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, 83, 314-320.
- Groot, A.D. de (1974). The problem of evaluating national educational systems. In H.F. Crombag & Dato N. de Gruyter (Eds.), *Contemporary issues in educational testing* (pp. 9-27). Den Haag: Mouton.
- Groot, A.D. de (1978). Wat neemt de leerling mee van onderwijs? Gedragsrepertoires, programma's, kennis-en-vaardigheden. In *Handboek voor de onderwijspraktijk: Deel 2. Ontwerpen van onderwijs* (pp. 2.3.Gro.A.1-2.3.Gro.A.23). Deventer: Van Loghum Slaterus.
- Groot, A.D. de (1980). Over leerervaringen en leerdoelen. In *Handboek voor de onderwijspraktijk: Deel 2. Ontwerpen van onderwijs* (pp. 2.3.Gro.B.1-2.3.Gro.B.18). Deventer: Van Loghum Slaterus.
- Groot, A.D. de (1982a). Commentaar. In L. van der Kamp en M. van der Kamp (Red.), *Methodologie van onderwijsresearch. (Bijdragen tot de onderwijsresearchdagen 1982; 1)* (pp. 111-126). Lisse: Swets & Zeitlinger.
- Groot, A.D. de (1982b). *Academie en Forum: over hoger onderwijs en wetenschap*. Meppel: Boom.
- Groot, A.D. de (1986). *Over intuïtie*. (Heymans Bulletin nr. HB-85-768-EX). Groningen: Rijksuniversiteit, Psychologische Instituten Heymans.
- Haan, H. de, van Heur, R., Kornmann, K., te Riele, H., & Rijlaarsdam, F. (1985). *Ontwerpen 1. Jaarboek 85/86*. Delft: Technische Hogeschool, Tussenafdeling van het Industrieel Ontwerpen.
- Haanstra, F. (1976). *Van idee naar instrument*. Amsterdam: Kohnstamm Instituut voor Onderwijsresearch.
- Haanstra, F. (1983). Onderzoek naar de resultaten van een educatieve tentoonstelling. In M. van der Kamp, F. Haanstra en J. Oostwoud Wijdenes (Red.), *Kijk op kunstzinnige vorming* (pp. 163-169). Purmerend: Muusses.
- Kamp, M. van der, & Oostwoud Wijdenes, J. (1982). Op zoek naar leerervaringen met het Learner Report. In: M. van der Kamp en L.J. van der Kamp (Red.), *Methodologie van onderwijsresearch* (pp. 75-82). Lisse: Swets & Zeitlinger.
- Kamp, M. van der (1984). Self-assessment as a tool for the evaluation of learning outcomes. *Studies in Educational Evaluation*, 10, 265-272.
- Kamp, M. van der (1980). *Wat neemt de leerling mee van kunstzinnige vorming?* (SVO-reeks 29). 's-Gravenhage: Staatsuitgeverij.
- Kat, E. de, & Riemersma, F.S.J. (1978). *Het projekt 'School-Samenleving': een beschrijving van een jaar projektonderwijs aan het Haags Montessori Lyceum. Verslag over het schoolgericht onderzoek-middenschool 1977-1978*. Amsterdam: Research Instituut voor Toegepaste Psychologie.

- Kesteren, B.J. van (1983). *Ik heb geleerd dat het niet waar is dat ik mijn verlegenheid niet kan veranderen*. (Heymans Bulletin nr. HB-83-664-SW). Groningen: Rijksuniversiteit, Vakgroep Persoonlijkheidspsychologie.
- Kreeft, H., & Rijlaarsdam, G. (1980). Zelfevaluatie in het moedertaalonderwijs. *Levende Talen*, 352, 397-410.
- Maandag, D. (1984). *Constructie van instrumenten voor het meten van niet-technisch-instrumentele vaardigheden*. Haren: RION.
- Meerum Terwogt-Kouwenhoven, K. (1983). Wat hebben eerstejaars geleerd tijdens de propedeuse? In G. de Zeeuw, W. Hofstee en J. Vastenhouw: *Funderend onderzoek van het onderwijs en onderwijsleerprocessen (Bijdragen tot de onderwijsresearchdagen 1983; 6)* (pp. 149-158). Lisse: Swets & Zeitlinger.
- Meijer, K. (1982). 'Learner Reports' over leereffekten van stages. In C. van Dorp en A. Pilot (Red.), *Aspecten van Onderwijsresearch* (pp. 51-61). Lisse: Swets & Zeitlinger.
- Min, M., & Nooter, N. (1982). *Theater in het onderwijs: een oplossing? Van Learner Report tot vragenlijst*. Groningen: Rijksuniversiteit, Werkgroep Onderwijspsychologie.
- Molen, H.T. van der (1985). *Hulp als onderwijs. Effecten van cursussen voor verlegen mensen*. Groningen: Wolters-Noordhoff.
- Nieuwenhuis, L. (1984a). *Participerend leren in het MEAO: Verslag van een onderzoek naar buitenschools leren*. Groningen: Rijksuniversiteit, RION.
- Nieuwenhuis, L. (1984b). *Participerend leren in het MDGO: Verslag van een onderzoek naar stages in verzorging en civiele dienst*. Groningen: Rijksuniversiteit, RION.
- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Oostwoud Wijdenes, J. (1982). *Drama in het voortgezet onderwijs*. Amsterdam: Stichting Centrum voor Onderwijsresearch.
- Raalte, G. van (1981). *Ballet in de basisschool. Stageverslag*. Amsterdam: Stichting voor Onderwijsresearch.
- Raalte, G. van (1983). Ballet in de basisschool. In M. van der Kamp, F. Haanstra en J. Oostwoud Wijdenes (Red.), *Kijk op kunstzinnige vorming* (pp. 170-174). Purmerend: Muusses.
- Rijlaarsdam, G. (1984). Wat leren leerlingen ervan volgens eigen zeggen? Effecten van leerlingenrespons, gemeten aan zelfrapportages. In H. Blok, G. Rijlaarsdam, J.B. Hoeksma en K. de Glopper (Red.), *Moedertaalonderwijs in onderzoek. Recente onderzoeksresultaten gepresenteerd op het VIOT-congres 1984* (pp. 17-32). Amsterdam: Stichting Centrum voor Onderwijsresearch.
- Roefs, A.J.M. (1980). *Evaluatie van een cursus leerlingbegeleiding. Bruikbaarheid van de Learner Report methode*. Groningen: Rijksuniversiteit, Werkgroep Onderwijspsychologie.
- Schouteten, C.J.H. (1984). De vergeten praktikumdoelen. In G. van den Berg e.a., *Forumconferentie SO₃. Amsterdam, 21 mei en 4 juni 1984*. (pp. 13-42). Arnhem: Werkgroep Scheikunde-Onderwijs: Onderzoek en Ontwikkeling.
- Shrauger, J.S., & Osberg, T.M. (1981). The relative accuracy of self-prediction and judgments by others in psychological assessment. *Psychological Bulletin*, 90, 322-351.
- Smith, E.R., & Miller, F.D. (1978). Limits on perception of cognitive processes: a reply to Nisbett and Wilson. *Psychological Review*, 85, 355-362.
- Vedder, J. (1984). *Oriëntatie op het beroep van leraar*. Lisse: Swets & Zeitlinger.
- White, P. (1980). Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bem. *Psychological Review*, 87, 105-112.

Manuscript ontvangen 27-7-1987.

Definitieve versie ontvangen 16-12-1988.

Vergrijzing in het Onderwijs?

Een verklaring voor de ontwikkeling van de leeftijd van leerkrachten

J. de Voogd en G. den Broeder*

ABSTRACT

The development since 1975 in the age structure of teachers in secondary education and vocational colleges has been analysed by type of school. This is compared with the development in the age structure of the total labour force. Projections are made for teachers until 2000. The influence of reduction in teacher employment, as a consequence of the decreasing number of students, upon the mean age of teachers is established. The mean age will continue to rise with about one year in the period 1985-2000. Consequences for the salary costs are discussed.

1. INLEIDING

Al enkele jaren wordt gewezen op de mogelijkheid dat het lerarenbestand in het voortgezet onderwijs zal verouderen. De veroudering wordt veelal in negatieve termen beoordeeld omdat kwaliteitsverlies van het onderwijs en verminderde aandacht voor onderwijsvernieuwing wordt gevreesd (ARVO-I, 1984). De empirische basis voor of tegen de veronderstelde vergrijzing is tot heden nauwelijks geleverd. Dit artikel tracht daarin te voorzien. Als bron is gebruik gemaakt van de sinds 1977 jaarlijks door het CBS uitgevoerde Integrale Personeelstelling Onderwijs (IPTO). Deze bestrijkt ruim 90% van het door het Ministerie van O & W en L & V bekostigde voortgezet (VO, 10 schoolsoorten) en hoger beroepsonderwijs (HBO, 7 schoolsoorten). Onder de leerkrachten (het onderwijzende personeel) is ook de (adjunct-)directie van de scholen begrepen. De inhoud van dit artikel kwam tot stand in aansluiting op een onderzoek waarin de salariskosten van het huidige stelsel van het voortgezet onderwijs worden vergeleken met die van een nieuw stelsel waarin de basisvorming is opgenomen (De Voogd, 1987).

De opzet van dit artikel is als volgt. Allereerst worden per schoolsoort de ontwikkelingen in de leeftijdsstructuur van leerkrachten in het verleden beschreven. Daarna wordt met behulp van regressie-analyse gezocht naar een verklaring van deze ontwikkeling. Vervolgens worden voorspellingen over de toekomstige leeftijdsontwikkeling gedaan. De gevolgen van de te verwachten leeftijdsontwikkeling worden ten slotte kort besproken.

2. DE ONTWIKKELING VAN DE LEEFTIJD VAN LEERKRACHTEN IN DE PERIODE 1974 (1977)-1985/'86

Voor het verleden staan gegevens tot en met het schooljaar 1985/'86 ter beschikking.

Tabel 1 geeft de ontwikkeling van de gemiddelde leeftijd per schoolsoort weer, vanaf 1977 of eerder.

*De auteurs zijn respectievelijk werkzaam als adjunct-hoofd van de afdeling Arbeidsmarkt van het Nederlands Economisch Instituut en als wetenschappelijk directeur van Magnana Mu Publishing and Research, beide te Rotterdam.

Tabel 1. De ontwikkeling van de gemiddelde leeftijd van leerkrachten VO en HBO schooljaren 1974-1985.

Schoolsoort	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	
AVO	38,7	38,4	38,4	38,4	38,5	38,7	38,7	38,9	39,0	39,3	36,6	39,9	Algem. Voorg. O.
LTO	39,6	39,0	39,1	39,2	39,3	39,4	39,2	39,2	39,3	39,2	39,5	39,8	Lager Techn. O.
MTO	41,8	41,7	42,1	42,2	42,3	42,2	41,8	41,8	41,6	41,5	41,5	41,6	Midd. Techn. O.
HuNO*	37,1	36,4	36,3	36,3	36,4	36,6	36,8	37,2	37,3	37,4	37,7	38,1	Huish. en Nijv. O.
LAO				37,1	36,7	36,6	36,4	36,1	35,9	35,7	35,6	35,9	Lager Agrar. O.
MAO				40,0	39,8	39,4	39,1	39,2	38,7	38,1	37,9	37,8	Middelb. Agrar. O.
LMNO				43,0	43,3	43,5	43,1	43,1	42,6	42,1	42,5	42,9	Lager en Midd. Naut. O.
MSPO				37,3	37,2	37,1	37,4	37,7	38,6	39,0	40,2	42,0	Midd. Soc. Ped. O.
KLOS				39,5	39,9	40,3	39,8	40,0	40,2	40,8	42,5	45,1	Opl. v. Kleuterl.
LMEO				37,5	37,6	37,8	37,7	37,6	37,7	37,6	37,7	37,7	L. en Mid. Econ. O.
VO				38,2	38,3	38,5	38,5	38,6	38,8	38,8	39,1	39,1	Voortgez. Onderw.
HTO	44,1	43,1	44,1	44,1	44,4	44,3	44,1	44,2	44,2	44,1	44,0	44,1	Hoger Techn. O.
PA(BO)	40,8	40,6	40,2	40,3	41,1	41,4	41,6	41,9	42,6	42,7	42,8	43,5	Pedag. Academie
HAO				40,1	39,9	40,0	39,9	40,4	40,5	40,2	39,6	39,4	Hoger Agrar. O.
HNO				45,5	45,3	45,1	44,9	45,4	45,3	43,9	44,1	44,8	Hoger Naut. O.
HSPO				38,4	38,7	38,6	38,8	39,1	39,3	39,9	40,2	40,6	Hoger Soc. Ped. O.
HEAO/HMO				40,8	40,8	40,3	40,4	40,3	40,5	40,4	40,0	40,0	Hoger Econ. O.
KUO				42,2	42,3	42,4	42,4	42,4	42,2	42,2	42,4	42,4	Kunst O.
HBO				41,1	41,2	41,1	41,2	41,4	41,6	41,6	41,7	41,9	Hoger Ber. O.

* Bevat voor een klein deel Hoger Onderwijs

Bron: IPTO, CBS.

Tabel 2 geeft inzicht in de ontwikkeling van de relatieve leeftijdsverdeling, eveneens per schoolsoort¹⁾.

In tabel 3 is de ontwikkeling van de leeftijdsverdeling afgezet tegen die van de totale werkzame bevolking en van verwante beroeps categorieën. De leeftijdsgemiddelden zijn geraamd met behulp van klassengemiddelden per leeftijdscategorie.

Uit tabel 1 is af te leiden dat zich in de meeste schoolsoorten vanaf 1977 een proces van lichte veroudering heeft voorgedaan, dat in de jaren tachtig een zekere versnelling ondergaat. De gemiddelde leeftijd in het VO steeg continu van 38,3 in 1977 tot 39,3 jaar in 1985 en in het HBO van 41,1 tot 41,9 jaar. Uitzonderingen op dit beeld vormen het MTO, het HTO, het LAO, het MAO, het NO, het LMEO en het KUO waar nauwelijks van veroudering of zelfs van een daling van de gemiddelde leeftijd sprake is.

Tabel 2. Het percentage van de jonge (35—) en de oude (55+) leeftijdsgroepen in het bestand leerkrachten 1974-1985, per schoolsoort.

		1974	1977	1980	1983	1985
AVO	35—	44	44	41	35	31
	55+	12	10	8	7	6
LTO	35—	37	39	38	35	32
	55+	10	9	9	7	6
MTO	35—	25	23	25	24	23
	55+	12	11	11	10	9
HuNO	35—	51	53	50	44	39
	55+	10	7	6	5	4
LAO	35—		50	53	54	53
	55+		10	7	4	4
MAO	35—		41	43	43	43
	55+		14	12	8	6
LMNO	35—		23	22	24	21
	55+		16	13	11	14
MSPO	35—		46	44	33	13
	55+		5	5	5	8
KLOS	35—		39	36	26	12
	55+		10	9	8	8
LMEO	35—		47	45	42	41
	55+		9	7	5	4
VO	35—		44	41	37	34
	55+		9	8	7	6
HTO	35—	21	19	19	14	14
	55+	18	17	17	16	15
PA(BO)	35—	38	37	26	16	12
	55+	17	13	12	11	9
HAO	35—		38	38	29	32
	55+		12	11	9	7
HNO	35—		17	15	17	17
	55+		20	16	17	20
HSPO	35—		44	41	28	22
	55+		7	8	7	7
HBO	35—		35	34	33	28
	55+		12	12	10	7
KUO	35—		29	27	22	19
	55+		16	15	12	12
HBO	35—		34	30	23	20
	55+		13	12	11	10

Er zijn structurele verschillen in de leeftijdsopbouw tussen schoolsoorten. Deze hangen samen met het niveau van de vereiste bevoegdheden en (wellicht) ook met het feit dat voor bepaalde leraarsberoepen buiten het onderwijs eerst werkervaring wordt opgedaan. Zo is men in het HBO gemiddeld ouder dan in het VO. Ook speelt de geslachtsopbouw een rol: leraressen hebben gemiddeld een lagere leeftijd dan de leraren. In schoolsoorten waar weinig leraressen voorkomen (Het TO en het NO bijvoorbeeld) is de gemiddelde leeftijd dan ook hoger dan in schoolsoorten waar veel vrouwen aanwezig zijn (het HuNO en het SPO).

Voor enkele schoolsoorten zijn ook gegevens vóór 1977 beschikbaar. Die duiden erop dat het proces van veroudering niet permanent is geweest. Vooral in het kwantitatief zeer belangrijke AVO is in het begin van de jaren zeventig sprake geweest van een zekere verjonging.

Tabel 2 geeft meer inzicht in de veranderingen die zich in de leeftijdsopbouw hebben voorgedaan. Vanaf 1977 is er sprake van een vrij aanzienlijke daling van het aandeel van jongeren (35—) in het lerarenbestand. Dit geldt voor vrijwel alle schoolsoorten: uitzonderingen zijn slechts het MTO, het LMNO, het LAO en het MAO. Ook het aandeel van de ouderen (55+) is in de meeste schoolsoorten vanaf het midden van de jaren zeventig (en waarschijnlijk reeds eerder) gedaald. Per saldo leverden deze ontwikkelingen desondanks voor de meeste schoolsoorten een stijging van de gemiddelde leeftijd op. Hierbij is het aandeel van de middencategorie van 35-55-jarigen sterk gegroeid; voor het VO en HBO tesamen van 0,48 in 1977 tot 0,61 in 1985. Uit meer gedetailleerde gegevens blijkt dat de leeftijdsgroep jonger dan 25 jaar vrijwel verdwenen is.

In hoeverre is er nu sprake van een uitzonderlijke leeftijdsontwikkeling in het onderwijs? Uit tabel 3 blijkt dat het lerarenbestand in het VO en het HBO gemiddeld wat ouder is dan de totale werkzame bevolking. Hetzelfde geldt in vergelijking met de hogere beroepsgroepen en in vergelijking met andere leerkrachten (vooral onderwijzers). Het aandeel van jongeren (35—) in het bestand leerkrachten VO en HBO is beduidend lager dan in de overige zojuist genoemde categorieën. De ouderen (55+) zijn in ongeveer gelijke mate vertegenwoordigd bij leraren als in de rest van de werkende bevolking.

Tabel 3. De ontwikkeling van de leeftijd van werkzame leerkrachten vergeleken met die van (vergelijkbare categorieën in) de beroepsbevolking.

		1975	1977	1979	1981	1983	1985
Gemiddelde leeftijd	- beroepsbevolking	36,5	36,5	36,3	36,1	36,5	36,4
	- hogere beroepsgroepen*	37,3	36,8	36,6	36,7	37,5	37,7
	- leerkrachten**	36,6	35,7	36,2	36,8	37,7	37,9
	- leerkrachten VO ±	38,1	38,2	38,5	38,6	38,8	39,3
	- leerkrachten HBO		40,8	41,1	41,4	41,6	41,9
Percentage jonger dan 35 jaar	- beroepsbevolking	51,0	51,5	52,0	52,5	50,0	49,7
	- hogere beroepsgroepen**	49,2	52,1	52,5	51,9	47,0	45,3
	- leerkrachten**	51,1	55,5	53,8	50,8	45,0	43,0
	- leerkrachten VO ±	45,0	43,9	42,5	40,1	37,0	33,7
	- leerkrachten HBO ±	34,5	34,2	32,0	27,5	22,9	19,6
Percentage ouder dan 54 jaar	- beroepsbevolking	10,9	10,4	9,7	8,9	8,8	8,0
	- hogere beroepsgroepen*	9,9	9,5	8,3	7,8	8,9	8,0
	- leerkrachten**	8,1	7,3	6,9	6,7	6,7	5,4
	- leerkrachten VO ±	10,3	9,2	8,7	7,6	6,6	5,6
	- leerkrachten HBO ±	13,3	12,5	12,4	12,5	10,6	9,8

* Wetenschappelijke e.d. vakspecialisten, kunstenaars, beleidsvoerende en hogere functies.

** Code 13 beroepsgroep CBS, inclusief onderwijzers e.d.

Bronnen: CBS, IPTO en AKT

De leraren VO en HBO zijn vanaf 1977 tot 1985 gemiddeld 1,1 jaar ouder geworden. Dit wijkt slechts weinig af van de ontwikkeling bij die beroepscategorieën waartoe zij kunnen worden gerekend. De totale werkzame bevolking is in dezelfde periode gemiddeld genomen nauwelijks in leeftijd veranderd. Daarbij speelt een rol dat het aandeel van (jongere) vrouwen in de beroepsbevolking drastisch is gestegen. Wel is het aandeel van de jongeren bij leerkrachten in het recente verleden sterker gedaald dan bij vergelijkbare categorieën in de werkzame bevolking.

3. ENKELE ACHTERGRONDEN VAN DE LEEFTIJDSONTWIKKELING

Het hiervoor beschreven lichte verouderingsproces van het bestand leerkrachten VO & HBO vraagt om een nadere toelichting. Welke factoren beïnvloeden de leeftijdsstructuur van het leerkrachtenbestand? Gedacht kan worden aan een aantal zaken:

- de algemene ontwikkeling in de leeftijdsstructuur van de potentiële beroepsbevolking;
- de veranderde deelneming van vrouwen aan beroepsarbeid;
- de ontwikkeling van de vraag naar (nieuwe) leerkrachten in VO & HBO;
- de leeftijd waarop men als leraar werkzaam wordt;
- wijzigingen in de leeftijd bij degenen die uit het onderwijs vertrekken, met name onder invloed van de instelling van de VUT en de regeling taakverlichting voor ouderen.

Wat kan over deze factoren worden gezegd? Een algemene veroudering van de potentiële beroepsbevolking zal zijn weerslag kunnen hebben op de veroudering van het lerarenbestand. Zojuist hebben we geconstateerd dat de gemiddelde leeftijd van de werkzame bevolking in zijn totaliteit tot 1985 nauwelijks is veranderd. Dit is mede het gevolg van het stijgende aandeel van vrouwen in de werkzame bevolking: van 26,8% in 1975 tot 33,6% in 1985. De bevolking van 15-64-jarigen vertoont immers wel een zekere veroudering. Zo daalde het aandeel van de 35-ers van 50,8% in 1976 tot 48,6% in 1986 en steeg het aandeel

Tabel 4. De ontwikkeling van de vraag naar leerkrachten in arbeidsjaren per schoolsoort, indexcijfers (1977 = 100), 1975-1985.

	1975	1976	1977	1977 n=	1978	1979	1980	1981	1982	1983	1984	1985
AVO	94	97	100	41.821	102	102	103	103	104	102	99	95
LTO	96	97	100	13.632	102	104	109	111	114	111	106	102
MTO	95	96	100	2.644	104	108	116	123	134	135	138	141
HTO	99	98	100	1.477	103	105	111	114	119	119	123	122
HuNO*	97	97	100	11.466	100	98	97	96	96	94	95	89
PA(BO)*	92	98	100	2.009	91	83	78	70	64	65	78	74
LAO			100	1.633	107	113	116	122	132	142	152	153
MAO			100	612	116	126	135	142	154	165	173	181
HAO			100	297	107	115	119	127	133	139	151	164
LMNO			100	236	104	103	102	107	115	144	149	144
HNO			100	262	101	101	98	98	87	95	96	97
MSPO*			100	538	114	132	161	167	218	197	62	30
HSPO			100	1.710	110	130	128	133	129	145	153	163
KLOS*			100	653	94	92	98	100	104	104	13	9
LMEO			100	4.151	103	110	114	122	127	130	129	141
HEO			100	384	115	137	155	176	185	226	284	316
KUO			100	1.116	109	113	114	118	128	125	132	132
Totaal			100	84.540	102	104	105	107	109	108	105	103

* De ontwikkelingen zijn onderhevig geweest aan herstructureringen.

Bron: IPTO, CBS

van de 55⁺ers van 14,0% tot 14,6%. Deze lichte veroudering van de bevolking is voor de totale beroepsbevolking gecompenseerd door de toenemende deelneming van vrouwen aan beroepsarbeid. De gemiddelde leeftijd van vrouwen in de beroepsbevolking kan voor 1985 worden berekend op ongeveer 33,6 jaar tegen 38,0 jaar bij mannen.

Hoe ligt dit bij leerkrachten? Bij hen is slechts in beperkte mate sprake geweest van een toenemend aandeel van vrouwen. Van 1977 tot 1985 steeg dit in het VO van 28,1% tot 28,5%, maar in het HBO van 20,6% tot 28,4%²⁾. De beperkte toename van vrouwen heeft vooral in het BHO een rem op het verouderingsproces aangebracht en in zekere mate ook in het LTO, het MTO, het LAO, het LMNO en het LMEO e.d. waar stijgingen van het aandeel van vrouwen van vier tot zeven procentpunten zijn te constateren in de periode 1977-1985³⁾.

De leeftijdsveranderingen van het lerarenbestand zijn ook op een andere wijze te benaderen. Indien er geen instroom, uitstroom en sterfte bestond zou de gemiddelde leeftijd jaarlijks met 1 jaar stijgen. De verandering in leeftijd is het gevolg van leeftijdsverschillen tussen instroom en uitstroom: de nieuw aangestelde leerkrachten zijn gemiddeld jonger dan de vertrekkende leerkrachten. De verhouding tussen het aantal aangestelden en vertrokken leerkrachten is echter vooral in de jaren tachtig bij veel schoolsoorten gedaald. Dit hangt samen met de verminderde groei c.q. de daling van het aantal arbeidsplaatsen (de vraag). De ontwikkeling van de vraag is per schoolsoort uit tabel 4 af te lezen.

In het onderwijs als geheel is vanaf 1977 tot 1982 sprake geweest van groei van de vraag naar leerkrachten, terwijl na 1982 – als gevolg van bezuinigingen en de leerlingendaling – de vraag daalt. Per schoolsoort zijn de ontwikkelingen zeer uiteenlopend. Vooral het HBO heeft een sterke groei doorgemaakt, evenals bepaalde lagere en middelbare vormen van beroepsonderwijs (LMNO, AO, MTO, LMEO). Er lijkt een verband te bestaan tussen de ontwikkeling van de gemiddelde leeftijd in een schoolsoort en de vraagontwikkeling: juist daar waar de vraag sterk is gegroeid, is de gemiddelde leeftijd stabiel gebleven of gedaald, terwijl het tegenovergestelde het geval is als de vraag is gedaald. Een daling van de vraag zal naar verwachting (TEASE) ook leiden tot (een toename van de) werkloosheid onder leerkrachten. Dit zou een leeftijdsverhogend effect kunnen hebben op de instroom van abiturienten, maar de cijfers doen dit niet vermoeden: het leeftijdsverschil tussen de leeftijd van aanstelling en afstuderen is nagenoeg constant gebleven op ca. 3 jaar. Van de werkloosheidsomvang als verklarende variabele voor de leeftijdsstructuur is dan ook niet veel te verwachten, temeer daar deze variabele een veelheid aan ontwikkelingen weerspiegelt.

De gemiddelde leeftijd van nieuw aangestelde leerkrachten is in het hier geanalyseerde onderwijs gestegen van 32,0 jaar in 1978 tot 34,7 jaar in 1985, terwijl die van vertrekkenden fluctueerde rond 38 à 39 jaar. Wel zij aangetekend dat deze cijfers slechts indicatief zijn omdat ze niet gecorrigeerd zijn voor de mobiliteit tussen de schoolsoorten die vrij omvangrijk is.

De gemiddelde leeftijd van aangestelde leerkrachten ligt ongeveer 3 jaar hoger dan die van abiturienten van lerarenopleidingen. De laatste steeg van 28,8 jaar in 1979 tot 31,0 jaar in 1986. Het is daarom denkbaar dat de stijging van de gemiddelde leeftijd van nieuw aangestelde leerkrachten te maken heeft met een stijging van het gemiddelde opleidingsniveau. Het aandeel van eerstegraads leerkrachten is het laatste decennium sterk toegenomen: van 32% in 1977 tot 40% in 1985. Het verouderingsproces is dus versterkt doordat de gemiddelde leeftijd van nieuw aangestelde leerkrachten enigszins is gestegen.

De (vrijwillige) VUT-regeling voor onderwijzend personeel is ten dele gelijk aan – en gelijktijdig ingevoerd met – die voor overheidspersoneel, en wel vanaf 1979. Nadat enkele malen verlagingen van de uittredingsleeftijd zijn toegepast, ligt deze vanaf 1984 op 59-

Tabel 5. De fractie van 55-60- en 60-65-jarigen in het personeelsbestand, enkele jaren, per schoolsoort.

	1977		1981		1985	
	55-<60	60-<65	55-<60	60-<65	55-<60	60-<65
AVO	.051	.040	.051	.026	.055	.008
LTO	.055	.034	.060	.024	.055	.008
MTO	.070	.041	.082	.030	.078	.011
HTO	.097	.063	.110	.059	.113	.038
HuNO	.047	.028	.042	.019	.034	.005
PA/PABO	.067	.060	.070	.044	.078	.015
LAO	.052	.042	.040	.024	.034	.005
MAO	.085	.046	.069	.045	.049	.010
HAO	.078	.036	.070	.041	.058	.013
LMNO	.093	.058	.093	.042	.118	.023
HNO	.097	.094	.128	.059	.164	.031
MSPO	.031	.014	.036	.010	.066	.012
HSPO	.041	.023	.053	.020	.050	.016
KLOS	.054	.042	.058	.031	.056	.017
LMEO	.050	.032	.043	.023	.034	.006
HEO	.063	.053	.074	.027	.048	.023
KUO	.087	.061	.090	.051	.090	.030

Bron: IPTO, CBS

jarige leeftijd. Voor zover daarvan gebruik is gemaakt, moet de VUT een verjongende invloed op de leeftijdsopbouw hebben gehad. Tabel 5 geeft daarover indicaties.

In alle schoolsoorten neemt het – overigens reeds vrij geringe aandeel – van de 60+ers na 1979 drastisch af, hetgeen er op wijst dat van de VUT inderdaad op vrij grote schaal gebruik wordt gemaakt. Pas vanaf 1984 mag een (geringe) invloed verwacht worden op het aandeel van de 55-60-jarigen. Uit de hier gepresenteerde cijfers is echter niet af te leiden in welke mate de 59-jarigen van de VUT-regeling gebruik maakten.

4. EEN MODEL VOOR DE ONTWIKKELING VAN DE LEEFTIJDSTRUCTUUR VAN HET LEERKRACHTENBESTAND.

De ontwikkeling van de leeftijdsstructuur van het bestand leerkrachten kan in principe worden verklaard uit een aantal factoren dat in het voorgaande is besproken. Theoretisch gezien zou de beste benadering zijn dat een stromenmodel wordt opgesteld. Daarbij zou de instroom kunnen worden gerelateerd aan het potentiële bestand van degenen die over een bewijs van bekwaamheid beschikken en de uitstroom aan het werkzame bestand (per leeftijdscategorie). Het theoretische model kan diverse vormen aannemen. Plausibel is dat de in- en uitstroomfracties afhankelijk zijn van de arbeidsmarktsituatie, zowel binnen als buiten het onderwijs, van de relatieve loonverhoudingen tussen het onderwijs en andere sectoren, van onderwijsinterne aspecten (zoals de kwaliteit van de arbeid en de arbeidsomstandigheden) en van overheidsbeleid (direct VUT, indirect: arbeidsduurverkortings). Ook is aannemelijk dat er verschillen naar geslacht en type leerkracht (sommige leraren hebben de keuze tussen een baan als leraar of buiten het onderwijs, anderen hebben deze gezien hun opleiding in veel mindere mate) zullen zijn.

Deze benadering kan hier niet worden gevolgd, aangezien de gegevens voor de te verklaren stroomgrootheden naar leeftijd niet beschikbaar zijn. Daarom is gekozen voor een tamelijk eenvoudig regressiemodel, waarbij het leeftijdsgemiddelde van het bestand als de te verklaren variabele is genomen. De navolgende vergelijking is te beschouwen als een

herleide vormvergelijking van het theoretische model.

De verklarende variabelen die empirisch bleken te voldoen, zijn als volgt gedefinieerd:

- dV%: de relatieve verandering van de vraag naar leerkrachten t.o.v. het voorgaande jaar, in procenten;
- VUT: het aantal jaren waarmee de pensioneringsleeftijd wordt vervroegd op basis van de VUT;
- FEM: het aandeel van vrouwen, uitgedrukt in personen, in de tewerkstelling;
- TAAK: de mate waarin in bepaalde jaren aan oudere leerkrachten taakverlichting is toegekend door de overheid, 0 in de jaren t/m 1983/4, 1 in het jaar 1984/5 en 2 in 1985/6;
- c: constante = te schatten gemiddelde leeftijd van mannelijke leerkrachten bij gelijkblijvende vraag, ongeacht VUT en TAAK.

De volgende factoren lijken theoretisch eveneens van invloed, maar bleken empirisch niet te voldoen:

- Het opleidingsniveau van de tewerkstelling, geoperationaliseerd als het aandeel eerstegraads leerkrachten in de tewerkstelling. Gemiddeld zijn deze enigszins ouder dan hun collega's van lagere graadniveaus. Als gevolg van arbeidsmarkt- en onderwijskundige ontwikkelingen is in het verleden van een stijgend aandeel eerstegraders sprake geweest, ook in de tweede- en derdegraadssectoren. Aannemelijk is dat deze ontwikkeling voorlopig door zal gaan.
- Het leeftijdsgemiddelde van abiturienten van lerarenopleidingen. Zoals we hiervoor hebben gezien is dit stijgende. Ten dele wordt dit veroorzaakt door de opgetreden algemene studieduurverlenging, ten dele doordat meer voor de langer durende studies van hogere graadniveaus is gekozen en ten dele doordat reeds werkzame personen op latere leeftijd alsnog aan een deeltijd(vervolg)opleiding als leraar deelnamen. Mogelijk veroorzaakte dit laatste, vrij frequent voorkomende verschijnsel, dat de invloed van deze variabele niet kon worden vastgesteld.
- De gemiddelde leeftijd van de beroepsbevolking (zie tabel 3). Hoewel het aannemelijk is dat een algemene leeftijdsontwikkeling ook bij leerkrachten tot uitdrukking komt, kon dit niet worden vastgesteld. De geringe variatie in de waarden van deze variabele in het verleden kan daaraan debet zijn.
- Een variabele die de mate van verwantschap van het leerkrachtenbestand tussen schoolsoorten weergeeft.

Zoals gesteld, is er sprake van een vrij grote mobiliteit van leerkrachten tussen schoolsoorten. Derhalve zal de leeftijdsontwikkeling van de ene schoolsoort van invloed kunnen zijn op die van andere. Het leerkrachtenbestand is heterogeen naar kwaliteit en in dit opzicht kan er van een meer of minder grote verwantschap tussen schoolsoorten worden gesproken. Het is echter niet geheel duidelijk hoe deze verwantschap precies ligt. Om tot betere resultaten te komen zouden de stromen van leerkrachten tussen schoolsoorten in extenso beschreven moeten worden, waarvoor het beschikbare datamateriaal echter ontoereikend was.

Aan te nemen valt dat de geringe lengte van de tijdreeksen er debet aan is, dat van een aantal van de in aanmerking komende verklarende factoren geen invloed kan worden vastgesteld. Dit betekent dat de met het model gemaakte voorspellingen met de nodige voorzichtigheid dienen te worden bekeken. Variabelen die in de periode 1978-1985 constant of trendmatig waren, en daardoor misschien ten onrechte buiten het model vielen, zouden immers in 1985-2000 een ander verloop kunnen gaan vertonen.

Verondersteld wordt, dat de gemiddelde leeftijd met een verdeelde vertraging reageert op de vraagmutatie. Na herschrijving luidt de te schatten regressie-vergelijking als volgt:

$$GLT = \alpha GLT_{-1} + (1-\alpha) \beta dV\% + \mu (FEM - \alpha FEM_{-1}) + \tau (VUT - \alpha VUT_{-1}) + \delta (TAAK -$$

$$\alpha \text{TAAK}_{-1}) + (1-\alpha) c$$

Gezien de beperkte lengte van de tijdreeksen was het niet mogelijk voor alle coëfficiënten een per schoolsoort verschillende schatting te maken. Daarom is aangenomen dat de invloed op de gemiddelde leeftijd van leerkrachten van het aandeel vrouwen, de VUT-leeftijd en de taakverlichting voor ouderen voor alle schoolsoorten gelijk is, terwijl de constante term (d.w.z. de verwachte gemiddelde leeftijd voor mannelijke leerkrachten ongerekend VUT en taakverlichting), alsmede de mate en snelheid van reactie op de vraagmutatie per schoolsoort, kunnen verschillen.

Met betrekking tot de reactie op de vraagmutatie zijn bovendien ter vermijding van implausibele uitschieters de schoolsoorten zowel voor α als voor β gebundeld in vier clusters, op basis van de uitkomsten bij volledig vrije schatting.

Gezien de bovengenoemde restricties op de coëfficiënten dienden de vergelijkingen voor de schoolsoorten simultaan te worden geschat. Als te minimaliseren doelstellingsfunctie is

Tabel 6. Schattingsresultaten 1978-1985

Schoolsoort	R ²	DW	SE	Constante
AVO	0,98	0,83	0,07	42,2 (1,3)
LTO	0,69	2,92	0,11	41,1 (1,2)
MTO	0,81	1,63	0,14	42,2 (1,1)
HuNO	0,96	1,34	0,12	41,8 (1,5)
LAO	0,91	2,21	0,11	37,2 (0,7)
MAO	0,90	1,81	0,24	34,1 (3,4)
LMNO	0,74	2,22	0,23	44,3 (0,7)
MSPO	0,93	1,64	0,47	51,9 (5,2)
KLOS	0,96	1,87	0,26	44,0 (1,0)
LMEO	- 1,41	1,85	0,11	40,2 (1,2)
HTO	- 0,96	1,60	0,16	44,9 (1,1)
PA	0,95	2,43	0,17	54,0 (5,5)
HAO	0,16	1,08	0,36	41,6 (0,7)
HNO	0,27	1,84	0,49	44,7 (0,5)
HSPO	0,91	1,09	0,22	50,2 (4,4)
HEAO	0,67	2,26	0,16	44,0 (1,3)
KUO	- 0,75	2,07	0,13	44,4 (1,1)
α (vertraging) =	0,961 (0,017) 0,908 (0,032) 0,859 (0,036) 0,825 (0,041)	MAO, MSPO, PA, HSPO AVO, LTO, MTO, HuNO, LMEO, HTO, HEAO, KUO LAO, LMNO, KLOS, HAO HNO		
β (vraag) =	- 0,106 (0,032) - 0,142 (0,075) - 0,205 (0,052) - 0,234 (0,066)	KLOS, PA, HSPO, KUO AVO, LTO, MTO, HuNO, MAO, LMEO, HTO LAO, LMNO, MSPO, HAO, HEAO HNO		
μ (vrouwen) =	- 5,099 (1,654)			
τ (VUT) =	- 0,038 (0,028)			
δ (TAAK) =	0,163 (0,056)			

R² = kwadraat van de meervoudige correlatie-coëfficiënt.

DW = Durbin-Watson statistic (autocorrelatie van de residuen)

SE = standaardafwijking van de residuen

Tussen haakjes achter de geschatte coëfficiënten staan de bijbehorende standaardfouten vermeld.

Alle statistics (behalve DW) zijn gecorrigeerd voor vrijheidsgraden.

hierbij gehanteerd het totaal van de 17 kwadratensommen van de residuen per schoolsoort⁴⁾.

De uitkomsten van de schattingsprocedure staan vermeld in tabel 6. De resultaten zijn niet voor alle schoolsoorten bevredigend. Voor LMEO, HTO en KUO is de standaardafwijking van de residuen groter dan die van de gemiddelde leeftijd zelf (dus een negatieve R^2)⁵⁾. De residuen bij AVO, HAO en HSPO vertonen positieve en die bij LTO negatieve autocorrelatie. De schattingen van de constante termen voor PA, MSPO, HSPO en MAO zijn tamelijk onzeker. Over het algemeen zijn de resultaten in statistisch opzicht echter bevredigend.

Opvallend is, dat de gemiddelde leeftijd van vrouwelijke leerkrachten ceteris paribus 5 jaar lager is dan van mannelijke leerkrachten. Het vervroegen van de uittredingsleeftijd heeft de vergrijzing in het voortgezet onderwijs nauwelijks verminderd en is zelfs meer dan ongedaan gemaakt door de ingevoerde taakverlichting voor oudere leerkrachten.

De constante termen weerspiegelen dat leerkrachten in de PA en het SPO naar verhouding aanzienlijk ouder zijn. In het MAO en het LAO wordt juist meer van jongere leerkrachten gebruik gemaakt.

5. VOORUITBEREKENINGEN TOT HET JAAR 2000; ENKELE IMPLICATIES

De gemiddelde leeftijd is tot het jaar 2000 geprognostiseerd met behulp van de in tabel 6 weergegeven regressie-resultaten. De voorspelde leeftijdsontwikkeling is weergegeven in tabel 7⁶⁾.

De toekomstige ontwikkeling van de vraag en het aandeel van vrouwen in de tewerkstelling is ontleend aan TEASE-85 (De Voogd, 1986). Er is tevens nagegaan welk effect kan uitgaan van de verdere verlaging van de VUT-leeftijd, die tegelijkertijd een beperking van de regeling voor arbeidsduurverkortung (taakverlichting) voor ouderen inhoudt (Concept van Wet, 1987).

Zoals al eerder is opgemerkt, dienen de prognoses met de nodige voorzichtigheid te worden bezien, omdat de kans aanwezig is dat als gevolg van de geringe lengte van de schattingsperiode (8 jaar) belangrijke verklarende variabelen in het model ontbreken. Anderzijds kon van een aantal variabelen – met name de vraagverandering en de man/vrouwverhouding – de invloed wel worden vastgesteld, zodat de hier gepresenteerde voorspelling in principe een betere indicatie geeft dan het alternatief: een simpele trendextrapolatie.

De minste veroudering – of zelfs verjonging – lijkt het AO te wachten te staan, terwijl de sterkste veroudering zal gaan optreden bij NO, PABO, HSPO en HEAO. De verlaging van de VUT-leeftijd tot 57 jaar bij volledige uittreding zal volgens deze uitkomsten op lange termijn een verlaging van de gemiddelde leeftijd van ongeveer 0,24 jaar tot gevolg kunnen hebben. Benadrukt zij daarbij echter dat dit niet meer dan een ruwe schatting kan zijn, omdat allerlei nieuwe regels in verband met de herbezetting (die op kostenneutraliteit gericht zijn) niet in het model konden worden opgenomen.

De berekende verschillen in leeftijdsontwikkeling zijn voorts vooral toe te schrijven aan de verschillen in vraagontwikkeling en in mate van 'feminisering'. De vrij sterke vraagdalingen die zich bij veel schoolsoorten zullen voordoen, worden, in hun effect op de veroudering, gemitigeerd door het veelal toenemende aandeel van vrouwen, dat – in personen uitgedrukt – nog enigszins groter is dan uit tabel 7 voor full-time equivalenten kan worden afgelezen.

We kunnen de voorspellingen ook plaatsen tegen de achtergrond van vooruitberekeningen die het CPB (Op de Beke, 1987) voor de beroepsbevolking heeft gemaakt.

Uit tabel 8 blijkt dat de beroepsbevolking sterker zal verouderen: een stijging van de gemiddelde leeftijd van ruim 2 jaar tegen slechts bijna 1 jaar bij leraren van 1985-2000. Dit

verschil is in zoverre plausibel dat de leeftijdsstructuur van leerkrachten afwijkt van die van de gehele beroepsbevolking. Wordt bij de laatste de veroudering meestal veroorzaakt door de sterke vermindering van het aandeel jongeren, zo is zulks bij leerkrachten veel minder het geval aangezien bij hen het aandeel van de jongeren al vrij laag is.

Welke implicaties hebben deze resultaten? Er mag worden verwacht dat de vastgestelde tendens van afname van jongeren in het leerkrachtenbestand zich, afgezien van bijzondere effecten van het nieuwe VUT-beleid, zal voortzetten, zij het dat in die schoolsoorten waar men er in slaagt meer vrouwen aan te trekken de veroudering beduidend kan worden tegengegaan. De middelbare leeftijdscategorie van 25-55 jaar zal blijven groeien en ook enigszins verouderen. De groep 60+ers heeft reeds thans een zeer gering aandeel in het bestand en zal – zolang de VUT in stand blijft – erg klein blijven.

Verdere verlaging van de VUT-leeftijd volgens het onlangs ingevoerde beleid zal zeker het aandeel van de groep 55-60-jarigen in het bestand verlagen, maar op de gemiddelde leeftijd slechts een beperkte invloed hebben.

Er is voorts een effect op de loonkosten van het onderwijs te verwachten. Gezien de structuur van salariëring in het onderwijs, die in sterke mate door anciënniteit en door leeftijd wordt bepaald, zullen de kosten per arbeidsplaats toenemen. Dit geldt althans indien we ervan uitgaan dat de leeftijd-salarisprofielen exogeen zijn. In werkelijkheid zijn de salarisschalen voor leraren afgeleid van die van ambtenaren. Aannemende dat ook bij deze laatste van een zekere veroudering sprake zal zijn, is nog de vraag of de toekomstige economische ontwikkeling en/of die van het overheidsbudget niet tot structuurwijzigingen in de inkomensverhoudingen naar leeftijd zullen leiden. Dat van de veranderingen in de leeftijdsstructuur een bepaalde invloed op de loonkosten uitgaat, is voor de totale beroepsbevolking reeds enkele malen aangetoond (o.m. Van Reyn, 1984). Nemen we aan dat de beloningsstructuur naar leeftijd voor leerkrachten niet zal veranderen, dan kan het effect van de veroudering op de loonkosten van leerkrachten voor het VO worden geraamd (De Voogd, 1987). Tabel 9 geeft daarin inzicht.

Als gevolg van de veroudering zullen de loonkosten in het onderwijs iets minder dalen dan het aantal arbeidsplaatsen. Dit laatste neemt in de periode 1985-2000 nog met ruim 25% af, terwijl de loonkosten met ongeveer 22% zullen dalen. Bij de hier gepresenteerde gegevens over de loonkostenontwikkeling is overigens afgezien van de invloed van andere factoren. Zo bestaat er onzekerheid over het effect van de niet-lineaire relaties tussen de vraag en de leerlingenaantallen bij de daling van de laatste (Interdepartementale Werkgroep, 1986) op de loonkosten. Ook zijn de overgangsrechten die ontstonden bij herstructurering van de salarissen van het onderwijspersoneel in 1985 niet in de cijfers van tabel 9 betrokken. Deze laatste zullen het eerste decennium waarschijnlijk eveneens tot een verhoging van de kosten per arbeidsplaats leiden (De Voogd, 1987). Tenslotte is afgezien van eventuele effecten van de veroudering op de te verwachten verhoging van de bijscholingskosten.

Noten

- 1) Meer gedetailleerde gegevens zijn op verzoek door de auteurs te leveren.
- 2) Dit is niet hetzelfde als de ontwikkeling van het aandeel in de werkgelegenheid in arbeidsjaren. De genoemde cijfers betreffen hier aantallen personen.
- 3) Dit wil overigens niet zeggen dat vrouwen meer gaan lesgeven in bijvoorbeeld de technische vakken. Uit ander NEI-onderzoek blijkt dat er geen sprake van is – en voorlopig ook niet zal zijn – dat vrouwen een groter aandeel krijgen in die leraarsberoepen die als 'mannelijk' te boek staan. Zie (De Voogd, 1986).
- 4) Dit betekent dat er geen rekening is gehouden met verschillen in varianties tussen de residuen van verschillende schoolsoorten. Op voorhand is er ook geen reden om aan te nemen dat voor de ene schoolsoort het toeval een grotere rol speelt dan voor de andere. Tabel 6 geeft (uiteraard) voor de specifieke periode 1978-1985 wel verschillen in standaardfouten (SE) te zien. Om te kunnen

Tabel 7. Vooruitberekende gemiddelde leeftijd leerkrachten.

Schoolsoor- ten	Feitelijke waarde		Bij ongewijzigd beleid			Bij verlaging VUT-leeftijd		Vraag in 2000		Aandeel vrouwen in tewerkstel- ling*	
	1978	1985	1990	1995	2000	2000	t.o.v. 1985*	1985	2000		
AVO	38,50	39,88	40,43	40,47	40,38	40,15	0,79	0,203			0,264
LTO	39,27	39,77	40,38	40,57	40,63	40,39	0,69	0,078			0,058
MTO	42,34	41,59	41,80	42,13	42,15	41,91	0,74	0,043			0,026
HuNO	36,43	38,10	38,78	38,83	38,66	38,42	0,55	0,580			0,657
LAO	36,66	35,93	36,47	36,48	36,49	36,25	0,79	0,121			0,130
MAO	39,77	37,83	36,85	35,72	34,73	34,25	0,77	0,092			0,257
LMNO	43,29	42,88	44,01	44,38	44,34	44,10	0,61	0,058			0,039
MSPO	37,18	42,04	•	•	•	•	•	0,509		•	•
KLOS	39,92	43,05	•	•	•	•	•	0,527		•	•
LME0	37,62	37,71	38,23	38,50	38,49	38,25	0,74	0,226			0,261
VO	38,3	39,3	39,1	40,0	39,7	39,5	0,75	0,223			0,252
HTO	44,36	44,13	44,20	44,44	44,61	44,37	1,16	0,035			0,041
PA(BO)	41,05	43,48	45,24	46,48	47,46	47,22	0,73	0,177			0,199
HAO	39,90	39,37	39,77	40,57	40,57	40,33	0,88	0,111			0,172
HNO	45,28	44,83	45,13	45,38	45,46	45,22	1,42	0,026			0,013
HSPO	38,70	40,63	41,94	43,01	43,86	43,62	0,79	0,331			0,362
HEAO	40,76	39,96	40,84	41,95	42,55	42,32	0,73	0,167			0,152
KUO	42,33	42,44	42,77	42,92	43,04	42,80	1,00	0,226			0,229
HBO	41,2	41,9	42,7	43,5	44,0	43,7	0,90	0,179			0,197
VO+HBO	38,6	39,6	40,3	40,5	40,5	40,3	0,77	0,217			0,244

* Bron: Voorspellingen TEASE-85 (De Voogd, 1987). Het betreft hier gegevens berekend naar het aantal volledige arbeidsplaatsen m.i.v. niet in IPTO opgenomen delen van het onderwijs.

Tabel 8. De te verwachten gemiddelde leeftijd van de beroepsbevolking (15-64 jaar) en de werkzame bevolking in het VO en het HBO, enkele jaren.

	1977	1985	1990	1995	2000
Beroepsbevolking**	35,9*	35,7*	36,0	37,0	38,3
Leraren VO	38,2	39,3	39,9	40,0	39,9
Leraren HBO	40,8	41,9	42,7	43,5	44,0
Leraren VO + HBO	38,5	39,6	40,3	40,5	40,5

* Van de werkzame bevolking resp. 36,5 en 36,4

** Berekend op basis van Op de Beke, 1987

Tabel 9. De kosten per arbeidsplaats in duizenden guldens (in prijzen van 1985)*

Schoolsoort	1985	2000 (bij verlaging VUT)
AVO	76,8	79,8
LTO	69,6	71,1
MTO	81,2	84,9
HuNO	68,6	68,5
LAO	64,5	65,1
MAO	75,6	66,2
LMNO	78,4	83,9
MSPO	82,9	92,2
LMEO	72,0	75,2
VO	74,4	76,7

* Voor het HBO zijn geen berekeningen gemaakt. Zie voor de methodiek ook: De Voogd, 1987, waar de verdeling over functies en salarisschalen wordt behandeld.

vaststellen of dit systematisch is – zodat de gemaakte veronderstelling van gelijke varianties dient te worden verworpen – lijkt de schattingsperiode echter (andermaal) te kort. Dit laatste geldt a fortiori ook voor het kunnen vaststellen van een eventuele samenhang tussen de residuen van verschillende schoolsoorten. Hierbij zij opgemerkt, dat de toevoeging van de variabelen TAAK en VUT de samenhang tussen residuen die er anders in de laatste jaren van de schattingsperiode zou zijn geweest, verklaren en (dus) weergeven.

- 5) Hieruit mag niet de conclusie worden getrokken dat voor een schoolsoort waar de R^2 negatief is, het model beter niet toegepast zou kunnen worden. Bij een aantal schoolsoorten dat ver uitstijgt boven het aantal waarnemingen in de tijd is het veeleer te verwachten dat één of enkele malen het 'toeval' tegengesteld is aan de invloed van de verklarende variabelen, zodat het lijkt of deze invloed afwezig is. Er is echter geen reden om aan te nemen dat dit voor de betrokken schoolsoort(en) in de toekomst ook het geval zal zijn.
- 6) De hier vermelde gegevens wijken om redenen van definitieverschillen wat betreft de schoolsoorten in het VO af (De Voogd, 1987). Bovendien is aldaar een simpeler verklaringmodel gehanteerd, waarbij o.m. van de verdeling naar geslacht is geabstraheerd.

GERAADPLEEGDE LITERATUUR

- ARVO-I (1984), Adviesraad voor de eerste fase van het voortgezet onderwijs, Immobilititeit en de gevolgen voor het onderwijs, Zeist.
- Op de Beke, J.M.J. (1987), Herziening van het trendmatige arbeidsaanbod 1985-2000, interne notitie CPB, nr. 10, 16 april 1987, 's-Gravenhage.
- Centraal Bureau voor de Statistiek (CBS), (diverse jaargangen), Arbeidskrachtentellingen, Staatsuitgeverij, 's-Gravenhage.
- Centraal Bureau voor de Statistiek (CBS), (diverse jaargangen), Maandstatistiek van de Bevolking, Staatsuitgeverij, 's-Gravenhage.

- Centraal Bureau voor de Statistiek (CBS, 1984), Instroom, herintreding en doorstroom van leerkrachten in het gewoon lager onderwijs, het buitengewoon onderwijs en het voortgezet onderwijs, 's-Gravenhage.
- Centraal Planbureau (1986), De onderwijsprognose 1986, leerlingen, schoolverlaters en bevolking naar opleidingsniveau tot 2000, 's-Gravenhage.
- Concept van Wet van houdende regels met betrekking tot de doorstroming en verlaging van de uittredingsleeftijd van onderwijspersoneel (Wet bevordering doorstroming onderwijspersoneel), 21-12-1987.
- Interdepartementale Werkgroep (1986), Heroverweging 1985, Deelrapport nr. 82, Demografische verandering en de ontwikkeling van de collectieve sector, 's-Gravenhage.
- Van Reyn, H. (1984), Verdiende lonen versus regelingsloon; de ontwikkeling in de jaren 1972-1983, in: Supplement bij de sociaal-economische maandstatistiek, no. 10, blz. 18-28, CBS, 's-Gravenhage.
- De Voogd, J., Den Broeder, G., Stijnen, H. (1986), De capaciteit van de universitaire en de nieuwe lerarenopleidingen, TEASE-85, NEI, Rotterdam.
- De Voogd, J., Stijnen, H., De salariskosten van een nieuw stelsel van voortgezet onderwijs; de basisvorming in het voortgezet onderwijs, NEI, Rotterdam.
- Weisbeek, J. (1985), Vergrijzing, ontgroening en collectieve uitgaven, SCP, Rijswijk.

Manuscript ontvangen 23-6-1987

Definitieve versie ontvangen 19-10-1988

Notities en Commentaren

EVALUATIE VAN DE BASISVORMING COMMENTAAR OP HET KADER VOOR HET UIVOERINGSPLAN

Wijnand Th.J.G. Hoebe

RION, Instituut voor Onderwijsresearch, Rijksuniversiteit Groningen

In juli 1988 is het zesde advies van de Coördinatiecommissie Evaluatieplan Voortgezet Onderwijs (CCE) aan de minister en de staatssecretaris van Onderwijs en Wetenschappen aangeboden. Daarin wordt geadviseerd over het kader voor het uitvoeringsplan Evaluatie Invoering Basisvorming. Een belangrijk onderdeel van het advies betreft het kader voor het uitvoeringsplan dat is geschreven door prof.dr. J.L. Peschar. De Commissie neemt het kaderplan als geheel voor haar rekening en geeft tegelijkertijd vorm aan de eigen professionele verantwoordelijkheid van de auteur door het onder zijn naam te presenteren. Voor een kritisch commentaar op een discussie over de inhoud van dit plan heeft een dergelijke constructie het voordeel dat de gesprekspartner duidelijk is. Wie in discussie wil gaan over de voornemens voor de Evaluatie Basisvorming, gaat in discussie met Peschar. Of de CCE zich aangesproken voelt, zal moeten blijken.

Het kaderplan verdient een kritische discussie. Het verdient deze discussie op de eerste plaats omdat het de invoering van een belangrijke onderwijsvernieuwing betreft. Belangrijker echter is dat het een gedegen en goed onderbouwd evaluatieplan is dat als een voorbeeld kan worden gezien van hoe grootschalige beleidsevaluatie in het onderwijs met eenduidige effectcriteria moet worden aangepakt. Vraagstellingen, criteria en design van het evaluatieonderzoek zijn helder verwoord, goed beargumenteerd, en leiden tot de verwachting dat het onderzoek in staat zal zijn antwoord te geven op de vraag of het beleid inzake basisvorming het beoogde resultaat zal hebben gehad. Naast de evaluatie van het onderwijsvoorrangsbeleid is dit het tweede voorbeeld van een dergelijke evaluatie. Na de op zijn zachtst gezegd gebrekkige evaluaties van de grote vernieuwingsprogramma's in de jaren zeventig zijn we met de evaluatie van resultaten van onderwijsbeleid op de goede weg.

De hier volgende kritische opmerkingen moeten dan ook worden gewogen tegen de achtergrond van een principiële instemming met het evaluatieplan. Het is goed, maar het zou beter kunnen.

Mijn belangrijkste kritiekpunten hebben betrekking op de reductie van vraagstellingen met betrekking tot met basisvorming samenhangende schoolkenmerken (vormgeving van en ontwikkelingen binnen de basisvorming) tot beschrijvende in plaats van evaluatieve vraagstelling, op een te vrezen sluipende criteriumverschuiving van leerprestaties naar schoolloopbanen, en op de organisatorische scheiding tussen dataverzameling, analyse, en rapportage. Op de twee eerstgenoemde kritiekpunten wordt dieper ingegaan, omdat ze de doelstellingen en de opbrengsten van de evaluatie raken. Het laatste punt doet slechts twijfel ontstaan over de haalbaarheid van de voorgestelde tijdsplanning. Omdat hiermee de voorgenomen parlementaire besluitvorming in het najaar van 1997 in het geding komt, is dit punt van zodanig belang dat daar kort op wordt ingegaan.

Voorgesteld wordt de analyses van de gegevens door vier teams te laten plaatsvinden die aan de slag kunnen op de basis van een gegevensbestand dat beschikbaar wordt gesteld door degenen die de gegevens van de leerlingcohorten verzamelen. De resultaten van de

vier analyses worden gecombineerd in een eindrapport welke eindrapportage in één hand wordt gehouden. Met name de analyses en de eindrapportages zijn telkens in een zeer kort tijdbestek gepland. Voor deze zes activiteiten zullen afzonderlijke contracten gesloten worden. Ervaring met de planning en uitvoering van redelijk omvangrijke onderzoeksprojecten leert dat ook hier de wet van Murphy van toepassing is: er gaat altijd wel iets mis; de uitvoering verloopt altijd anders dan in de planning is voorzien. Om overeengekomen termijnen toch te kunnen halen moet vaak veel worden geïmproviseerd. De voorgestelde uitvoering van de evaluatie basisvorming is op een grotere schaal en aanzienlijk ingewikkelder dan het onderwijsonderzoek dat we tot dusverre zijn gewend. Het beroep op het improvisatietalent van de uitvoerders zal dienovereenkomstig groter zijn. De organisatorische en contractueel vastgelegde scheiding tussen methodologisch op elkaar volgende activiteiten reduceert het aantal improvisatiemogelijkheden aanzienlijk. Dienengevolge lijkt het onwaarschijnlijk dat het parlement in het najaar van 1997 daadwerkelijk de beschikking zal hebben over de eindrapportage van het eerste cohort, en de rapportages na vier, respectievelijk na drie jaar van het tweede, respectievelijk het derde cohort.

Gegevens op school- en klasseniveau

Verwacht mag worden dat een verklaring van de met basisvorming bereikte effecten na controle voor instroomkenmerken van leerlingen gezocht moet worden in variabelen op school- en klasseniveau. Mijn voornaamste bezwaar tegen het voorliggende evaluatieplan is, dat ondanks alle vergelijkende vraagformuleringen in die richting, de gegevensverzameling binnen de vergelijkende evaluatie van zodanig globale aard zal zijn dat hiervan weinig effectenverklaring kan worden verwacht. Naast het voornemen om rekening te houden met verschillen in schooltypen wordt voor de vergelijkende evaluatie voorgesteld de nadruk te leggen op prestatiegericht beleid van de school en op de organisatorische condities voor gestructureerd onderwijs en voor de maximalisering van het leren (pagina 60). Wat er in feite op klasse- en leraarniveau wordt gedaan aan vorderingenregistratie en feedback, of het prestatiegerichte schoolbeleid wordt vertaald in dienovereenkomstige attitudes van leraren, en hoe dit op leerlingniveau doorwerkt in huiswerkpatronen en andere effectieve tijdsbesteding, dreigt daarmee buiten de vergelijkende evaluatie te vallen. Voor dergelijke gegevens wordt verwezen naar de dieptestudies op schoolniveau die zijn voorgenomen in het kader van de niet vergelijkende, beschrijvende vraagstellingen, waarbinnen tevens de specifieke ervaringen met vormgeving en ontwikkeling van basisvorming worden beschreven.

De gevolgen van deze voornemens zijn dat waar het eigenlijk om gaat in het onderwijs op scholen buiten het gezichtsveld van de vergelijkende evaluatie dreigt te vallen. In plaats daarvan wordt de nadruk gelegd op variabelen die ver van het feitelijk gedrag van leraren en leerlingen verwijderd zijn, en dan nog wel voor ingekorte versies van de daarvoor in aanmerking komende meetinstrumenten. Dit soort keuzes heeft gevolgen voor de stabiliteit van de meting van schoolkenmerken: gekozen wordt voor kenmerken en voor meetprocedures die de risico's van onbetrouwbaarheid en ongeldigheid van de metingen verhogen. De afstand in een eventuele causale keten tussen de gekozen schoolkenmerken en de effectvariabelen en de introductie van betrouwbaarheids- en geldigheidsrisico's zijn de voornaamste argumenten voor de verwachting dat deze kenmerken weinig verschillen in effecten zullen kunnen verklaren. Dat is temeer teleurstellend omdat rekening moet worden gehouden met de mogelijkheid dat ook scholen de effectcriteria zullen realiseren die niet, althans niet in de eerste fase van het evaluatieonderzoek naar basisvorming overgaan. Het zal zeer moeilijk worden om binnen de gegevens van het evaluatieonderzoek een verklaring voor een dergelijk verschijnsel te vinden.

De suggestie om in het kader van dieptestudies schoolkenmerken intensiever en uitgebreider te onderzoeken om met name de causale mechanismen sterker te onderbouwen (pagina 61) kan in dit verband niet ernstig worden genomen. De dieptestudies vinden immers plaats op scholen die naar basisvorming overgaan. Zij hebben als voornaamste doelstelling ervaringen te beschrijven met de vormgeving van en de ontwikkeling naar basisvorming. Het ontbreken van een vergelijkend perspectief maakt het onderbouwen van causale mechanismen, daar waar het gaat om de invloed van schoolkenmerken op de effectcriteria problematisch. Daarmee worden de kansen minimaal dat we uit het evaluatieonderzoek iets leren waarmee scholen hun voordeel kunnen doen om hun ontwikkelingen naar basisvorming zo effectief mogelijk te kunnen maken.

Met dit laatste punt wordt een kernprobleem geraakt van evaluatie van onderwijsprogramma's. Uiteraard behoort een dergelijke evaluatie niet gericht te zijn op het aantonen van de sinds jaar en dag bekende effecten van (instroom)kenmerken van leerlingen. Evaluatie van onderwijsprogramma's dient erop te zijn gericht dat effecten van onderwijs worden aangetoond onder controle van leerlingkenmerken. Bij het aantonen van effecten van onderwijs moet onderscheid worden gemaakt tussen effecten van beoogde onderwijsveranderingen en effecten van onderwijskenmerken die in beginsel niet door de beoogde veranderingen worden geraakt. Op die manier worden effecten van de beoogde verandering geïsoleerd. Dergelijke effecten kunnen bovendien alleen uit vergelijkende evaluatie blijken. In de vergelijkende evaluatie wordt alleen het formele kenmerk van de beoogde verandering, het al dan niet overgaan naar basisvorming, meegenomen, terwijl bij de keuze van andere onderwijskenmerken de voorkeur wordt gegeven aan kenmerken die in eventuele causale ketens ver van de effectcriteria verwijderd zijn. Daarentegen wordt aanzienlijke tijd en energie gestoken in de verzameling van longitudinale gegevens op leerlingniveau. Met name vanuit de doelstelling van de evaluatie kan op deze afweging van prioriteiten worden afgedongen. Met het voorgestelde evaluatieplan wordt een schat aan gegevens verzameld waarin een toekomstige generatie schoolloopbaanonderzoekers zijn tanden kan zetten. Het blijft echter de vraag in hoeverre de resultaten van het onderzoek bruikbaar zullen zijn voor praktijk en beleid. Indien de effectenverschillen tussen scholen die naar de basisvorming zijn overgegaan, en andere scholen gering zullen blijken te zijn, mag van het voorgestelde evaluatieonderzoek weinig gedetailleerd inzicht worden verwacht in causale mechanismen die wel tot gunstige effecten zullen leiden.

Effectcriteria

Bij de invoering van basisvorming wordt voorgesteld om de mate waarin leerlingen eindtermen realiseren te meten aan de hand van toetsen in de veertien vakken op de twee niveaus. Bij de invoering wordt er tevens vanuit gegaan dat niet alle leerlingen drie jaar over basisvorming zullen doen, maar sommigen vier of zelfs vijf jaar.

In het evaluatieplan wordt voorgesteld om leerlingen na drie jaar vergelijkende toetsen voor te leggen op vier centrale vakken en om in een volgende periode van drie jaar twee loopbaanmetingen te doen bij dezelfde leerlingen. De loopbaanmetingen zijn uitdrukkelijk niet bedoeld om leerprestaties te meten, maar om vast te stellen welk niveau leerlingen hebben bereikt (voor leerlingen die langer over basisvorming mogen doen, wordt vastgesteld hoe snel ze op welk niveau de toetsen in veertien vakken hebben afgelegd; voor leerlingen die reeds aan het werk zijn, wordt informatie over de eerste werkkring verzameld; voor overige leerlingen worden schoolloopbaangegevens vastgelegd, waarbij het behalen van het VWO-eindexamen alleen zal kunnen worden vastgesteld van de leerlingen die dit examen in een rechte lijn behalen).

De keuze voor vergelijkende toetsen op vier centrale vakken na drie jaar wordt vooral beargumenteerd met behulp van de vergelijkingsdoelstelling: vergelijking tussen scholen

die wel en die niet naar basisvorming zijn overgegaan; en vergelijking in de tijd, onder andere met een nu reeds vast te stellen baseline. De veertien vakken basisvorming zullen niet aan alle leerlingen worden onderwezen; de vier kernvakken daarentegen wel. Deze kernvakken betreffen: Nederlands, Engels, Wiskunde en Biologie. Niet beargumenteerd wordt de beslissing om de snelheid waarmee leerlingen een bepaald niveau op de veertien vakken voor basisvorming hebben behaald, allen maar te meten bij die leerlingen die langer over basisvorming mogen doen. Van de overige leerlingen mag immers niet worden verwacht dat ze allen op tenminste het laagste niveau voldoende scores op de toetsen voor de veertien vakken. Evenmin is onderzoek voorzien waarin kan worden nagegaan in hoeverre scores op de vergelijkende toetsen van de vier kernvakken als predictoren kunnen worden gezien voor de snelheid waarmee leerlingen een bepaald niveau basisvorming bereiken.

Er worden dus twee groepen van criteria geïntroduceerd die geen van beide kunnen worden gerelateerd aan het bereikte niveau van basisvorming en aan de snelheid waarmee dat niveau wordt bereikt. Op zichzelf valt de keuze voor deze twee groepen van criteria te billijken vanwege de noodzaak voor een vergelijkende evaluatie. Niet te billijken valt dat wordt afgezien van de verhoudingsgewijs geringe investering die zal zijn gemoeid met het nagaan van de relaties tussen beide groepen van criteria en de doelstellingen van basisvorming, zoals operationeel gemaakt met snelheid en niveau van bereikte basisvorming. Op basis van het voorgestelde evaluatieplan hebben scholen met en scholen zonder basisvorming te veel mogelijkheden om de uitkomsten van het evaluatieonderzoek te beïnvloeden door zich te concentreren op onderwijs in de vier genoemde kernvakken. Daarmee zouden zowel het vigerende onderwijs als het met basisvorming beoogde onderwijs op een onaanvaardbare manier worden verengd.

ontvangen 18-10-1988

OVER DE EVALUATIE VAN DE BASISVORMING. EEN REACTIE OP HOEBEN.

Jules L. Peschar¹⁾

Sociologisch Instituut, Rijks Universiteit Groningen.

Het is verheugend dat er aandacht wordt besteed aan de voorgenomen evaluatie van de geplande Basisvorming in het voortgezet onderwijs. Allereerst is het van belang dat reeds bij de introductie van een onderwijsinnovatie de discussie over mogelijke alternatieven van evaluatie worden gevoerd. Dit is te meer van belang, omdat in het voorontwerp van wet een expliciete verwijzing naar de evaluatie wordt gemaakt. In de tweede plaats blijken de voorstellen voor de evaluatie van grotere onderwijsprojecten over het algemeen beperkt te circuleren binnen het ambtelijke (ministeries) en semi-ambtelijke (SVO, CCOO, CITO enz) circuit. Zo is het Kader voor de evaluatie van de Basisvorming integraal onderdeel van het ambtelijke zesde advies van de CoördinatieCommissie Evaluatieplan Voortgezet Onderwijs CCE aan de minister²⁾. Alleen degenen die – min of meer toevallig – er mee in aanraking komen, kunnen zich nog een oordeel aanmeten voordat de werkelijke evaluatie van start gaat. Een meer publieke discussie vooraf is daarom een goede zaak. Er is immers sprake van een betrekkelijk unieke situatie waar het evaluatieplan wel, maar het evaluandum nog niet bestaat. Hierna ga ik eerst kort in op het kader voor het evaluatieplan, om daarna het commentaar van Hoeben te bespreken. Vervolgens komen nog afsluitende opmerkingen aan de orde.

Oude Boteringestraat 23, 9712 GC Groningen

Het kader van de evaluatie

Voor een goed begrip van het kader waarbinnen de evaluatie van de Basisvorming zich afspeelt, is een iets ruimere beschrijving nodig dan Hoebe zich gunt. Een aantal korte opmerkingen is in dit verband van belang voor een goed begrip van mijn reactie.

a. De evaluatie van de Basisvorming komt niet geheel uit de lucht vallen, maar ligt rechtstreeks in het verlengde van de destijds voorgenomen evaluaties van het Voortgezet Basis Onderwijs. In feite is de commissie, die daarvoor in 1985 was ingesteld (de CoördinatieCommissie Evaluatieplan Voortgezet Onderwijs, CCE), in de loop van 1987 geruisloos omgevormd tot de verantwoordelijke commissie die de evaluatie van de Basisvorming zou coördineren. Dit houdt in dat sterk werd aangesloten bij reeds opgedane inzichten van de evaluatie van VBaO-experimenten.

b. De kern van het evaluatieplan voor de Basisvorming is het vergelijkend perspectief. Om de effecten van de Basisvorming te kunnen beoordelen is vergelijking nodig met referentiegroepen in het bestaande systeem. Omdat de invoering van de Basisvorming wordt uitgesmeerd over een periode van 10 jaar, moeten bovendien vergelijkingen op verschillende tijdstippen uitmaken of – gaande de invoering – geleidelijk aan eventuele verschillen in de loop van de tijd stabiliseren of juist groter worden. Zodoende kunnen effecten in verschillende stadia van invoering worden beoordeeld. Bovenop dit raster van vergelijkingen kan ook nog de vergelijking binnen de Basisvormingsscholen worden gelegd. In het huidige plan wordt daaraan geen prioriteit gegeven, omdat dergelijke gezichtspunten minder een evaluatief dan een sturend karakter hebben. Het gaat hierbij om de vraag of de Basisvorming zich in de gewenste richting ontwikkelt, zonder vergelijking met een referentiegroep.

c. Het evaluatiekader heeft niet uitsluitend betrekking op de eerste drie leerjaren in het voortgezet onderwijs, hoewel daar de Basisvorming programmatisch wordt ingevoerd. Omdat er grote groepen leerlingen zijn die om een of andere reden vertraagd alle toetsen Basisvorming zullen hebben voltooid, moet een evaluatie zich over langere tijd uitstrekken. Twee centrale vragen moeten daarbij worden opgelost. Allereerst is er de keuze van het criterium, dat een vergelijking tussen Basisvorming en het bestaande systeem moet toestaan. In de evaluatienota wordt beargumenteerd waarom curriculumonafhankelijke toetsen aan het eind van de derde klas de voorkeur verdienen. Ik kom hier verderop nog op terug. In de tweede plaats moeten dergelijke toetsresultaten worden gerelateerd aan de (cognitieve en sociale) kenmerken van de leerlinggroepen op de verschillende scholen (de opmerkingen terzake van Hoebe moeten dus op een vervelend misverstand berusten, juist omdat daar in de evaluatienota nogal wat aandacht aan wordt besteed). De bovenstaande uitgangspunten leiden ertoe dat over individuele loopbanen van leerlingen moet kunnen worden beschikt, waarbij cognitieve begin- en eindmetingen van doorslaggevend belang zijn. Omdat dergelijke gegevens niet regulier worden verzameld, is nieuw onderzoek noodzakelijk.

d. Een belangrijk deel van de evaluatiedoelstelling is aan te geven in hoeverre er verschillen tussen Basisvormingsscholen en die uit het bestaande systeem bestaan en waaraan deze kunnen worden toegeschreven. Dit impliceert dat er naast de leerlinggegevens ook informatie over de schoolkenmerken en -curricula moet worden verzameld. Om een juiste schatting van de verschillende effecten op leerling-, school- en curriculumniveau te verkrijgen, wordt de toepassing van multi-level analyses voorzien; voor een deel wordt de benodigde programmatuur op dit moment in opdracht van de CCE en SVO ontwikkeld.

Het evaluatiekader voor de Basisvorming omvat de bovenstaande elementen. Gedurende de komende tien jaar wordt een stelsel van drie verschillende 'dakpan'-cohorten leerlingen gevolgd in hun schoolloopbaan. Toetsen aan het eind van de derde klas en

loopbaangegevens in het vierde en zesde jaar vormen de belangrijkste effectmaten, gerelateerd aan de instroomkenmerken. Hieraan gekoppeld is de verzameling van school- en curriculumkenmerken, die overigens ook onderwerp van aparte analyse zijn. Hiermee kunnen de gestelde vragen naar verschillen tussen het bestaande en het voorgenomen systeem van Basisvorming worden geëvalueerd, apart voor verschillende schooltypen, cognitieve niveaus en sociale groepen. Zoals Hoeben terecht stelt is hiermee een complex onderzoeksprogramma geschetst.

Het is misschien hier eveneens van belang te vermelden dat in de voorgenomen evaluatie van de Basisvorming er vanuit wordt gegaan dat een zeer groot deel van de dataverzameling zal kunnen worden gecombineerd met een vrijwel gelijktijdige evaluatie van het onderwijsvoorrangsbeleid OVB in het voortgezet onderwijs.

Het commentaar van Hoeben

Tegen de boven geschetste achtergrond kan het commentaar van Hoeben beter worden geplaatst: het heeft betrekking op a. de gegevens op school- en klassenniveau, b. de gekozen effectmaten en c. de organisatorische structuur.

a. De naar voren gebrachte bezwaren lijken me van drieërlei aard te zijn. In de eerste plaats zou te weinig aandacht worden besteed aan een gedetailleerde 'account' van feitelijke gedragingen en tijdsbestedingen van leerkracht en leerling. Hiermee zou alleen oppervlakkige informatie beschikbaar komen, die geen enkel doel zou dienen.

Het probleem is natuurlijk algemener. Uit verricht vooronderzoek door het RION blijkt dat er voor het voortgezet onderwijs betrekkelijk weinig concrete aanknopingspunten voor variabelen zijn, die in het school- en klasgebeuren voldoende discrimineren en een zekere voorspellende waarde hebben.³⁾ Dat maakt het voor 'gewoon' verklarend onderzoek net zo min gemakkelijk als voor evaluatieonderzoek. De vraag is echter of dan maar alles, wat mogelijkerwijs van belang zou kunnen zijn, zou moeten worden meegenomen, ook wanneer daar geen duidelijke causale mechanismen bij gedacht kunnen worden (die empirisch enige referentie hebben). Mijn afweging zou hierbij zijn, dat alleen variabelen worden meegenomen, die enige verklarende bijdrage kunnen leveren. Om een antwoord te kunnen geven op de gestelde evaluatievragen wordt het onderzoek al complex genoeg. Vandaar dat de vraag, waardoor de eventuele verschillen tussen Basisvorming en bestaand systeem zijn te verklaren, wordt opgenomen in de dieptestudies naar schoolprocessen.

Hierop richt zich Hoebens tweede punt van kritiek. Omdat deze dieptestudies zich alleen zouden richten op de Basisvormingsscholen en een beschrijvende doelstelling kennen, zouden deze te weinig effectief zijn. Daar is veel voor te zeggen. In het kader voor de evaluatie wordt dit dan ook gedaan. Op pag 61 wordt een verklarende doelstelling aan de dieptestudies 'schoolprocessen' toegewezen. Bij de planning van deze onderzoeken wordt voorgesteld om vooral naar de uitbijters te kijken, die uit de grootschalige analyse met globale indicatoren naar voren komen (pag 89). Voor de eerste jaren zijn deze inderdaad alleen voorzien bij de beginnende Basisvormingsscholen, maar het zou zeker aan te bevelen zijn om ook contrastgroepen uit het huidige systeem op te nemen. De ervaringen op basis van het eerste cohort zouden vervolgens kunnen worden gebruikt bij de – dan nog te starten – twee volgende cohorten. Deze vraag moet dus wel degelijk serieus worden genomen. Het onderzoek ernaar zou daarom bijvoorbeeld in het overheidsdeel van het SVO-budget zijn onder te brengen.

In de derde plaats meent Hoeben dat een dergelijke evaluatie niet gericht behoort te zijn op het 'aantonen van de sinds jaar en dag bekende effecten van (instroom)kenmerken van leerlingen'. Dat lijkt me maar gedeeltelijk waar te zijn. De doelstelling van de Basisvorming is immers om een deel van dergelijke effecten te verminderen en daarom zijn

enkele evaluatievragen in die richting geformuleerd (pag 33 en volgende). Maar Hoeben heeft gelijk wanneer hij stelt dat bij de evaluatie van onderwijsprogramma's de leerlingkenmerken dienen te worden gecontroleerd. Daarover bestaat geen verschil van mening en het wordt dan ook in verschillende evaluatievragen geformuleerd (o.a. op pag. 34, 38, 42 en 46).

b. Hoeben is ook weinig gelukkig met de gekozen effectmaten, namelijk een viertal cognitieve toetsen aan het eind van de derde klas, alsmede verdere schoolloopbaangegevens in het vierde en het zesde jaar. De cognitieve toetsen voor de vakken Nederlands, Engels, Wiskunde en Biologie worden curriculum-onafhankelijk geconstrueerd door het CITO, zodat een vergelijking tussen bestaand systeem en de Basisvormingscholen mogelijk wordt. Daar hebben ook de centrale evaluatievragen betrekking op. De opmerkingen van Hoeben steken in het hart van elke evaluatie, waar verschillende doelen met elkaar worden vergeleken (in het kaderplan pag. 20-24). Omdat er een criterium wordt gekozen, dat voor twee doelstellingen inhoudsvalide moet zijn, blijft er altijd een zekere discrepantie aanwezig; een fundamenteel dilemma in vergelijkend onderzoek. Nietemin is Hoebens suggestie de moeite waard: er kan wel degelijk worden nagegaan in hoeverre de scores op de 4 toetsen correleren met de scores op de 14 toetsen van de Basisvorming. Deze informatie is zeker van belang.

Over de predictieve validiteit van de vier toetsen zou ik daarentegen uiterst voorzichtig willen zijn. In de meeste gevallen worden de 14 Basisvormingstoetsen eerder afgelegd dan de vier toetsen voor het evaluatieonderzoek, hetgeen een causale uitspraak onmogelijk maakt. Ten aanzien van de alternatieve criteria snelheid en niveau van Basisvorming worden vanzelfsprekend gegevens verzameld, maar deze criteria kunnen wellicht expliciet als doelstelling binnen de Basisvorming worden geformuleerd. Gegeven het gekozen vergelijkende uitgangspunt, zou ik dit echter niet tot de kern van de evaluatie rekenen.

Dit sluit aan bij de bezorgdheid van Hoeben over een 'onaanvaardbare' verenging van de doelen van de Basisvorming tot het hoog scoren op de vier toetsen die gekozen zijn als evaluatiecriterium. Deze opmerking mag op twee punten worden gerelativeerd. In de eerste plaats zijn het niet de scholen, die worden geëvalueerd en het valt dan ook niet goed in te zien waarom die zich zouden laten beïnvloeden om juist op deze vier vakken alle aandacht te richten. Van een 'vertekening' van de evaluatie is in dit opzicht dan ook geen sprake. In de tweede plaats past een dergelijke argument ook op de situatie aan het eind van de lagere school, waar scores op een tweetal vakken worden gehanteerd als voorspeller van het voortgezet onderwijs. Worden daarmee de doeleinden van het Basisonderwijs nadelig beïnvloed?

c. Met betrekking tot de organisatorische structuur van de evaluatie van de Basisvorming is een apart TOR nummer te vullen. Ik volsta met enkele kanttekeningen bij het commentaar van Hoeben. Uiteraard is de evaluatie van zoiets als de invoering van de Basisvorming in het Nederlandse onderwijs een ambitieuze onderneming.

Het betreft organisatorisch complex onderzoek, waarbij vele belanghebbenden – zowel inhoudelijk als materieel – betrokken zijn. De opdrachtgever voor de evaluatie is de minister van onderwijs en wetenschappen. Deze wordt geadviseerd door de genoemde CoördinatieCommissie Evaluatieplan Voortgezet Onderwijs CCE, die weer bestaat uit vertegenwoordigers van CITO, CBS, SVO en onderwijskoepels, verder aangevuld met waarnemers van het ministerie. Het geadviseerde onderzoek wordt door de minister ten uitvoering opgedragen aan SVO; voor zover het toetsconstructie betreft aan het CITO. Het CBS is betrokken in de CCE vanwege haar ervaring met grootschalig cohortonderzoek (het SMVO en het SLVO cohort) en verkrijgt hierdoor een 'trekkingsrecht' op onderzoeksgelden van de CCE. Wanneer ook nog bedacht wordt dat tegelijkertijd wordt samengewerkt met de evaluatie van het onderwijsvoorrrangsbeleid – waarvan de onder-

zoeksgelden via SVO grotendeels naar RION en ITS lopen – dan wordt duidelijk dat het bestuurlijk kader weinig overzichtelijk is. In een dergelijke situatie is het niet ondenkbaar dat er op een bepaald moment een belangenconflict zou kunnen gaan spelen, waarbij de evaluatie op een cruciaal tijdstip wordt geblokkeerd. Vandaar dat er gekozen is voor een modulair systeem, waarbij verschillende uitvoerende onderzoeksinstituten bij de evaluatie worden betrokken en de bevoegdheden duidelijk zijn afgebakend. Zodoende zou de dataverzameling in één hand kunnen worden gehouden. Enkele analysegroepen zouden hierna parallel aan elkaar aan de verschillende evaluatievragen kunnen werken, zodat in een eindfase de gevraagde informatie door de coördinator bij elkaar kan worden gebracht. Aan een dergelijke aanpak zitten natuurlijk risico's, zoals overigens bij elk krap tijdschema.

Terecht wijst Hoebe erop dat er altijd wel wat mis gaat. Dat zal ongetwijfeld ook hier het geval zijn, maar deze constatering is minder belangrijk dan de mogelijke oplossing. Waar het bij een dergelijk grootschalige onderneming om gaat, is dat er voldoende flexibiliteit is. Het punt is nu dat deze op persoonlijk niveau vaak wel aanwezig is, maar op institutioneel niveau meestal niet. Personen kunnen effectiever op hun ambities en verantwoordelijkheden worden aangesproken; bij organisaties is dat veel onduidelijker, doordat er gemakkelijk competentiekwesties kunnen gaan spelen. Hoewel de realiseerbaarheid van de evaluatie door de modularisering naar mijn mening is toegenomen, blijft het antwoord op de vraag naar wie er – gegeven de complexe structuur – uiteindelijk verantwoordelijk is voor de afstemming, voorslagnog onbevredigend.

Afsluiting

De evaluatie van de basisvorming is een zeer complex geheel, zoals ook de invoering van de basisvorming zelf. Hoebe heeft er terecht op gewezen dat enkele vraagstellingen nog kunnen worden aangescherpt: gegeven het tijdspad moet er zeker ruimte zijn voor een dergelijke detaillering van de evaluatievraagstellingen. Hoewel er kennelijk over de hoofdlijnen geen verschil van mening is, zijn met name de organisatorische voorwaarden een punt van zorg. Toch moet er in dit opzicht vooral van buitenlandse ervaringen worden geprofiteerd. Ook zonder al te grote illusies te koesteren, verwijs ik naar wat Coleman reeds in de jaren 1965/66 (en later in 1981 opnieuw) realiseerde. Zijn onderzoek was effectief opgezet, alle fasen waren nauwkeurig gepland, van te voren werd proefgedraaid op gesimuleerde data en de analysemethoden waren van te voren gekozen, waardoor in een korte tijd een afgewogen rapport kon verschijnen. Dat moet in Nederland toch ook mogelijk zijn?

¹⁾ Van 1-9-86 tot 1-10-88 coördinator voor de evaluatie van de Basisvorming.

²⁾ De evaluatienota is inmiddels afzonderlijk door SVO gepubliceerd en aldaar te verkrijgen.

³⁾ Zie hiervoor R.J. Bosker & E. Lugthart (1987), *Schoolkenmerken en designs: eerste verkenningen*. Groningen, RION; alsmede R.J. Bosker, E. Lugthart & P. Roeders (1988), *Schoolkenmerken en designs: verdere algemene verkenningen*, Groningen, RION.

Boekbesprekingen

M. P. C. van der Werf (1988). *Het schoolwerkplan in het basisonderwijs: ontwikkeling, implementatie en opbrengst*.

(Proefschrift R.U. Groningen) Amsterdam/Lisse: Swets & Zeitlinger. f 35,-.

'De opkomst en neergang van het verschijnsel schoolwerkplan' zou een journalistieke ondertitel kunnen zijn voor het proefschrift waarop Greetje van der Werf onlangs promoveerde.

Haar onderzoek levert een empirische bijdrage aan de discussie over schoolwerkplanontwikkeling die al zo'n 15 jaar speelt in het Nederlandse onderwijs. Eerdere rapportages over het onder leiding van Van der Werf uitgevoerde ESP-onderzoek ('Evaluatie van schoolwerkplanontwikkelingsprojecten in het basisonderwijs'; 1980-1983) zijn overigens al duidelijk van invloed geweest op het beleid inzake schoolwerkplanontwikkeling.

In het proefschrift stelt de auteur zich ten doel 'om de op verschillende momenten heersende theoretische en de daaruit voortvloeiende prescriptieve opvattingen over het nut van schoolwerkplannen voor onderwijsvernieuwingen te toetsen aan onderzoeksresultaten'.

Het proefschrift kent de volgende opbouw.

In het eerste deel wordt beschreven hoe de opvattingen over schoolwerkplanontwikkeling zich in de jaren zeventig en tachtig hebben ontwikkeld. Van der Werf signaleert daarbij een aandachtsverschuiving van een procesgerichte naar een produktgerichte benadering. In de eerste aanpak (zoals onder meer gepropageerd door de ICB in 1977) werd het ontwikkelproces van een schoolwerkplan nadrukkelijk beschouwd als een middel tot inhoudelijke vernieuwingen op schoolniveau. In de tweede benadering (zoals onder meer verwoord door de ARBO in 1983) is die koppeling minder direct: er wordt onderscheid gemaakt tussen enerzijds vernieuwingsprocessen en anderzijds het vervaardigen en benutten van een schoolwerkplan-document. Beide benaderingen geven de auteur aanleiding tot het formuleren van onderzoeksvragen.

In het tweede deel van het proefschrift staat centraal de vraag naar de relatie tussen de *ontwikkeling* van schoolwerkplannen en het realiseren van onderwijsvernieuwingen. Ter beantwoording van die vraag worden de resultaten van het ESP-onderzoek benut. Daaruit blijkt dat de verwachtingen in het algemeen niet zijn uitgekomen. Enkele meer specifieke conclusies: (a) er blijkt geen sprake van een planmatig ontwikkelproces; (b) de plannen blijken geen documenten op schoolniveau, maar veeleer te typeren als een bundeling van samenvattingen van de op school gebruikte methodes; (c) de relatie tussen het schoolwerkplan op papier en de klaspraktijk is niet erg hecht.

Centraal in het derde deel staat de relatie tussen het *gebruik* van schoolwerkplannen en onderwijsvernieuwingen. Op basis van de resultaten van secundaire analyses van gegevens uit een vervolgonderzoek wordt nagegaan in hoeverre de verwachting is uitgekomen dat scholen onderwijsvernieuwingen realiseren door van hun schoolwerkplan daadwerkelijk gebruik te maken. Ook deze verwachting wordt in belangrijke mate ontzenuwd. Om te beginnen blijkt dat de organisatorische en onderwijskundige afspraken op schoolniveau doorgaans slechts matig zijn uitgewerkt in de schoolwerkplannen (een bevestiging van de ESP-conclusies). Vervolgens blijkt dat de schoolwerkplannen slechts matig worden gebruikt: vooral als voorlichtingsmiddel aan derden en weinig als hulpmiddel bij het didactisch handelen van leerkrachten. Op grond daarvan ligt het voor de hand dat over de invloed van (het gebruik van) schoolwerkplannen op leerlingprestaties nog nauwelijks duidelijke uitspraken te doen zijn.

Tot slot van het proefschrift volgt een nabeschuiving over de consequenties van de beide onderzoeken en worden aanbevelingen voor vervolgonderzoek geformuleerd.

Het proefschrift getuigt van een nauwgezette onderzoeksaanpak en is beknopt en over het algemeen helder geschreven.

Voor de discussie over schoolwerkplanontwikkeling, waarover zoveel normatieve beweringen en informele waarde-oordelen te beluisteren vallen, is het nuttig dat meer systematisch geordende en empirische gegevens gepresenteerd worden. Dat die gegevens vooral een ontnuchterende werking hebben op de soms fors opgeblazen verwachtingen, zullen weinigen betreuren.

Enigszins teleurstellend acht ik de in de nabeschuiving geschetste theoretische en praktische opbrengst van het onderzoek. Wellicht is dat in de hand gewerkt doordat Van der Werf zelf de verwachtingen over de ontwikkeling, implementatie en opbrengst van schoolwerkplannen wel erg sterk heeft aangezet. Door allerlei afzonderlijke wensen, verlangens en criteria uit zowel de onderwijspraktijk, het onderwijsbeleid als de onderwijswetenschappen bijeen te zetten en te presenteren als één samenhangend en afgerond geheel, is de werkelijkheid enigszins geweld aangedaan. Als samenvatting van de vele verwachtingen is de uiteenzetting in hoofdstuk 3 en 4 wel geschikt. Het is echter wat geforceerd om te suggereren dat 'men' (wie eigenlijk?) destijds al deze verwachtingen tegelijkertijd koesterde. Met name de op zich interessante verbinding met de recente denkbeelden over effectieve scholen en effectief onderwijzen is riskant. De auteur beoordeelt nu, achteraf, de toenmalige schoolwerkplan-activiteiten aan de hand van criteria, die nog niet goed uitgekristalliseerd zijn en die destijds zeker niet gehanteerd zijn door allerlei betrokkenen in de schoolpraktijk, het onderwijsbeleid en de verzorgingsstructuur. In die zin moesten de resultaten haast wel tegenvallen. Het in de loop van het proefschrift steeds sterker wordende accent op micro-effecten (didactisch handelen en in het verlengde daarvan ook leerlingprestaties) doet weinig recht aan de functies die velen oorspronkelijk toekenden aan schoolwerkplanontwikkeling. De beperkte wijze waarop de gebruiksgegevens over dat microniveau verzameld zijn, namelijk louter via schriftelijke vragenlijsten, maakt het bovendien moeilijk om tot valide en betrouwbare uitspraken te komen.

Ook praktisch en beleidsmatig zijn de overwegingen en aanbevelingen weinig concreet. Na de doorgaans negatieve onderzoeksuitkomsten komt de auteur, m.i. enigszins verrassend, tot de slotsom dat het schoolwerkplan een uitstekend middel is voor de implementatie van nieuwe curricula en voor de concretisering van de kenmerken van effectieve scholen. Over de voorwaarden waaronder deze functies van het schoolwerkplan tot hun recht zouden kunnen komen, is Van der Werf echter rijkelijk vaag. Ik heb de indruk dat een concretisering van dergelijke denkbeelden toch gebaat zou zijn bij een meer nadrukkelijk onderscheid tussen enerzijds schoolwerkplanontwikkeling als een proces van organisatie-ontwikkeling, waarin op teamniveau bepaalde, noodzakelijk geachte afspraken gemaakt worden (over basisvisie, taakverdeling e.d.), en anderzijds het vervaardigen en benutten van een schoolwerkplandocument waarin op microniveau voor allerlei vakken afspraken uitgewerkt zijn. Meer dan voorheen zouden kosten-baten afwegingen (waarin vooral tijd een belangrijke variabele dient te zijn) daarbij een rol moeten spelen.

Hopelijk zal het proefschrift van Van der Werf leiden tot meer realistische en precieze discussies over schoolwerkplanontwikkeling. Zeker is in ieder geval dat vervolgonderzoek nuttig kan zijn. Wellicht kan dat ertoe bijdragen dat de neergang van het fenomeen schoolwerkplan omgebogen wordt in een meer perspectiefrijke richting.

Jan J. H. van den Akker
Faculteit Toegepaste Onderwijskunde
Universiteit Twente,
Postbus 217, 7500 AE Enschede

ERRATUM

Tijdschrift voor onderwijsresearch, Vol. 13, nr. 6, p. 307

Tabel 3 dient als volgt te worden gelezen:

Tabel 3: Resultaten van de LLTM-toetsing voor Strategie-groep 2.

Item type	Itemnr	q_{i1}	q_{i1}	$\hat{b}_i$	$\hat{b}_i^+$	$\hat{b}_i - \hat{b}_i^+$
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
" "	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			

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. 13, nr. 1

Februari 1989

Inhoudsopgave

De meting van wetenschappelijke productiviteit <i>W.J. van der Linden</i>	56
Congressen en symposia	60
Ontvangen publicaties	63

De meting van wetenschappelijke productiviteit

W.J. van der Linden
Universiteit Twente

In een recente aflevering van het *VOR Bulletin* is door Ax en Creemers (1988) een poging ondernomen om voor de onderwijskunde normen voor de meting van de wetenschappelijke productiviteit op te stellen. Toen ik kennis nam van hun voorstel voor puntentoekening, brandde de pen al in mijn hand. Nu Van der Sijde (1988) inmiddels aan de hand van een beschrijving van een tweetal (gefingeerde) onderzoekers heeft laten zien dat het voorstel zo lek is als een mandje, kan ik helemaal niet nalaten te reageren.

Het voorstel van Ax en Creemers

Aan het voorstel ligt een eenvoudig recept ten grondslag: stel een lijst op van *alle* denkbare producten die een onderwijskundige ooit zou kunnen produceren. Aan ieder product geven men een vast aantal punten. Moet er iemand beoordeeld worden, tel dan alle behaalde punten op. En klaar is Kees.

Punten van kritiek

Op het voorstel is nogal wat kritiek mogelijk. Ik geef deze puntsgewijs:

1. In de eerste plaats vindt er volstrekt geen beoordeling van de kwaliteit van de producten plaats. Ieder rapport, lezing, boek, symposium, enz., krijgt een vast aantal punten. Of het goed was of slecht, door vakgenoten gekraakt werd of enthousiast ontvangen, diepgaand of oppervlakkig, gebaseerd is op onderzoek of slechts met een gedachte speelde, het doet er allemaal niets toe. Een onderzoeker hoeft zijn product slechts als zodanig aan te merken, er een kftje om te doen met een jaartal erop, en daar rinkelt de kassa al. Ax en Creemers voelen dit ook wel aan. Daarom pleiten ze in een slotalinea voor een aanvullende kwalitatieve beoordeling door een panel van gerenommeerde vakgenoten. Alsof dit evenwel uitvoerbaar is! De lezer stelle zich eens een panel voor dat een vergelijking tussen de vier SVO-instituten in Nederland moet maken en hiertoe een oordeel uit moet spreken over de kwaliteit van het wetenschappelijke werk achter ieder van de honderden, zo niet duizenden producten van deze instituten. Hoeveel zullen ze daadwerkelijk kunnen lezen? Hoe groot moet het panel zijn om iedere publicatie werkelijk te kunnen begrijpen en op waarde in te schatten? Wat moet het panel met al die lezingen, voordrachten en symposia waar het zelf niet bij is geweest?
2. De lijst met mogelijk producten die Ax en Creemers presenteren, kent geen grenzen. Er mocht eens iemand uit de boot vallen die vijf minuten met zijn vak bezig is geweest en dat niet kan verantwoorden op een manier die in de meer gevestigde wetenschappen bruikbaar is. Op deze wijze treffen we in de lijst punten aan voor: "begeleiding van promovendi" (Je zal maar hoogleraar zijn; al die promovendi die mooi onderzoek zitten te doen, terwijl je er zelf geen tijd voor hebt); "deelname aan de begeleidingscommissies etc" (En dan nog al die vergaderingen); "ZWO-rapporten" (Deze organisatie heet tegenwoordig NWO en vraagt geen onderzoeksverslag in rapportvorm; dat doet die andere); en "boek + b.v. opleiding/onderwijs" (Wat dat ook moge zijn???)
3. Tenslotte is er die moeilijke vraag van het aantal punten. Vertel mij uw puntensysteem en ik zal zeggen wat uw vooroordelen zijn. Wat dit betreft is het voorstel van Ax en Creemers een open boek: een rapport voor een opdrachtgever krijgt 6 punten, tegen een dissertatie 8 (Dat de eerste soms gebaseerd is op eenvoudig survey-onderzoek van 12

maanden en de laatste altijd op dissertatie-onderzoek van minimaal 4 jaar, mag geen rol spelen); een congresbijdrage in een bundel krijgt evenveel punten als een tijdschriftartikel (Hoeveel bijdragen aan ORD-bundels zouden de beoordelingsprocedure van een tijdschrift overleven?); deelname aan een begeleidingscommissie e.d. krijgt evenveel punten als een praktisch tijdschriftartikel (Goed dat er een vergadercultuur is); aan een artikel in een Nederlands tijdschrift wordt in principe hetzelfde aantal punten toegekend als in een internationaal tijdschrift (Hoeveel maal groter is het aanbod waar het *American Journal of Educational Research* zijn artikelen uit selecteert dan dat van dit tijdschrift?). Een verwant probleem is dat ieder product in dezelfde categorie evenveel punten krijgt. Van sommige dissertaties hoor je echter nooit meer iets, terwijl andere het in de vorm van artikelen tot acceptatie in de literatuur brengen. Rapporten voor opdrachtgevers kunnen gebaseerd zijn op innovatief onderzoek of op een routine-survey, 3 maanden of 3 jaar werk omvatten, als eindrapport of als een reeks rapporten over deelonderwerpen uitgebracht worden; het maakt allemaal niets uit, ieder rapport krijgt 6 punten.

De gevolgen

Van der Sijde heeft in een reactie laten zien wat de gevolgen zijn als we alles van punten voorzien en de onderzoeker vrijlaten zijn producten zelf als zodanig aan te merken. In één van de door hem omschreven gevallen publiceert een onderzoeker hetzelfde stand-van-zaken paper als congrespresentatie, als artikel in door hemzelf geredigeerde proceedings, als hoofdstuk in een eigen boek en vertaalde hij het ook nog eens voor opname in een buitenlands boek. De andere onderzoeker gebruikte hetzelfde empirische onderzoek voor een ORD-presentatie, voor publicatie in een ORD-bundel, als artikel in *Pedagogische Studiën*, als vertaling in een Engelstalig tijdschrift, als voorlichtend artikel in een vaktijdschrift en als SVO-rapport. Beide onderzoekers kunnen bij Ax en Creemers hun inspanning voor een vorstelrijk aantal punten verzilveren en zijn veel beter af dan hun ingetogen collega die hetzelfde onderzoek in slechts 1 tijdschriftartikel zou publiceren.

Wat is er nu precies zo erg aan deze voorbeelden? Dat is het feit dat dezelfde onderzoeksopbrengst veelvuldig met punten beloond kan worden, terwijl – afhankelijk van het medium – de kwaliteit van het onderzoek ook nog eens nauwelijks beoordeeld hoeft te zijn. Let wel, het gaat hier dus om het veelvuldig *belonen* en niet om het veelvuldig *publiceren* van hetzelfde onderzoek. Voor het laatste kunnen geldige redenen bestaan; ik kom daar aanstonds nog op terug. Wat uitgesloten moet worden, is dat onderzoekers en instellingen langs deze weg, zonder enige kwaliteitsbeoordeling, bijna onbegrensd hun eigen beloning kunnen opvoeren.

Een beter voorstel

Een voorstel voor een puntensysteem dient gebaseerd te zijn op de volgende uitgangspunten:

1. Alleen publicaties die vooraf door vakgenoten op kwaliteit beoordeeld zijn, komen voor punten in aanmerking. Beoordeling achteraf van een bulk publicaties, rijp en groen, door een panel van vakgenoten heeft geen zin.
2. De beoordeling moet volgens de regelen der kunst geschieden, d.w.z. (a) de beoordeelaars dienen experts te zijn op het gebied van de onderzoeksvraagstelling, (b) de beoordelingen moeten anoniem zijn, (c) er dienen meerdere onafhankelijke beoordeelaars te zijn, en (d) er moet een beroepsprocedure mogelijk zijn.
3. Dezelfde publicatie mag niet meerdere malen met punten beloond worden.

Dames en Heren, er bestaat een beoordelingssysteem dat op deze wijze werkt. Het is reeds onder ons en verricht dagelijks zijn geruisloze arbeid. Telkens als een onderzoeker een publikatie af heeft, is het bereid deze zonder enige kosten te beoordelen. Ik doel

natuurlijk op: HET WETENSCHAPPELIJKE TIJDSCHRIFT. Alle andere media voldoen niet aan belangrijke voorwaarden die hierboven genoemd zijn. Boeken worden door uitgevers voornamelijk beoordeeld op de vraag of er een groot lezerspubliek voor is. De beoordeling voor opname in proceedings (vergelijk de ORD-bundels) gebeurt door een redacteur en voldoet niet aan het tweede uitgangspunt. Rapporten voor opdrachtgevers worden nooit beoordeeld; de enige vorm van beoordeling is die van een offerte vooraf. Voor alle andere publicatiemediën geldt a fortiori hetzelfde. Het wanhopigst zijn in dit opzicht de interne rapporten, waarover niet zelden de auteur zelf beslist of ze verschijnen of niet.

Mijn voorstel is gebaseerd op het volgende eenvoudige systeem:

1. Ken alleen punten toe aan artikelen in tijdschriften die volgens de bovengenoemde criteria werken.
2. Baseer het aantal punten op het niveau van het tijdschrift, het aantal pagina's en eventuele co-auteurschappen.

Spelregels voor de bepaling van het aantal punten kunnen ontleend worden aan de ervaring in andere wetenschapsgebieden, waar al langer met een dergelijke opzet gewerkt wordt. Bijvoorbeeld het niveau van een tijdschrift kan bepaald worden door zaken te wegen als de gerenommeerdheid van de redactieleden, het aantal beoordelaars per manuscript, de acceptatieratio, het wervingsgebied van auteurs, terwijl het aantal pagina's afhangt van de gemiddelde hoeveelheid tekst per pagina, de gemiddelde lengte van een artikel in het tijdschrift, e.d.

Omdat redacties erop toezien dat geaccepteerde artikelen niet eerder in andere tijdschriften gepubliceerd werden en de auteursrechten altijd via een contract aan de uitgever overgedragen moesten worden, is publicatie van hetzelfde artikel in meerdere tijdschriften alleen mogelijk in de vorm van bewuste fraude. Hiermee is tevens aan het derde uitgangspunt voldaan.

Geen ongewenste effecten?

Zijn er geen nadelen verbonden aan het accent dat in het bovenstaande op het artikel als eenheid van publicatie wordt gelegd? Er valt immers nogal eens wat verzet tegen "dat gepubliceer in tijdschriften" te beluisteren. We laten de gebruikelijke argumenten de revue passeren:

1. Publiceren in andere media heeft toch ook zin? Zeker wel, de ideale onderzoeker publiceert ook in andere vorm, maar voor een ander doel. Congrespapers dienen voor discussie met vakgenoten. Als het onderzoek nog niet af is, zijn het nog geen zelfstandige producten van wetenschappelijke arbeid; is het wel af, dan dient het manuscript van een (ingezonden of nog in te zenden) artikel als paper. Interne rapporten hebben dezelfde functie; het zijn prepublicaties van artikelen die onderzoekers elkaar toesturen om de *publication lag* te reduceren. Maar hoe zit het dan met rapportage van onderzoek naar opdrachtgevers als SVO? Ook hier kiest de efficiënte onderzoeker voor het artikel als de eenheid van publicatie. Opdrachtonderzoek kan voortreffelijk gerapporteerd worden in de vorm van een of meer samenhangende artikelen (waarvan sommige reeds verschenen zijn en andere gelijktijdig naar tijdschriften ingestuurd worden) voorzien van een inleiding met conclusies en aanbevelingen. Voor SVO zitten hier alleen maar voordelen aan: de cultuur van dikke, ongelezen rapporten verdwijnt, terwijl tijdschriften kosteloos de service van *post hoc* beoordeling verlenen, terwijl thans iedere vorm van beoordeling ontbreekt. Voor dissertaties geldt eigenlijk hetzelfde. Een proefschrift waarvan de inhoud niet vroeg of laat in de tijdschriften terecht komt, onthoudt de wetenschappelijke wereld belangrijke informatie of is verdacht. Gelukkig verandert ook hier de cultuur, zij het in de onderwijskunde langzamer dan in de meer gevestigde gebieden. Het promoveren op een

reeks gebundelde artikelen is een efficiënte vorm en zal steeds populairder worden nu promoties overwegend via een aanstelling als aio of oio plaats gaan vinden. Nogmaals, de ideale onderzoeker kiest van meet af aan voor de omvang en het niveau van het tijdschriftartikel als eenheid van publicatie, brengt deze uit als congrespaper of interne publicatie en bundelt ze met een aparte inleiding tot rapport aan een opdrachtgever of dissertatie. Het gedeelte hiervan dat waarlijk de tijdschriften haalt, heeft een kwaliteitsbeoordeling door vakgenoot-deskundigen doorstaan en kan als betrouwbare meting van de wetenschappelijke productiviteit van de onderzoeker gebruikt worden.

2. Tijdschriften zijn toch altijd vooringenomen tegen bepaalde soorten onderzoek; zo geven ze toch altijd de voorkeur aan experimenteel boven toegepast onderzoek? Dit is onjuist, voor alle soorten onderzoek zijn er tijdschriften, of dit nu toegepast, experimenteel, theoretisch, beleidsondersteunend of technologisch van aard is. En als er voor sommige soorten onderzoek geen tijdschriften zijn, of als er te weinig zijn waardoor kwalitatief goed onderzoek niet of slechts met grote vertraging gepubliceerd kan worden, dan beginnen we er toch gewoon een? Naar de mate waarin de redactie gerenommeerd is, de beoordelingen onafhankelijk en anoniem zijn, meerdere beoordelaars ingeschakeld worden en er auteurs met een goede staat van dienst op af komen, wordt het vanzelf een tijdschrift met een hoge puntenwaardering.

3. Sommige soorten onderzoek zijn te zeer gebonden aan de Nederlandse situatie; is het niet onrechtvaardig dat ze daardoor nooit de hogere puntenwaardering van internationale tijdschriften halen? Alhoewel er uitzonderingen zijn, is de vooronderstelling dat internationale tijdschriften een hogere puntenwaardering verdienen in zijn algemeenheid terecht. Dit is niet zo omdat ze Engelstalig zijn, maar omdat ze zowel hun redactieleden en beoordelaars als hun aanbod aan artikelen uit een veel groter wervingsgebied putten en dus een gunstigere selectieratio hebben dan lokale tijdschriften. De veronderstelling dat ze geen onderzoek accepteren dat aan de Nederlandse situatie gebonden is, is onterecht. Er zit ook altijd iets paradoxaals aan het gebruik van dit argument: degenen die het hanteren, grijpen zelf voor hun onderzoek maar wat graag naar literatuur met resultaten uit buitenlandse projecten. De lezer van het *Tijdschrift voor Onderwijsresearch* moet de laatste paar afleveringen maar eens doornemen en tellen naar hoeveel buitenlandse publikaties verwezen wordt. Wat wij in het Nederlandse onderwijs onderzoeken, is in principe net zo interessant voor een Amerikaan, Duitser of Engelsman als omgekeerd. Als het om puur beschrijvend onderzoek gaat, zijn er zelfs gespecialiseerde internationale tijdschriften op het gebied van de vergelijkende onderwijskunde beschikbaar.

Conclusie

Met puntenstelsels zoals voorgesteld door Ax en Creemers, die alles maar honoreren ongeacht de kwaliteit en hoe vaak het ook gepubliceerd wordt, zijn we er niet. In ons wetenschapsgebied moeten we de beoordeling van onze publicaties ter hand laten nemen door deskundigen op het gebied van de onderzoeksvraagstelling. Tijdschriften zijn hiervoor bij uitstek geschikte kanalen. De meer gevestigde wetenschapsgebieden zijn ons in dit opzicht reeds voorgegaan. Het wordt tijd dat de onderwijskunde het wiel ook eens uitvindt.

REFERENTIES

- Ax, J. & Creemers, B. (1988). Kwaliteitsbepaling wetenschappelijke productie in de onderwijskunde. *VOR Bulletin*, 12, 118-121.
 Van der Sijde, P. (1988). Output: kwaliteit of kwantiteit. *VOR Bulletin*, 12 294-295.

Congressen en symposia

**Third European conference for research on learning and instruction.
Madrid, september 4-7, 1989**

EARLI was founded to promote research on learning and instruction in Europe. It supports the systematic exchange and discussion of ideas and results among European researchers. It also encourages the exchange with non-European researchers. The main interests of EARLI lie on processes of learning, development, and instruction with focus on a microanalytical approach emphasizing the psychological and educational nature of these processes. Research on learning and instruction should describe and explain the processes and provide possibilities to improve practice in this domain.

Main objectives of the conference

The conference will provide opportunities to:

- exchange and discuss results of ongoing research on learning and instruction, especially in Europe;
- review the state-of-the-art with regard with central themes in learning and instruction in Europe;
- discuss new trends, approaches, methodologies, concepts, etc. in this domain;
- promote co-operation between researchers.

Main themes

- LEARNING AND INSTRUCTION IN SPECIFIC KNOWLEDGE DOMAINS
- MOTIVATIONAL AND EMOTIONAL FACTORS IN LEARNING
- DEVELOPMENTAL PROCESSES IN LEARNING
- LEARNING IN NATURAL SETTINGS
- CLASSROOM LEARNING PROCESSES
- KNOWLEDGE ACQUISITION, METACOGNITION AND PROBLEM SOLVING
- PARENTAL INVOLVEMENT IN LEARNING
- ADULT LEARNING AND CONTINUING EDUCATION
- MULTICULTURAL LEARNING
- LEARNING AND INSTRUCTION WITH NEW MEDIA AND INFORMATION TECHNOLOGIES
- LEARNING TO LEARN
- LEARNING ENVIRONMENTS FOR SPECIAL EDUCATIONAL NEEDS
- INSTRUCTIONAL AND SOCIAL INTERACTIONS IN THE CLASSROOM
- SECOND LANGUAGE LEARNING

MAILING ADDRESS

Program Committe

Mario Carretero
Facultad de Psicología
UNIVERSIDAD AUTONOMA
DE MADRID
Cantoblanco
28094 MADRID - SPAIN
Teléf. 397 50 46 and 397 46 40
Telex: 27810

Accomodation

Viajes Ecuador
(Miss Cristina)
Calle San Bernardo, 20
28015 MADRID - SPAIN
Teléf. 522 07 95
Telex: 27476

3RD JOINT MEETING

ONDERZOEKSTHEMAGROEP ONDERWIJSLEERPROCESSEN (OTG)
ARBEITSGRUPPE FÜR EMPIRISCHE PÄDAGOGISCHE FORSCHUNG (AEPF)
FACHGRUPPE FÜR PÄDAGOGISCHE PSYCHOLOGIE (FPP)

APRIL 23-25, 1989 UNIVERSITY OF TWENTE, THE NETHERLANDS

The Joint meeting of OTG, AEPF, and FPP provides an opportunity for Dutch, German, Austrian, Swiss, and Belgian educational researchers to share ideas, experiences, research findings, courseware and other educational software. The Joint Meeting will be set up in such a way to assure maximal information exchange between researchers. Space and time in the program will be available for informal discussions.

The main themes for the Twente Meeting are:

- LEARNING ENVIRONMENTS
- COGNITION, EMOTION, AND MOTIVATION
- ADULT LEARNING
- INSTRUCTIONAL MODELS FOR INTELLIGENT TUTORING
- SCHOOLEFFECTIVENESS
- LEARNING TO LEARN

DEADLINES OF TWENTE MEETING:

- Participation Reply for all types of participation February 1, 1989, mailed to:
OTG Onderwijsleerprocessen
Department of Education
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
- For those who only want to attend the Twente Meeting reply is allowed till March 1, 1989.
- Notice of acceptance to authors march 15, 1989.
- Program with summaries of presentations to participants April 8, 1989.

For more information:
OTG Onderwijsleerprocessen
Dr. J. M. Pieters
University of Twente
Department of Education
P.O. Box 217
7500 AE Enschede
The Netherlands
(X-31)-53-893594

Ontvangen publicaties

- De Vries, B. (1988). *Het leven en de leer. Een studie naar de verbinding van leren en werken in de stage*. Proefschrift. Nijmegen: ITS.
- Ven, J. van de, Vrieze, G. & Claessen, J. (1988). *Grensverkenners in de basisschool. Een onderzoek naar volledige inzetbaarheid van voormalige LO- en KO-leerkrachten in het basisonderwijs*. Nijmegen: ITS.
- Vries, B. de (1988). *Het regionale voorlichtingsnetwerk. De samenwerking tussen regionale organen voor het leerlingwezen en bureau's voor studie- en beroepskeuzevoorlichting*. Nijmegen: ITS.
- Kuijk, J. van, Vierke, H. & Claessens, J. (1988). *Deeltijdarbeid: een verdeeld genoegen? Een onderzoek naar aard, omvang en gevolgen van deeltijdarbeid, duobanen en ADV in het basisonderwijs*. Nijmegen: ITS.
- Esch, W. van (1988). *Schoolbesturen en de vrijheid van onderwijs. Bestuurlijke en organisatorische factoren van invloed op de benutting van beleidsruimte door bijzondere schoolbesturen*. Nijmegen: ITS.
- Kerkhoff, A. (1988). *Taalvaardigheid en schoolsucces*. Proefschrift. Lisse: Swets & Zeitlinger.

S&Z

**Taalvaardigheid
en schoolsucces**

Anne Kerkhoff

**Onderwijs-
kansen
van
allochtonen**

252 blz. f 43,50
isbn 90 265 0959 6

Dit boek beschrijft een onderzoek naar de relatie tussen taalvaardigheid in het Nederlands van allochtone en autochtone basisschoolleerlingen en hun schoolsucces. Onderwijsproblemen van allochtonen worden niet gezien als problemen die voor hen uniek zijn. Gebleken is dat allochtone leerlingen wel lagere schoolresultaten behalen, maar dat dit niet direct aan de taalachtergrond ligt. Het sociaal-economisch milieu speelt een grote rol.

ook verkrijgbaar in de boekhandel.

SWETS & ZEITLINGER B.V.

Heereweg 347, 2161 CA Lisse, tel. 02521-35111

Psychologisch Onderzoek van het Aanvankelijk Rekenen

Jacobijn Sandberg, Christiaan Hamaker
Universiteit van Amsterdam

Jos Beishuizen
Vrije Universiteit Amsterdam

SAMENVATTING

Dit artikel schetst een beeld van binnen- en buitenlands onderzoek op het gebied van het aanvankelijk rekenen en is bedoeld als een inleiding op het onderzoek dat in dit themanummer besproken wordt. De nadruk ligt op psychologisch onderzoek, waarbij zowel aandacht wordt besteed aan traditioneel, experimenteel onderzoek als aan onderzoek binnen Cognitive Science waarbij de aandacht sterk gericht is op het verloop van cognitieve processen.

ABSTRACT

This article sketches an outline of contemporary research in the field of early mathematics, and is meant as an introduction to the research that is presented in this issue. Emphasis is placed on psychological research, both within the experimental tradition and within the field of Cognitive Science. The main focus is on the course and development of cognitive processes.

Het Nederlandse onderzoek op het gebied van het aanvankelijk rekenen heeft de laatste jaren een grote vlucht genomen. Op verschillende plaatsen in Nederland houden onderzoekers zich bezig met de aard en ontwikkeling van rekenvaardigheid. De aspecten waar de verschillende onderzoekers zich op richten zijn zeer divers van aard: getalbegrip, metacognitie, technisch rekenen, probleem oplossen, etc. Het doel van onderzoek verschilt ook nogal. Sommige onderzoekers richten zich primair op het verwerven van inzicht in de aard en ontwikkeling van rekenvaardigheid, andere onderzoekers richten zich voornamelijk op het verbeteren van het rekenonderwijs. Nadruk op het verbeteren van het onderwijs legt ook het onderzoek dat uitgaat van de handelingspsychologische stroming die voornamelijk in de Sovjet-Unie ontwikkeld is. Het didactisch gerichte onderzoek blijft in dit themanummer buiten beschouwing. De aandacht valt op onderzoek van leer- en ontwikkelingsprocessen bij het aanvankelijk rekenen. Dit heeft dikwijls wel didactische consequenties, maar die staan in het onderzoek niet voorop.

In dit onderzoeksveld zijn twee stromingen te onderscheiden: het experimentele onderzoek waarin vooral met kwantitatieve gegevens wordt gewerkt, en het procesgerichte onderzoek waarin veel gebruik gemaakt wordt van kwalitatieve gegevens. Vaak vullen beide benaderingen elkaar aan en in veel onderzoek worden beide benaderingen gecombineerd. Hoewel beide benaderingen elkaar niet uitsluiten en de scheidslijn niet eenduidig is, biedt de tweedeling experimenteel versus procesgericht voldoende houvast om de lezer te introduceren in het onderzoeksgebied. Dit inleidende artikel neemt de stand van zaken op binnen deze twee stromingen in het onderzoek van aanvankelijk rekenen. Het wordt besloten met een discussie waarin een visie op toekomstig onderzoek wordt geschetst.

adres: J.A.C. Sandberg, Universiteit van Amsterdam, Vakgroep Sociaal Wetenschappelijke Informatica, Herengracht 196, 1016 BS Amsterdam

DE EXPERIMENTELE RICHTING

De experimentele benadering van het leren rekenen wordt gekenmerkt door het gebruik van kwantitatieve methoden voor het meten van prestaties.

Veel experimenteel onderzoek heeft zich in de laatste jaren gericht op de vraag hoe kinderen de 'rekenfeiten' leren. De rekenfeiten zijn in de eerste plaats de elementaire optellingen met een som kleiner dan 10. Bij het onderzoek naar het leren van de rekenfeiten is veelvuldig gebruik gemaakt van reactie- of oplostijden.

Dit onderzoek heeft aan het licht gebracht dat kinderen over het algemeen de volgende stadia doorlopen bij het leren van rekenfeiten. In het eerste stadium worden de op te tellen getallen gepresenteerd, bijvoorbeeld door het opsteken van vingers, en vervolgens wordt de som berekend door de gecombineerde representatie te tellen. In een volgend stadium wordt een snellere procedure gevolgd. In navolging van de ontdekkers, Groen en Parkman (1972), wordt deze procedure aangeduid als 'minimum addend'. De procedure kan als volgt worden voorgesteld: bij presentatie van de opgave wordt eerst vastgesteld welk van de twee getallen het grootst is, dan wordt een teller op die waarde gezet, en ten slotte wordt de som berekend door vanaf die waarde door te tellen met een aantal stapjes gelijk aan het andere, kleinste getal. De tijd benodigd voor deze procedure is lineair afhankelijk van het aantal door te tellen stapjes en dat is gelijk aan het kleinste van de op te tellen getallen. Vandaar de naam 'minimum addend.'

Wanneer kinderen ouder worden is de overgang bij alle rekenfeiten naar een oplossingsmethode waarbij de antwoorden direct uit het geheugen worden opgehaald normaal. Betekent dit dat de snelheid waarmee de rekenfeiten uit het geheugen worden opgediept voor alle feiten dezelfde is? Dat is niet het geval. Er blijven systematische verschillen in oplostijd tussen rekenfeiten bestaan. Bovendien blijft de 'minimum addend' steeds een van de beste voorspellers. Maar de extra-tijd die nodig is voor elke toename van de 'minimum addend' met 1 wordt wel veel kleiner (20 msec. bij volwassenen).

Hoe is dat te verklaren? Er zijn twee verschillende verklaringen voor gegeven. Groen en Parkman gaven een verklaring die men zou kunnen aanduiden als het 'twee-processen model'. Zij gaan ervan uit dat de tijd die in de geheugenstrategie nodig is voor het zoekproces onafhankelijk is van de grootte van de getallen in de opgave. Maar de geheugenstrategie wordt niet bij alle trials gevolgd, omdat de automatisering van de rekenfeiten bij de meeste personen niet volledig is. Bij sommige trials zullen sommige proefpersonen het antwoord niet in het geheugen ophalen, maar uitrekenen. Als men onderzoeksgegevens over een aantal mensen combineert, combineert men de oplostijden voor geheugentrials en berekeningstrials. De geleidelijke toename van de gemiddelde oplostijd met het groter worden van de getallen in de opgave wordt dan verklaard door de toename van de fractie berekeningstrials in de geaggregeerde data.

Directe steun voor dit twee-processen model werd kort geleden geleverd door Siegler (1987). Hij vroeg kinderen na elke opgave of zij het antwoord in hun geheugen hadden gevonden of het hadden uitgerekend. De kwantitatieve gegevens over deze twee categorieën trials, geheugentrials en rekentrials, verschillen sterk. De gemiddelde oplostijd bij geheugentrials is veel korter dan bij rekentrials en het aantal fouten is bij geheugentrials veel groter dan bij rekentrials.

Ashcraft (1982) heeft een andere verklaring geponeerd voor het verschijnsel dat ook bij volwassenen de oplostijd afhankelijk is van de getalgrootte. Hij stelt dat de rekenfeiten in het geheugen gerepresenteerd zijn als een netwerk. Men kan zich dit als volgt voorstellen. De opgave, bijvoorbeeld $6 + 3 =$, activeert de knopen '6' en '3' in dit netwerk, en deze activatie verspreidt zich naar met '6' en '3' geassocieerde knopen, bijvoorbeeld '9', '12', en

'18'. Bij voldoende oefening is de associatie van ' $6 + 3$ ' met '9' sterker dan die met '12', '18', etc. en wordt '9' als antwoord gegeven.

Binnen dit netwerk-model worden verschillen in gemiddelde reactietijd verklaard door te veronderstellen dat in het deel van het netwerk waar zich de kleine getallen bevinden de verbindingen tussen knopen selectiever zijn, d.w.z. dat bij een opgave als ' $2 + 3$ ' incorrecte geassocieerde knopen minder geactiveerd worden dan bij ' $4 + 5$ ' en dat daardoor het antwoord sneller gegeven kan worden. Grotere selectiviteit van verbindingen ontstaat door meer oefening. Er is alle reden om aan te nemen dat in een mensenleven de verbindingen tussen kleine getallen vaker en vanaf jongere leeftijd geoefend worden dan die tussen grotere getallen (Hamann & Ashcraft, 1986).

Naast de verwerving van 'rekenfeiten' staat het leren optellen en aftrekken tot 100 de laatste jaren sterk in de belangstelling van onderzoekers. Centraal in het laatste onderzoek is de vraag naar de ontwikkeling van rekenprocedures, met name ook bij tientalpassering. Illustratief voor dit onderzoek zijn de artikelen van Van der Heijden en van Beishuizen, Beishuizen en Felix in dit nummer.

Een ander populair onderzoeksgebied is de redactieopgave. Voor het onderzoek naar de factoren die van invloed zijn op de moeilijkheid van redactieopgaven wordt een andere onderzoeksmethode gebruikt, nl. regressieanalyse van de p-waarde van opgave-typen met als predictoren verschillende opgavenkenmerken. Men laat een flink aantal personen een reeks opgaven oplossen en bepaalt dan de p-waarde van elk type opgave. Deze methode wordt hoofdzakelijk gebruikt om de invloed na te gaan die systematische veranderingen in de opgave hebben op de moeilijkheid ervan.

In een eerste ronde van onderzoek analyseerden Jerman en medewerkers (o.a. Jerman & Mirman, 1974) de voorspellende waarde van allerlei kenmerken van redactieopgaven voor de p-waarden. Zij gebruikten hiervoor heterogene verzamelingen opgaven. De kenmerken die zij als predictoren gebruikten betroffen zowel taalkundige aspecten van de opgaven (aantal woorden, syntactische complexiteit) als wiskundige aspecten (aard en aantal van de benodigde rekenkundige bewerkingen, getalgrootte). Dit onderzoek leverde geen duidelijke resultaten. De vermoedelijke reden hiervoor is dat de onderzochte kenmerken alle betrekkelijk uiterlijke kenmerken zijn, terwijl uit later onderzoek duidelijk is geworden, dat de moeilijkheidsgraad van een opgave in de eerste plaats wordt bepaald door zijn onderliggende semantische structuur. Daardoor kunnen opgaven die op uiterlijke kenmerken niet van elkaar te onderscheiden zijn toch sterk in moeilijkheid verschillen.

De tweede ronde van onderzoek heeft zich in hoofdzaak beperkt tot een zeer beperkte groep opgaven. Een kernpublicatie is van de hand van Riley, Greeno en Heller (1983). Het door hen gepresenteerde schema van opgavetypen wordt in brede kring gehanteerd. Zij onderzochten optel- en aftrekopgaven van vier verschillende semantische typen: verandering-, combinatie-, vergelijkings- en gelijkmakingsopgaven. Het 'sometje' wordt in deze vier typen op verschillende manieren ingekleed. Een veranderingsopgave is: 'Jan had 4 knikkers. Hij kreeg er 3 bij. Hoeveel heeft hij er nu?' Een combinatieopgave is: 'Jan heeft 4 knikkers, Piet heeft er 3. Hoeveel hebben zij samen?' Een vergelijkingsopgave is: 'Jan heeft 4 knikkers. Piet heeft er 3 meer. Hoeveel knikkers heeft Piet?' En een gelijkmakingsopgave is: 'Jan heeft 4 knikkers. Als hij er 3 bij krijgt heeft hij er evenveel als Piet.' Hoeveel knikkers heeft Piet? Binnen ieder van deze vier typen kan men nog verschillende subtypen construeren door het kiezen van de onbekende. Bij veranderingsopgaven kan dat bijvoorbeeld de eindverzameling zijn, zoals in het hierboven gegeven voorbeeld, maar het kan ook de veranderingverzameling zijn, zoals in 'Jan had eerst 4 knikkers. Nu heeft hij er 7. Hoeveel heeft hij erbij gekregen?', of de beginverzameling zoals in 'Jan heeft er 3 knikkers bijgekregen. Nu heeft hij er 7. Hoeveel knikkers had hij eerst?'.

Deze ronde heeft gedetailleerde kennis opgeleverd over de factoren die de moeilijkheid van opgaven in deze klasse van opgaven bepaalden. Ten eerste: de semantische structuur is van groot belang; grofweg kan men zeggen dat veranderings- en combinatieopgaven gemakkelijker zijn dan vergelijkingsopgaven. Ten tweede: de plaats van de onbekende is zo mogelijk van nog groter belang. Een enkel voorbeeld: de hierboven als voorbeeld gegeven combinatieopgaven had in het onderzoek van Tamburino (1980; gegevens vermeld in Riley et al.) een p-waarde van 0.83, maar het andere type combinatieopgave waarin een van de deelverzamelingen de onbekende is ('Jan heeft 4 knikkers. Samen met Piet heeft hij er 7. Hoeveel knikkers heeft Piet?') had een p-waarde van slechts 0.18. Tenslotte zijn ook de gekozen bewoordingen (zie bv. De Corte, Verschaffel & De Win, 1985) van belang, maar deze invloed is toch relatief gering in vergelijking met de eerder genoemde twee factoren.

DE PROCESGERICHTE BENADERING

Het onderzoek van Riley, Greeno en Heller markeert de overgang van de experimentele richting naar de procesgerichte benadering. Hun werk is namelijk niet alleen kwantitatief van aard, maar schetst bovendien een kwalitatief model van de kennis die het oplossen van verschillende opgaven vereist en de ontwikkelingsstadia waarin deze kennis wordt verworven. De procesgerichte cognitieve benadering richt zich op de bestudering van de aard en ontwikkeling van cognitieve processen. Het voornaamste probleem daarbij is dat dergelijke processen zich niet lenen tot directe inspectie en zeker geen direct meetbare variabelen kennen. Daarom is het 'cognitieve proces' ook lange tijd geweerd als voorwerp van wetenschappelijk onderzoek. Alleen direct observeerbare verschijnselen konden zich verheugen in de belangstelling van de wetenschap, maar het menselijk denken moest helaas buiten beschouwing blijven. Toch zijn er methoden ontwikkeld om enigszins zicht te krijgen op cognitieve processen. De belangrijkste is de hardop-denken methode. Vraag mensen die je een probleem laat oplossen of een taak laat uitvoeren om dit hardop denkend te doen.

Aan deze methode kleven nogal wat bezwaren. Niet alle processen lenen zich voor verwoording. Vooral waar non-verbaal materiaal een rol speelt, is het moeilijk de juiste woorden te vinden. Bovendien laat een hardop-denken protocol altijd slechts het topje van de ijsberg zien. Allerlei processen spelen zich onbewust af, zodat alleen tussentijdse produkten in het werkgeheugen terechtkomen en verwoord worden. Het analyseren van protocollen is dan ook een gevaarlijk karwei. Aangezien het protocol per definitie onvolledig is, moet altijd ten dele geïnfereerd worden wat zich heeft afgespeeld. Om zoveel mogelijk houvast te hebben wordt de onderzoeker geacht van te voren een model te ontwikkelen van het proces dat bestudeerd wordt en op basis van dit model voorspellingen te doen over wat voor type uitspraken er in welke volgorde verwacht worden in een protocol dat de neerslag vormt van dit proces. Ideaal zou zijn wanneer het van te voren opgestelde model dat in feite een operationele theorie van een proces vormt, geïmplementeerd zou kunnen worden in de vorm van een computer-programma dat het proces simuleert en tussentijds protocol-uitspraken genereert. Nu is de situatie vaak verre van ideaal. Vaak weet men niet wat te verwachten en is een model verre van volledig zodat het zich niet leent voor implementatie. En zelfs in het ideale geval blijft het een probleem te beoordelen hoe de protocol-data zich verhouden tot het model.

Hoe wordt bepaald of en in hoeverre model en data overeenkomen? Zo langzamerhand worden wel technieken ontwikkeld die het mogelijk maken model en data volgens bepaalde criteria te vergelijken, maar deze technieken worden zeker niet op grote schaal toegepast. Bovendien staat de cognitieve richting niet een zeer ver ontwikkeld arsenaal aan

bewerkingstechnieken ter beschikking zoals voor de experimentele richting wel het geval is (statistiek/psychometrie). De 'cognitive science' methodologie staat nog in de kinderschoenen. Jammer is dat men vaak geneigd is daardoor weinig waarde te hechten aan het kwalitatieve onderzoek dat op dit terrein plaatsvindt. In de meeste gevallen zal het verzamelen en interpreteren van protocollen op zich niet kunnen leiden tot voldoende betrouwbare uitspraken over de bestudeerde processen. De kwalitatieve data die het hardop denken oplevert zullen dan ook ondersteund moeten worden door kwantitatieve gegevens die op andere wijze worden verzameld.

Waarom heeft de belangstelling voor het verloop van denkprocessen geleid? Allereerst is duidelijk geworden dat veel ogenschijnlijk onverklaarbare fouten wel degelijk op systematische wijze tot stand komen. De antwoorden die kinderen geven op een toets blijken op zich onvoldoende voor het achterhalen van de systematiek waarop foute antwoorden berusten. Deze bevinding heeft consequenties voor de inrichting van het hedendaagse rekenonderwijs. Niet alleen wordt de prestatie op een toets beschouwd als indicatie van het kennisniveau van de leerling, maar bovendien wordt getracht te achterhalen door individuele gesprekken welke kennis leerlingen ontberen en welke misvattingen er bij leerlingen bestaan.

Bovendien blijkt dat zelfs goede antwoorden op toetsvragen niet noodzakelijkerwijs inzicht in het gebied in kwestie vertegenwoordigen. Mensen, niet alleen kinderen, blijken vaak allerlei problemen te lijf te gaan met oppervlakkige kennis die in sommige gevallen wel tot goede resultaten leidt (Jansweijer, Wielinga & Elshout, 1986, Barnard & Sandberg, 1988). Deze bevinding heeft een aantal implicaties. Welke kennis is precies nodig voor het oplossen van vakinhoudelijke problemen? Hoe is de verhouding tussen kennis van het domein en kennis van algemene probleemoplossingsstrategieën? Een tijd lang is het vertrouwen in het ontwikkelen van algemene probleemoplossingsvaardigheden nogal groot geweest. De gebrekkige vaardigheid van kinderen in het oplossen van vakinhoudelijke problemen werd eerder geweten aan gebrekkige strategieën dan aan gebrekkige vakinhoudelijke kennis. Wat leerlingen vooral zouden moeten leren is eerst te denken en dan pas te doen. Vooral het analyseren van een probleem bleef vaak achterwege (oriëntatie-fase). Maar onderzoek naar het ontwikkelen van efficiënte oplossingsstrategieën heeft uitgewezen dat dergelijke strategieën niet veel uitkomst bieden wanneer de noodzakelijke vakinhoudelijke kennis ontbreekt. Hoe bepaal je waarop je je moet oriënteren wanneer je onbekend bent met de feiten, begrippen en relaties die het vakdomein bepalen? Duidelijk is inmiddels dat beide soorten kennis, vakinhoudelijk en algemeen strategisch, een even grote rol spelen in het probleemoplossen. Zonder algemene strategieën kan kennis niet worden toegepast, maar zonder vakinhoudelijke kennis kunnen de strategieën niet worden gebruikt.

Binnen het rekenonderzoek wordt protocol-analyse gebruikt om te achterhalen hoe leerlingen in verschillende stadia van de ontwikkeling rekenen. Een probleem daarbij is dat jonge kinderen de opdracht hardop te denken tijdens het rekenen of niet begrijpen of niet kunnen combineren met het rekenen zelf. Op verschillende manieren wordt dit probleem ondervangen. Bijvoorbeeld door het kind te vragen de rol van leerkracht te spelen en aan de proefleider uit te leggen hoe een opgave gemaakt moet worden, of door de leerling onmiddellijk na afloop van het oplossingsproces te vragen: hoe heb je dat uitgerekend. Beide oplossingen leveren interessante protocollen. Vooral binnen de domeinen van redactieopgaven en open beweringen heeft protocol-analyse vaste grond onder de voeten gekregen (DeCorte & Verschaffel, 1985, Carpenter & Moser, 1984, Lindvall & Tamburino, 1981, Barnard & Sandberg, 1988).

Dergelijk onderzoek heeft veel aan het licht gebracht over de wijze waarop leerlingen opgaven beschouwen en vormt een welkome aanvulling op studies die kwantitatieve gegevens over de moeilijkheidsgraad van verschillende typen redactieopgaven, respectieve-

lijk open beweringen leveren. Onderzoek van Barnard en Sandberg (1988) wees uit dat het niet voldoende is leerlingen een stipsummentoets af te nemen om te achterhalen hoe zij dergelijke opgaven oplossen. Individuele interviews toonden bijvoorbeeld aan dat sommige leerlingen geneigd zijn in bepaalde gevallen van leesrichting te veranderen (van rechts naar links in plaats van omgekeerd) en dan tot een andere oplossing komen. Bovendien bleek uit de interviews dat juiste antwoorden niet altijd gebaseerd zijn op inzicht en kennis. Sommige leerlingen maken zogenaamd moeilijke opgaven goed door zich te baseren op oppervlakte kenmerken en toevallig de goede operatie te kiezen. In het geval van open beweringen leverden interviews allerlei gegevens op die via een normale toets-afname niet te achterhalen zijn.

Met name op het gebied van de redactieopgaven zijn modellen ontwikkeld die de kennis die leerlingen in verschillende stadia van de ontwikkeling bezitten, beschrijven en illustreren hoe deze kennis bij individuele opgaven wordt toegepast (Riley, Greeno & Heller, 1983, Kintsch & Greeno, 1985). De modellen beschrijven verschillende stadia in de ontwikkeling en de typen kennis die bij de stadia horen. Volgens Riley, Greeno en Heller (1983) veronderstelt het oplossen van redactieopgaven verschillende typen kennis. Kennis van de semantiek van redactieopgaven is vastgelegd in een beperkt aantal schema's. Op basis van het lezen en begrijpen van de opgavetekst wordt de opgave geïnterpreteerd in termen van één van deze schema's. Vervolgens wordt dit conceptuele schema gekoppeld aan een zogenaamd actie-schema waarbij de semantische elementen gelieerd worden aan rekenkundige procedures. Fouten kunnen in verschillende fasen van het oplosproces optreden. De aanvankelijke interpretatie van de opgave kan onjuist zijn, het gekozen representatie-schema kan onjuist zijn en tot slot kan de koppeling tussen representatie-schema en rekenkundige bewerking problemen opleveren. De modellen beschrijven hoe leerlingen in verschillende stadia de opgave-tekst interpreteren, welk schema vervolgens wordt gekozen en hoe dit schema wordt gekoppeld aan een rekenkundige bewerking. De modellen beschrijven mogelijke interne representaties, oplossingsstrategieën en de ontstaanswijzen van fouten.

Uitgangspunt van de modellen is de veronderstelling dat de moeilijke opgaven cognitieve operaties vereisen die pas laat in de ontwikkeling verworven worden. Een van deze operaties is het interpreteren van opgaven in termen van een deel-geheel schema. Pas hierdoor zou het mogelijk worden veranderingsopgaven waarin de beginverzameling onbekend is op te lossen.

Hamaker, Hoeks, Ekkebus en Van Kampen (in dit nummer) maakten gebruik van deze veronderstelling in hun onderzoek naar mogelijke verschillen tussen kinderen met leermoeilijkheden en normaal lerende kinderen. Meer specifiek toetsen zij de gedachte dat kinderen met leermoeilijkheden wel de fundamentele cognitieve vaardigheden verwerven die nodig zijn voor het oplossen van eenvoudige opgaven, maar niet de vaardigheden waarmee de meer complexe opgaven opgelost moeten worden.

Het onderzoek naar wat redactieopgaven moeilijk maakt heeft in het algemeen naast kwantitatieve methoden ook op grote schaal kwalitatieve methoden toegepast. Het werk van De Corte en Verschaffel is hiervan een goed voorbeeld (bijv. De Corte, Verschaffel & De Win, 1985, De Corte & Verschaffel, 1987). Behalve kwalitatieve gegevens leveren de modellen van Riley et al. indirect ook kwantitatieve data die getoetst kunnen worden aan de empirie. Op basis van de modellen wordt namelijk de moeilijkheidsgraad van opgaven vastgesteld: elk model bepaalt welke opgaven wel en welke opgaven niet correct worden opgelost met de gegeven kennis. De door de modellen voorspelde moeilijkheidsgraad kan vervolgens worden vergeleken met empirische gegevens over de moeilijkheidsgraad van verschillende typen opgaven.

Een tweede terrein waar modelbouw en implementatie sterk ontwikkeld is is het terrein

van de 'Repair-theorie'. Deze oorspronkelijk door Brown en VanLehn (1980) ontwikkelde theorie gaat uit van de aanname dat een leerling niet zo maar opgeeft wanneer het probleemoplossingsproces stopt, dat wil zeggen wanneer de leerling in een impasse raakt. Leerlingen zullen in tegenstelling tot een computer-programma zoeken naar middelen om uit een impasse te raken. Dergelijke middelen worden door Brown en VanLehn 'repairs' genoemd. Een repair is een lapmiddel, een noodgreep om uit een impasse te raken en leidt meestal tot een incorrecte oplossing. Brown en VanLehn hebben hun repair-theorie geïllustreerd aan de hand van verticaal aftrekken op papier. Door uit het bestaande algoritme voor dit domein stapjes weg te laten wisten zij impasses te genereren die werden opgeheven met repairs als: trek het kleinste van het grootste getal af, in het geval je eigenlijk zou moeten lenen, maar niet weet hoe dat gaat. Een computer-programma genereert incomplete procedures voor het aftrekken door stappen uit de correcte procedure weg te laten. De impasses die dan ontstaan worden opgeheven met verschillende repairs die willekeurig worden gekozen. De bedoeling is dat het programma uiteindelijk alle empirisch gevonden foute antwoorden kan genereren zonder fouten te genereren die in de praktijk nooit voorkomen en volgens het oordeel van leerkrachten ook onwaarschijnlijk zijn. De uitwerking van de repair-theorie en het daaraan verbonden empirisch onderzoek vormt een goede illustratie van wat Kempen (1987) beschrijft als het kunstmatige intelligentie paradigma. Het computer-programma wordt beschouwd als een operationalisatie van de theorie en de 'produkten' van het programma worden vergeleken met de oplossingen van kinderen bij aftreksommen. Steeds vaker wordt de computer ingeschakeld bij dit type onderzoek. Soms alleen als diagnostisch hulpmiddel, waarbij leerlingen via het scherm opgaven krijgen aangeboden en hun antwoord moeten intikken. Het voordeel is dat bij deze werkwijze alle gegevens van de leerlingen ook onmiddellijk geadministreerd worden. Soms is de rol van de computer veel uitgebreider en wordt een programma ontwikkeld als operationalisatie van een bepaald cognitief proces.

DISCUSSIE

In het bovenstaande is een overzicht gegeven van de experimentele benadering met nadruk op kwantitatieve methoden en de procesgerichte benadering met nadruk op kwalitatieve methoden. In drie opzichten is een zekere convergentie in de onderzoeksresultaten te bespeuren:

- De voorgestelde procesmodellen worden steeds genuanceerder en ingewikkelder.
- Er ontstaat een steeds genuanceerder kijk op de aard van de ontwikkeling van rekenvaardigheden.
- Het arsenaal aan onderzoeks- en analysemethoden wordt steeds uitgebreider.

Op elk van deze drie conclusies wordt hieronder nader ingegaan. In de uitwerkingen zijn suggesties voor verder onderzoek opgenomen.

De oorspronkelijke simpele telmodellen van Groen en Parkman hebben plaatsgemaakt voor meer genuanceerde en complexe modellen die in drie richtingen zijn uitgebreid:

- Naast tellen (met een verhogen of verlagen) zijn nu ook andere strategieën geïdentificeerd. Baroody, Ginsburg en Waxman (1983) wezen erop dat leerlingen spontaan allerlei strategieën 'uitvinden' die min of meer afwijken van de onderwezen aanpak. Steinberg (1985) en Hamann en Ashcraft (1985) verzamelden middels retrospectieve interviews een scala aan rekenstrategieën bij leerlingen op de basisschool. In deze strategieën wordt niet alleen gebruik gemaakt van algemene telprocedures, maar ook van opgave-specifieke alternatieve werkwijzen, waarbij dikwijls een beroep gedaan wordt op in het geheugen aanwezige kennis van het rekendomein. Beishuizen (1985) constateerde bij

basisschoolleerlingen een opmerkelijke discrepantie tussen de onderwezen strategie en de door leerlingen toegepaste strategieën (zie Beishuizen, Beishuizen, & Felix in dit nummer).

- Het is duidelijk dat diverse kennisbronnen in het semantisch geheugen geraadpleegd worden bij het oplossen van een rekensom. Baroody (1985) noemde in dit verband (a) semantische kennis van het tientallig stelsel (bijvoorbeeld de communicatieve eigenschap), (b) representatie van de volgorde der natuurlijke getallen, (c) procedurele kennis, en (d) associaties van deel-geheel relaties. Andere auteurs hebben vergelijkbare indelingen gemaakt.
- Er wordt meer aandacht geschonken aan het optreden van rekenfouten en de ontstaanswijze ervan. Hiertoe hebben met name Brown en Burton (1978) en Brown en VanLehn (1980) met hun repair-theorie belangrijke bijdragen geleverd.

Dit alles levert een aantal onderzoeksvragen op. In de eerste plaats is niet duidelijk hoe kennis van getalsfeiten georganiseerd is. De tot nu gedane voorstellen van met name Ashcraft en Battaglia (1978) zijn nog niet erg uitgewerkt. Het gaat hier om een betrekkelijk klein corpus van geassocieerde feiten dat snel en vrij 'random' toegankelijk moet zijn. Nader onderzoek zou meer over de structuur van deze kennis aan het licht moeten brengen. Met name verdient het aanbeveling om na te gaan of connectionistische modellen in dit verband een bruikbaar beschrijvingskader bieden.

In de tweede plaats is de vraag van belang hoe de controle in procesmodellen verloopt waarin zowel rekenstrategieën worden toegepast als raadpleging van het geheugen plaatsvindt. Er is een trend waarneembaar van lineaire naar pandemonium modellen. Young (1985) suggereerde dat de rekenaar diverse interpretatiedomeinen ter beschikking heeft (getallenlijn, abacus, etc.). Er zijn, aldus Young, allerlei wezentjes in de weer die op de uitkijk staan en opletten of er iets van hun gading langs komt. Er is geen opzichter die toeziet op de ordelijke afhandeling van de rekenopgave. Alle ingeschakelde wezentjes schrijven hun resultaten op een gemeenschappelijk 'mededelingenbord', waar ieder geacht wordt van elkaars bevindingen op de hoogte te blijven. Lawler (1981) kwam tot vergelijkbare voorstellen op basis van een gedetailleerde studie van de ontwikkeling van rekenvaardigheden bij zijn dochter. Er is verder onderzoek in deze richting nodig. Hierdoor wordt het wellicht mogelijk verbindingen te leggen tussen enerzijds de theorievorming op het terrein van hoofdrekenen, cijferend rekenen, etc., en anderzijds het terrein van de redactiesommen. Verder is hierdoor ook meer aandacht mogelijk voor de netelige kwestie van de relatie tussen inzicht en begrip enerzijds en de toepassing van rekenalgorithmen anderzijds, een kwestie die in Nederland al tot menig debat heeft geleid.

De gesignaleerde trend van lineaire naar pandemonium modellen moet gezien worden in het licht van een algemene discussie: kan het menselijk denken bestudeerd worden zonder de werking van de hersenen als basis te kiezen of moet juist het bestuderen van de aard van de hersenen leiden tot het verklaren van het menselijk denken? Aanhangers van het eerstgenoemde standpunt maken meestal gebruik van lineaire, sequentiële modellen die stap voor stap tot een oplossing van een probleem raken. Aanhangers van het tweede standpunt bepleiten het gebruik van parallelle computers die in analogie met het menselijk brein enorme hoeveelheden informatie parallel, dat wil zeggen op hetzelfde moment, kunnen verwerken.

Tot nu toe hebben de connectionisten, zoals de bepleiters van parallellisme worden genoemd, nog niet overtuigend aangetoond dat hun simpele netwerkmodellen in staat zijn complex gedrag te simuleren. Het is niet ondenkbaar dat sommige cognitieve processen een parallel karakter vertonen (bijvoorbeeld waarneming) en andere een meer sequentieel (zoals taalproductie). Interessant is dat waar beide stromingen zich voorheen nogal

onafhankelijk van elkaar bewogen er nu meer sprake is van onderlinge samenwerking.

De strijd tussen 'parallel' en 'sequentieel' raakt nog een ander essentieel geschilpunt. De modellen die de connectionisten voorstellen bevatten geen 'diepe kennis'. Onder diepe kennis wordt werkelijk inzicht in verschijnselen en hun relaties verstaan. In connectionistische modellen is het zo dat arbitraire relaties tussen knopen in het netwerk en aan deze relaties verbonden sterktes bepalen welke relatie/knoop zal worden geactiveerd. De sterktes van de relaties zouden afhangen van het succes waarmee relaties in bepaalde situaties zijn gebruikt. Op zich kunnen dergelijke netwerken zich adequaat gedragen. De 8-puzzle, de welbekende puzzle waarbij je de cijfers 1 tot en met 8 op volgorde moet zien te zetten door telkens een cijfer via een open vakje te verplaatsen, bijvoorbeeld, kan via een netwerkmodel worden opgelost (Steels, 1988), maar dat wil niet zeggen dat het model over 'kennis' ten aanzien van de 8-puzzle beschikt. Het model heeft geen weet van getallen en hun betekenis, noch van mogelijke oplossingsstrategieën. Tegenstanders van connectionistische modellen voeren dan ook als belangrijk argument aan dat mensen in tegenstelling tot de voorgestelde modellen wel degelijk over diepe kennis beschikken die het probleemoplossen begeleidt en stuurt. Huns inziens schieten parallelle modellen te kort als theoretisch model voor menselijke cognitie. De hierboven geschetste geschilpunten overstijgen natuurlijk het terrein van het rekenen in belangrijke mate. Aardig is om te zien dat deze in het algemeen belangrijk geachte vraagstukken zich in het klein vertonen in dilemma's binnen het rekenonderzoek, zoals de keuze tussen lineaire en pandemonium modellen en het onderscheid tussen diepe en oppervlakte-kennis.

Het terrein van de ontwikkeling van rekenvaardigheden kent nog vele witte plekken. Belangrijke vragen betreffen de ontwikkeling van expertise en de lange-termijn ontwikkeling van rekenkennis, -vaardigheden en -inzicht.

Op grond van onderzoek naar het leren oplossen van problemen (zie Anderson, 1985) wordt verwacht dat leerlingen in de loop van hun ontwikkeling algemene heuristische strategieën verruilen voor domeinspecifieke algoritmische strategieën, die uiteindelijk worden verkort tot automatismen. Nader onderzoek is nodig om na te gaan of dergelijke ontwikkelingen van beginner tot expert zich ook voordoen in het rekendomein.

De lange-termijn ontwikkeling van rekenkennis, -vaardigheden, en -inzicht is wellicht theoretisch en methodologisch goed te benaderen met de 'rule-assessment'-aanpak van Siegler en Klahr (1982). Deze aanpak omvat een methode om de validiteit van stadia-modellen van cognitieve ontwikkeling te toetsen. Siegler en Klahr (1982) demonstreerden de 'rule-assessment'-aanpak met een analyse van de manier waarop kinderen tussen 4 en 8 jaar de balanstak van Inhelder en Piaget (1958) uitvoerden. Beishuizen, Beishuizen en Felix (dit nummer) passen de aanpak toe bij het analyseren van de ontwikkeling van rekenkennis en -vaardigheden (optellen en aftrekken in het domein van 20 tot 100). De 'rule-assessment'-aanpak van Siegler en Klahr kan wellicht worden aangevuld met Goldsteins (1982) 'genetic graph' waarin voorstellen worden gedaan voor de manier waarop de overgang tussen expertise-stadia plaats kan vinden. De genetic graph beschrijft hoe kennis zich kan ontwikkelen door bepaalde relaties te veronderstellen tussen nieuwe en oude kennis. Een voorbeeld van zo'n relatie is generalisatie. Bij dit alles moet speciale aandacht worden besteed aan de relatie tussen interne sturing (spontane ontwikkeling) en externe beïnvloeding via het onderwijs. Dit onderzoek kan resulteren in een 'atlas voor de ontwikkeling van rekenkennis', waarin specifieke wegen worden geschetst voor de ontwikkeling van verschillende rekenvaardigheden onder verschillende onderwijsomstandigheden.

Leidt dit alles tot nieuwe vormen van onderwijs, waarbij de leerling in een op maat gesneden leeromgeving door zelfstandige exploratie met de voor hem of haar uitgekende instrumenten een eigen weg gaat? O'Shea, Evertsz, Henessy, Floyd, Fox, en Elsom-Cook

(1988) hebben voorstellen gedaan voor het ontwerp van een 'intelligent arithmetic tutor' die een dergelijke leeromgeving biedt. Een andere richting vormt het ontwerp van coöperatieve leersituaties waarin kleine groepjes van leerlingen samen met de docent de 'kunst' van het rekenen ontdekken via onderlinge demonstratie en discussie. Het toegenomen inzicht in de ontwikkeling van cognitieve vaardigheden en de functies van 'leermeester' en 'gezel' (zie Collins, Brown, & Newman, in druk; Schoenfeld, 1985) maakt de kans op succes van deze onderwijsvorm groter dan men zou verwachten op grond van vrij vruchteloze pogingen in het recente verleden (zie bijvoorbeeld Vedder, 1985) waarbij geen diepgaande analyse van de vaardigheid in kwestie plaatsvond.

De geschetste uitbreidingen van de procesmodellen vereisen nieuwe statistische technieken om cognitieve groei te meten, bijvoorbeeld toepassing van een aangepaste versie van het Rasch model, waarbij het mogelijk is op grond van toetsprestaties van leerlingen vast te stellen of veronderstelde ontwikkelingsstadia inderdaad afzonderlijk van elkaar en in vaste volgorde in werkelijkheid bestaan. De toets dient te zijn samengesteld uit per stadium specifieke opgaven die zijn ontleend aan een voor de vaardigheid in kwestie homogene itemverzameling (Wilson, 1985).

Bovendien bestaat behoefte aan uitbreiding van (gestandaardiseerde) methoden om cognitieve processen te meten die ten grondslag liggen aan het oplossen van rekenopgaven. Op dit moment worden diverse vormen van kwalitatieve analyse gebezigd: stapjes opschrijven, hardop denken, verbaal rapport achteraf, registratie van oogbewegingen. Aan deze methoden kleven allerlei voor- en nadelen. Het betrouwbaar gebruik van de technieken is gebonden aan strikte voorwaarden. Het onderzoek is gebaat bij verdere uitwerking van deze kwalitatieve methoden zodat betrouwbare en valide data over de ontwikkeling en toepassing van rekenvaardigheden kunnen worden verworven.

Tenslotte zijn recentelijk nieuwe technieken ontwikkeld om procesmodellen te implementeren en testen, bijvoorbeeld SOAR (Laird, Newell, & Rosenbloom, 1987). Dergelijke technieken zijn bedoeld om lerende systemen te ontwikkelen en zijn afkomstig uit het gebied van 'machine learning', wat een terrein is binnen de kunstmatige intelligentie. De voornaamste tot nu toe gebruikte leertechniek is inductie. Via inductie wordt uit een aantal specifieke voorbeelden een algemene regel afgeleid. Toepassing van deze techniek en andere in het rekendomein kan nieuwe inzichten opleveren en een uitbreiding betekenen van het onderzoeksinstrumentarium.

REFERENTIES

- Anderson, J.R. (1985). *Cognitive Psychology and Its Applications* (2nd Ed.). New York: W.H. Freeman & Co.
- Ashcraft, M.H. (1982). The development of mental arithmetic: a chronometric approach. *Developmental Review*, 2, 213-236.
- Ashcraft, M.H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology*, 4, 5, 527-538.
- Barnard, Y.F. & Sandberg, J.A.C. (1988). Applying artificial intelligence insights in a CAI program for 'open sentence' mathematical problems in primary schools. *Instructional Science*, 17, 263-276.
- Baroody, A.J. (1985). Mastery of basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education*, 16, 2, 83-98.
- Baroody, A.J., Ginsburg, H.P., & Waxman, B. (1983). Children's use of mathematical structure. *Journal for Research in Mathematics Education*, 14, 3, 156-168.
- Beishuizen, M. (1985). Evaluation of the use of structured materials in the teaching of primary mathematics. In B.S. Alloway & G.M. Mills (Eds.), *New directions in education and training*

- technology - Aspects of educational technology*, Vol. 18. London: Kogan Page.
- Brown, J.S., & Burton, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-192.
- Brown, J.S., & VanLehn, K. (1980) Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.
- Carpenter, T.P., & Moser, J.M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15, 179-202.
- Collins, A., Brown, J.S., & Newman, S.E. (in druk). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- De Corte, E. (1984). Kwalitatieve gegevens in onderwijsonderzoek. In L.F.W. De Klerk & A.M.P. Knoers (Eds.), *Onderwijspsychologisch onderzoek* (pp. 6-26). Lisse: Swets & Zeitlinger.
- De Corte, E., & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *Journal for Research of Mathematical Behavior*, 4, 3-21.
- De Corte, E., & Verschaffel, L. (1987). Oogbewegingen van eersteklassers tijdens het oplossen van redactie-opgaven. *Pedagogische Studiën*, 64, 137-149.
- De Corte, E., & Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, 77, 460-470.
- Goldstein, I. (1982). The genetic graph: A representation for the evolution of procedural knowledge. In D. Sleeman & J.S. Brown (Eds.), *Intelligent Tutoring Systems*. New York: Academic Press.
- Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review*, 79, 329-343.
- Hamann, M.S., & Ashcraft, M.H. (1985). Simple and complex mental addition across development. *Journal of Experimental Child Psychology*, 40, 49-72.
- Hamann, M.S., & Ashcraft, M.H. (1986). Textbook presentation of the basic addition facts. *Cognition and Instruction*, 3, 173-192.
- Inhelder, B., & Piaget, J. (1985). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Jansweijer, W.N.H., Elshout, J.J. M Wielinga, B.J. (1986). The expertise of novice problem solvers. In proceedings ECAI'86, Brighton.
- Jerman, M.E., & Mirman, S. (1974). Linguistic and computational variables in problem solving in elementary mathematics. *Educational Studies in Mathematics*, 5, 317-362.
- Kempen, G.A.M. (1987). *Natuurlijke Taal en Kunstmatige Intelligentie*. Groningen: Wolters-Noordhoff.
- Kintsch, W., & Greeno, J.G. (1958). Understanding and solving arithmetic word problems. *Psychological review*, 92, 109-129.
- Laird, J.E., Newell, A., & Rosenbloom, P.S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33, 1-64.
- Lawler, R.W. (1981). The progressive construction of mind. *Cognitive Science*, 5, 1, 1-31.
- Lindvall, C.M., & Tamburino, J.L. (1981). Information processing capacities used by kindergarten children when solving simple arithmetic story problems. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, C.A.
- Nesher, P. (1982). Levels of description in the analysis of addition and subtraction word problems. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.) *Addition and subtraction: A cognitive perspective* (pp. 25-38). Hillsdale, NJ: Erlbaum.
- O'Shea, T., Evertsz, R., Hennesy, S., Floyd, A., Fox, M., & Elsom-Cook M. (1988). Design choices for an intelligent arithmetic tutor. In J. Self (Ed.), *Artificial Intelligence and Human Learning*. London: Chapman & Hall.
- Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem solving ability. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.
- Schoenfeld, A.H. (1985). *Mathematical problem solving*. New York: Academic Press.
- Siegler, R.S., & Klahr, D. (1982). When do children learn? The Relationship between Existing

- Knowledge and the Acquisition of New Knowledge. In R. Glaser (Ed.), *Advances in Instructional Psychology, Vol. 2* Hillsdale, NJ: Lawrence Erlbaum Associates.
- Siegler, R.S. (1987). The perils of averaging data over strategies: an example from children's arithmetic. *Journal of Experimental Psychology: General*, 116, 250-264.
- Steels, L. (1988). Steps towards common sense. In Proceedings ECAI'88, München.
- Steinberg, R.M. (1985). Instructions on derived facts strategies in addition and subtraction. *Journal for Research in Mathematics Education*, 16, 5, 337-355.
- Vedder, P. (1985). *Cooperative Learning*. Den Haag: Stichting voor Onderzoek van het Onderwijs.
- Wilson, M.A. (1985). *Measuring Stages of Growth: A Psychometric Model of Hierarchical Development*. Hawthorn, Victoria: Australian Council for Educational Research.
- Young, R.M. (1985). Multi-agent model of children's understanding of arithmetic. Paper gepresenteerd op de conferentie 'Cognitive processes in mathematics and mathematics learning'. Keele: University of Keele.

Manuscript ontvangen 4-10-1988

Definitieve versie ontvangen 3-2-1989.

Complexiteit van Rekenopgaven en kinderen met leermoeilijkheden

Christiaan Hamaker, Jan Hoeks, Michel Ekkebus en Ineke van Kampen.
Vakgroepen Psychonomie en Ontwikkelingspsychologie, Universiteit van Amsterdam.

ABSTRACT

In two experiments we tested two contrasting views of arithmetic learning difficulties. One view holds that children with learning difficulties have a general developmental lag, and this view predicts that the profile of performance on assorted arithmetic tasks will show no specific strengths or weaknesses. The other view holds that children with arithmetic learning difficulties have weaknesses in some arithmetic tasks but not in others. We applied Van Meel's (1968) version of the specific weakness position to the domain of arithmetic. Van Meel supposes that these children are weak in complex tasks, because they tend to restrict their temporal perspective. In both experiments we compared a group of children with learning difficulties with a younger group of normal learning children. In only one out of six operationalizations of task complexity did we find evidence supporting the specificity position. This effect is attributed to salience of misleading cues rather than to information overload due to complexity of the task. Because we did not replicate data from the research literature on other operationalizations of complexity, we conclude that support for the specificity view is weak, although it can not yet be discarded entirely.

INLEIDING

Veel kinderen lopen in het onderwijs achterstand op bij het leren rekenen. Is deze achterstand algemeen, dat wil zeggen, is de achterstand even groot bij alle onderdelen van het rekenen? Of is de achterstand specifiek, dat wil zeggen, bestaat er naast grote achterstanden van deze kinderen bij sommige rekentaken, een geringere of zelfs in het geheel geen achterstand op andere rekentaken? In ons onderzoek trachten wij deze vragen te beantwoorden door rekenprestaties van leerlingen uit het gewoon basisonderwijs en uit het l.o.m. (leer- en opvoedingsmoeilijkheden) -onderwijs met elkaar te vergelijken.

In de onderzoeksliteratuur over de structuur van cognitieve vaardigheden bij kinderen met leermoeilijkheden (LM kinderen) treft men tegengestelde standpunten aan. Sommigen menen dat kinderen met leermoeilijkheden 'cognitief normaal' zijn (Russell & Ginsburg, 1984), d.w.z. er is bij deze leerlingen alleen een vertraagde cognitieve ontwikkeling, maar het vaardigheidsprofiel vertoont geen bijzondere zwakke plekken. Ginsburg is een vertegenwoordiger van dit standpunt (Ginsburg, 1972; Allardice & Ginsburg, 1983; Ginsburg & Russell, 1984). We verwijzen in wat volgt naar deze opvatting als de 'algemene achterstandshypothese'.

Andere onderzoekers menen dat leerlingen met leermoeilijkheden *specifieke* cognitieve zwaktes vertonen. Vertegenwoordigers van deze groep vindt men vaak bij onderzoekers die leermoeilijkheden vanuit de neurologie benaderen, zoals Rourke (Rourke, Bakker, Fisk, & Strang, 1983), maar ook bij anderen, zoals Van Meel (1968). We zullen naar deze opvatting verwijzen als de 'specificiteitshypothese'.

De specificiteitshypothese is een algemene veronderstelling die pas onderzoekbaar

wordt wanneer tevens wordt aangegeven welke specifieke zwakte kinderen met leermoeilijkheden zouden hebben in het profiel van cognitieve vaardigheden. In ons onderzoek zijn we uitgegaan van de opvatting van Van Meel (1968), omdat deze goed leek aan te sluiten bij enkele resultaten van Russell en Ginsburg (1984).

Van Meel vat zijn opvatting over de specifieke zwakte van kinderen met leermoeilijkheden als volgt samen: 'De meest algemene karakteristiek van cognitie bij deze kinderen zou kunnen worden geformuleerd door nadruk te leggen op de *inkrimping van het temporele perspectief*. Het cognitieve systeem is afgestemd op snelle beslissingen op korte termijn' (Van Meel, 1968, p. 182, cursivering in het origineel). Deze algemene typering van de cognitieve zwakte van l.o.m.-leerlingen steunt op specifieke onderzoeksresultaten. Op de Stroop taak bleken de l.o.m.-leerlingen moeite te hebben met het negeren van irrelevante informatie en bij een categorisatietask hadden deze kinderen de neiging op globale kenmerken af te gaan en brede categorieën te vormen.

Onderzoeksresultaten van Russell en Ginsburg (1984) brachten ons op de veronderstelling dat sommige typen redactieopgaven sterker te lijden zouden hebben van de door Van Meel beschreven cognitieve zwakte dan andere typen. Russell en Ginsburg onderzochten de rekenprestaties van een groep vierde klassers ('fourth graders') uit het gewoon lager onderwijs met een rekenachterstand van tenminste één jaar en vergeleken deze prestaties met die van normaal lerende derde en vierde klassers. Zij namen bij deze drie groepen 14 rekentaken af, en vonden bij twaalf taken geen significante achterstand van de rekenachterstand groep ten opzichte van de derde klassers. Op twee taken waren de prestaties wel significant slechter dan de normaal lerende derde klassers: (a) kennis van rekenfeiten en (b) redactieopgaven. We bespreken hier alleen de voor onze vraagstelling relevante resultaten van de redactieopgaven taak. Het is overigens opmerkelijk dat deze resultaten van Russell en Ginsburg steun lijken te bieden voor de specificiteitshypothese, maar dat zij zelf desondanks geen reden zien hun conclusie dat deze kinderen cognitief normaal zijn te herzien.

Russell en Ginsburg vonden bij analyse van de resultaten per opgave dat het geconstateerde prestatieverschil op de redactieopgaventest vooral moest worden toegeschreven aan twee van de acht gebruikte opgaven. De ene is een *indirecte* opgave, dat is een opgave waarvan de bewoordingen (vaak een zogenaamd sleutelwoord, zoals 'samen', 'eraf', 'minder') verleiden tot het uitvoeren van een verkeerde rekenkundige operatie. De andere is een opgave *met irrelevante numerieke informatie*. Cruickshank (1948) vond in zijn onderzoek met 'mentally retarded' kinderen een soortgelijke specifieke uitval bij opgaven met irrelevante informatie.

De bevindingen van Russell en Ginsburg (1984) en van Cruickshank (1948) vormen een (vooral bij Russell en Ginsburg onbedoelde) ondersteuning van de opvatting van Van Meel. Redactieopgaven met irrelevante informatie hebben met de Stroop test gemeen, dat opvallend aanwezige informatie niet gebruikt moet worden. Indirecte opgaven hebben met de categorisatie taak gemeen dat een globale verwerking van de informatie in de opgave onvoldoende is voor een goede taakuitvoering. Naar analogie van Van Meel's 'inkrimping van het temporeel perspectief' formuleren wij de hypothese dat kinderen met leermoeilijkheden vooral problemen zullen hebben met redactieopgaven met een hoge informatieverwerkingslast, dat zijn opgaven die vereisen dat *relatief veel informatie tegelijkertijd en in onderlinge samenhang* beschouwd wordt.

Deze hypothese wordt in ons onderzoek getoetst met variantie analyse. De specificiteits-hypothese leidt tot de verwachting van een interactie van de variabelen leerlingtype en informatieverwerkingslast. Bij kinderen met leermoeilijkheden wordt een groter verschil in prestatieniveau verwacht tussen taken met een geringe en taken met een grote informatieverwerkingslast dan bij normaal lerenden. Men dient echter in gedachte te

houden dat onze oorspronkelijke vraagstelling een keuze tussen twee alternatieve zienswijzen behelst. Dat betekent dat een significante interactie in de verwachte richting geïnterpreteerd zal worden als steun voor de specificiteitshypothese, terwijl het uitblijven van een significant resultaat zal worden opgevat als steun voor de algemene achterstandshypothese.

De variabele informatieverwerkingslast hebben wij in de hier gerapporteerde experimenten op uiteenlopende wijzen geoperationaliseerd. Deze operationalisaties komen bij de afzonderlijke experimenten aan de orde. Overigens hebben de beide experimenten eenzelfde design, waarvan één aspect het verdient hier afzonderlijk vermeld te worden. We kozen voor een design waarin een groep normaal lerende kinderen en een groep l.o.m.-leerlingen van ongeveer *gelijk rekenprestatieniveau* met elkaar vergeleken worden. Uiteraard komt men dan uit op een groep normaal lerenden met lagere gemiddelde kalenderleeftijd dan de l.o.m.-groep. Deze opzet werd gekozen op methodologische gronden. Kiest men twee groepen van gelijke leeftijd dan is te verwachten dat de l.o.m.-groep over de hele linie minder goed zal presteren dan de normaal lerende groep. De prestaties van de l.o.m.-groep liggen dan dus in een ander deel van de meetschaal dan de prestaties van de groep waarmee zij vergeleken worden. Dat betekent dat een eventueel te vinden statistische interactie niet eenduidig interpreteerbaar zou zijn, omdat men niet zeker kan zijn dat (on)gelijke scoreverschillen in verschillende gebieden van de meetschaal ook wijzen op (on)gelijke vaardigheidsverschillen. De gekozen opzet garandeert dat de prestaties van de te vergelijken groepen zich bevinden in hetzelfde gebied van de meetschaal.

Naast het genoemde voordeel heeft dit design natuurlijk ook nadelen. De voornaamte is dat niet kan worden nagegaan hoe groot de rekenachterstand van l.o.m.-leerlingen is ten opzichte van normaal lerende leeftijdgenoten. Het gekozen design is echter te verkiezen bij de door ons onderzochte hypothese.

Voor het onderzoek werd binnen de groep l.o.m.-kinderen geen verdere selectie gemaakt. Dit ondanks de gangbare mening dat de groep kinderen met rekenmoeilijkheden heterogeen is (Badian, 1983; Rourke & Strang, 1984). De voornaamste reden voor onze werkwijze is dat andere onderzoekers de hiervoor besproken interacties aantreffen in zeer uiteenlopende en heterogene populaties. Cruickshank (1984) deed zijn onderzoek onder 'mentally retarded' kinderen; Russell en Ginsburg (1984) bij kinderen met een rekenachterstand in het gewoon lager onderwijs. Er was derhalve voor ons geen reden aan te nemen dat binnen de groep l.o.m.-kinderen subgroepen zouden bestaan waarvan sommige wel en andere niet de gezochte interactie zouden vertonen.

EXPERIMENT 1

Twee operationalisaties van de onafhankelijke variabele 'complexiteit' in Experiment 1 werden direct ontleend aan het onderzoek van Russell en Ginsburg. Het zijn (a) directe versus indirecte opgaven; in directe opgave is de correcte rekenkundige bewerking in overeenstemming met het in de opgave aanwezige sleutelwoord; bij indirecte opgaven is dit niet het geval; en (b) de aan- of afwezigheid van irrelevante numerieke informatie in de opgaven. Hieraan voegden wij nog een derde operationalisatie toe: (c) de enkel- of meervoudigheid van de opgave; enkelvoudige opgaven kunnen worden opgelost door één rekenkundige bewerking; bij meervoudige moeten twee bewerkingen worden uitgevoerd. Evenals bij opgaven met irrelevante informatie is in meervoudige opgaven een extra gegeven in de opgave aanwezig. Anders dan bij opgaven met irrelevante informatie moet dit extra gegeven in meervoudige opgaven juist wel gebruikt worden. Voorbeelden van deze variabelen worden gegeven in de Materialen sectie.

Proefpersonen

Aan het onderzoek werd deelgenomen door 52 12/13-jarige leerlingen van twee l.o.m.-scholen in Arnhem en in Wormerveer, en 42 vijfde groep leerlingen en 49 zesde groep leerlingen op twee basisscholen in de omgeving van Arnhem. Aan deze kinderen werd een rekentoets voorgelegd bestaande uit 32 opgaven. De opgaven waren ontleend aan de Schiedamse Reken Test (Heesen, Strelitski & Van der Wissel, 1963). De l.o.m.-leerlingen behaalden op deze toets ongeveer even hoge scores als de zesde groep leerlingen ($M = 29.13$, $s = 2.27$ voor de l.o.m.-leerlingen, $M = 29.08$, $s = 2.03$ voor de zesde groepers), terwijl de vijfde groepers aanmerkelijk slechter presteerden ($M = 23.50$, $s = 2.57$). Op grond van deze resultaten werd besloten in het eigenlijke onderzoek de resultaten van de l.o.m.-leerlingen te vergelijken met die van de zesde groepers. De onderzoeksgroep bestaat daardoor uit 52 l.o.m.-leerlingen met een gemiddelde leeftijd van 12.58 jaar en 49 zesde groep leerlingen met een gemiddelde leeftijd van 9.30 jaar. De leeftijdsverspreiding in de l.o.m.-groep ($s = 0.57$) is zoals verwacht mocht worden iets groter dan in de basisschool groep ($s = 0.41$).

Materialen

De experimentele taak bestond uit een verzameling van 88 redactieopgaven. Deze verzameling omvatte 32 optelopgaven, 32 aftrekopgaven, 16 vermenigvuldigopgaven en 8 deelopgaven.

In de optelopgaven zijn drie dichotome operationalisaties van het concept 'informatieverwerkingslast' verwerkt. Deze dichotomieën zijn (a) directe versus indirecte opgaven, (b) opgaven zonder irrelevante informatie versus opgaven met irrelevante informatie, en (c) enkelvoudige versus meervoudige opgaven. Bij elke mogelijke combinatie van waarden op de dichotomieën kunnen opgaven worden geconstrueerd. Enkele voorbeelden illustreren dit. Een directe, enkelvoudige opgave zonder irrelevante informatie is: 'Karel had 11 knikkers. Toen hij naar school liep vond hij er nog 3 bij. Hoeveel knikkers heeft hij samen?' Een indirecte, meervoudige opgave met irrelevante informatie is: 'Anita heeft wat appels. Brigitte heeft 4 appels minder dan Anita. Anita geeft er 3 weg en eet er 7 op. Nu heeft ze er 12. Hoeveel appels had Anita eerst?' Deze opgave is indirect omdat de gebruikte sleutelwoorden ('geeft... weg' en 'eet... op') geassocieerd zijn met aftrekken, terwijl de opgave door optellen moet worden opgelost. De beide andere kenmerken van de opgave behoeven geen nadere toelichting.

Omdat alle combinaties van waarden op de dichotomieën realiseerbaar zijn, kunnen in totaal $2 \times 2 \times 2 = 8$ soorten optelopgaven geconstrueerd worden. Van elk van deze typen werden 4 exemplaren gemaakt, zodat we op een totaal van 32 optelopgaven kwamen. De verzameling aftrekopgaven werd op geheel analoge wijze geconstrueerd. We maakten twee soorten vermenigvuldigopgaven, acht zonder irrelevante informatie en acht met irrelevante informatie. De acht deelsommen tenslotte waren opgaven zonder irrelevante informatie.

Bij de constructie van de opgaven werd de bekende indeling van Riley, Greeno en Heller (1983) als richtlijn gebruikt. Onze opgavenverzameling bevat 28 veranderings-, 13 combinatie- en 23 vergelijkingsopgaven. Deze zijn gelijkelijk over de deelcategorieën verdeeld. De enige uitzondering is de verdeling van de typen over opgaven zonder en met relevante informatie. Bij opgaven zonder relevante informatie is het veranderingstype met 18 opgaven enigszins oververtegenwoordigd en zijn de combinatie- en vergelijkingstypen met resp. 5 en 9 opgaven licht ondervertegenwoordigd.

De opgaven werden in gerandomiseerde volgorde geplaatst zodanig dat tussen twee opgaven van hetzelfde type nooit minder dan zes opgaven van andere typen staan. De opgaven werden in drie testboekjes afgedrukt, die resp. 30, 30 en 28 opgaven bevatten.

Procedure

De opgaven zijn klassikaal aangeboden waarbij één van de auteurs als proefleider optrad. Het onderzoek vond plaats in 3 zittingen. Aan het begin van iedere zitting verzocht de proefleider de leerlingen hun naam, school en klas op het testboekje te schrijven. De leerlingen werd verteld dat zij kladpapier mochten gebruiken. Per zitting was ongeveer drie kwartier beschikbaar.

De antwoorden van de leerlingen zijn dichotoom gescoord. Alle antwoorden op optel- en aftrekopgaven die één hoger of lager zijn dan het juiste antwoord werden beschouwd als rekentechnische fouten en daarom als 'correct' gescoord. De reden hiervoor is dat het ons niet gaat om kennis van getalfeiten of rekenvaardigheid, maar om het vermogen van leerlingen de verbale informatie in een redactieopgave juist te interpreteren en te komen tot de keuze van de correcte rekenoperatie.

Voor de analyse van de uitkomsten werden de scores behaald op de vier opgaven van elk type optel- of aftrekopgave bij elkaar opgeteld, evenals de scores op de acht opgaven van de twee typen vermenigvuldigopgaven en de deelopgaven. Dat levert 8 deelscores voor optelopgaven, 8 deelscores voor aftrekopgaven, 2 deelscores voor vermenigvuldigen, en 1 deelscore voor delen.

De deelscores voor optel- en aftrekopgaven werden geanalyseerd met behulp van variantie-analyse voor een $2 \times 2 \times 2 \times 2$ factorieel design met herhaalde metingen op vier van de vijf factoren. De factor zonder herhaalde meting is leerlingtype met twee niveaus, gewoon lager onderwijs en l.o.m.-onderwijs. De vier factoren met herhaalde metingen zijn optellen versus aftrekken en de drie operationalisaties van 'informatieverwerkingslast'. De deelscores voor vermenigvuldigen werden geanalyseerd met een 2×2 variantie-analyse met herhaalde metingen op één van de factoren. De factor zonder herhaalde meting is wederom leerlingtype, de factor met herhaalde meting is de aanwezigheid versus afwezigheid van irrelevante informatie. De deelscores voor delen werden geanalyseerd met een t-toets voor onafhankelijke steekproeven. De statistische berekeningen werden uitgevoerd met behulp van het programmapakket BMDP.

Resultaten

Tabel 1 vermeldt de resultaten voor de optel- en aftrekopgaven. Onze hoofdhypothese veronderstelt dat het verschil tussen normaal lerenden en l.o.m.-leerlingen groter is op opgaven met een hoge informatieverwerkingslast dan op opgaven met een lage informatieverwerkingslast. Dat betekent dat we in de analyse geïnteresseerd zijn in het vinden van statistische interactie tussen leerlinggroep en de informatieverwerkingslast variabelen. In de variantieanalyse was geen van deze interactietermen significant op 5%-niveau. Dit is het belangrijkste resultaat van Experiment 1, omdat we hiermee geen steun gevonden hebben voor de veronderstelde specifieke zwakte van l.o.m.-leerlingen op redactieopgaven met een hoge informatieverwerkingslast.

Russell en Ginsburg vonden nog een andere interactie: de groep met rekenachterstand lag verder achter op hun leeftijdsgenoten op het gebied van de redactieopgaven dan op de meeste andere rekentaken. Wij hebben onze groepen gematched op gewone rekentaken; als in onze groep hetzelfde prestatiepatroon als bij Russell en Ginsburg bestaat, zou de l.o.m.-groep het dus op de redactieopgaventaak als geheel slechter moeten doen dan de groep normaal lerenden. Er is echter geen hoofdeffect van leerlingtype ($F(1,99) < 1$). De gemiddelde deelscore van de l.o.m.-leerlingen is 3.12 en van de basisschool-leerlingen 3.17. Ook hier ontbreekt dus ieder spoor van een interactiepatroon. Dat is opvallend, omdat Russell en Ginsburg (1984) in hun onderzoek constateerden dat vierdeklassers met rekenachterstand niet slechter presteerden dan normaal lerende derdeklassers op standaard rekentaken, maar wel significant slechter waren op een redactieopgaven test.

Tabel 1. Gemiddelden en standaard deviaties voor zesde groepers van de basisschool en l.o.m.-leerlingen op optel- en aftrekopgaven.

		Optellen			
		Zonder irrelevante info.		Met irrelevante info.	
		Basisschool ^b	L.o.m. ^c	Basisschool	l.o.m
Directe opgaven					
Enkelvoudig	M	3.90	3.85	3.06	2.92
	sd	0.42	0.36	0.72	0.86
Meervoudig	M	3.84	3.87	2.79	2.63
	sd	0.47	0.34	1.12	0.95
Indirecte opgaven					
Enkelvoudig	M	3.75	3.60	2.34	2.17
	sd	0.43	0.60	1.07	1.23
Meervoudig	M	3.24	3.50	2.37	2.63
	sd	0.88	0.80	1.18	1.08

		Aftrekken			
		Zonder irrelevante info.		Met irrelevante info.	
		Basisschool	L.o.m.	Basisschool	l.o.m
Directe opgaven					
Enkelvoudig	M	3.80	3.83	3.24	2.87
	sd	0.54	0.38	0.75	1.05
Meervoudig	M	3.63	3.56	2.61	2.58
	sd	0.53	0.64	1.04	1.14
Indirecte opgaven					
Enkelvoudig	M	3.78	3.69	2.96	3.04
	sd	0.51	0.61	0.86	0.88
Meervoudig	M	3.18	3.04	2.27	2.23
	sd	0.85	0.95	1.19	1.20

^a De maximum score voor elke deeltaak is 4.00^b n = 49 ^c n = 52

We vermelden hier ook kort de resultaten met betrekking tot de experimentele dichotomieën, hoewel deze niet van belang zijn voor de centrale onderzoeksvraag van het experiment. Er was geen significant verschil in moeilijkheid tussen optel- en aftrekopgaven ($F(1,99) < 1$). De drie experimentele dichotomieën leverden wel ieder een significant hoofdeffect op. Directe opgaven zijn makkelijker dan indirecte ($F(1,99) = 54.80$, $p < .01$; $M = 3.31$ voor directe opgaven, $M = 2.99$ voor indirecte). Opgaven zonder irrelevante informatie zijn makkelijker dan opgaven met irrelevante informatie ($F(1,99) = 327.47$ $p < .01$; $M = 3.62$ voor opgaven zonder irrelevante informatie en $M = 2.67$ voor opgaven met irrelevante informatie). Enkelvoudige opgaven zijn makkelijker dan meervoudige opgaven ($F(1,99) = 62.01$, $p < .01$; $M = 3.29$ voor enkelvoudige opgaven en $M = 2.99$ voor meervoudige opgaven).

Bij de analyse van vermenigvuldig- en deelopgaven werden geen verschillen tussen de groepen gevonden en geen andere significante effecten.

Discussie

In dit eerste experiment werd geen enkele bevestiging gevonden voor de interacties die gerapporteerd zijn door Cruickshank (1948) en Russell en Ginsburg (1984). Bij het overzien van de gegevens zoals die in Tabel 1 zijn gepresenteerd valt vooral de grote overeenkomst op tussen het prestatiepatroon van de l.o.m.-leerlingen en dat van de leerlingen van de basisschool.

Op de experimentele taak hebben we te maken met mogelijke plafondeffecten. Op de makkelijke deeltaken (de directe opgaven zonder irrelevante informatie, en de enkelvoudige indirecte opgaven zonder irrelevante informatie) worden deelscores behaald van rond de 90% van de maximum score of hoger. Op deze deeltaken zijn de standaard deviaties ook kleiner (0.34 - 0.64) dan op de overige deeltaken (0.72 - 1.23), hetgeen eveneens wijst op een plafondeffect.

Het is niet geheel uit te sluiten dat dit plafondeffect onze resultaten vertekent. De volgende redenering maakt dit echter wel uiterst onwaarschijnlijk. Beschouwen we voor een ogenblik alleen de gegevens voor opgaven met irrelevante informatie (de rechterhelft van Tabel 1). We hebben dan te maken met een verzameling deelscores zonder plafondeffect (gemiddelden variërend van 2.17 tot 3.24), waarin de effecten van twee van onze operationalisaties van 'informatieverwerkingslast' werkzaam zijn: directheid versus indirectheid van de opgaven, en enkelvoudige versus meervoudige opgaven. Voor deze deelverzameling van onderzoeksuitkomsten leidt de Van Meel hypothese even goed tot een voorspelde interactie als voor de gehele dataverzameling. Inspectie van de data in Tabel 1 leidt tot de conclusie dat ook binnen deze deelverzameling van onderzoeksuitkomsten geen interactiepatroon wordt aangetroffen.

EXPERIMENT 2

De operationalisaties van de onafhankelijke variabele 'complexiteit' bewegen zich in Experiment 2 wat verder van de taken van Russell en Ginsburg dan de operationalisaties van Experiment 1. Ten eerste werden in Experiment 2 naast redactieopgaven ook formulesommen gebruikt (b.v. $4 = . - 3$). In formulesommen is de betekenis van de som verpakt in rekenkundige symbolen, in plaats van in woorden zoals bij de redactiesom. Derhalve zou men kunnen zeggen dat de formulesom behoort tot het domein van de formele rekenkennis, en de redactiesom tot het domein van informele rekenkennis (vgl. Allardice & Ginsburg, 1983). Allardice en Ginsburg geven aan dat kinderen met moeilijkheden bij het leren rekenen vaak wel de onderliggende informele kennis bezitten, maar moeite hebben met het verbinden van deze informele kennis met de formele kennis die in het schoolse rekenen wordt aangeboden. De formulesom vraagt van deze leerlingen dat zij de formele kennis activeren en paraat houden tijdens het oplossen van de opgave. We veronderstellen dat de formulesom daardoor een grotere informatieverwerkingslast vormt dan de redactieopgave. Formulesommen zullen dus ook moeilijker zijn dan redactieopgaven.

Als verdere operationalisatie van informatieverwerkingslast werd in Experiment 2 de uit te voeren handeling gevarieerd. In de traditionele formule- of redactiesom staat één onbekende, die uit de gegeven getallen berekend kan worden. De hierbij uit te voeren handeling hebben we 'opereren' genoemd. Daarnaast hebben we twee andere vormen aan de leerlingen aangeboden, één met twee onbekenden en één met nul onbekenden. In de vorm met twee onbekenden kan geen antwoord berekend worden, maar wel is het mogelijk aan te geven welke van de twee onbekenden het grootste is. We hebben de uit te voeren handeling aangeduid als 'redeneren'. In de vorm met nul onbekenden wordt een antwoord gegeven en moet de leerling aangeven of dit antwoord juist of onjuist is. De uit te voeren handeling hebben we 'evalueren' genoemd. We veronderstellen dat de informatieverwerkingslast groter is naarmate het aantal onbekenden groter is. Dat leidt weer tot de verwachting dat opereer-opgaven moeilijker zullen zijn dan evalueer-opgaven, en redeneer-opgaven weer moeilijker dan opereer-opgaven.

Tenslotte werd binnen iedere deeltaak een onderscheid gemaakt tussen makkelijke en

moeilijke opgaven. Dit onderscheid is deels gebaseerd op de directheid of indirectheid van de opgaven, maar daarnaast ook op de plaats van het '='-teken in de formulesommen.

Proefpersonen

Aan Experiment 2 werd deelgenomen door 112 kinderen. Deze kinderen waren afkomstig van een l.o.m.-school en de vierde, vijfde en zesde groepen van een basisschool in Amsterdam Zuidoost. Aan de deelnemende kinderen werd een voortoets afgenomen bestaande uit 28 formulesommen (gewone optel- en aftreksommen en zgn. stipsommen). Omdat in Experiment 2 jongere kinderen als proefpersonen betrokken werden dan in Experiment 1 kon niet dezelfde toets als voortoets gebruikt worden. De opgaven voor deze toets werden eveneens ontleend aan bestaande rekentoetsen. De resultaten van deze voortoets werden gebruikt voor het selecteren van gematchte groepen voor het onderzoek. Van 25 kinderen van de l.o.m.-school en 57 kinderen van de basisschool waren de verzamelde gegevens volledig. Van de overige kinderen ontbraken de gegevens wegens afwezigheid bij één of meer van de klassikale afnames van de taken.

Het bleek mogelijk twee gematchte groepen samen te stellen van ieder 20 leerlingen, en met een leeftijdsverschil van ruim 16 maanden. Van de l.o.m.-school werden vijf zeer laag presterende leerlingen buiten de analyse gehouden, en van de basisschool waren het vooral boven gemiddeld presterende kinderen die bij de matching uit de boot vielen. De groep basisschool leerlingen had een gemiddelde leeftijd van 100.18 maanden met een standaard deviatie van 10.50 maanden; het gemiddelde van de l.o.m.-groep was 116.45 maanden met een standaard deviatie van 11.6.

Materialen

Uitgangspunt bij de constructie van de opgaven voor dit experiment waren de 14 typen formulesommen van Sandberg en Barnard (1985). Twaalf typen formulesommen ontstaan door in de basisformule (' $a + b = c$ ') de operatie (optellen versus aftrekken), de positie van het '='-teken (rechts versus links) en de plaats van de onbekende (3 mogelijke posities) te variëren. Sandberg & Barnard gebruikten voor de typen ' $. - b = c$ ' en ' $c = . - b$ ' nog twee varianten, één waarin $b > c$ en één waarin $b < c$. De taak formulesommen-opereren bestond uit 14 opgaven, één van elk van de 14 typen. In deze taak, en in alle andere taken, werden getallen onder de 10 gebruikt. Alleen in de evalueertaken werden soms onjuiste antwoorden boven de 10 gegeven.

Binnen de taak formulesommen-opereren is een verdeling gemaakt in makkelijke en moeilijke opgaven gebaseerd op gegevens uit de literatuur (voor details zie van Kampen, 1987). We stelden op grond van deze gegevens dat de moeilijkheid van formulesommen bepaald wordt door de plaats van het '='-teken en door de functie van de onbekende. Opgaven met het '='-teken links zijn moeilijker dan die met het '='-teken rechts. Optelopgaven waarin de opteller bepaald moet worden zijn moeilijker dan opgaven waarin de som gevonden moet worden; en opgaven waarin het opteltal moet worden gevonden zijn weer moeilijker dan die waarin de opteller de onbekende is. Mutatis mutandis geldt hetzelfde voor aftrekopgaven.

Op grond hiervan besloten we de ceasuur tussen makkelijk en moeilijk als volgt te leggen: Makkelijk zijn alle opgaven waarin de som (resp.) het verschil bepaald moet worden, ongeacht de plaats van het '='-teken; en verder de opgaven waar de opteller (resp. aftrekker) moet worden bepaald, maar waarin het '='-teken op de gewone plaats staat. De 6 makkelijke opgaven zijn dan ' $a + b = .$ ', ' $a - b = .$ ', ' $a + . = c$ ', ' $a - . = c$ ', ' $. = a + b$ ' en ' $. = a - b$ '.

De overige 8 opgaven zijn de moeilijke opgaven.

Het onderscheid makkelijk-moeilijk valt bijna samen met het onderscheid direct-

indirect. Van de zes makkelijke opgaven zijn er vier direct en twee indirect; van de acht moeilijke opgaven is er één direct en zijn er zeven indirect.

De taak redactieopgaven-opereren bestond eveneens uit 14 opgaven. Iedere opgave is een vertaling van een formulesom in een redactiesom van het veranderings-type (Riley et al., 1983). Bij de formulesom ' $8 = . + 2$ ' behoort bijvoorbeeld de redactieopgave: 'Carla heeft nu 8 stripboeken. Ze kreeg vanmorgen van moeder 2 stripboeken. Hoeveel stripboeken had ze eerst?' De indeling in makkelijke en moeilijke opgaven in de taak redactieopgaven-opereren loopt parallel met die in de taak formulesommen-opereren.

De taak formulesommen-evalueren bevatte 24 opgaven. De 14 typen formulesommen vormden wederom het uitgangspunt. Alle 14 typen werden eenmaal met een correct antwoord aangeboden. Daarnaast werden 10 typen ook eenmaal met een onjuist antwoord aangeboden. Als onjuiste antwoorden werden die antwoorden gekozen die bij Sandberg en Barnard het meest frequent voorkwamen. Alle gegeven onjuiste antwoorden berustten derhalve op denkfouten. We gebruikten geen onjuiste antwoorden gebaseerd op reken-technische fouten.

Voor de taak redactieopgaven-evalueren werden weer directe vertalingen van de formulesommen in veranderingsopgaven gebruikt. De presentatievorm was hier:

Brian had wat stickers.
Hij geeft er 3 aan Bob.
Hij heeft dan zelf nog 5 stickers over.
Hoeveel stickers had hij eerst?

Hij had eerst 8 stickers.

Bij de redactieopgaven werden dezelfde foute antwoorden gegeven als bij de formulesommen. Dit kan van invloed geweest zijn op de moeilijkheid van deze taak, omdat sommige foute antwoorden die bij redactieopgaven veel voorkomen (met name het geven van één van de gegeven getallen als antwoord) nu in deze taak ontbreken omdat zij bij formulesommen niet of zelden voorkomen.

Het uitgangspunt bij de constructie van de redeneertaken is weer de verzameling typen formulesommen. In redeneeropgaven staan twee onbekenden; aangezien men hiervoor drie verschillende paren gegevens kan nemen (a en b, a en c, of b en c) zijn er hier 12 typen opgaven. De vraag die de kinderen bij redeneer-opgaven moeten beantwoorden is steeds welke van de twee onbekenden de grootste is. Bij enkele van de 12 typen is dit echter niet te bepalen. Twee van deze typen hebben we niet gebruikt (' $. + . = c$ ' en ' $c = . + .$ '). Daardoor komt het aantal opgaven in de redeneertaken op 10. Dit type opgave is geïntroduceerd in de Kwantiwijzer (Van den Berg & Van Eerde, 1983). Op grond van soortgelijke overwegingen als bij de andere taken werden makkelijke en moeilijke redeneeropgaven onderscheiden. De makkelijke zijn ' $a + . = .$ ', ' $. + b = .$ ', ' $. = a + .$ ', ' $. = . + b$ ' en ' $. - b = .$ '. De moeilijke zijn ' $a - . = .$ ', ' $. - . = c$ ', ' $c = . - .$ ', ' $. = a - .$ ' en ' $. = . - b$ '.

Voor de taak formulesommen-redeneren werd steeds de ene onbekende vervangen door een rode vlek, de andere door een blauwe vlek, en de vraag luidde dan 'Welk getal is groter, het getal onder de rode vlek, het getal onder de blauwe vlek, of zijn ze even groot?'

De taak redactieopgaven-redeneren werd geconstrueerd door de 10 opgaven van formulesommen-redeneren te vertalen in veranderingsverhaaltjes. Een voorbeeld is de vertaling van ' $9 - . = .$ ': 'Dennis heeft 9 koekjes. Hij deelt wat koekjes uit. Heeft hij nu meer of minder koekjes? Of heeft hij er nog evenveel?' De vraag in deze opgave is geen directe vertaling van de vraag bij de corresponderende formulesom. Een directe vertaling zou

echter erg verwarrend worden, nl. 'Heeft hij nu minder koekjes over dan hij heeft uitgedeeld, heeft hij meer over dan hij heeft uitgedeeld of heeft hij evenveel uitgedeeld als hij nu over heeft?' We meenden dat antwoorden van kinderen op dergelijke vragen niet informatief zouden zijn. Voor vijf van de tien formulesommen is een soortgelijke alternatieve formulering gekozen, in de andere vijf gevallen was een directe vertaling van de vraag uit de formulesom in de redactieopgave wel mogelijk.

Procedure

De opgaven werden per taak tot een toetsboekje gebundeld. De taken werden klassikaal afgenomen waarbij één van de auteurs als proefleidster fungeerde. Per dag werd in een klas niet meer dan één taak afgenomen. Bij de afnames las de proefleidster de redactieopgaven voor; kinderen konden kiezen of zij de opgaven zelf wilden lezen en in hun eigen tempo werken, dan wel naar de voorgelezen opgaven wilden luisteren en in het tempo van de proefleidster werken.

De gegeven antwoorden werden goed-fout gescored, waarbij alleen het exact juiste antwoord goed gerekend werd. Dit laatste was gedwongen, omdat bij de gebruikte kleine getallen soms niet vermeden kan worden dat een antwoord dat 1 meer of minder is dan het goede antwoord tegelijk ook één van de gegeven getallen is. De toegekende scores per opgave werden per deeltaak opgeteld en vervolgens omgerekend tot een proportie goede antwoorden.

De deeltaak-scores worden geanalyseerd met een $2 \times 2 \times 2 \times 3$ variantie analyse met herhaalde metingen op drie van de vier factoren. De factor zonder herhaalde metingen is leerlingtype (basisschool versus l.o.m.). De factoren met herhaalde metingen zijn opgavetype (formulesom versus redactieopgave), moeilijkheid (makkelijk versus moeilijk) en taakvorm (opereren, evalueren, en redeneren). De berekeningen werden uitgevoerd met behulp van het programmapakket BMDP.

Resultaten

Tabel 2 vat de uitkomsten van Experiment 2 samen.

Uit de resultaten van de variantieanalyse blijkt dat verschillende interacties van opgavenkenmerken met de factor leerlingtype significant zijn. De interactie van moeilijkheid en leerlingtype is significant ($F(1, 38) = 4.84, p < .05$). Uit Tabel 2 is af te lezen dat bij bijna alle deeltaken het verschil tussen basisschool en l.o.m. groter is bij de moeilijke dan bij de makkelijke opgaven.

Vervolgens is per deeltaak nagegaan of binnen die deeltaak de interactie van leerlingtype met moeilijkheid significant was. Voor formulesommen-opereren en formulesommen-evalueren is dit het geval, $F(1, 38) = 4.21, p < .05$, resp. $F(1, 38) = 16.85, p < .01$. Voor redactieopgaven-opereren benadert de interactie statistische significantie, $F(1, 38) = 3.56, p < .0668$.

De interactie taakvorm $\times$ opgavetype $\times$ leerlingtype is significant ($F(2, 76) = 3.16, p < .05$). Op sommige deeltaken is er, ongeacht de moeilijkheid, een groter prestatieverschil tussen basisschool en l.o.m. dan op andere deeltaken. Tabel 3 vat de uitkomsten samen. De grootste verschillen tussen basisschool en l.o.m. komen voor bij formulesommen-opereren, formulesommen-evalueren en redactieopgaven-redeneren.

We hebben voor deze interactie geen interpretatie. Zo'n interpretatie zou moeten verklaren waarom opereren en evalueren juist bij formulesommen en redeneren juist bij redactieopgaven voor l.o.m.-kinderen moeilijk is. Uit onze hypothese valt een dergelijke verklaring niet af te leiden.

Evenals in experiment 1 zijn alle hoofdeffecten van de taakfactoren significant. Bovendien zijn alle onderlinge interacties tussen de taakfactoren significant. Dit betekent

Tabel 2. Gemiddelden en standaard deviaties op de verschillende deeltaken van basisschool leerlingen en l.o.m.-leerlingen in experiment 2.

		Formulesommen		Redactieopgaven	
		Basisschool ^b	L.o.m. ^c	Basisschool	L.o.m.
Opereren					
Makkelijke	M	.908	.825	.808	.817
	sd	.127	.166	.218	.161
Moeilijke	M	.569	.363	.619	.494
	sd	.196	.194	.201	.245
Evalueren					
Makkelijke	M	.792	.788	.817	.771
	sd	.166	.168	.147	.148
Moeilijke	M	.725	.538	.763	.667
	sd	.175	.139	.165	.191
Redeneren					
Makkelijke	M	.740	.670	.820	.730
	sd	.252	.262	.128	.198
Moeilijke	M	.370	.390	.830	.690
	sd	.198	.210	.163	.220

^aDe maximum score voor elke deeltaak is 1.000^bn = 20 ^cn = 20

bijvoorbeeld dat er een prestatieverschil is tussen formulesommen en redactiesommen; dat dit verschil niet even groot is voor de drie taakvormen; en dat het bovendien per deeltaak niet even groot is voor makkelijke als voor moeilijke opgaven. We laten een systematische bespreking van deze effecten hier achterwege omdat zij niet van direct belang zijn voor onze onderzoeksvraag.

Tabel 3. Gemiddelde scores per deeltaak en verschillen tussen de gemiddelden van basisschool leerlingen en l.o.m.-leerlingen

Taakvorm	Basis	Formulesommen		Basis	Redactieopgaven	
		L.o.m.	Vershil		L.o.m.	Vershil
Opereren	.73	.59	.14	.71	.66	.05
Evalueren	.76	.66	.10	.79	.72	.07
Redeneren	.56	.53	.03	.83	.71	.12

Discussie

In Experiment 2 werd het begrip complexiteit geoperationaliseerd door middel van de variabelen opgavetype (formulesom versus redactieopgave), taakvorm en moeilijkheid. Bij de variabelen die in Experiment 2 het begrip complexiteit vertegenwoordigen vonden we twee interacties met leerlingtype. De eerste was de interactie van leerlingtype met moeilijkheid van de opgaven. De analyse wees uit dat deze interactie zich voordeed bij twee van de zes deeltaken van Experiment 2, nl. formulesommen-opereren en -evalueren.

Bij de overige vier deeltaken trad geen interactie op. Wat betreft de taak formulesommen-redeneren kan hierover het volgende worden opgemerkt. De moeilijke opgaven werden maar juist boven kansniveau beantwoord. Deze vragen zijn in wezen driekeuzevra-

gen omdat gekozen kan worden uit de antwoorden 'het getal onder de blauwe vlek is groter', 'het getal onder de rode vlek is groter' en 'ze zijn even groot'. De gevonden gemiddeldes van .370 en .390 liggen maar juist boven het kansniveau van .333. Een dergelijk bodemeffect laat geen ruimte voor het vinden van een interactie.

Twee andere taken waarbij geen interactie optrad, redactieopgaven-evalueren en -redeneren, hebben een eigenschap gemeen: het zijn de taken waarop de moeilijke opgaven relatief het gemakkelijkst zijn. Met name bij redactieopgaven-redeneren is de p-waarde van de 'moeilijke' opgaven nauwelijks lager dan die van de 'makkelijke'. Hetzelfde geldt ook voor redactie-opgaven-evalueren, zij het in mindere mate. Waar het verschil in moeilijkheidsgraad klein is, is het optreden van een significante interactie natuurlijk ook niet waarschijnlijk. Bij de taak redactieopgaven-opereren trad een interactiepatroon op in de gemiddelde scores, dat echter niet significant bleek.

De tweede interactie was die van taakvorm en opgavetype met leerlingtype. Deze interactie konden wij niet zinvol interpreteren.

Onze conclusie luidt daarom dat in Experiment 2 in de formulesommen steun gevonden is voor de specificiteitshypothese op de dimensie makkelijk/moeilijk. In de redactieopgaven en op de andere operationalisaties van informatieverwerkingslast werden geen interacties gevonden.

ALGEMENE DISCUSSIE

In twee deelonderzoeken werd nagegaan of l.o.m.-leerlingen specifieke moeilijkheden ondervinden bij het oplossen van redactieopgaven. Deze vraagstelling werd door ons toegespitst op de vraag of l.o.m.-leerlingen een groter prestatieverschil ten opzichte van leerlingen in het regulier basisonderwijs vertonen bij complexe opgaven dan bij eenvoudige opgaven. In Experiment 1 werd geen steun gevonden voor deze veronderstelling. In Experiment 2 daarentegen werd wel gevonden dat l.o.m.-leerlingen bij moeilijke opgaven meer problemen ondervinden dan leerlingen in het regulier basisonderwijs. De resultaten van de beide experimenten lijken met elkaar in tegenspraak, te meer omdat de variabele moeilijk/makkelijk in Experiment 2 hoog gecorreleerd is met de variabele direct/indirect waarbij in Experiment 1 geen interactie gevonden werd. Het doel van deze discussie is voor deze tegenstrijdigheid een verklaring te suggereren.

Er bestaat een karakteristiek verschil tussen het patroon van onjuiste antwoorden bij indirecte formulesommen en indirecte redactieopgaven. Dit verschil vormt een aanwijzing voor de moeilijkheid die kinderen bij deze soorten opgaven ondervinden. De typische fout bij formulesommen is dat de verkeerde operatie wordt uitgevoerd (Sandberg & Barnard, 1985). Bij een opgave zoals $8 = . - 3$ wordt bijna altijd als fout antwoord het getal '5' gegeven. Bij indirecte redactieopgaven komt dit type antwoord ook voor, maar het is, in ieder geval bij jonge kinderen, minder frequent dan het antwoorden met één van de gegeven getallen (zie bijv. De Corte, Verschaffel, & De Win, 1985). Dit wijst erop dat het operatieteken in indirecte formulesommen een krachtiger invloed heeft op de keuze van de rekenoperatie dan het sleutelwoord in de indirecte redactieopgaven.

We baseren hierop onze interpretatie van de resultaten van het onderzoek. We vermoeden dat het de *dwingendheid* van het misleidende operatieteken in indirecte formulesommen geweest is dat de interacties in Experiment 2 heeft veroorzaakt. Het hier geïntroduceerde begrip dwingendheid kan in deze context worden omschreven als de sterkte van de associatie tussen een symbool (operatieteken of woord) en een rekenkundige operatie.

Hoe verhoudt zich dit begrip dwingendheid tot het begrip complexiteit dat in het onderzoek centraal stond? Is het op te vatten als een deelaspect van complexiteit, zoals

meervoudigheid en de aanwezigheid van irrelevante informatie deelaspecten van complexiteit zijn? Complexiteit hebben we opgevat als 'wat een grote informatieverwerkingslast met zich meebrengt'. De invloed van complexiteit ligt daarmee bij de informatieverwerkingscapaciteit. Zo opgevat impliceerde onze hypothese dat de grenzen van de informatieverwerkingscapaciteit bij l.o.m.-kinderen eerder bereikt worden dan die van normaal lerende kinderen. Bij dwingendheid van misleidende aanwijzingen daarentegen speelt informatieverwerkingscapaciteit niet de hoofdrol. Het is nu veeleer het vermogen van kinderen zelf een structuur op te leggen aan de binnenkomende informatie die een doorslaggevende rol speelt. Alleen wanneer vanuit eigen kennis de binnenkomende informatie geschilderd wordt in wat belangrijk en wat onbelangrijk is, kan een kind weerstand bieden aan de misleidende aanwijzing.

Met deze interpretatie zijn we weer bij Van Meel (1968) terug. Deze baseerde zijn conclusie op onder meer de Stroop test waarin het vermogen om relevant van irrelevant te scheiden ook een belangrijker factor lijkt dan informatieverwerkings-capaciteit.

De resultaten van Ginsburg en Russell (1984) hebben wij niet kunnen repliceren. In Experiment 1 gebruikten we operationalisaties van complexiteit die direct van hun redactieopgaventaak waren afgeleid, maar vonden geen spoor van een interactie. We kunnen hieruit slechts concluderen dat nader onderzoek op dit punt nodig is. Daarbij zal het noodzakelijk zijn zorgvuldiger dan wij in dit onderzoek gedaan hebben te controleren voor de samenstelling van de onderzoeksgroep en de inhoud van het onderwijs. Vooralsnog concluderen wij uit onze eigen gegevens dat de specificiteitshypothese slechts in geringe mate empirische steun heeft gekregen. De specificiteitshypothese kan niet categorisch worden afgewezen, maar evenmin is er sprake van een duidelijke en betrouwbaar te meten cognitieve zwakte bij kinderen met leerachterstanden in het rekenen.

LITERATUUR

- Allardice, B., & Ginsburg, H.P. (1983). Children's psychological difficulties in mathematics. In H.P. Ginsburg (Ed.) *The development of mathematical thinking* (pp. 319-350). New York: Academic Press.
- Badian, N.A. (1983). Dyscalculia and nonverbal disorders of learning. In H.R. Myklebust (Ed.), *Progress in learning disabilities*. Vol. V. (pp. 235-264). New York: Grune & Stratton.
- Berg, W. van den, & Eerde, D. van (1983). *De Kwantiwijzer* (4e experimentele versie). Rotterdam: vakgroep OS en SB, EUR.
- Cruickshank, W. (1948). Arithmetic ability of mentally retarded children: Ability to differentiate extraneous material from needed arithmetic fact. *Journal of Educational Research*, 42, 161-170.
- De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, 77, 460-470.
- Ginsburg, H.P. (1972). *The myth of the deprived child*. Englewood Cliffs, NJ: Prentice-Hall.
- Heesen, H., Strelitski, D., & Van der Wissel, A. (1963). *Schiedamse Rekentest*. Groningen: Wolters-Noordhoff.
- Kampen, C.L.A. van (1987). *Zijn stipsommen en redactiesommen analoge somvormen? Een vergelijkend onderzoek bij kinderen met en kinderen zonder leermoeilijkheden*. Doctoraalwerkstuk Faculteit Psychologie. Universiteit van Amsterdam.
- Meel, J.M. van (1968). *Bedreigd denken*. Groningen: Wolters-Noordhoff.
- Riley, S., Greeno, J.G., & Heller, J.I. (1983). Development of problem solving abilities in arithmetic. In H. Ginsburg (Ed.). *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.

- Rourke, B.P., Bakker, D.J., Fisk, J.L., & Strang, J.D. (1983). *Child neuropsychology: An introduction to theory, research and clinical practice*. New York: Guilford.
- Rourke, B.P., & Strang, J.D. (1984). Subtypes of reading and arithmetical disabilities: A neuropsychological analysis. In M. Rutter (Ed.), *Developmental Neuropsychiatry* (pp. 473-488). Edinburgh: Churchill Livingstone.
- Russell, R.L., & Ginsburg, H.P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition and Instruction, 1*, 217-244.
- Sandberg, J.A.C., & Barnard, Y.F. (1985). Probleemoplossen binnen het aanvankelijk rekenen: een analyse van oplossingsprocessen bij open beweringen. In S. Dijkstra & P. Span (red.) *Leerprocessen en instructie*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 22-8-1988

Definitieve versie ontvangen 2-2-1989.

Leren optellen en aftrekken: een longitudinaal onderzoek

J.J. Beishuizen*, M. Beishuizen** en E.J.H.M. Felix*

ABSTRACT

Based on the assumption that (1) addition and subtraction are rule-based cognitive skills, and (2) that the acquisition of cognitive skills is a result of a stage-like development process, a 'genetic model' was proposed in which emergence of addition and subtraction skills is described in six stages. The stages were chosen on theoretical and empirical grounds. In a longitudinal experiment 16 second grade primary school students were followed during 11 months. Data on their addition and subtraction performance were gathered with a diagnostic computer program. The students appeared to work consistently according to the strategies available at the level of expertise. The proposed genetic model accounted for most of the performance levels displayed by the students. The conclusion seems warranted that instruction should be adapted to the actual level of expertise of the student. By carefully providing appropriate problems students can be brought to recognize the shortcomings of their strategies. This insight may stimulate the students to restructure their procedural and declarative knowledge on a higher level of expertise.

INLEIDING

Het uitvoeren van optel- en aftrekoperaties in het domein van de natuurlijke getallen kan worden beschouwd als een regelgeleide cognitieve handeling, waarbij zowel declaratieve als procedurele kennis een rol speelt. Diverse auteurs (Anderson, 1982; Siegler & Klahr, 1982; Case, 1982; Case & Bereiter, 1984) stellen dat de verwerving van cognitieve vaardigheden verloopt via stadia, waarbinnen relatief stabiele complexen van regels worden toegepast. De ontwikkeling wordt mogelijk gemaakt door verbetering of uitbreiding van de interne representatie van probleem en doel. Als deze stellingen juist zijn, dan dient de docent te weten welke complexen van regels op de diverse niveaus van expertise voorkomen en hoe de overgang van het ene niveau naar het andere plaatsvindt. Goldstein (1982) heeft op basis van deze uitgangspunten zijn 'genetisch model' van de wording van een expert in het Hunt-the-Wumpus spel ontworpen. Dit genetisch model beschrijft de opbouw, bewerking en uitbreiding van kennis en procedures als resultaat van spelervaring. Siegler en Klahr (1982) toonden in een uitvoerig onderzoek naar de balanstak van Inhelder en Piaget (1958) aan dat proefpersonen (ppn.) in een beginstadium van het leerproces nog niet alle kenmerken van de probleemsituatie encoderen. Training was in hun experimentele situatie effectief wanneer eerst encoding van een ontbrekend kenmerk van de opgave werd bewerkstelligd en vervolgens de overgang naar

* Vrije Universiteit Amsterdam, Vakgroep Psychonomie

**Rijksuniversiteit Leiden, Vakgroep Onderwijsstudies

Dit onderzoek was mogelijk dankzij een subsidie van het Instituut voor Onderzoek van het Onderwijs, projectnummer 6618.

Adres: J.J. Beishuizen, Vrije Universiteit, Vakgroep Psychonomie, De Boelelaan 1111, 1081 HV Amsterdam.

een volgend expertiseniveau werd ingezet. Case en Bereiter (1984) somden op basis van de geschetste inzichten een aantal eigenschappen op waaraan een leeromgeving moet voldoen: (1) De leeromgeving moet de pp. laten zien dat zijn of haar huidige expertiseniveau nog niet toereikend is voor alle typen opgaven binnen het domein. (2) De leeromgeving moet de relevante aspecten van de probleemsituatie, die de pp. nog niet encodeert, saillant maken. (3) De belasting van het werkgeheugen moet bij elke stap minimaal blijven. (4) Er moet voldoende oefening geboden worden om de pp. de gelegenheid te geven de te leren regel te automatiseren.

Ook binnen het cognitief-psychologisch onderzoek naar leren rekenen worden ontwikkelingsstadia onderscheiden, waarin kinderen oplossingsstrategieën op een hoger niveau gaan hanteren (Carpenter, 1985). Aanvankelijk worden informele (tel)strategieën gebruikt, die zich later verkorten en ook aanpassen aan verschillende probleemttypen. Nog weer later verdwijnt bij veel kinderen deze variatie en gaat het mechanisch toepassen van één strategie overheersen, ongeacht het probleemttype (Carpenter, 1985; De Corte en Verschaffel, 1983). Het genoemde onderzoek beperkt zich tot nu toe tot optellen en aftrekken met kleine getallen tussen 0 en 20, veelal met probleemopgaven ingekleed als redactiesommen.

Oplossingsprocedures bij optellen en aftrekken

In de inleiding tot dit themanummer is ingegaan op de procesmodellen die in de literatuur over het leren optellen en aftrekken worden onderscheiden. M. Beishuizen (1986) onderzocht bij Nederlandse ppn. uit groep 3 en 4 van de basisschool de manier waarop uit het hoofd gerekend wordt. Hij kwam daarbij tot een zestal rekenprocedures, die worden geschetst in Tabel 1. De strategieën werden geregistreerd bij leerlingen die in de fase van verwerving van optel- en aftrekvaardigheden tussen 20 en 100 verkeerden (groep vier basisschool). De 'handige' procedures zoals omkering van de termen bij optellen of het oplossen van de som via een rond tiental kwamen bij deze leerlingen nog niet voor.

Het hoofdonderscheid in bovengenoemde strategieën betreft het verschil tussen de 1010-strategie en de G10-strategie (zie Tabel 1 voor voorbeelden). De G10-strategie wordt door veel rekendidactici aanbevolen en door veel rekenmethoden voorgeschreven als een effectieve en efficiënte werkwijze (M. Beishuizen & Van Mulken, 1988). De G10-strategie werkt als een algoritme bij elk somtype op dezelfde manier. Voorwaarde is echter uitbreiding van (declaratieve) kennis van de telrij tot 100, niet alleen met 'mooie' tientallen (10, 20, 30, 40, enz. en 60, 50, 40, 30, enz.), maar ook met 'samengestelde' G10-sprongen (14, 24, 34, 44, enz. en 66, 56, 46, 36, enz.). Dit laatste geeft veel meer aanleer-moeilijkheden (Fuson, Richards & Briars, 1982). In de onderwijspraktijk blijken veel ppn. de 1010-strategie te prefereren, hoewel deze strategie bij sommen met tientalpassering tot problemen leidt (zie Tabel 1 voor een bekend voorbeeld). Waarschijnlijk verkiezen leerlingen de 1010-strategie allereerst omdat de 1010-strategie in de aanleerfase gemakkelijker lijkt door de sterke overeenkomst met het opereren op kleine getallen (vergelijk Cijferend Rekenen in Tabel 1). Maar ook omdat de bovengenoemde uitbreiding van de telrij met G10-sprongen hier niet nodig is. De 1010-strategie wordt uitgevoerd door eerst op basis van kennis van tientallen en eenheden de termen in de opgave te splitsen en vervolgens de in het domein van 0 tot 20 geleerde optel- en aftrekprocedures toe te passen. De 1010-strategie vormt in feite een toepassing van de heuristische methode van oplossen door analogie waarbij de karakteristieken van de opgave (optellen of aftrekken, wel of geen tientalpassering, kleine of grote getallen) een veel grotere invloed op het oplossingsproces hebben dan bij de G10-strategie. Terwijl de G10-strategie specifiek toepasbaar is op optel- en aftreksommen vinden we een variant van de 1010-strategie ook terug bij het vermenigvuldigen.

Tabel 1. Door M. Beishuizen (1986) onderscheiden oplossingsprocedures bij optel- en aftreksommen in het getallengebied tussen 20 en 100.

T =	Tellen: $24 + 12 = \dots$ "24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, <u>36</u> "
S =	Structureren: $24 + 12 = \dots$ " $24 + 6 = 30$, $30 + 6 = \underline{36}$ "
cfr =	Cijferend rekenen (mechanische, te snelle verkorting van 1010): $24 + 12 = \dots$ " $2 + 1 = 3$, $4 + = 6$, dus: <u>36</u> "
1010 =	getallen splitsen, Tientallen bij/van elkaar, daarna eenheden: $46 + 23 = \dots$ " $40 + 20 = 60$, $6 + 3 = 9$, samen $60 + 9 = \underline{69}$ " $42 - 15 = \dots$ " $40 - 10 = 30$, $2 - 5 = ?$ " (dikwijls fout: " $5 - 2 = 3$, samen $30 + 3 = \underline{33}$) (goed is: $30 > 20$, $12 - 5 = 7$, samen $20 + 7 = \underline{27}$ of 10t-aanpassing, zie onder)
10t =	gelijk aan 1010, maar eenheden met tussenstap toevoegen: $46 + 23 = \dots$ " $40 + 20 = 60$, $60 + 6 = 66$, $66 + 3 = \underline{69}$ " $42 - 15 = \dots$ " $40 - 10 = 30$, $30 + 2 = 32$, $32 - 5 = \underline{27}$ "
G10 =	eerste Getal niet splitsen, tientallen erbij/eraf, daarna eenheden: $46 + 23 = \dots$ " $46 + 20 = 66$, $66 + 3 = \underline{69}$ " $42 - 15 = \dots$ " $42 - 10 = 32$, $32 - 5 = \underline{27}$ "

In het onderzoek van M. Beishuizen (1986) werden de oplossingsprocedures geregistreerd via navragen. In ons onderzoek werd gebruik gemaakt van een computerprogramma dat op grond van de stapjes die een pp. intypte vaststelde volgens welke strategie de pp. te werk ging. Onder 'Materiaal' is een toelichting op de werking van het programma te vinden.

Een genetisch model voor het leren optellen en aftrekken

Zoals in de Inleiding gesteld dient een docent, die uitgaat van een stadia-gewijze ontwikkeling van cognitieve vaardigheden, inzicht te hebben in de regels die ppn. achtereenvolgens leren en de volgorde waarin de regels worden geleerd en toegepast. Hierboven werd voor het domein van het leren optellen en aftrekken in het getallengebied 20 tot 100 een zestal rekenstrategieën geschetst. De vraag is nu in welke volgorde deze regels worden geleerd en toegepast en hoe de overgang tussen de regels plaatsvindt. Met andere woorden: hoe ziet het genetisch model voor het leren optellen en aftrekken eruit?

Op grond van a priori overwegingen kan een scheiding worden gemaakt tussen Tellen en Structureren enerzijds en de Cijferend Rekenen-, 1010-, 10t- en G10-strategieën anderzijds. Tellen als strategie en Structureren als gebruik maken van 'number facts' worden algemeen beschouwd als de basis voor het kunnen optellen en aftrekken (Groen & Parkman, 1972; zie ook het inleidend artikel tot dit themanummer), waarop de overige strategieën (combinaties van declaratieve en procedurele kennis) voortbouwen.

Cijferend Rekenen wordt opgevat als een 1010-strategie, waarbij een te snelle verkorting heeft plaatsgevonden (Van Parreren, 1975). De vorm van Cijferend Rekenen, waarop hier wordt gedoeld, wordt waargenomen in groep 4 van de basisschool, ná de introductie van G10 of 1010 en vóór het moment waarop voor het eerst met pen en papier onder elkaar wordt gerekend. Het gaat hier dus niet om de door veel volwassenen gepraktiseerde generalisatie van schriftelijk rekenen naar hoofdrekenen, maar om een afgeleide van de 1010-strategie. Voor het typeren van deze procedure als verkorting van de 1010-strategie gelden de volgende argumenten.

Bij het uitvoeren van de strategie wordt de kardinale betekenis van de cijfers in een getal

niet benut, en reduceert de rekenaar de opgave tot het combineren van twee basiscombinaties (zie Tabel 1). In de opgave $24 + 13 = \dots$ worden de 2 van 24 en de 1 van 12 niet opgevat als '20', resp. '10' (de kardinale betekenis van beide cijfers) maar als '2', resp. '1'. Deze aanpak leidt bij opgaven zonder tientalpassering tot snelle oplossingen. Bij opgaven met tientalpassering ontstaat een onoverkomelijke impasse (zie in Tabel 1 het voorbeeld ' $42 - 15 = \dots$ '). Inzicht in de kardinale betekenis van de cijfers in een getal treedt op de voorgrond in de volwaardige 1010-strategie (M. Beishuizen & Van Mulken, 1988). In de ontwikkelingslijn wordt Cijferend Rekenen aan het eind van een doodlopende zijweg vanuit 1010 geplaatst als een ongewenste, vroegtijdige verkorting van de 1010-strategie, waarop de rekenaar gemakkelijk gefixeerd raakt.

Voor de bepaling van een hiërarchie van de 1010-, 10t- en G10-strategie gelden de volgende argumenten. De 1010-strategie heeft een geringere reikwijdte dan de G10-strategie. Aftreksommen met tientalpassering vormen een probleem voor ppn. die de 1010-strategie volgen. Hoewel ppn. aanvankelijk dikwijls de 1010-strategie prefereren, zullen ze uiteindelijk met de met inzicht verworven G10-strategie meer succes hebben. We zien dan ook dat de G10-strategie verder wordt geautomatiseerd en sneller wordt uitgevoerd naarmate de ppn. ouder worden. De 10t-strategie neemt een tussenpositie in. M. Beishuizen & Van Mulken (1988) schrijven hierover: De in de G10-strategie benodigde 10-sprongen (46, 56, 66, etc., zie Tabel 1) 'vormen nog onbekend terrein. Deze moeten in het begin echt worden 'geconstrueerd' op begripsniveau via analogie. Of zoals wij dikwijls in de beginfase observeerden via de 1010-procedure met tussenstap of '10t': $25 + 12 =$ via $20 + 10 = 30$, $30 + 5 = 35$ en $35 + 2 = 37$ (waarbij de ppn. dikwijls verbaliseren, dat zij de eenheden van het eerste getal 'even opzij' doen en daarna onmiddellijk 'weer erbij').' (p. 34). De 10t-strategie kan dus worden opgevat als een opstap tot de G10-strategie. Dit hoeft echter niet te gebeuren. De 10t-strategie kan ook een eindstadium vormen, als een vervolg op de 1010-strategie. De impasse, waarin ppn. met de 1010-strategie terecht komen bij aftreksommen met tientalpassering, kan worden doorbroken door na de afhandeling van de tientallen eerst de eenheden van het aftrektal bij het tussenresultaat op te tellen en vervolgens de eenheden van de aftrekker van dit tussenresultaat af te halen (zie Tabel 1 voor een voorbeeld). Hiermee stappen ppn. met een 1010-achtergrond geleidelijk over naar de 10t- of G10-strategie, waarin het resultaat van een deeloperatie uitgangspunt vormt voor de volgende bewerking.

Een en ander leidt tot de postulering van de volgende algemene ontwikkelingslijn voor het optellen en aftrekken:

Tellen - Structureren - 1010 - 10t - G10.

| |
Cijferend Rekenen

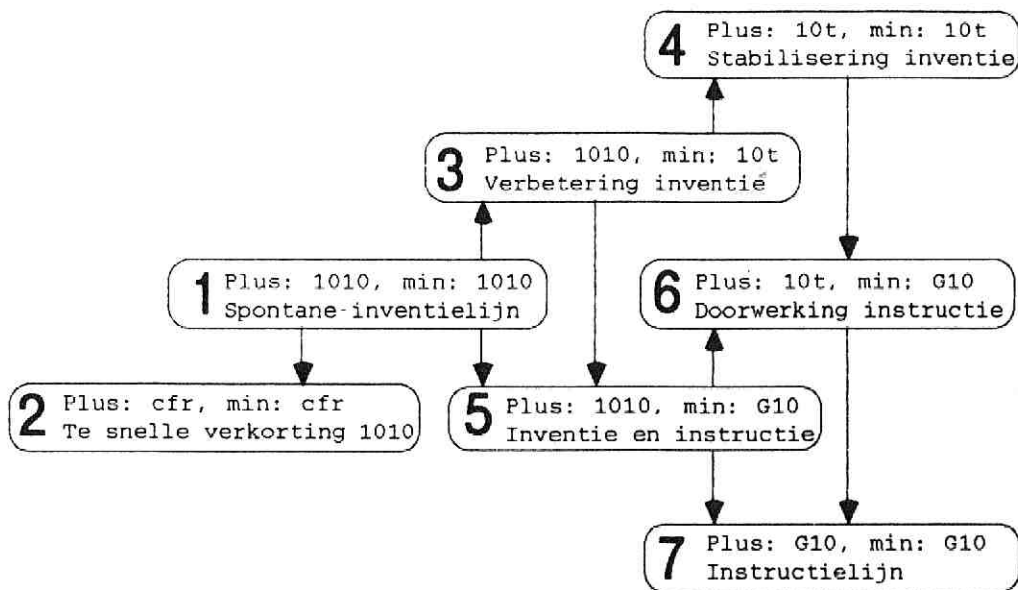
Deze algemene lijn schetst de volgorde waarin rekenstrategieën verondersteld worden voor te komen bij ppn. die expertise opbouwen bij het optellen en aftrekken in het getallengebied 20-100. Wanneer een pp. verandert van strategie in de richting van G10 spreken we in het kader van deze lijn van vooruitgang. Verandert de pp. in de richting van Tellen dan spreken we van terugval. Terugval kan zich voordoen wanneer een strategie nog niet is gestabiliseerd in een 'sterk schema' (Anderson, 1984). Strategieën, die nog vrij intuïtief en onsystematisch worden toegepast (Baroody & Ginsburg, 1986), zullen gemakkelijk (tijdelijk) worden opgegeven, wanneer de opgaven moeilijker worden.

Bovenstaande algemene ontwikkelingslijn moet nader worden uitgewerkt. Het is niet zo dat leerlingen in de geschetste volgorde de strategieën op alle somtypen toepassen. Er is steeds sprake van een deelverzameling van strategieën die bij bepaalde somtypen gevolgd

worden. De keuze van strategieën wordt zowel beïnvloed door het onderwijs dat de leerling ontvangt als door spontane inventies van de leerling zelf. Omdat beide invloeden in sterkte kunnen variëren kunnen verschillende ontwikkelingslijnen ontstaan. Tabel 2 bevat een voorstel voor een genetisch model waarin deze ontwikkelingslijnen zijn opgenomen. De volgorde waarin de zeven onderscheiden stadia kunnen worden doorlopen is niet lineair! Figuur 1 schetst de mogelijke overgangen tussen de stadia. De zeven stadia kunnen alle een eindstadium in de ontwikkeling zijn. Met uitzondering van stadium 2 en stadium 7 veronderstellen we dat de stadia ook als tussenstadium kunnen fungeren in het genetisch model.

Tabel 2. Zeven stadia binnen het algemene genetische model.

Stadium	Optellen	Aftrekken	Toelichting
1	1010	1010	Spontane inventielijn
2	cfr	cfr	Te snelle verkorting van 1010
3	1010	10t	Verbetering van spontane inventie
4	10t	10t	Stabilisering inventie
5	1010	G10	Spontane inventie en instructie
6	10t	G10	Doorwerking instructie
7	G10	G10	Instructielijn



Figuur 1. Genetisch model voor het leren optellen en aftrekken in het getallengebied 20-100. De cijfers verwijzen naar de in Tabel 2 beschreven stadia.

Tenslotte is het ook mogelijk om elk van de zes stadia als een beginstadium te zien. Een leerling hoeft bijv. niet via de voorliggende stadia in stadium 7 aan te komen. Instructie kan ook een rechtstreekse plaatsing in stadium 7 bewerkstelligen. De status van dit model is hypothetisch. Het lopende onderzoek moet uitwijzen of het model empirisch ondersteund kan worden, door na te gaan of de veronderstelde combinaties van strategieën bij optel- en aftreksommen in de praktijk voorkomen. De keuze van de zeven stadia is overigens hoofdzakelijk gebaseerd op empirisch onderzoek dat in Leiden heeft plaatsgevonden (M. Beishuizen, 1986). Het model kan als volgt worden toegelicht.

Vooraf wordt gesteld dat we aannemen dat de 1010-strategie in het algemeen het

resultaat is van spontane inventie en de G10-strategie voortkomt uit al of niet succesvolle instructie. Deze aanname is gebaseerd op de overwegende G10-instructie in Nederlandse rekenmethoden, terwijl de 1010-strategie weinig instructie-aandacht krijgt. Omdat de 1010-strategie veel voorkomt onder Nederlandse leerlingen (M. Beishuizen, 1986) moeten we aannemen dat eigen vindingrijkheid hieraan, in weerwil van de vigerende G10-instructie, ten grondslag ligt. Stadium 1 (1010 bij optel- en aftreksommen) kan voorkomen wanneer een pp. de G10-instructie negeert en zijn of haar eigen 1010-strategie consistent toepast, met de reeds vermelde problemen bij aftreksommen met tientalpassering. Stadium 2 beschrijft de toestand die kan ontstaan bij te snelle verkorting van de 1010-strategie bij optellen en aftrekken tot Cijferend Rekenen. Stadium 3 schetst de toestand van ppn. die hun spontane 1010-strategie hebben verbeterd, zodat de aftreksommen met tientalpassering goed afgehandeld kunnen worden. Stadium 4 laat de doorwerking van de succesvol gebleken 10t-strategie van de aftreksommen naar de optelsommen zien. Stadium 5 geeft weer dat sommige ppn. hun eigen 1010-strategie zullen blijven volgen bij optelsommen en voor aftreksommen de aangeleerde G10-strategie toepassen. Stadium 6 toont dat bij sommige ppn. de G10-strategie, die wordt toegepast bij aftreksommen, invloed heeft op de 1010-strategie, die bij optelsommen wordt gehanteerd. Onder invloed van de G10-strategie wordt de 1010-strategie 'opgewerkt' tot 10t-strategie. Stadium 7 (G10 bij optellen en aftrekken) wordt verwacht wanneer de G10-instructie effectief is geweest.

In de bovenstaande presentatie van het genetisch model en de onderscheiden stadia is nog in het midden gelaten op welke manier de overgangen tussen de niveaus plaatsvinden, met andere woorden: Hoe ontdekt de leerling dat een bepaalde strategie ontoereikend is en hoe wordt de stap naar een volgend niveau gezet? Een compleet beeld van de verantwoordelijke leermechanismen is nog niet te geven. Twee 'change agents' worden in ieder geval verondersteld. Het eerste leermechanisme betreft de overgang van de 1010-strategie naar een hoger niveau. De aftreksommen met tientalpassering plaatsen de leerling, die de 1010-strategie volgt, voor een impasse. Verwacht wordt dat dit somtype aanleiding zal geven tot een herstructurering van de gevolgde rekenstrategie waaruit dan vervolgens de 10t-strategie voortkomt (spontaan ontdekt of geïnduceerd uit de G10-strategie) of de G10-strategie (rechtstreeks overgenomen van de docent).

Als tweede leermechanisme in het genetisch model wordt generalisatie genoemd van een succesvolle strategie bij de moeilijker aftreksommen naar de relatief eenvoudige optelsommen. Dit mechanisme bewerkstelligt de overgang van stadium 3 naar 4, van stadium 5 naar 7 en van stadium 6 naar 7.

De inhoud van het postuleerde genetische model stoelt op de in Nederland bestaande onderwijspraktijk op het terrein van het leren optellen en aftrekken. Hoewel de inhoud van het model derhalve een domeinspecifiek karakter heeft, is de onderliggende hypothese van de stadia-gewijze ontwikkeling en de daarbij vigerende leermechanismen van algemeen cognitief-psychologische aard.

Vraagstelling longitudinaal onderzoek

Ter toetsing van het geschetste genetisch model werd in een longitudinaal onderzoek nagegaan of ppn. consistent rekenstrategieën hanteerden volgens de onderscheiden stadia, en of de ontwikkeling van de rekenstrategieën van de ppn. in overeenstemming was met het genetisch model.

De mate van consistentie bij het toepassen van rekenstrategieën werd getoetst door binnen elk van zes aangeboden somtypen (optelsommen zonder tientalpassering, met een uitkomst op een tiental, met tientalpassering; aftreksommen zonder tientalpassering, met

een uitkomst op een tiental, met tientalpassering) na te gaan of kleine sommen (bij optelsommen: beide termen tussen 10 en 20; bij aftreksommen: aftrektal tussen 20 en 30; aftrekker tussen 10 en 20) op dezelfde manier werden opgelost als grote sommen (beide termen tussen 20 en 40).

De validiteit van het genetisch model werd op twee manieren onderzocht. In de eerste plaats werd nagegaan in welke mate de onderscheiden stadia teruggevonden konden worden in de profielen die door het diagnostisch programma werden vastgesteld. Voorts werd van de waargenomen *stadia-overgangen* nagegaan in hoeverre ze overeen kwamen met de in het genetisch model opgenomen ontwikkelingslijnen, waarbij vooral na de presentatie van minssommen met tientalpassering stadia-overgangen verwacht werden.

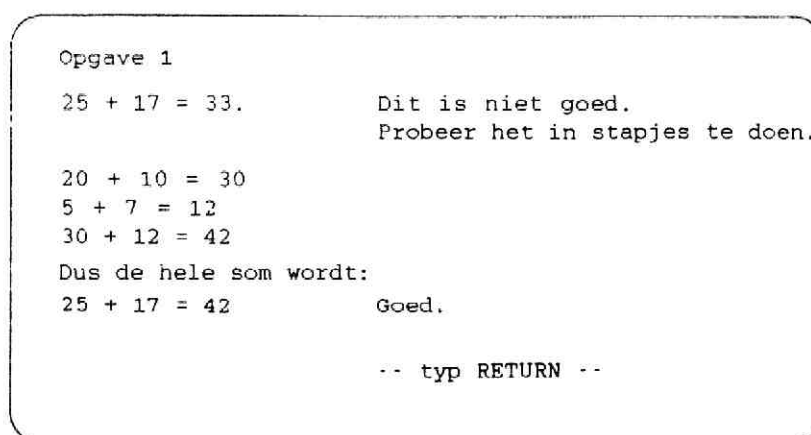
METHODE

Proefpersonen

Voor de selectie van proefpersonen (ppn.) werd in de vierde groep van vier basisscholen in Amsterdam een collectieve schriftelijke voortoets afgenomen. Met de vier proefscholen wordt vanuit de vakgroep Psychonomie al lange tijd samengewerkt. Op elk van de vier proefscholen wordt een traditionele rekenmethode toegepast. De voortoets bevatte per somtype vijf opgaven. De opgaven werden in willekeurige volgorde aangeboden. De ppn. kregen opdracht om in 20 minuten zoveel mogelijk van de 30 sommen foutloos te maken. Op elk van de vier scholen werden de ppn. uitgekozen die een hoge score haalden op de voortoets. Tijdens de eerste zitting met het diagnostisch programma werden de ppn. geïdentificeerd die gebruik maakten van telstrategieën. Deze ppn. werden niet verder in het longitudinale onderzoek betrokken. In de loop van het onderzoek vielen vier van de resterende 20 ppn. om uiteenlopende redenen af, zodat van 16 ppn. een volledig protocol kon worden vastgelegd.

Materiaal

De ppn. werkten individueel aan een computerprogramma. Het programma voert een eenvoudige dialoog met de leerling. Na de presentatie van een opgave en nadat de leerling een antwoord heeft gegeven stelt het programma de vraag: 'Geef aan hoe je het antwoord gevonden hebt door stapjes in te typen.'. Deze vraag wordt bij een fout antwoord altijd gesteld en bij een goed antwoord in twee van de drie gevallen. Deze beperking in het navragen bij goed beantwoorde opgaven is ingevoerd om te voorkomen dat de leerlingen alle aangeboden opgaven uitvoerig moeten afhandelen, waardoor het aantal te maken sommen terugloopt en ook de motivatie negatief beïnvloed wordt. Na een vraag om stapjes typt de leerling de gevolgde stapjes in. Figuur 2 geeft een karakteristiek schermbeeld uit het programma. Op basis van de ingetypte stapjes bepaalt het programma (1) welke strategie de leerling heeft gerapporteerd, en (2) welke fouten zijn gemaakt bij het uitvoeren van de strategie. Het programma gaat bij het stellen van de diagnose uit van een lijst van strategieën (Tabel 1 met varianten) en een lijst van foutsoorten. Het programma gaat voor zes somtypen (optellen en aftrekken; uitkomst onder tiental, op tiental, over tiental heen) de gevolgde strategie met daarbij optredende fouten na bij het oplossen van sommen met kleine getallen en sommen met grote getallen. Per somtype (in volgorde van gemakkelijk naar moeilijk) worden zes opgaven aangeboden, drie opgaven met kleine getallen en drie opgaven met grote getallen. De leerling ontvangt tijdens het intypen van de stapjes feedback bij syntactisch onjuist opgegeven tussenstapjes (bijvoorbeeld het veel voorkomende ' $12+10=22+3=25$ '), die onmiddellijk verbeterd moeten worden. Over de juistheid



Figuur 2. Karakteristiek schermbeeld diagnostisch computerprogramma.

van het antwoord op een opgave wordt uitsluitend goed/fout feedback gegeven. In eerder onderzoek (Mannaerts, 1987) is vastgesteld dat leerlingen het programma goed leren bedienen. Bovendien is het programma goed in staat om de gevolgde strategieën te identificeren.

Procedure

Tussen februari en oktober 1987 werkten de 16 ppn. gedurende zeven sessies, steeds met tussenpozen van vier weken, aan het rekenprogramma. Voordat de ppn. met het programma aan de slag gingen werden ze tijdens een oefensessie vertrouwd gemaakt met de bediening van de computer en het rekenprogramma. De ppn. werkten tijdens de zeven sessies onder toezicht van een proefleider.

Tijdens de eerste zes sessies werd het aanbod van de sommen voortdurend afgestemd op het niveau van elke afzonderlijke leerling. Per sessie is daardoor bij enkele leerlingen niet voor elk van de zes somtypen de gevolgde strategie vastgesteld. De data van de eerste drie sessies en van de volgende drie sessies zijn ten behoeve van de analyse samengevoegd, zodat per twee maanden een compleet profiel van het gebruik van strategieën over de zes somtypen kon worden samengesteld. Eventuele strategieveranderingen binnen elk van de twee perioden van twee maanden zijn hierdoor noodgedwongen genegeerd. In de profielen is per somtype voor de kleine en de grote sommen de meest gebruikte strategie in de analyse betrokken. Kwamen twee strategieën even vaak voor, dan werd de hoogste in de hiërarchie van de algemene ontwikkelingslijn voor nadere analyse gekozen. Tabel 3 geeft een voorbeeld van een profiel.

Tabel 3. Voorbeeld van een profiel. Waargenomen strategieën bij zes somtypen (kleine en grote sommen). Voor elk van de 16 ppn. zijn drie van deze profielen bepaald.

In totaal is dus $16 \times 3 \times 6 \times 2 = 576$ maal een strategie vastgesteld.

	Optellen			Aftrekken		
	Zonder ttp*	op tiental	met ttp	zonder ttp	op tiental	met ttp
Kleine sommen	1010	1010	1010	10t	10t	10t
Grote sommen	1010	1010	1010	10t	10t	10t

* ttp betekent: tientalpassering.

RESULTATEN

De analyse van de data op consistentie bij het toepassen van de strategieën, aanwezigheid van de onderscheiden stadia en de validiteit van het genetisch model werd zonder significantietoetsing uitgevoerd. De omvang van de dataset was hiertoe niet toereikend. In het vervolg van het onderzoek is een uitgebreide replicatie van het experiment met een groter aantal ppn. en intensievere meting van het prestatieniveau voorzien.

Consistentie bij het toepassen van strategieën

Voor elk van de 16 ppn. zijn drie profielen beschikbaar waarin voor elk somtype de gevolgde strategie bij kleine en grote sommen is vastgelegd. Om de consistentie bij het toepassen van strategieën na te gaan werd *binnen elk somtype* het oplossingsgedrag bij de kleine en grote sommen met elkaar vergeleken. Soms werd een somtype niet aangeboden, omdat de pp. eerder nog niet tot het maken van opgaven van het betreffende somtype in staat bleek. Soms werden geen stapjes nagevraagd, omdat de som juist was beantwoord (zie boven). Van de 288 (16 ppn. maal zes somtypen maal drie profielen) metingen bleven 75 voor analyse over. In deze gevallen werd 64 keer (85%) bij de kleine en grote sommen binnen een bepaald somtype telkens dezelfde strategie gevolgd. In 11 gevallen was er sprake van strategie-verandering, meestal een achteruitgang (8 keer). We kunnen dus concluderen dat de ppn. vrij consistent in het gebruik van strategieën waren. Bij de overgang naar grote sommen binnen een somtype werd soms een 'lagere' strategie gekozen, wellicht omdat de pp. een zekere voorzichtigheid prefereerde bij het aanpakken van de grotere som.

De aanwezigheid van de onderscheiden stadia

Van de 48 geregistreerde profielen (drie voor elk van de 16 ppn.) leenden zich 34 voor analyse op het voorkomen van de in Tabel 2 onderscheiden stadia. De overige 14 profielen vielen af omdat er ondanks de zorgvuldige selectie van de ppn. toch teveel telstrategieën in voorkwamen of omdat door het grote aantal juiste antwoorden te weinig nagevraagd is naar de gevolgde strategie. De aanwezigheid van de onderscheiden stadia werd getoetst door de toegepaste strategieën bij de optelsommen en aftreksommen met tientalpassering met elkaar te vergelijken. Bij deze somtypen werden de meeste strategieveranderingen verwacht, met name bij ppn. die de voor aftreksommen met tientalpassering ontoereikende 1010-strategie volgden. Juist bij deze somtypen zouden de ppn. kunnen ontdekken dat hun expertiseniveau nog ontoereikend was voor alle typen opgaven binnen het domein (Case & Bereiter, 1984).

In Tabel 4 is voor de optelsommen en aftreksommen met tientalpassering het gebruik van rekenstrategieën opgesomd. Tabel 4 laat zien dat 30 van de waargenomen strategiecombinaties bij optel- en aftreksommen met tientalpassering ondergebracht kunnen worden in de onderscheiden stadia.

De validiteit van het genetisch model

Ter toetsing van de validiteit van het genetisch model werd nagegaan hoeveel strategieveranderingen in overeenstemming waren met het genetisch model. Bij de 16 ppn. kon voor elk van de zes somtypen tweemaal een strategieverandering worden geconstateerd: tussen profiel 1 en profiel 2 en tussen profiel 2 en profiel 3. Van de 192 mogelijke strategieveranderingen waren 132 geschikt voor analyse. Bij de overige veranderingen werden ofwel telstrategieën gevolgd of ontbraken gegevens over de gevolgde strategie wegens juiste antwoorden. Tabel 5 geeft een overzicht van het aantal vooruitgangen, het aantal achteruitgangen en het aantal malen dat de strategie ongewijzigd bleef.

Tabel 4. Overzicht van gebruikte strategieën bij optel- en aftreksommen met tientalpassering.

Stadium	Strategie bij sommen met tientalpassering		Aantal
	Optellen	Aftrekken	
1	1010	1010	5
2	cfr	cfr	4
3	1010	10t	8
4	10t	10t	7
5	1010	G10	4
6	10t	G10	-
7	G10	G10	2
-	G10	10t	1
-	cfr	10t	1
-	1010	cfr	1
-	10t	cfr	1
Totaal			34

Tabel 5. Overzicht strategieovergangen. Op de diagonaal staan de overgangen zonder verandering. Onder de diagonaal staan overgangen met vooruitgang. Boven de diagonaal staan de overgangen met achteruitgang.

	Van:	cfr	1010	10t	G10
Naar: cfr		16	7	-	-
1010		12	30	7	4
10t		4	11	15	7
G10		2	10	6	1

Van 132 overgangen waren 62 stabiel (geen nieuwe strategie). Bij 45 overgangen was sprake van vooruitgang in termen van de algemene ontwikkelingslijn, in 25 gevallen werd teruggang geconstateerd. Van de 12 maal dat de G10-strategie in profiel 1 of profiel 2 werd gevolgd, was slechts in één geval een stabiele vorm bereikt. Bij de overige 11 gevallen viel de pp. in het tweede of derde profiel terug naar 1010 of 10t.

Bij de 45 gevallen van vooruitgang was 29 maal sprake van een vooruitgang van één stap in termen van de algemene ontwikkelingslijn. Een vooruitgang van drie stappen (van Cijferend Rekenen naar G10) kwam tweemaal voor.

Bij de 25 gevallen van achteruitgang werd 21 maal één stap teruggezet en 4 maal twee stappen. Terugval van G10 naar Cijferend Rekenen kwam niet voor.

Vervolgens is nagegaan of strategieveranderingen vaker optraden na sommen met tientalpassering (met name de aftreksommen) dan na sommen zonder tientalpassering. Ook werd de invloed van het al of niet goed opgelost hebben van een som op de strategiekeuze bij de daarop volgende som onderzocht. Op grond van het genetische model werd verondersteld dat vooral sommen, waarin de leerling in een impasse terecht kwam, aanleiding zouden geven tot strategieverandering. Derhalve werden meer strategieveranderingen verwacht na fout gemaakte sommen dan na goed gemaakte sommen en eveneens meer overgangen na sommen met tientalpassering (met name de aftreksommen) dan na sommen zonder tientalpassering. Uit Tabel 6 is af te leiden dat na sommen met tientalpassering iets vaker strategieverandering optrad (59% verandering en 41% stabiel) dan na sommen zonder tientalpassering (50% verandering en 50% stabiel). Bij de aftreksommen met tientalpassering leidden 11 van de 16 fout beantwoorde sommen tot

Tabel 6. Overzicht strategieveranderingen per somtype, uitgesplitst naar resultaat vorige opgave (goed of fout).

	Vooruit		Stabiel		Achteruit	
	Na goed	Na fout	Na goed	Na fout	Na goed	Na fout
Sommen zonder ttp*	11	14	33	11	14	5
Sommen met ttp	6	14	8	10	5	1
Totaal	17	28	41	21	19	6

*ttp betekent: tientalpassering

strategiewijziging, terwijl bij de optelsommen met tientalpassering juist de goed beantwoorde opgaven dikwijls (8 van de 12 gevallen) tot strategiewijziging leidden. In het algemeen werd na fout opgeloste sommen vaker van strategie gewisseld (62% verandering en 38% stabiel) dan na goed opgeloste sommen (47% verandering en 53% stabiel).

DISCUSSIE

De data geven steun aan het voorgestelde genetisch model van de ontwikkeling van rekenstrategieën voor het optellen en aftrekken in het getallengebied van 20 tot 100. De ppn. werkten consistent volgens de gekozen strategieën. De momenten waarop de ontoereikendheid van de strategie bleek vormden de aanleiding tot het veranderen van de strategie. Op elk van deze conclusies wordt hieronder nader ingegaan.

Een groot deel van de geanalyseerde profielen was onder te brengen in de onderscheiden stadia van het genetisch model. Vooral de lijn van stadium 1 naar 2 naar 3, waarin de 1010-strategie geleidelijk wordt vervangen door de 10t-strategie, werd in de protocollen zichtbaar.

De ppn. pasten consistent in de verschillende stadia van het genetisch model een gekozen strategie toe, zo bleek uit de vergelijking van de wijze waarop kleine en grote sommen van een bepaald somtype werden opgelost. Daarmee wordt steun gegeven aan de opvatting dat optellen en aftrekken regelgeleide cognitieve activiteiten zijn. Er zijn zelfs redenen om aan te nemen dat er sprake is van sterke dominantie van regels bij het oplossen van optel- en aftreksommen. De Cijferend Rekenen-strategie is een voorbeeld van een simpele aanpak, die rechtstreeks voortvloeit uit de elementaire kennis van eenheden en ronde tientallen, waarmee de leerling is uitgerust als voor het eerst kennis gemaakt wordt met getallen uit het gebied van 20 tot 100. De Cijferend Rekenen-strategie komt voort uit de 1010-strategie, en is in feite een te snelle verkorting waarbij de ronde tientallen als eenheden worden opgevat en de opgave gereduceerd wordt tot de combinatie van twee bewerkingen uit het vertrouwde getallengebied van 0 tot 20.

Toch blijkt de herhaalde confrontatie met niet oplosbare opgaven (in ons geval de aftreksommen met tientalpassering) bij veel ppn. de ontwikkeling in beweging te zetten. De ontoereikende 1010-strategie wordt veranderd in de 10t-aanpak of wellicht vervangen door een G10-procedure. Welke omstandigheden naast de confrontatie met de tekortkomingen van de huidige aanpak een herstructurering van de procedurele en declaratieve kennis teweeg brengen, zal nader onderzocht moeten worden.

REFERENTIES

- Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Anderson, R.C. (1984). Some reflections on the acquisition of knowledge. *Educational Researcher*, 13, 9, 5-10.
- Baroody, A.L., en Ginsburg, H.P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.). *Conceptual and procedural knowledge: the case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum Ass.
- Beishuizen, M. (1986). Vervolgonderzoek: Invloeden van leermiddelen op de uitvoering van rekenhandelingen. In S. Dijkstra & P. Span (Eds.). *Leerprocessen en instructie*. Lisse: Swets & Zeitlinger.
- Beishuizen, M. & Van Mulken, F. (1988). Twee veelgebruikte oplossingsmanieren bij hoofdrekenen. *Panama-post*, 6, 3, 32-37.
- Carpenter, Th. P. (1985). Learning to add and subtract: an exercise in problem solving. In: A. Silver (Ed.). *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives*. Hillsdale, NJ: Lawrence Erlbaum Ass.
- Case, R. (1982). *Intellectual development: a systematic reinterpretation*. New York: Academic Press.
- Case, R., & Bereiter, C. (1984). From behaviourism to cognitive behaviourism to cognitive development: Steps in the evolution of instructional design. *Instructional Science*, 13, 141-158.
- Corte, E. de, & Verschaffel, L. (1983). Onderzoek van onderwijsleerprocessen en onderwijspraktijk. *Pedagogische Studiën*, 60, 153-164.
- Fuson, K.C., Richards, J., & Briars, D.J. (1982). The acquisition and elaboration of the number word sequence. In: C. Brainerd (Ed.), *Children's Logical and Mathematical Cognition (Progress in Cognitive Development, Vol. 1)*. Heidelberg: Springer-Verlag.
- Goldstein, I. (1982). The genetic graph: a representation for the evolution of procedural knowledge. In: D. Sleeman & J.S. Brown (Eds.). *Intelligent Tutoring Systems*. New York: Academic Press.
- Groen, G.J., & Parkman, J.M. (1972). A chronometric analyses of simple addition. *Psychological Review*, 79, 4, 329-343.
- Heijden, M. van der, & Beishuizen, M. (1986). Diagnostisering van individuele verschillen in het optellen en aftrekken van basisschoolleerlingen. In: A. Reints, & P. Span (Eds.). *Differentiatie in het onderwijs*. Lisse: Swets & Zeitlinger.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Mannaerts, A.A.J. (1987). Een instrument voor de diagnose van rekengedrag. *Nederlands Tijdschrift voor de Psychologie*, 42, 7, 386-391.
- Siegler, R.S., & Klahr, D. (1982). When do children learn? The relationship between existing knowledge and the acquisition of new knowledge. In: R. Glaser (Ed.). *Advances in Instructional Psychology*, Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Ass.
- Parreren, C.F. van (1975). *Psychologie van het leren, I*. Deventer: Van Loghum Slaterus.

Manuscript ontvangen 22-8-1988

Definitieve versie ontvangen 14-2-1989

Veranderingsdiagnostiek bij handig rekenen

M.K. van der Heijden

Pedologisch Instituut en Vakgroep Onderwijsstudies

ABSTRACT

When students think first about the most efficient way to solve arithmetic problems, before executing certain arithmetic procedures, one can speak of metacognitive activities. We used elementary arithmetic problems of the form $15+18+2=$ and $14+17-14=$ to assess the degree of these metacognitive activities in 148 third grade students. We helped students if they did not exhibit these metacognitive activities spontaneously, first with a general and next with a taskspecific instruction. On the basis of this helping procedure we developed a taskspecific measure of learning potential. The results show significant correlations between this metacognitive learning potential measure and the arithmetic scores, the schoolnotes and the RAVEN Progressive Matrices Scores.

INLEIDING

Bij het optellen en aftrekken tot 100 wordt gesproken van 'handig rekenen' of 'flexibel eigenschapsrekenen' indien leerlingen opgaven als $17+18+2$ en $15+19-15$ niet zuiver algoritmisch aanpakken, maar, voordat ze met de uitvoerende rekenhandelingen beginnen, de opgave eerst analyseren om te bezien wat de meest efficiënte oplossingsweg is. Het belang van het onderwijzen van dit handig- en flexibel eigenschapsrekenen wordt momenteel in de rekendidaktische vakliteratuur nogal benadrukt (Treffers e.a., 1987). Span (1974) gebruikt de term 'aanpakgedrag' om aan te duiden hoe een leerling met een taak omgaat. Van der Heijden (1986) spreekt indien een leerling rekenopgaven handig aanpakt van een 'probleemanalyserende aanpak'. Zowel uit handelingspsychologisch (Menčinskaja, 1973) als uit cognitief psychologisch (Brown, 1987) onderzoek blijkt dat leerlingen met leerproblemen tekorten in het aanpakgedrag vertonen. In de cognitieve psychologie spreekt men van tekorten op metacognitief gebied. Volgens van Parreren (1988) kan en moet een dergelijk aanpakgedrag onderwezen worden omdat het de denkontwikkeling van de leerlingen bevordert (ontwikkeland onderwijs). Binnen de cognitieve psychologie is het onderwijzen van metacognitie een belangrijk onderzoeksthema (Brown, 1987; Beukhof & Simons, 1986).

In een constaterend onderzoek (Van der Heijden, 1986) werd het belang van de hierboven aangeduide handige aanpak bij het hoofdrekenen tot 100 aangetoond. Bij tweede klas leerlingen werden significante correlaties ($p \leq .003$) van respectievelijk .45, .40 en .35 gevonden tussen de probleemanalyserende aanpak en de rekenprestaties, het schoolcijfer voor rekenen en het RAVEN IQ. In dit artikel staan vragen over de onderwijsbaarheid van deze handige aanpak centraal. We kiezen hierbij, om redenen waarop nog zal worden ingegaan, voor een diagnostische invalshoek.

Hieronder gaan we aan de hand van twee parallelle onderzoeksstrategieën, namelijk de veranderingsdiagnostiek en de onderzoeksstrategie van het formerend experiment, in op de vraagstelling van het onderzoek.

Bij een *formerend experiment* (Kossakowski & Lompscher, 1986) wordt geprobeerd de gewenste eindhandelingsstructuur (Van Parreren, 1969) door middel van het onderwijzen van de leerling tot stand te brengen. Hiertoe dient deze eindhandelingsstructuur eerst op theoretisch niveau geanalyseerd te worden. We beschouwen daarom de eindhandelingsstructuur waar het bij de handige aanpak om gaat wat nader. Om te beginnen kan worden vastgesteld dat bij deze aanpak sprake is van oriënterende handelingen: de leerling laat bij een opgave als $17+18+2$ in eerste instantie uitvoerende handelingen achterwege en beziet wat de meest efficiënte oplossingsweg is. Het gaat bij deze oriëntering om het actualiseren van het geleerde in een concrete taaksituatie (Van der Heijden, 1986): door middel van een analyse van de opgave actualiseert de leerling kennis van de eigenschappen der getallen en op grond daarvan komt de leerling tot het inzicht dat het (in de hierboven genoemde voorbeelden) handig is eerst de 2 bij de 18 te doen respectievelijk eerst 15-15 te doen in plaats van vooraan te beginnen met rekenen. Bij de handige aanpak kunnen dus twee soorten oriënterend handelen onderscheiden worden:

- algemeen oriënterend handelen dat wil zeggen het in eerste instantie achterwege laten van uitvoerende handelingen, het aandachtig bekijken en analyseren van de opgavesituatie.
 - vakspecifiek oriënterend handelen waarbij in de opgave regelmatigheden opgemerkt worden op grond van inzicht in de eigenschappen van het tientallige systeem en inzicht in het optellen en aftrekken, en waarin tot een bepaalde handige aanpak 'besloten' wordt.
- Op grond van deze analyse werd in dit onderzoek aan de leerlingen achtereenvolgens twee soorten hulp gegeven bij het oriënterend handelen: *algemene hulp* en *vakspecifieke hulp* (zie: Tabel 2). Op grond van de mate waarin de leerlingen van deze hulp profiteren werd een veranderingsmaat gedefinieerd.

Witzlack (1977) spreekt van *veranderingsdiagnostiek* indien er op twee tijdstippen een meting wordt verricht en er tussen deze twee tijdstippen aan de leerling hulp of onderwijs wordt gegeven. Deze diagnostiseringsstrategie wordt ook aangeduid met de term leerbaarheids- of leergeschiktheidsdiagnostiek, ook spreekt men, naar Vygotskij, over het onderzoeken van de *zône* van de naaste ontwikkeling. In de klinische praktijk wordt gesproken van *limit-testing*.

In dit veranderingsdiagnostisch onderzoek willen we de volgende vragen beantwoorden:

- In welke mate profiteren leerlingen van algemene respectievelijk van vakspecifieke hulp om tot een handige aanpak te komen?
- Wat is het verband tussen de mate waarin leerlingen van de geboden hulp profiteren en de zelfstandige rekenprestaties?
- Wat is het verband tussen de mate waarin leerlingen van de geboden hulp profiteren en het abstracte denkniveau?

Deze vragen zijn relevant voor de controverse over het belang van algemene versus vakspecifieke strategieën (Beukhof & Simons, 1986; Brown, 1987). De vragen zijn tevens van belang in verband met de begripsverheldering van het begrip '*zône* van de naaste ontwikkeling'. Rubinstein (1983) stelt dat bij het vaststellen van de *zône* van de naaste ontwikkeling eerst algemene oriënterende en organiserende en pas daarna taakspecifieke hulp dient te worden gegeven. Bij de meeste leertests, en dit geldt zowel voor tests die werken met schoolse als niet-schoolse taken (zie b.v.: Ruijsenaars & Hamers, 1986 & 1988) is echter uitsluitend sprake van taak- c.q. vakspecifieke hulp. Daarnaast zijn de bovengenoemde vragen relevant voor de schoolpraktijk. Omdat zowel tekorten wat betreft de oriëntering en metacognitie als een beperkte leergeschiktheid of leerbaarheid worden gezien als belangrijke kenmerken van leerlingen met leerproblemen (Menčinskaja, 1973; Rubinstein, 1983; Brown, 1987) wordt er voor gepleit metacognitieve toetsen en

leertests op te nemen in het toelatingsonderzoek voor het speciaal onderwijs (Ruijsenaars & Hamers, 1986; Wolters, 1988). Validiteitsonderzoek met leertests laat echter nogal wisselende resultaten zien (Ruijsenaars & Hamers, 1986 & 1988). In het kader van het bepalen van de constructvaliditeit van de veranderingsdiagnostiek bij handig rekenen zijn wij geïnteresseerd in het verband tussen de door ons gedefinieerde veranderingsmaat en het abstracte denkniveau bij basisschoolleerlingen. We kozen voor dit criterium omdat het niveau van abstraheren en generaliseren als een belangrijk differentiaaldiagnostisch criterium wordt beschouwd bij de verwijzing naar LOM- en MLK-scholen (Rubinstein, 1983; Van der Heijden, 1989).

METHODE

Proefpersonen

Het onderzoek werd in januari/februari 1986 uitgevoerd bij 145 vijfde groep leerlingen uit 7 schoolklassen afkomstig uit 6 verschillende scholen (dat wil zeggen 2 klassen waren parallelklassen). Uit deze klassen werden alle leerlingen onderzocht. De gebruikte rekenmethodes waren Taltaal, Operator Rekenen en Naar Zelfstandig Rekenen. Aan het onderzoek namen 75 jongens en 70 meisjes deel. De gemiddelde leeftijd der leerlingen was 8;9 jaar (min. 7;4 - max. 10;1).

Materiaal en procedure

Er werd een *individuele diagnostische rekentest* afgenomen bestaande uit 15 optel- en aftrekopgaven met getallen van twee cijfers onder de 100, met en zonder tientalpassering. Acht van de vijftien opgaven waren handig-reken-opgaven. Bij deze 8 opgaven werd op grond van retrospectie bepaald of de leerling een handige strategie toepaste. Tabel 1 geeft een overzicht van deze 8 opgaven en de scoringscriteria voor de probleemanalyserende aanpak. Tevens zijn in deze tabel per opgave de vakspecifieke kennisinhouden vermeld die geactualiseerd dienen te worden in het kader van de handige aanpak. De overige zeven opgaven waren 'gewone' opgaven van het type $a \pm b =$

Tabel 1. Handig rekenen: opgaven, scoringscriteria en vakspecifieke kennisinhouden.

Opgave	Criterium probleemanalyserende aanpak	Vakspecifieke kennisinhoud
34 + 22 34 + 23	leidt antwoord af uit vorige opgave	inzicht in optellen en aftrekken
58 - 35 58 - 34	leidt antwoord af uit vorige opgave	
17 - 14 + 14 15 + 19 - 15	laat termen tegen elkaar wegvallen laat termen tegen elkaar wegvallen	inzicht in optellen en aftrekken kennis van de commutatieve wet
17 + 18 + 2 15 + 17 + 5	herstructureert, doet eerst 18 + 2 herstructureert, doet eerst 15 + 5	inzicht in de tientallige structuur van het getallensysteem kennis van de commutatieve wet
27 + 15 42 - 15	gebruikt de relatie optellen/aftrekken	inzicht in optellen en aftrekken
12 + 37 49 - 37	gebruikt de relatie optellen/aftrekken	inzicht in de relatie optellen/aftrekken

Omdat we (ook) geïnteresseerd waren in de zelfstandige rekenprestaties en de spontane aanpak van leerlingen en om voortijdige beïnvloeding te voorkomen werd in eerste instantie geen enkele hulp geboden bij het maken van de 15 opgaven. Nadat alle 15 opgaven zo met de leerlingen waren doorgenomen (zie Van der Heijden, 1986, voor een beschrijving van de procedure) werd door de onderzoeker bij die opgaven waar de leerlingen spontaan geen handige strategie hadden toegepast eerst algemene en vervolgens vakspecifieke hulp geboden. Een en ander gebeurde op strict gestandaardiseerde wijze volgens de in Tabel 2 beschreven procedure.

Tabel 2. Veranderingsdiagnostische procedure

Opgaven	Algemene hulp	Vakspecifieke hulp
34 + 22 34 + 23 58 - 35 58 - 34	<i>Kan je deze som ook sneller en handiger doen?</i>	Ik zal je een beetje helpen. Deze sommen lijken erg op elkaar. Hier (<i>aanwijzen</i>) staat 34 erbij 23 en kijk nu eens naar de vorige som, hier (<i>aanwijzen</i>) staat 34 erbij 22. Kan je nu heel snel en handig weten wat er uit deze (<i>wijs aan</i>) komt? (<i>max. 1 min.</i>). <i>Indien de leerling de handige strategie niet toepast:</i> Hier moet er 1 meer bijgedaan worden, dus komt er ook 1 meer uit het antwoord op deze som: 56, erbij 1, is 57.
17 - 14 + 14 15 + 19 - 15	<i>Kan je deze som ook sneller en handiger doen?</i>	Ik zal je een beetje helpen. Hier staat (<i>aanwijzen</i>) 17 eraf, 14, en dan weer erbij 14. Eerst gaat er 14 af en dan komt er ook weer 14 bij. Kan je nu heel snel en handig weten wat er uit de som komt? <i>Indien de leerling de handige strategie niet toepast:</i> 17 eraf, 14 en dan weer 14 erbij, blijft 17. Als je hetzelfde er eerst af en dan er weer bij doet blijft het gelijk. Er komt dus 17 uit deze som.
17 + 18 + 2 15 + 17 + 5	<i>Kan je deze som ook sneller en handiger doen?</i>	Ik zal je een beetje helpen. Welke twee getallen van deze drie kan je heel gemakkelijk en handig het eerst bij elkaar doen? <i>Indien de leerling de handige strategie niet toepast:</i> 18 en 2 kan je handig het eerst samen doen, want 18 erbij 2 is 20, dan doe je 20 erbij 17, is 37.
27 + 15 42 - 15 12 + 37 49 - 37	<i>Kan je deze som ook sneller en handiger doen?</i>	Ik zal je een beetje helpen. Deze sommen hebben iets met elkaar te maken. Kijk maar: hier (<i>aanwijzen</i>) hebben we uitgerekend dat 27 en 15 samen bij elkaar even veel is als 42. En hier (<i>aanwijzen</i>) staat weer 42, maar nu gaat er 15 van af. Kan je nu handig weten wat er dan overblijft? <i>Indien de leerling de handige strategie niet toepast:</i> 27 (<i>opschrijven of aanwijzen</i>), want 42 (<i>aanwijzen in 12</i>) is evenveel als 27 en 15 samen.

Na het geven van de hulp werd door het stellen van een retrospectievraag telkens nagegaan of de leerling van de gegeven hulp had geprofiteerd, dat wil zeggen alsnog de handige strategie had toegepast. Op grond van deze procedure werden de volgende

variabelen geoperationaliseerd.

In aansluiting bij het eerder gedane onderzoek werd als mate van *probleemanalyserende* (dwz. *handige*) *aanpak* genomen het aantal keer dat door een leerling spontaan een handige aanpak werd toegepast (min. 0 - max. 8).

Als maat voor het profiteren van *algemene respectievelijk vakspecifieke hulp* werd genomen het aantal keren dat geprofiteerd werd van algemene respectievelijk vakspecifieke hulp gedeeld door het aantal opgaven dat niet al spontaan handig werd aangepakt (min. 0 - max. 1).

In het kader van de veranderingsdiagnostiek wilden we de twee bovenstaande maten combineren tot één veranderings- c.q. leerbaarheidsmaat. Er van uitgaande dat een leerling die aan de algemene hulp (zie tabel 2) genoeg heeft om tot een handige aanpak te komen, leerbaarder is dan een leerling aan wie ook nog vakspecifieke hulp gegeven moet worden, namen we als veranderingsmaat 2x het aantal keer dat geprofiteerd werd van algemene hulp plus 1x het aantal keer dat geprofiteerd werd van vakspecifieke hulp gedeeld door het aantal opgaven dat niet al spontaan handig werd opgelost (min. 0 - max. 2).

Als maten van de *rekenprestaties* namen we het aantal spontaan - dat wil zeggen voordat de hulp werd gegeven - goed gemaakte opgaven van de diagnostische rekentest (min. 0 - max. 15) en de totale tijd die de leerling nodig had voor het maken van de opgaven (in seconden). Daarnaast werden in verband met de ecologische validiteit de rapportcijfers voor rekenen van de leerkrachten vekregen.

Als maat voor het niveau van abstraheren en generaliseren werd de score op RAVEN's Coloured Progressive Matrices genomen. Volgens Raven (1960) moet deze test niet zozeer als een test voor de algemene intelligentie beschouwd worden (het verbale aspect ontbreekt), maar als een test voor het abstracte denkniveau.

Door praktische omstandigheden werden de schoolcijfers slechts van 5 klassen (N=90) verkregen en werd de RAVEN maar in 3 klassen (N=69) afgenomen. De desbetreffende analyses werden over het maximaal beschikbare aantal proefpersonen verricht.

RESULTATEN

Tabel 3 geeft als achtergrondgegevens de correlaties tussen de prestatievariabelen.

Tabel 3. Rangcorrelaties (Spearman's Rho) prestatievariabelen

	Indiv. test aantal goed	Indiv. test opl. tijd	School- cijfer
Indiv. test opl. tijd	- 0.45***		
schoolcijfer	0.60***	- 0.59***	
Raven IQ	0.30**	- 0.24*	0.53***

*** d.w.z. $p < .001$; ** d.w.z. $.01 > p > .001$, * d.w.z. $.05 \geq p > .01$

Uit de tabel blijkt dat sprake is van significante correlaties in de verwachte richting tussen het aantal goed gemaakte opgaven (gemiddelde: 11.9; $sd=2.9$), de oplossingstijd (gemiddelde: 186 sec.; $sd=123$), het schoolcijfer voor rekenen (gemiddelde: 6.7; $sd=1$) en het Raven IQ (gemiddelde: 28; $sd=4.9$).

Tabel 4 geeft de statistische gegevens van de procesvariabelen. Uit de tabel blijkt dat erg weinig leerlingen spontaan een handige aanpak vertonen. Voorts blijkt dat de statistische eigenschappen van de variabelen profiteren van algemene en

Tabel 4. Statistische gegevens der kwalitatieve variabelen

	Probl. anal. aanpak (spontaan)	profiteren van alg. hulp	profiteren van vakspec. hulp	veranderingsmaat
gem.	1.67	0.30	0.36	0.95
st. dev.	1.43	0.23	0.21	0.45

profiteren van vakspecifieke hulp met elkaar vergelijkbaar zijn.

Als we de losse opgaven als uitgangspunt nemen levert het onderzoek de volgende resultaten op:

21% van alle opgaven wordt spontaan handig opgelost

26% van alle opgaven wordt na algemene hulp handig aangepakt

27% van alle opgaven wordt na specifieke hulp handig aangepakt

25% van alle opgaven wordt uiteindelijk niet handig aangepakt.

We kunnen ook uitgaan van het aantal niet handig opgeloste opgaven dat telkens overblijft. Dan blijkt dat:

– bij 33% van de niet spontaan handig opgeloste opgaven algemene hulp effectief is

– bij 52% van de na algemene hulp nog niet handig opgeloste opgaven vakspecifieke hulp effectief is.

Tabel 5 bevat de rangcorrelatie coëfficiënten van Spearman tussen de proces- en de prestatievariabelen.

Tabel 5. Rangcorrelaties (Spearman's Rho) tussen de kwalitatieve en de prestatievariabelen

	Probl. anal. aanpak (spontaan)	profiteren van alg. hulp	profiteren van spec. hulp	veranderingsmaat
<i>Indiv. test</i>				
aantal goed	0.13	0.25**	0.08	0.31***
opl. tijd	- 0.41***	- 0.33***	- 0.16*	- 0.40***
<i>schoolcijfer</i>	0.48***	0.39***	0.21*	0.52***
<i>Raven IQ</i>	0.32**	0.34**	0.01	0.36**

Opvallend is het ontbreken van een verband tussen de probleemanalyserende aanpak en het aantal goed gemaakte opgaven. Daarnaast blijkt er sprake te zijn van duidelijk significante samenhangen tussen de prestatievariabelen en het profiteren van algemene hulp. Dit is niet het geval bij de variabele profiteren van vakspecifieke hulp. De door ons gedefinieerde veranderings- c.q. leerbaarheidsmaat hangt eveneens duidelijk significant samen met zowel de rekenprestaties als de mate van abstract denken.

DISCUSSIE

In dit onderzoek wordt bevestigd wat we al eerder (Van der Heijden, 1986) vonden: spontaan handig rekenen komt nogal weinig voor: van de 8 handig-reken-opgaven lost de gemiddelde leerling er nog geen 2 handig op. Tot en met dit onderzoek is onderzoek verricht bij leerlingen die les kregen met 6 verschillende rekenmethoden. Deze onderzoeksresultaten ondersteunen dan ook, in elk geval voor wat betreft Nederland, de stellingname van Van Parreren (1983, p.10): 'In het algemeen kunnen wij stellen dat een van de meest verbreide tekortkomingen van onderwijs over de hele wereld is, dat de onderwijsgevende,

of het onderwijsprogramma, zich direct richt op het instrueren van de uitvoering van de taak, en de fase van oriëntering, die vooraf hoort te gaan, verwaarloost of negeert. Op die manier worden te veel 'foefjes' aangeleerd, krijgt de leerling te veel kant-en-klare recepten in te prenten....".

In overeenstemming met het eerder verrichte onderzoek bij vierde groep leerlingen worden ook bij de vijfde-groepers significante correlaties gevonden tussen de spontane handige aanpak, het schoolcijfer voor rekenen en het Raven IQ.

In tegenstelling tot eerder onderzoek wordt in dit onderzoek geen verband gevonden tussen de handige aanpak en het aantal goed gemaakte opgaven. Een mogelijke verklaring hiervoor zou kunnen zijn dat veel zwakke rekenaars in dit onderzoek nog tellend rekenden. Voor deze leerlingen levert een handige aanpak veel voordelen op (geringere belasting van het korte termijngeheugen). Veel van deze zwakke rekenaars pakten de opgaven wel handig aan maar maakten (desondanks) veel telfouten zodat ze vaak tot onjuiste antwoorden kwamen.

Dit onderzoek onderstreept het belang van algemene strategieën: bij éénderde van de niet spontaan handig opgeloste opgaven instigeert alleen al een algemeen oriënterende opmerking van de onderzoeker ('Kan je deze som ook sneller en handiger doen?') een handige aanpak. Het direct onderwijzen van een dergelijk algemeen strategisch handelen los van vakspecifieke kennisinhouden lijkt echter weinig zinvol omdat eveneens in dit onderzoek wordt bevestigd dat het bij het aanpakgedrag gaat om het actualiseren van het geleerde in een bepaalde taaksituatie: bij ruim de helft van de na de algemene hulp niet handig opgeloste opgaven blijkt vakspecifieke hulp effectief.

Tevens blijkt dat goede rekenaars meer profiteren van algemene hulp dan zwakke rekenaars hetgeen waarschijnlijk verklaard kan worden doordat goede rekenaars vakspecifieke rekenkennis op een meer adequate wijze georganiseerd en dus beter beschikbaar hebben in hun handelingsrepertoire of doordat bij zwakke rekenaars bepaalde vakspecifieke kennis ontbreekt.

Er zijn geen duidelijke verbanden gevonden tussen de prestatievariabelen en het profiteren van vakspecifieke hulp. Het lijkt niet verstandig om aan deze bevinding veel consequenties te verbinden omdat alleen vakspecifieke hulp werd geboden indien de algemene hulp geen effect sorteerde. Aangezien het profiteren van algemene hulp wel correleert met de prestatievariabelen, kunnen de resultaten bij het profiteren van vakspecifieke hulp veroorzaakt zijn door een restriction of range effect.

De door ons in het kader van de veranderingsdiagnostiek geconstrueerde veranderings-c.q. leerbaarheidsmaat correleert zeer duidelijk significant met de als criterium gekozen prestatievariabelen. Door de berekeningswijze van de veranderingsmaat (2x profiteren van algemene hulp plus 1x profiteren van specifieke hulp) kan men geen conclusies trekken op basis van een onderlinge vergelijking tussen de drie veranderingsdiagnostische maten in Tabel 5. Wel kan het volgende geconcludeerd worden: rekenaars die veel opgaven goed maken vertonen in dit onderzoek spontaan niet vaker een handige rekenaanpak dan zwakke rekenaars maar leren deze aanpak wel sneller te gebruiken. Daarnaast bezit de leerbaarheidsmaat een redelijke ecologische validiteit: leerkrachten verdisconteren in hun rekenbeoordeling de leerbaarheid wat betreft handig rekenen van de leerlingen in ongeveer dezelfde mate als de rekentestprestaties. Tenslotte blijkt uit dit onderzoek dat basisschool-leerlingen met een relatief hoog abstract denkniveau een grotere leerbaarheid wat betreft handig rekenen vertonen dan leerlingen met een relatief laag abstract denkniveau. Deze resultaten geven aanleiding tot verder onderzoek naar de differentiële validiteit (Ippel, 1982) van deze vorm van vakspecifieke veranderingsdiagnostiek van het oriënterend handelen bij leerlingen met leerproblemen op LOM- en MLK-scholen.

LITERATUUR

- Beukhof, G., & Simons, R.J. (1986). *German and Dutch Research on Learning and Instruction*. Den Haag: S.V.O.
- Brown, A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F.E. Weinert & R.H. Kluwe (ed.). *Metacognition, Motivation and Understanding* (p. 117-141). Hillsdale, NJ: LEA.
- Heijden, M.K. van der (1986). Aanpak, kennis en reflectie in het optellen en aftrekken van tweede klas lagere school leerlingen. In S. Dijkstra & P. Span (red.). *Leerprocessen en instructie* (p.145-161). Lisse: Swets & Zeitlinger.
- Heijden, M.K. van der (1989). *Onderzoek van LOM- en ML-leerlingen door middel van het psychologisch/pedagogisch diagnostisch experiment*. Paper gepresenteerd op de conferentie Handelingstheorie in de praktijk. Leusden, 8 februari 1989.
- Ippel, M.J. (1982). Diagnostiek vanuit een veranderd scholingsconcept. In J.J.F. Schroots (red.). *Diagnostiek in het onderwijs*. Lisse: Swets & Zetlinger.
- Kossakowski, A., & Lompscher, W. (1986). Das Tätigkeitskonzept in der Psychologie der D.D.R.. In Survey 1. Internationaler Kongress zur Tätigkeitstheorie. Berlin: Hochschule der Künste.
- Menčinskaja, N.A. (1973). Leermoeilijkheden als psychologisch probleem. *Pedagogische Studiën*, 50, p. 217-229.
- Parreren, C.F. van (1969). *Psychologie van het leren I*. Deventer: Van Loghum Slaterus.
- Parreren, C.F. van (1983). *Leren door handelen*. Apeldoorn: Van Walraven.
- Parreren, C.F. van (1988). *Ontwikkeland onderwijs*. Leuven/Amersfoort: Acco.
- Raven, J.C. (1960). *Guide to using The Coloured Progressive Matrices*. London: Lewis & Co.
- Rubinstein, S.J. (1983). *Psychologie des Hilfsschulkindes*. Berlin: Volk und Wissen Volkseigener Verlag.
- Ruijsenaars, A.J.J.M., & Hamers, J.H.M. (1986). Leertests in het toelatingsonderzoek. *Tijdschrift voor Orthopedagogiek*, 6, 314-325.
- Ruijsenaars, A.J.J.M., & Hamers, J.H.M. (1988). *Instructiegericht diagnostiek*. Paper gepresenteerd op de ORD'88 te Leuven.
- Span, P. (1974) Cognitieve stijl en aanpak van de leerstof. In Parreren, C.F. van & Peeck, J. (red.). *Informatie over leren en onderwijzen*. Groningen: Tjeenk Willink.
- Treffers, A., Moor, E. de & Feijs, E. (1987). Proeve van een nationaal programma voor het reken-wiskunde onderwijs op de basisschool (1) en (2). *Tijdschrift voor nascholing en onderzoek van het reken-wiskunde-onderwijs*, 6(1/2), 7-28 en 24-31.
- Witzlack, G. (1977). *Grundlagen der Psychodiagnostik*. Berlin: Volk und Wissen Volkseigener Verlag.
- Wolters, M.A., Kort, F.A. de & Loon, C.A. van (1988). Reflectie, een zinvol criterium voor doorverwijzing? *Tijdschrift voor Orthopedagogiek*, 9, 393-403.

Manuscript ontvangen 22-8-1988

Definitieve versie ontvangen 1-2-1989

Vakinhoudelijke Kennis en Strategische Vaardigheden bij het Oplossen van Rekenopgaven

M. Bloemkolk

vakgroep PA & O, sectie Onderwijskunde, Rijksuniversiteit Groningen.

ABSTRACT

It is still a major issue whether metacomponents such as planning, monitoring and checking can be viewed meaningfully as separate skills or whether they are just a manifestation of extensive knowledge. On the basis of discrepancies between scores on two tests (tapping the same knowledge of addition and subtraction but differing in expected use of strategic skills), it is argued that it takes more than just declarative or procedural knowledge to solve arithmetic problems. It is hypothesized that training in strategic skills will be especially useful for those children that have a certain amount of domainspecific knowledge but have trouble in applying that knowledge.

PROBLEEMSTELLING

Een kenmerk van beginnende probleemoplossers is dat zij te snel van start gaan en niet alle relevante gegevens uit de opgave in ogenschouw nemen bij het kiezen van een oplossingsweg. Vaak richten ze zich op oppervlakkige kenmerken uit de opgave (Beishuizen, 1983).

Dit blijkt bijvoorbeeld uit onderzoek naar het oplossen van natuurkundeopgaven (De Jong, 1986) en van grammaticaopgaven (Breuker en Van der Linden, 1983). De Jong spreekt in dit verband van een 'kick-and-rush' strategie, Breuker en Van der Linden noemen het een 'doe eerst, denk dan' strategie. Voor dit falen bij het oplossen van problemen worden vaak twee oorzaken genoemd. De eerste is een gebrekkige strategie. Een strategie kan worden gezien als een plan dat de aanpak van een probleem organiseert en de volgorde van stappen in het oplossingsproces bepaalt. Een goede strategie voorkomt dat essentiële stappen in dat proces, zoals bijvoorbeeld controle, worden overgeslagen (De Jong, 1986). De tweede oorzaak is onvoldoende vakinhoudelijke kennis (declaratief en/of procedureel).

Ook in onderzoek naar het vinden van een goede oplossingsweg (bewerking) voor stipsommen treffen we deze twee problemen aan. Stipsommen zijn opgaven waarbij de onbekende op een voor kinderen minder gewone plaats staat, bijvoorbeeld de opgave $..-56=20$. In tegenstelling tot een 'gewone' formule-opgave zoals $20+56=..$ is bij stipsommen de juiste oplossingsweg (de uit te voeren bewerking) niet direct gegeven. Zo kan de opgave $..-56=20$ niet worden opgelost door 20 van 56 af te trekken, maar moeten beide getallen bij elkaar worden opgeteld.

Het vinden van een juiste oplossingsweg voor stipsommen levert veel kinderen problemen op (De Corte en Verschaffel 1980, 1983; Bloemkolk en Koster, 1984; Sandberg en Barnard, 1985). Deze problemen kunnen hun oorzaak hebben in een foute aanpakstrategie. Het kind kan bijvoorbeeld de opgave slechts oppervlakkig lezen en na een blik op de

stipsom $..-56=20$ op basis van de '-' in de opgave kunnen besluiten dat het moet aftrekken. Als het vervolgens de uitkomst niet controleert, dan zal het tot het foute antwoord '36' komen.

Het kan echter ook zijn dat het kind de opgave wel goed leest, maar deze niet begrijpt. Probleem is dan niet een foute aanpakstrategie, maar een gebrekkige kennisbasis. Lezen kinderen de opgaven wel van links naar rechts? Wordt een juiste betekenis toegekend aan de in de opgave gebruikte symbolen? Het lijkt vaak alsof kinderen denken dat een som met een '-' teken erin altijd moet worden opgelost door het kleinste gegeven getal van het grootste gegeven getal af te trekken. Ook blijken kinderen vaak verkeerde opvattingen te hebben over het 'is gelijk' symbool. Bij het oplossen van redactieopgaven spelen soortgelijke problemen (De Corte en Verschaffel 1980, 1983; Bloemkolk en Koster, 1984).

Aanvankelijk lag het accent bij het onderzoek van probleemoplossend gedrag op algemene, domein-onafhankelijke strategieën. Achterliggende gedachte was dat beginners (vaak kinderen) dergelijke strategieën niet zouden bezitten en experts (vaak volwassenen) wel. De laatste jaren is steeds meer aandacht ontstaan voor vakinhoudelijke kennis als verklaring voor verschillen in prestaties. Tegelijkertijd is twijfel ontstaan over de betekenis van strategische vaardigheden. Kunnen die gezien worden als afzonderlijke vaardigheden of zijn ze niets anders dan een manifestatie van uitgebreide (domeinspecifieke) kennis? (Chi, 1981; Chipman, 1985).

Chi (1981, 1984) toonde aan, dat indien het verschil in domeinspecifieke kennis tussen kinderen en volwassenen wegvalt ook de verschillen in prestaties binnen dat domein verdwijnen. Chi stelt dat jonge kinderen vaak wel over strategieën beschikken, maar deze alleen niet toe kunnen passen omdat bij hen de noodzakelijke vakinhoudelijke kennis ontbreekt. De oorzaak van het feit dat kinderen bijvoorbeeld hun uitkomsten niet controleren, zou naar haar mening niet zozeer gezocht moeten worden in het ontbreken van controleprocessen, maar veeleer in het ontbreken van (vakinhoudelijke) schemata die aangeven dat een gevonden uitkomst vreemd of onmogelijk is. Zou zo'n schema wel aanwezig zijn, dan zou een afwijkende uitkomst direct opvallen en aanleiding zijn de uitkomst en het oplossingsproces dat daartoe geleid heeft te controleren (Chi, 1984). Ook andere auteurs verklaren geobserveerde verschillen in aanpakgedrag in termen van verschillen in vakinhoudelijke kennis. Een kick-and-rush aanpak zou bijvoorbeeld samenhangen met het feit dat beginners de relevante cues in een probleem niet herkennen en daardoor aan de slag gaan op basis van een onjuiste interne probleem-representatie (Beishuizen, 1983).

Brown vindt echter (bijvoorbeeld Campione, Brown & Ferrara, 1982) dat hierin het gevaar schuilt van onderschatting van het belang van strategische aspecten. Het beschikken over vakinhoudelijke kennis betekent volgens haar niet dat er ook automatisch strategisch met die kennis wordt gehandeld.

Meer inzicht in de relatie tussen vakinhoudelijke kennis en algemene strategische vaardigheden kan worden verkregen door kinderen twee series opgaven voor te leggen waarvoor dezelfde domeinspecifieke kennis vereist is, maar waarbij het verwachte gebruik van strategische vaardigheden verschilt. Dit wordt hieronder toegelicht. Vervolgens kan de samenhang tussen de prestaties op beide toetsen worden nagegaan.

Gekozen is voor een serie vleksommen en een serie stipsommen. Vleksommen worden in de Kwantiwijzer, een handelingstheoretisch georiënteerd instrument voor rekendiagnostiek (Van den Berg en Van Eerde, 1983) gebruikt om de kennis van kinderen over optel- en aftrekopgaven in formulevorm achterhalen. In figuur 1 staat een voorbeeld van een vlekensom. De proefleider vraagt daarbij bijvoorbeeld aan het kind: "Onder welke vlek moet het grootste getal staan, onder deze (wijst op de eerste vlek) of onder deze? (wijst op de laatste vlek)". Het kind moet een antwoord op deze vraag geven, en tevens uitleggen



Figuur 1: Een vlekken-som.

waarom het denkt dat dit het goede antwoord is.

Overeenkomst tussen stipsommen en vlekssommen is dat het vinden van een juiste oplossingsweg voor een stipsom in principe dezelfde inhoudelijke kennis vereist als het vinden van het goede antwoord bij een vlekken-som. Bij beide typen opgaven gaat het om kennis van de relatie tussen optellen en aftrekken en van de symbolen '+', '-' en '='. Ook vereisen beide opgavetypen dat het kind de opgave goed analyseert en/of de oplossing controleert.

Het verschil betreft de mate waarin de opgaven het aanpakgedrag van een kind zullen beïnvloeden. Stipsommen lijken het meest op de opgaven die een kind in zijn gewone rekenboek tegenkomt. Het zijn formuleopgaven en er komen getallen in voor. Het 'normale' aanpakgedrag van het kind wordt hier niet beïnvloed. Bij vlekssommen daarentegen gebeurt dat wel. Vlekssommen zijn veel ongewoner voor de kinderen. Alleen daardoor al zullen ze deze opgaven beter bekijken. Het ontbreken van getallen in de vlekssommen dwingt de kinderen eveneens tot een meer dan oppervlakkige analyse. Variatie in het aanpakgedrag tussen kinderen wordt zo veel meer uitgeschakeld, waardoor bij de vlekssommen meer de nadruk ligt op alleen het kennisaspect dan bij de stipsommen. Daar zullen naast verschillen in vakinhoudelijke kennis, verschillen in aanpakgedrag een grote rol spelen bij het vinden van een goede oplossingsweg. Door de samenhang tussen prestaties op beide toetsen te onderzoeken, is het mogelijk na te gaan of vakinhoudelijke kennis een voldoende voorwaarde is voor goede prestaties op de stipsommentoets, of dat ook strategische vaardigheden een rol spelen. Heeft Chi gelijk, dan zal er een bijna perfecte relatie bestaan tussen beide toetsen. Heeft Brown gelijk, dan zal dit niet het geval zijn.

METHODE

Proefpersonen

Aan het onderzoek deden 47 kinderen mee, 31 uit groep 4 en 16 uit groep 5, afkomstig van 5 basisscholen in de provincie Groningen. Per klas is steeds een beperkt aantal kinderen geselecteerd. Dit gebeurde op basis van de score op een Cito-rekentoets (Cito, 1979) en het leerkrachtoordeel, om een redelijk verdeling van 'zwakke', 'gemiddelde' en 'goede' rekenaars te krijgen. De gemiddelde Citoscore (max. 10) voor de kinderen uit groep 4 was 5.3 (1.7), voor groep 5 was dit 5.6 (1.5).

Procedure

Het onderzoek is uitgevoerd aan het eind van het schooljaar. Alle kinderen van de deelnemende klassen maakten eerst de Cito-rekentoets en de stipsommentoets. Bij 47 kinderen is vervolgens individueel-mondeling de vlekentoets afgenomen.

Materiaal: de vlekentoets

De vlekentoets bestaat uit 12 drie- of vierkeuze vleksommen, en wordt individueel-mondeling afgenomen. Het kind leest de opgave eerst hardop voor. Verkeerd lezen van de opgave (bijvoorbeeld 'grootste' in plaats van 'kleinste') wordt gecorrigeerd. Het van rechts naar links lezen van de vlekensom wordt echter niet gecorrigeerd: dat betreft namelijk kennis en niet zozeer onzorgvuldig aanpakgedrag. In de opgaven zijn alle vlekken steeds even groot getekend. In de vlekken staan steeds de letters A, B en C. Over de somtypen $A-B=C$, $A=B+C$ en $A=B-C$ zijn per type vier vragen in de toets opgenomen. Er zijn geen vragen over het somtype $A+B=C$ opgenomen, omdat dit somtype wordt gebruikt in de instructie. Per opgavetype zijn twee vragen zodanig gesteld, dat het zonder meer invullen van getallen niet altijd tot het juiste antwoord zou leiden. (Zie bijvoorbeeld figuur 2.) Het invullen van bijvoorbeeld $5-1=4$ of van $7-5=2$ leidt tot een fout antwoord. Het kind moet hier de concrete voorbeelden overstijgen om het juiste antwoord te kunnen vinden.

$$\text{A} - \text{B} = \text{C}$$

Onder welke vlek staat het kleinste getal?

☐ Onder vlek A

☐ Onder vlek B

☐ Onder vlek C

☐ Dat kun je niet weten. Het kleinste getal kan onder vlek B of vlek C staan.

Figuur 2: Voorbeeldopgave uit de vlekentoets.

Elk goed antwoord op de vlekentoets levert 1 punt op. Op basis hiervan wordt een totaalscore berekend die kan variëren van 0 - 12. Daarnaast is per opgave gescoord wat voor redenering het kind gebruikte om het antwoord te beargumenteren. De interbeoordelaarsbetrouwbaarheid van deze scoring is goed: kappa is .81 (Cohen, 1960; Popping, 1983).

Voor de hele toets is Cronbach's alpha .78; voor de drie subschalen $A=B+C$, $A-B=C$ en $A=B-C$ (elk 4 items) zijn de alpha's respectievelijk .78, .64 en .72. Geconcludeerd kan worden dat de vlekentoets voldoende betrouwbaar is.

Er zijn verschillende indicaties voor de validiteit van de vlekentoets. De eerste is de relatie tussen de kwaliteit van de redenering op grond waarvan een kind aan het antwoord komt en de kans op een hoge score op de vlekentoets. Gokken of redeneren op grond van irrelevante aspecten (bijvoorbeeld 'die vlek is groter dan de andere', of 'je kunt het getal er nog doorheen zien') leidde in 31% van de gevallen tot een goed antwoord, ongeveer de toevalskans. Het zonder meer invullen van getallen in de opgave leverde in 61% van de gevallen een goed antwoord op. Een 'abstracte' redenering, zonder gebruik te maken van getallen, leidde in 87% van de gevallen tot een goed antwoord en het invullen van getallen in combinatie met een abstracte redenering was in 92% van de gevallen succesvol. Ook is er een relatie tussen schoolervaring en type redenering: gokken komt vaker voor bij kinderen uit groep 4, het redeneren zonder getallen vaker in groep 5. De kinderen zijn daarnaast redelijk consistent: 21 kinderen maakten zeven keer of vaker gebruik van dezelfde wijze van redeneren (Bloemkolk & Bruins, 1987b).

Van de vlekentoets gebruiken we hier alleen de totaalscore, die aangeeft bij hoeveel

vleksommen het goede antwoord is gegeven.

Materiaal: de stipsommentoets

De stipsommentoets bestaat uit 7 opgaven. De afgenomen opgaven zijn van het type $..=x+y$, $x=y+..$, $x=..+y$, en steeds twee van de typen $..-x=y$ en $x=..-y$. Bij de samenstelling van de toets is er vanuit gegaan dat sommige opgavetypen meer kennis van optel- en aftrekopgaven vereisen voor het vinden van een goede oplossingsweg dan andere. De toets bevat daarom geen opgaven van het type $x-..=y$ of $x=y-..$. Bij die opgaven kan een goede oplossingsweg namelijk ook worden gevonden op grond van onjuiste of onvolledige kennis, bijvoorbeeld: 'als er een min in de som staat moet ik het kleinste getal van het grootste getal aftrekken'. Bij $..-x=y$ of $x=..-y$ werkt zo'n aanpak niet, en deze opgaven blijken dan ook moeilijker te zijn (Bloemkolk & Koster, 1984; Sandberg & Barnard, 1985). De stipsommentoets is schriftelijk afgenomen. Kinderen moeten hun oplossing in de opgave invullen en vervolgens onder de som opschrijven hoe ze deze hebben opgelost. Dit laatste levert gegevens op over de gevolgde oplossingsweg.

Bij de stipsommentoets wordt gescoord of de gekozen oplossingsweg goed of fout is. Elke goede oplossingsweg levert 1 punt op. Dat gebeurt onafhankelijk van de vraag of de oplossingsweg goed of fout is uitgevoerd. Een voorbeeld: Een kind lost de opgave $..-20=56$ op door uit te rekenen $56+20=77$. Hoewel het antwoord (77) fout is, is de gevolgde oplossingsweg (optellen) goed. Bij de stipsommentoets wordt een totaalscore berekend: de oplossingswegscore, die loopt van 0 tot en met 7.

De procedure voor het scoren van de oplossingsweg is betrouwbaar; Cohen's kappa is .82 (Bloemkolk & Bruins, 1987a). Cronbach's alpha is .75 voor de stipsommentoets.

RESULTATEN: DE RELATIE VLEKKENTOETS - STIPSOMMENTOETS

In Tabel 1 staan de gemiddelde scores op de vlekentoets en de stipsommentoets.

Tabel 1. Gemiddelde scores en standaarddeviaties op de drie toetsen.

		vleksommen (max. 12)	opl.weg stipsommen (max. 7)
groep 4	(N=31)	7.0 (3.0)	4.1 (2.1)
groep 5	(N=16)	7.3 (2.1)	4.8 (1.8)
groep 4 + 5	(N=47)	7.1 (2.7)	4.3 (2.0)

De correlatie tussen vlekentoets en oplossingswegscore is .67. Dat beide toetsen onderling sterk samenhangen is geen verrassing. Er is echter een duidelijk en systematisch verschil tussen beide scores. Dit blijkt uit analyse van een kruistabel waarin prestaties op de vlekentoets worden afgezet tegen de oplossingswegscore voor de stipsommen. De scores op de twee toetsen zijn gedichotomiseerd in lage en hoge scores. Bij de vlekentoets was het criterium een score van 9 of hoger, bij de oplossingswegscore een score van 5 of hoger. Voor beide toetsen betekent dit een beheersingsnorm van ruim 70%. (Zie Tabel 2).

Er is een duidelijke discrepantie tussen prestaties op de vlekentoets en op de stipsommentoets. Van de 24 kinderen die hoog scoren op de vlekentoets, scoren er 12 laag op de stipsommentoets. Veel kinderen, die gezien hun prestaties op de vlekentoets weliswaar kennis hebben van de relatie tussen optellen en aftrekken en van de betekenis van de symbolen '+', '-' en '=', kiezen desondanks bij stipsommen toch foute

Tabel 2: Prestaties op de vlekentoets ($<9=0$; $\geq 9=1$) afgezet tegen de oplossingswegscore op de stipsommentoets ($<5=0$; $\geq 5=1$).

		vleksommen	
		0	1
opl.wegscore stipsommen:	0	19	12
	1	1	12

oplossingswegen. Daarentegen is er slechts 1 kind dat laag scoort op de vlekentoets en hoog op de stipsommen-toets. (Chi-kwadraat = 10.6, exact sign. = .002; toetsing m.b.v. FISHER (Verbeek, Kroonenberg en Kroonenberg, 1983).) Ook op itemniveau keert dit beeld terug. Tabel 3 geeft enkele voorbeelden.

Tabel 3: Kruistabellen van antwoorden op vleksommen en het vinden van een oplossingsweg op stipsommen. (0 = fout, 1 = goed.)

vleksom: $A - B = C$ Welk getal is groter, het getal onder vlek A of het getal onder B?			
... - 33 = 16		0	1
	0	7	24
	1	2	14
... - 12 = 25		0	1
	0	6	15
	1	3	23
vleksom: $A = B - C$ Onder welke vlek staat het grootste getal?			
21 = ... - 53		0	1
	0	20	13
	1	1	13
33 = ... - 16		0	1
	0	16	9
	1	5	18

We zien bijvoorbeeld dat van de 38 kinderen die bij de vlekensom $A-B=C$ zeggen dat het getal onder A groter moet zijn dan het getal onder B, 24 bij de stipsom $..-33=16$ desondanks een foute oplossingsweg kiezen. Door bijvoorbeeld 16 af te trekken van 33 lijken zij er dan vanuit te gaan dat 33 het grootste getal moet zijn.

Het is verder opmerkelijk dat deze kinderen, die als antwoord '17' invulden, in dat antwoord kennelijk geen aanleiding zagen hun oplossing te controleren en de opgave opnieuw te bekijken. Dit klopt niet met de veronderstelling van Chi, dat kinderen vanzelf zullen controleren als zij een uitkomst vinden die strijdig is met hun kennis (hier: het gegeven antwoord op de vlekensom).

Een mogelijke interpretatie van de tabellen 2 en 3 is dan ook dat domeinspecifieke kennis, zoals gemeten met de vlekentoets, een noodzakelijke maar niet voldoende voorwaarde is voor het kunnen vinden van een goede oplossingsweg voor stipsommen. In de probleemstelling is aangegeven dat het belangrijkste verschil tussen de stipsommen en de vleksommen het verwachte gebruik van strategische vaardigheden is. Bij de vleksommen worden kinderen door het ongewone karakter van de opgaven en door het ontbreken van getallen min of meer gedwongen tot een grondige oriëntatie op de opgave. Bij de

stipsommen, die veel gewoner lijken, zal dit minder het geval zijn. Het is dan ook plausibel om de oorzaak van het verschil in prestaties op beide toetsen te zoeken in het al dan niet gebruiken van strategische vaardigheden, zoals een goede oriëntatie op de opgave en/of controle van de oplossing. Kinderen die over voldoende kennis beschikken zullen een goede oplossingsweg vinden als ze strategisch handelen bij het oplossen van stipsommen. Bij kinderen die over dezelfde kennis beschikken, maar niet strategisch handelen en bijvoorbeeld volstaan met een oppervlakkige analyse van de opgave, zal dat niet het geval zijn.

DISCUSSIE

De hierboven gegeven verklaring van de resultaten betekent enerzijds dat Chi gelijk kan hebben met haar veronderstelling dat ook jonge kinderen beschikken over strategische vaardigheden. Immers: bij de vleksommen oriënteren de kinderen zich kennelijk wel nauwkeurig op de opgaven. Het betekent anderzijds dat de veronderstelling van Chi dat inhoudelijke kennis er vanzelf voor zorgt dat kinderen bijvoorbeeld controleren als zij een oplossing vinden die strijdig is met hun kennis, moet worden genuanceerd. In die zin kan de stelling van Brown worden ondersteund dat kennis noodzakelijk is, maar niet automatisch leidt tot strategisch gebruik daarvan. Kinderen moeten kennelijk leren wanneer ze deze vaardigheden moeten gebruiken.

Hierbij moet wel worden opgemerkt dat in dit onderzoek slechts een beperkt aantal kinderen is betrokken. De resultaten moeten dan ook met de nodige voorzichtigheid worden bekeken. Een tweede beperking is dat er geen analyse heeft plaatsgevonden van het aanpakgedrag van de kinderen bij het oplossen van de stipsommen. Ook zou meer aandacht kunnen worden besteed aan eventuele verschillen in aanpakgedrag tussen zwakke en goede rekenaars. Gezien het geringe aantal leerlingen is dit nu niet gebeurd.

Dat kinderen bij het oplossen van formuleopgaven met getallen kennelijk oppervlakkig te werk gaan hoeft overigens geen verbazing te wekken: bij bijna alle opgaven die kinderen in hun rekenmethodes krijgen aangeboden is deze aanpak immers succesvol. Daar betekent een '-' teken in een opgave bijna altijd dat het kleinste getal van het grootste getal moet worden afgetrokken om de oplossing te krijgen.

Op grond van de resultaten kan verwacht worden dat oefening in strategische vaardigheden een aanknopingspunt biedt voor het verbeteren van prestaties bij het (aanvankelijk) rekenen (vgl. Garofalo, 1987; Verschaffel, 1988). Het moet dan wel gaan om opgaven waarbij gebruik van deze vaardigheden nuttig is. Daarnaast leiden de resultaten tot de hypothese dat zo'n training met name succesvol zal zijn bij kinderen die reeds beschikken over de benodigde vakinhoudelijke kennis, maar problemen hebben met het toepassen daarvan.

NOOT

Dit onderzoek is mogelijk gemaakt door een subsidie van SVO (SVO project 6615)

LITERATUUR

- Beishuizen, J.J. (1983). Al doende leert men problemen oplossen. In Beishuizen, J.J., Hamaker, C., Hout Wolters, B. van & Koster, K.B. (red.) *Onderwijsleerprocessen* (p.75-78). Lisse: Swets en Zeitlinger.
- Berg, W. van den, & Eerde, D. van (1983). *De Kwantiwijzer* (4e experimentele versie). Erasmus

- Universiteit Rotterdam: vakgroep Onderwijssociologie en Onderwijsbeleid.
- Bloemkolk, M. & Bruins, R. (1987a). *Strategisch leren handelen bij het oplossen van rekenopgaven*. Deelrapport 1 van SVO-project 6615. Vakgroep IDOK, R.U. Groningen.
- Bloemkolk, M. & Bruins, R. (1987b). *De vlektoets: een maat voor inzicht in optel- en aftrekopgaven?* Paper Onderwijsresearchdagen 1987. Vakgroep IDOK, R.U. Groningen.
- Bloemkolk, M. & Koster, K.B. (1984). *Aanvankelijk rekenopgaven bij tweede- en derdeklassers: moeilijkheidsgraad en oplossingsstrategieën*. Paper Onderwijsresearchdagen 1984. Vakgroep IDOK, R.U. Groningen.
- Breuker, J. & Linden, E. van der (1983). Het oplossen van Franse grammatika-problemen. In: Beishuizen, J.J., Hamaker, C., Hout Wolters, B. van, & Koster, K.B. (red.) *Onderwijsleerprocessen* (p.113-122). Lisse: Swets en Zeitlinger.
- Campione, J.C., Brown, A.L. & Ferrara, R.A. (1982) Mental retardation and intelligence. In Sternberg, R.J. (ed.) *Handbook of human intelligence* (p.392-490). Cambridge: Cambridge University Press.
- CITO (1979). *Rekentoetsen leerjaar 2, 3 en 4 basisonderwijs*. Arnhem, CITO.
- Chi, M.T.H. (1981) Knowledge development and memory performance. In Friedman, M.P., Das, J.P., & O'Connor, N. (eds.) *Intelligence and learning* (p.221-229). New York: Plenum Press.
- Chi, M.T.H. (1984) Bereichsspezifisches Wissen und Metakognition. In Weinert, F.E., & Kluwe, R.H. (Hrsg.) *Metakognition, Motivation und Lernen* (p.211-232). Stuttgart: Kohlhammer.
- Chipman, S.F. (1985) Introduction to volume 2: Research trends and their implications. In Chipman, S.F., Segal, J.W. & Glaser, R. (eds.) *Thinking and learning skills*. (Vol. 2) Hillsdale: Erlbaum.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.
- Corte, E. de & L. Verschaffel (1980) Kwalitatief-psychologische analyse van het oplossen van aanvankelijke rekenopgaven bij 6- à 8-jarige basisschoolleerlingen. *Pedagogische Studiën*, 57, 383-396.
- Corte, E. de & L. Verschaffel (1983). Onderzoek van onderwijsleerprocessen en de onderwijspraktijk. *Pedagogische Studiën*, 60, (153-164).
- Corte, E. de & L. Verschaffel (1987). Oogbewegingen van eersteklassers tijdens het oplossen van redactie-opgaven. *Pedagogische Studiën*, 64, 137-149.
- Garofalo, J. (1987). Metacognition and school mathematics. *Arithmetic Teacher* (34) p.22-23.
- Jong, T. de (1986). *Kennis en het oplossen van vakinhoudelijke problemen*. Dissertatie T.H. Eindhoven.
- Popping, R. (1983). *Overeenstemmingsmaten voor nominale data*. Dissertatie, R.U. Groningen.
- Sandberg, J.A.C. en Y.F. Barnard (1985). *Probleemoplossen binnen het aanvankelijk rekenen: een analyse van oplossingsprocessen bij open beweringen*. Paper Onderwijsresearchdagen 1985. Utrecht: RU Utrecht.
- Verbeek, A., Kroonenberg, P.M., & Kroonenberg, S. (1983). *User's manual to FISHER*. R.U. Utrecht: Sociologisch instituut.
- Verschaffel, L. (1988). Enkele kanttekeningen bij de 'Proeve ...'. *Tijdschrift voor nascholing en onderzoek van het rekenwiskundeonderwijs* (6) nr.4, p.11-16.

Manuscript ontvangen 22-8-1988

Definitieve versie ontvangen 31-1-1989

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. 13, nr. 2

Mei 1989

Inhoudsopgave

Convocatie ledenvergadering VOR	120
Beknopt verslag verenigingsjaar 1988	121
De financiën van de VOR in 1988	122
Mededelingen	126
Congressen, Symposia, Cursussen, Trainingen	128
Evaluatie van kwaliteit of beoordeling van onderzoekers? <i>Wijnand Th.J.G. Hoebe</i>	131
Ontvangen publikaties	135

Convocatie ledenvergadering VOR

De ledenvergadering zal worden gehouden tijdens de Onderwijs Research dagen op 24 en 25 mei te Leiden. Tijd en plaats worden nader bekend gemaakt. Gezien het belang van agendapunt 6 rekent het bestuur op een grote opkomst van leden.

Agenda

1. Opening
2. Mededelingen/ingekomen stukken
3. Verslag vorige ledenvergadering (zie VOR-bulletin dec. '88)
4. Verslag verenigingsjaar 1988 (zie dit VOR-Bulletin)
5. Jaarrekening 1988 (zie dit VOR-bulletin)
6. OTG/ANGO-problematiek
7. W.v.t.t.k.
8. Sluiting

Beknopt verslag verenigingsjaar 1988

1. Bestuurssamenstelling

Het bestuur was in 1988 als volgt samengesteld:

a. Voorzitter	Prof. Dr. J. Claessen
b. Vice-voorzitter	Mevr. Dr. B. van Hout Wolters
c. Secretaris	Mevr. Dr. 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. Onderwijs Research Dagen

De Onderwijs Research Dagen vonden plaats op 30 en 31 mei in Leuven. Prof. Dr. R. Vandenberghe was voorzitter van de organisatiecommissie. Het aantal deelnemers bedroeg 650. Gast sprekers waren M. Fullan (University of Toronto) en G. Salomon (University of Arizona en Tel Aviv University). De bundels zullen in de loop van 1989 verschijnen.

3. Tijdschrift voor Onderwijs Research/VOR-bulletin

Ook in 1988 was Prof. Dr. B. Creemers redactiesecretaris van het Tijdschrift voor Onderwijs Research. Dr. P. van der Sijde was tot 1 november redactiesecretaris van het VOR-bulletin. Daarna is het secretariaat overgenomen door de secretaris van het VOR-bestuur. Het aantal in 1988 verschenen nummers van TOR/VOR-bulletin bedroeg 5. Het geplande zesde nummer verscheen in januari 1989.

4. VOR-aanmoedigingsprijs

De jury, onder leiding van Prof. Dr. W. van der Linden, kende de prijs van f 1.000,- toe aan Drs. J. F. A. Braster. De titel van het paper luidde: De Rijksinspectie en het Voortgezet Onderwijs. Het paper verscheen als artikel in het Tijdschrift voor Onderwijs Research van december 1988.

5. Gedragsregels voor onderwijsonderzoekers

De werkgroep gedragsregels, bestaande uit mevr. Drs. S. Blaauwboer, mevr. Dr. B. van Hout Wolters, Drs. K. Beintema en Drs. Ph. Groeneweg heeft in 1988 een concept-gedragscode gepresenteerd aan het VOR-bestuur. Na verwerking van de commentaren zal de code in de ledenvergadering van november 1989 worden vastgesteld.

6. Ledenwerving

Het aantal leden bedroeg per 1 januari 1989: 449.

7. Financiën

Zie kasverslag 1988 van de penningmeester van de VOR.

De financiën van de VOR in 1988

Het verslag over 1988 bestaat uit twee delen

I. Rekening en Verantwoording over 1988

II. Balans per 31 december 1988.

I. *Rekening en verantwoording over 1988*

Samengevat zien de inkomsten en uitgaven over 1988 er uit zoals is aangegeven in overzicht 1. Voor de verschillende posten is in drie opeenvolgende kolommen aangegeven welk bedrag was begroot voor 1988, wat het feitelijke bedrag in 1988 was en welk bedrag is begroot voor 1989.

Toelichting op overzicht

1. *Inkomsten*

- 110 *Contributie leden*: Het totaal aantal leden in 1988 bedroeg 460, waarvan 20 student-leden - tegen respectievelijk 490 en 20 in 1987. Hoewel er geen sprake is van een grotere betaaldiscipline van de leden zijn voor 1988 de contributie inkomsten ongeveer zoals was verwacht. Dit komt ondermeer doordat dankzij herhaalde aanmaningen toch nog aardig wat achterstallige contributie wordt betaald. De achteruitgang t.o.v. 1987 in aantal leden is vooral het gevolg van het verwijderen van een aantal leden met een te grote contributieachterstand (meer dan 1 jaar) in de loop van 1988.
- 130 *IJER*: In 1988 hebben 76 VOR-leden gebruik gemaakt van de mogelijkheid om tegen gereduceerd tarief (f 40,-) een abonnement op het International Journal of Educational Research (IJER) te nemen. Hiervan hebben in 1988 70 hun abonnement betaald. De garantiesubsidie van de SVO (in 1988 voor 75 en in 1989 voor 50 abonnementen, alsmede voor 'linguistic editing') leidt tot een klein exploitatiekort.
- 500 *SOR-bijdrage*: Ten behoeve van de jaarlijkse VOR-prijs ontvangt de VOR van de SOR een bijdrage welke gelijk is aan de door de VOR betaalde declaraties. In 1988 is door de SOR geen bijdrage overgemaakt, waardoor er op de SOR nog een vordering is van f 2.977,50 (over 1987) + f 1.000 voor de nog uit te betalen VOR prijs 1988.
- 700 *Royalties TOR*: De VOR ontvangt van Swets en Zeitlinger jaarlijks 10% van de inkomsten van de niet-VOR-abonnementen op het TOR. Deze zijn in 1988 ontvangen voor 1986 en 1987: f 2.675,17.
- 800 *SVO-subsidie*: De VOR ontving in 1988 van de SVO verschillende subsidies. Deze betroffen uitgaven in 1987. Ten behoeve van de TOR is er een jaarlijkse subsidie van f 5000,-, terwijl er voor het IJER, zoals hierboven al werd vermeld, voor 1988, een garantiesubsidie is van maximaal f 5.500,-. Het in 1988 ontvangen bedrag van f 9.370,50 betreft de bijdrage over 1987 betreffende het TOR en de subsidie voor 1987 voor het IJER.
- 910 *Rente*: De rente is wat minder dan begroot, door lagere rentepercentages.
- 920 *Overig*: In 1988 is o.a. een bedrag van f 200,- ontvangen voor het door de VOR geleverde ledenbestand op plaketketten.

2. *Uitgaven*

- 410 *Ledenwerving*: In 1988 is nog *f* 420 uitgegeven voor de ledenwerving in 1987.
- 430 *Werkgroepen*: Er zijn in 1988 geen werkgroepen geweest, die aan de VOR financiële ondersteuning hebben gevraagd.
- 440 *Colloquia*: Er zijn geen uitgaven voor colloquia geweest.
- 510 Voor 1988 is voor 444 leden, 20 studentleden en 50 in de loop van 1988 geroyeerde leden - wegens een contributieachterstand van meer dan 1 jaar - in totaal *f* 41.150,09 betaald aan Swets & Zeitlinger voor het TOR/VOR-bulletin.
- 520 *IJER*: In 1988 is door de VOR een bedrag van *f* 5.314,50 aan de uitgever van het IJER, Pergamon Press overgemaakt voor 150 abonnementen à £ 10,-.
- 600 *Bestuur*: De kosten in 1988 zijn iets hoger uitgevallen dan werd verwacht. Dit komt doordat ten laste van de post bestuur door de VOR *f* 500,- is betaald voor de inschrijvingskosten van twee deelnemers aan het ORD'87 symposium over de programmering van onderwijsonderzoek.
- 800 *Ledenadministratie*: In 1988 is meer uitgegeven dan was gepland, omdat de kosten voor de automatisering van de ledenadministratie in 1987 pas in 1988 zijn betaald. Het betaalde bedrag betreft naast de aanschaf van een PC (*f* 2.580,-) voornamelijk de aanmaak en verzending van de acceptgiro's voor de contributie, alsmede reproductie- en overige portiekosten.
- 900 *Diversen*: In 1988 zijn veel meer diverse kosten betaald dan was verwacht. Dat komt door twee uitgaven: a) *f* 2.862,- als garantiesubsidie voor de bundel over programmering en b) *f* 2.000,- voor de commissie 'beroepscode onderwijsonderzoekers'.

II. Balans per 31-12-1988

activa*)		passiva	
- saldi		- op korte termijn te betalen	
• gewone rekening	<i>f</i> 3.320,63	• afrekening ORD '88	<i>f</i> 2.320,-
• sterrekening	<i>f</i> 8.044,75	• ledenadministratie	<i>f</i> 800,-
• plusrekening	<i>f</i> 3,71	• abonnementen TOR	
• renterekening	<i>f</i> 1.548,63	rest 1988	<i>f</i> 4.000,-
- tegoed van SVO	<i>f</i> 10.000,00	• VOR-prijs 1988	<i>f</i> 1.000,-
- SOR bijdrage	<i>f</i> 3.977,50	• bestuur	<i>f</i> 750,-
- tegoed aan royalties	<i>f</i> 1.250,-	- eigen vermogen	<i>f</i> 20.995,22
- PC	<i>f</i> 1.720,-		
	<i>f</i> 29.865,22		<i>f</i> 29.865,22

*) exclusief: - de nog in voorraad aanwezige exemplaren van de bundel over Programmering,
- een tegoed aan achterstallige contributies van *f* 5.200,-.

Toelichting

- Het 'tegoed van SVO' betreft de bijdrage over 1988 van *f* 5.000,- voor het TOR, alsmede maximaal *f* 5.000,- ten behoeve van het collectieve abonnement van de VOR en linguïstic editing voor het IJER in 1988.
- De SOR-bijdrage betreft de door de VOR in 1987 betaalde declaraties vanwege de VOR-prijs 1985, 1986, en 1987, alsmede de nog te betalen declaraties voor de VOR-prijs 1988.
- De royalties betreffen de 10% die de VOR ontvangt van Swets & Zeitlinger over de niet-VOR-abonnementen op het TOR in 1988.
- De PC welke in 1987 is aangeschaft voor de ledenadministratie wordt over drie jaar afgeschreven. Voor 1988 is 2/3 van de waarde onder activa opgevoerd.
- De afrekening van *f* 2.320,- ORD '88 betreft de levering van 350 programmaboeken (*f* 1.750,-) voor toezending aan niet voor de ORD ingeschreven VOR-leden en *f* 1.570,-

voor een extra bijdrage van *f* 2,75 per ORD bundel door veranderde lay-out voorschriften. Van dit verschuldigde bedrag van *f* 3.320 wordt *f* 1.000 afgetrokken voor de door de VOR verleende diensten aan de ORD organisatie.

- Voor 1988 moet voor de ledenadministratie en financiële administratie (begin 1989) maximaal nog *f* 800,- worden betaald voor door de SCO gemaakte kosten voor de financiële en ledenadministratie.
- Abonnementen TOR: Aan Swets & Zeitlinger moeten voor 1988 nog voor enkele in de loop van 1988 lid geworden gewone en studentleden de verschuldigde abonnementskosten worden voldaan; in totaal verwacht *f* 4.000,-.
- Voor de VOR-prijs '88 moet nog maximaal *f* 1.000,- worden betaald.
- Voor het bestuur in 1988 gemaakte kosten moeten nog reiskosten van één van de bestuursleden en vergaderkosten worden betaald.
- Eigen vermogen: alles bij elkaar rest er op 31-12-1988 een reserve van *f* 20.995,22. In vergelijking met de reserve van *f* 19.578,81 op 31-12-1987 betekent dat een geringe toename. Deze reserve lijkt een goede buffer voor eventuele tegenvallers in de toekomst.

E. Roede (penningmeester)
maart 1989

*Overzicht Inkomsten en uitgaven over 1988
(aangevuld met begroting 1988 en 1989).*

<i>Inkomsten</i>	<i>begr. '88</i>	<i>feitelijk 1988</i>	<i>begr. '89</i>	<i>Uitgaven</i>	<i>begr. '88</i>	<i>feitelijk 1988</i>	<i>begr. '89</i>
100 Contributie				400 Activiteiten			
110 leden	45.150,-	45.943,20	45.150,-	410 ledenwerfing	500,-	420,-	1.000,-
120 student-leden	500,-	225,-	500,-	440 colloquia	1.500,-	-,-	1.500,-
130 IJER	3.000,-	2.800,-	4.000,-	450 ORD prog.boek	1.000,-	-,-	1.500,-
500 SOR-bijdrage	1.000,-	-,-	1.000,-	500 Tijdschriften			
				510 TOR/VOR bulletin			
700 Royalties	1.800,-	2.675,17	1.800,-	511 abonnementen	38.925,-	41.150,09	38.925,-
800 SVO subs.				512 redactie	4.000,-	9.000,-	4.000,-
810 TOR	5.000,-	5.000-	5.000-	515 VOR bulletin	1.500,-	-,-	1.500,-
820 IJER				520 IJER			
821 abonnement	3.000,-	4.370,50	2.000,-	521 abonnementen	6.000-	5.314,50	6.000,-
822 ling. edit.	2.500,-	-,-	2.500,-	522 ling. editing	2.500,-	2.805,-	2.500,-
900 Diversen				600 bestuur	2.000,-	2.284,60	2.000,-
910 Rente	600,-	506,46	600,-	700 VOR prijs	1.000,-	500,-	1.000,-
920 Overig	-,-	370,-	-,-	800 ledenadministratie	2.500,-	3.845,45	2.000,-
				900 diversen	625,-	5.018,15	625,-
Totaal	62.550,-	61.890,33	62.550,-		62.550,-	70.337,79	62.550,-

Mededelingen

VERKLARING INZAKE 'TWEEDE-GELDSTROOMONDERZOEK' BIJ SVO.

Uit kringen van de SVO-bestuursraad vernamen wij, hoogleraren actief in het onderzoek van het onderwijs, dat besloten is de structurering van het onderzoek op initiatief van de SVO vergaand te wijzigen. De kern van deze herstructurering blijkt te zijn:

- 1 opheffing van onderzoeksthema'sgroepen;
- 2 opheffing van beoordelingscommissie voor OTG-onderzoek en de ANGO;
- 3 instelling van een adviescommissie ter programmering van zogenoemd strategisch fundamenteel onderzoek; deze commissie zou zijn opgebouwd uit kringen die niet direct betrokken zijn bij onderwijsonderzoek;
- 4 vervanging van het 'ANGO-onderzoek' door een systeem van 'awards'.

Wij wensen bij deze onze grote ongerustheid uit te spreken over deze voornemens. Deze voorstellen betekenen dat onderwijsresearchers thans geen enkele mogelijkheid meer hebben om op eigen initiatief onderzoeksvoorstellen aan de SVO - als belangrijkste subsidiënt van onderwijsresearch - voor te leggen.

Deze omstandigheid, gekoppeld aan het in één klap wegvagen van een wetenschappelijke infrastructuur (de OTG's) en het in de plaats daarvan creëren van een programmeringsorgaan, waar onderwijswetenschappers grotendeels buiten worden gehouden, vormt een miskennis van de inbreng van het voor de SVO relevante professionele veld, die zijn weerga niet kent.

De in de voorstellen opgenomen procedurebeschrijvingen roepen het beeld op van een heilloze bureaucratie, waarbij de sturing van de programmering van onderwijsresearch sterker dan voorheen bepaald lijkt te worden door het SVO-bureau (wie bepaalt bijvoorbeeld waar de 'awards' heen gaan?).

Wij zijn de mening toegedaan, dat uitvoering van deze voorstellen het wegvallen van een 'tweede geldstroom' voor meer fundamentele onderwijsresearch betekent. Daarmee zou de onderwijskunde het enige wetenschapsgebied worden, waar voor de onderzoekers geen mogelijkheid bestaat eigen onderzoeksvoorstellen te formuleren die door collega-onderzoekers worden beoordeeld. Als de SVO op deze wijze voor 100% louter een subsidiënt van opdrachtresearch zou gaan worden, rest ons - indien deze voornemens niet worden teruggedraaid - niets ander dan met kracht te gaan streven naar een NWO-stichting voor fundamentele onderwijsresearch.

Ondertekend door;

prof. dr. M. Boekaerts;
prof. dr. J. F. M. Claessen;
prof. dr. B. P. M. Creemers;
prof. dr. S. Dijkstra;
prof. dr. W. K. B. Hofstee;
prof. dr. G. Kanselaar;
prof. dr. A. M. P. Knoers;
prof. dr. D. B. Koster;

prof. dr. J. M. G. Leune;
prof. dr. N. A. J. Lagerweij;
prof. dr. W. J. van der Linden;
prof. dr. G. W. Meijnen;
prof. dr. W. J. Nijhof
prof. dr. Tj. Plomp
prof. dr. J. Scheerens.

SVO

t.a.v. Dhr. H. Lodewijks

Swecelincplein 14

2517 CK 's Gravenhage

1 februari 1989

Het VOR-bestuur heeft ontstemd kennis genomen van het besluit van uw bestuur om in de toekomst geen gelden meer beschikbaar te stellen voor OTG/ANGO-onderzoek.

De bezwaren van het VOR-bestuur tegen dit besluit betreffen de volgende:

- de in de loop der jaren moeizaam opgebouwde OTG/ANGO structuur ten behoeve van de ontwikkeling van expertise van onderzoekers in de programmering en uitvoering van fundamenteel onderzoek wordt in één keer afgebroken. Ook de tot stand gekomen samenwerkingsverbanden gaan verloren;
- er is nu volstrekt geen ruimte meer voor onderzoek dat van onderaf wordt geïnitieerd. Daarentegen wordt de invloed van het beleid en van de SVO op de programmering van onderwijsonderzoek nog groter dan reeds het geval was. Kennelijk worden onderwijs-wetenschappers zelf niet meer serieus genomen;
- de op veel plaatsen tot stand gekomen en ook door de SVO gewenste integratie van geldstromen door middel van koppeling tussen voorwaardelijk gefinancierd onderzoek en OTG/ANGO-onderzoek wordt weer afgebroken;
- Het stimuleren en financieren van funderend onderzoek behoort expliciet tot de taakstelling van de SVO (zie ook de WOV). Strategisch beleidsonderzoek is geen funderend onderzoek en behoort dan ook niet te worden gefinancierd uit de voor funderend onderzoek bedoelde 20% van het SVO-budget.

Gezien bovenstaande dringt het VOR-bestuur sterk aan op herziening van de genomen beslissing ten aanzien van de OTG/ANGO-gelden. Als alternatief geeft het bestuur u de suggestie om overheveling van de 20% van het SVO-budget naar de NWO ten behoeve van de vorming van een werkgemeenschap voor onderwijsonderzoek te overwegen.

i.a.a. Min. O&W

OTG's

NWO

LOO

w.g. Dr. M.P.C. van der Werf

secretaris VOR-bestuur

AANVRAGEN SUBSIDIE TEN BEHOEVE VAN SYMPOSIA/CONGRESSEN

Op de begroting van de VOR staat jaarlijks een bedrag ten behoeve van activiteiten van de leden (symposia/congressen, studiedagen e.d.). Van deze subsidie-mogelijkheid wordt nauwelijks gebruik gemaakt. Naar aanleiding van de oproep in het VOR-bulletin van december 1988 is echter een verzoek binnengekomen van het Landelijk Overleg Studievaardigheden tot subsidiëring van de Landelijke dag Studievaardigheden 1989, die als thema heeft 'Economie van het Studeren'. Het VOR-bestuur heeft besloten een subsidie toe te kennen. Bij deze wil het bestuur de leden er nogmaals op attenderen dat ten beoeve van dergelijke activiteiten nog steeds de mogelijkheid bestaat subsidieverzoeken in te dienen. Hiertoe kunt u zich schriftelijk wenden tot de secretaris.

Congressen, Symposia, Cursussen, Trainingen.

HET PROJECT ONDERWIJS EN SOCIAAL MILIEU: EEN EINDBALANS

Congres ter afsluiting van het Rotterdamse O.S.M.-project (1969-1988)

donderdag 5 oktober 1989

Kriterion-zaal, Zalencentrum Engels,
Groothandelsgebouw, Stationsplein 45
te Rotterdam

Deelnemers aan het congres ontvangen de Eindbundel van het O.S.M.-project getiteld:

Het project onderwijs en sociaal milieu: een eindbalans

Deze bundel is geredigeerd door dr. Jan H. Slavenburg en Ton A. Peters; er zijn bijdragen in opgenomen, behalve van de beide redacteuren, van prof. dr. Bert P.M. Creemers, prof. dr. Han M. G. Leune, drs. Paul P. M. Leseman, drs. René R. P. Klaasman, drs. Leendert de Visser.

Na de opening van het congres door de voorzitter van het Bestuur van de Rotterdamse School Advies Dienst, drs. Hans J. Simons (wethouder Onderwijs Rotterdam) zullen als belangrijkste thema's aan de orde komen: de resultaten en opbrengsten van het O.S.M.-project en de betekenis hiervan voor het onderwijsvoorrrangsbeleid; deze resultaten/opbrengsten en aanbevelingen zullen gepresenteerd worden door dr. Jan H. Slavenburg (voormalig projectleider O.S.M., directeur Rotterdamse School Advies Dienst) en Ton A. Peters (voormalig projectleider O.S.M., plv. hoofd stafafdeling Onderwijs gemeente Rotterdam) en van commentaar worden voorzien door prof. dr. Wim G. Meijnen (onderwijs-socioloog) en drs. Jacques Wallage (lid Tweede Kamer).

Vervolgens zal er een forumdiscussie onder voorzitterschap van prof. dr. Han M. G. Leune (onderwijssocioloog) gehouden worden, waarna deze om $\pm$ 16.30 uur het congres afsluit.

De kosten voor deelname zullen f 95,- bedragen, inclusief de O.S.M.-eindbundel (die ter plaatse zal worden uitgereikt) en inclusief een lunch en koffie/thee.

U kunt zich inschrijven door f 95,- over te maken op postgirorekening 21.286 van de Rotterdamse School Advies Dienst, onder vermelding van 'O.S.M.-congres'. (Inschrijving is *uitsluitend* op deze wijze mogelijk.) Teruggave van het gestorte bedrag (bij eventuele verhindering) is NIET mogelijk.

Na ontvangst van uw overschrijving zenden wij u een bevestiging van uw inschrijving.

ONDERWIJS EN GEZIN

ACHTSTE ONDERWIJSSOCIOLOGISCHE CONFERENTIE

De conferentie zal plaatsvinden op 16 en 17 november 1989 in het conferentieoord 'De Blijde Werelt' te Lunteren.

De Stuurgroep Onderwijs sociologie - het interuniversitair samenwerkingsverband van de Nederlandse onderwijs sociologie, waaraan SISWO het inhoudelijk en organisatorisch kader biedt - heeft het initiatief genomen voor deze conferentie.

Evenals bij voorgaande conferenties zal het maximum aantal deelnemers 150 zijn. De conferentie is op de eerste plaats bedoeld voor onderwijs sociologen, maar ook anderen die zich bezig houden met onderzoek en theorievorming over de maatschappelijke achtergronden van het onderwijs zijn van harte welkom.

We denken daarbij o.a. aan onderzoekers die werkzaam zijn op het terrein van onderwijs en etnische minderheden en op het terrein van de gezinssociologie.

Deze conferentie heeft het karakter van een werkconferentie. Naast het bespreken van de conferentiebijdragen in werkgroepen, zullen een aantal plenaire inleidingen worden gehouden.

De conferentie zal worden afgesloten met een bijdrage en discussie over toekomstige ontwikkelingen.

De voorbereidingscommissie bestaat uit:

Drs. H. Kleijer (SISWO) voorzitter

Drs. C. Corver (SISWO) secretaris

Dr. J. Dronkers (KUB)

Dr. M. J. de Jong (EUR)

Dr. A. de Regt (U.v.A.)

Het adres van het secretariaat is:

SISWO

t.a.v. drs. C. Corver

Postbus 19079

1000 GB Amsterdam

Tel. 020 - 240075

STRATEGIC MANAGEMENT OF INFORMATION TECHNOLOGY IN HIGHER EDUCATION

Plaats: Europa Park hotel, Mövenpick, Trier, West-Duitsland

Data : zaterdag 26 en zondag 27 augustus 1989.

A unique management training

In the last decade growing attention has been paid to the important concept of strategic management for higher education institutions. It has become clear that new information technologies will have an enormous impact on the basic processes within higher education institutions.

These two processes, as well as their interrelationship, form the basic foci of attention for the training course 'Strategic management of Information Technology in Higher Education'. In this course both the process of strategic management and the various aspects and applications of information technology will be highlighted. Crucial issues to be discussed during this course are:

- * curriculum development and control,
- * sectorial and institutional reorganization processes,
- * goal formulation,
- * scenario-development,
- * resource management and
- * staff development.

Dates and Location

Saturday 26th and Sunday 27th August, 1989. Location: Europa Parkhotel, Mövenpick, Trier (FRG)

Fee

The costs of the course will be in Dutch guilders, Dfl. 560.- per person, covering participation, papers, lunches and dinner on Saturday evening.

Please remit the fee - in Dutch guilders only - as soon as we send you the invoice and before June 1st at the latest.

Registration

before **June 1th 1989** to:

TILBURG UNIVERSITY

Office for International Relations

attn.: drs. Sylvia G. M. Kokhuis, director

P.O. BOX 90153

5000 LE TILBURG

The Netherlands

tel.: 31-13-662236

telex 52426 KUB

telefax: 31-13-663019

Bitnet: B 180 Meulenbroek @ HTI KUB 5 BITN

Evaluatie van kwaliteit of beoordeling van onderzoekers?

Wijnand Th. J. G. Hoebe

De discussie in het VOR-bulletin over de kwaliteitsbepaling van de wetenschappelijke productie in de onderwijskunde gaat over (eventueel ongewenste effecten van) meetsystemen zonder dat de discussianten zich voldoende rekenschap geven van de doelstellingen en van de consequenties die de doelstellingen voor het systeem moeten hebben. Het voorstel namens het Landelijk Overleg Onderwijskunde van Ax in samenwerking met Creemers (1988) en het tegenvoorstel van Van der Linden (1989) kunnen zodoende nauwelijks rationeel met elkaar worden vergeleken.

Evaluatie en beoordeling

Het gaat mij om het door Hofstee (1987) gehanteerde onderscheid tussen evaluatie-in-engere-zin. Voor het gemak zal ik de termen evaluatie, respectievelijk beoordeling hanteren. Bij evaluatie wordt aan de unit of measurement (bijvoorbeeld onderzoekers, vakgroepen, instituten) iets anders (bijvoorbeeld een beleid of een programma) geëvalueerd; de unit of measurement is anoniem, althans in die zin dat de onderzoeksresultaten geen consequenties voor die unit als zodanig hebben. Bij beoordeling leiden de onderzoeksresultaten tot waardeoordelen over de unit of measurement als zodanig.

De consequenties van het begrip onderscheid zijn verstrekkend. Als de bedoeling voorligt onderzoekers of onderzoeksgroepen te beoordelen, dan dient zo'n beoordeling met de gebruikelijke waarborgen te zijn omkleed. Accenten in taakinterpretatie zijn dan een legitiem onderwerp van discussie. De mate waarin de beoordelingscategorieën te zamen dekkend zijn voor de bedoelingen van diverse actoren, vormt een centraal probleem. De beoordeelde personen of groepen zouden recht van beroep, of minstens op wederhoor dienen te hebben (Hofstee, 1987, pagina 4).

Wanneer de bedoeling voorligt aan de unit of measurement iets anders (bijvoorbeeld een beleid) te evalueren, dan wordt de opzet van zo'n evaluatie gedicteerd door een doelrationele context: de overheid beoogt bijvoorbeeld met voorwaardelijke financiering een bijdrage te leveren aan de kwaliteit van het onderzoek. Die kwaliteit blijkt uit wetenschappelijke productiviteit. De evaluatievraag is dan of voorwaardelijke financiering bijdraagt tot wetenschappelijke productiviteit.

In een evaluatie-context kan een hoge correlatie tussen een eigenlijk, maar ingewikkeld criterium en een oneigenlijke, afgeleide, maar eenvoudige indicator een argument zijn om de eenvoudige indicator als criterium te hanteren. In een beoordelingsperspectief zou een keuze voor de oneigenlijke, maar eenvoudige indicator leiden tot discriminatie op oneigenlijke gronden (Scriven, 1988).

Een onderzoeker of een onderzoeksgroep verbindt zich bovendien tot een bepaalde inspanning of taakstelling, niet tot een resultaat (bij voorbeeld een publiceerbaar artikel). Of de onderzoeksinspanning leidt tot een publiceerbaar artikel hangt vaak ook van andere factoren af. Het eventuele gegeven dat een onderzoeksinspanning niet gepaard gaat met een wetenschappelijke publicatie, kan in een evaluatieperspectief samen met andere van dergelijke gegevens worden genomen als indicatie voor de effectiviteit van een bepaald soort onderzoeksinspanningen; in een beoordelingsperspectief echter dient per te beoordelen geval te worden waargemaakt dat het achterwege blijven van een wetenschappelijke publicatie te wijten is aan degenen die de onderzoeksinspanning verrichten.

Het beoordelingsperspectief

Als Claessen (1988) schrijft over een maatlat 'teneinde kaf van koren te scheiden', en als Ax i.s.m. Creemers (1988) schrijft over consequenties 'in de sfeer van het alloceren van middelen' dan wordt duidelijk vanuit een beoordelingsperspectief over kwaliteitsbepaling gesproken. Ax en Creemers houden zich dan terecht bezig met legitieme accenten in taakinterpretatie en met de vraag of de beoordelingscategorieën te zamen dekkend zijn voor de bedoelingen van diverse actoren. Zij houden zich ten onrechte niet bezig met problemen van beroepsmogelijkheden in hun systeem.

Het voorstel van Van der Linden (1988) is zodanig geformuleerd dat het eveneens binnen een beoordelingsperspectief lijkt te passen: 'alleen publicaties die vooraf door vakgenoten op kwaliteit beoordeeld zijn, komen voor punten in aanmerking. (...) De beoordeling moet volgens de regelen der kunst geschieden'. Als vierde regel der kunst formuleert hij zelfs: 'Er moet een beroepsprocedure mogelijk zijn'. Daarna komt Van der Linden met zijn ei van Columbus: het wetenschappelijke tijdschrift als bestaand beoordelingssysteem dat op deze wijze werkt. De vraag is of Van der Linden vanuit een beoordelend perspectief daarmee een goede vondst heeft gedaan. Uiteraard zorgen redacties voor meerdere onafhankelijk van elkaar beoordelende experts. Omdat ook dergelijke experts blij kunnen geven van een parti pris, voert Van der Linden terecht de eis van een beroepsprocedure in. Welke tijdschriftredactie heeft achter een beroepsprocedure geïnstitutionaliseerd voor auteurs wier manuscripten zijn afgewezen? Welke redactie is bereid zich te laten corrigeren door een onafhankelijke beroepsinstantie? Het lijkt erop dat Van der Linden in de fout gaat, waar hij schrijft dat het wetenschappelijke tijdschrift werkt volgens het voor hem voorgestane beoordelingssysteem.

Daar komt bij dat Van der Linden op geen enkele manier legitieme accenten in taakinterpretatie aan de orde stelt. De door hem voorgestane publicatieprocedure is een voorbeeld hiervan. 'De ideale onderzoeker kiest van meet af aan voor de omvang en het niveau van het tijdschriftartikel als eenheid van publicatie', is een sympathiek uitgangspunt, dat ik zelf ook vaak ingang heb proberen te doen vinden. In het kader van legitieme accenten in taakinterpretatie moet echter de vraag worden beantwoord of elk onderzoek er zich toe leent om gerapporteerd te worden in de vorm van een of meer tijdschriftartikelen als eenheid van publicatie. Het zal vakgenoten niet moeilijk vallen om legio voorbeelden te vinden van onderzoek waarbij een dergelijke rapportagevorm zeer gewrongen en kunstmatig aandoet of zelfs onmogelijk is. Het zou me niet verbazen als een nadere analyse aanleiding zou geven tot de constatering dat er in dit opzicht sprake is van een systematisch verschil tussen onderzoeksgroepen (vakgroepen, instituten). De beantwoording van de vraag of Van der Linden's beoordelingscategorieën te zamen dekkend zijn voor de bedoelingen van diverse actoren, is mede afhankelijk van dit verschil. Legitieme praktijkgerichte accenten in taakinterpretatie waarop door Ax en Creemers wordt gewezen, lijken zich zodoende aan de beoordelingscategorieën van Van der Linden te onttrekken.

Vanuit een beoordelingsperspectief zou dus de voorkeur moeten worden gegeven aan de benadering van Ax en Creemers. Zij stellen accenten in taakinterpretatie aan de orde die in hun visie legitiem zijn. Zij bediscussiëren de vraag of hun beoordelingscategorieën te zamen dekkend zijn voor de bedoelingen van diverse actoren. Hun operationalisering van de criteriumvariabele in een kwantitatief beoordelingssysteem kan echter aan terechte kritiek worden onderworpen. Van der Sijde (1988) en Van der Linden (1989) wijzen op een aantal onwenselijke gevolgen. Het systeem lokt gedrag uit dat leidt tot een hoge productiescore, maar dat weinig te maken heeft met de bevordering van kwaliteit. Onderzoekers en instellingen kunnen hun kwaliteitsscore op een oneigenlijke manier maximaliseren. Het probleem lijkt te liggen in de simpele opsomming van toegekende punten voor elk onderscheiden product. Het Landelijk Overleg Onderwijskunde zou zich eens moeten buigen over de vraag of een compensatorische scoringsregel de voorkeur verdient. Een rapport of

een proefschrift dat tevens verschijnt in de vorm van één of meer samenhangende artikelen in een tijdschrift, hoeft dan niet dubbel te worden geteld. Een rapport of een proefschrift dat om een legitieme reden niet in deze vorm kan worden gepubliceerd, kan dan toch in de kwaliteitsbeoordeling worden betrokken. Een analoge redenering kan worden opgezet voor praktijkgerichte producten (leerplannen, toetsen, beleidsadviezen, e.d.). Als onderzoekers zijn we voldoende in staat om adequate compensatorische beoordelingssystemen uit te werken op het moment dat we overeenstemming hebben bereikt over legitieme accenten in taakinterpretatie en over de dekkendheid van de beoordelingscategorieën. Eventuele onvolkomenheden in het systeem kunnen alsnog worden gecorrigeerd in het kader van procedures voor beroep, c.q. voor hoor en wederhoor. Vanuit een beoordelingsperspectief zou de discussie derhalve dienen te gaan over accenten in taakinterpretatie, over de dekkendheid van beoordelingscategorieën en over beroepsprocedures.

Het evaluatieperspectief

Vanuit een evaluatieperspectief moet anders tegen de voorgestelde systemen worden aangekeken. Vanuit een evaluatieperspectief spelen niet meer doelstellingen als het scheiden van kaf en koren of het alloceren van middelen: in beide gevallen ervaren de units of measurement immers de consequenties aan den lijve. Vanuit een evaluatieperspectief is het van belang vast te stellen wat waaraan wordt geëvalueerd.

Wat wordt geëvalueerd is de vraag naar evaluatie-objecten die niet samenvallen met de units of measurement. Beleids- en programmaevaluatie zijn hierboven als voorbeelden genoemd. Het kan gaan om onderzoeksbeleid, waarbij sommige vakgroepen of instituten andere beleidsmaatregelen treffen dan andere. Het kan gaan om een vergelijking van vakgroepen met instituten die zich exclusief op onderzoek concentreren. Het kan gaan om een vergelijking van programma's met verschillende methodologische uitgangspunten. Het kan gaan om een vergelijking van opdrachtresearch en fundamenteel onderzoek. De voorbeelden hebben met elkaar gemeen dat de evaluatie leidt tot waardeoordelen over iets anders dan over de units of measurement als zodanig.

De vraag waaraan dergelijke programma's worden geëvalueerd, dient uiteraard te worden beantwoord met: aan de effecten ervan. In dat verband moeten verschillende effecten duidelijk van elkaar worden onderscheiden en afzonderlijk meetbaar gemaakt. Om een voorbeeld te noemen: wetenschappelijke productie zou duidelijk moeten worden onderscheiden van praktijkgerichte productie. Ongewenste effecten zouden eveneens afzonderlijk zichtbaar moeten worden gemaakt. De in discussies vaak genoemde oneigenlijke scores op kwaliteitsindicatoren zouden derhalve in een afzonderlijke operationalisering meetbaar moeten worden gemaakt.

Het voorstel van Van der Linden is uitstekend geschikt als een van de effectcriteria in een evaluatieperspectief. Iedereen is het er over eens dat althans de door hem operationeel gemaakte productie in wetenschappelijke tijdschriften een beoogd effect is. De discussie gaat er 'slechts' over of er nog andere beoogde effecten zijn. Het voorstel van Van der Linden kan derhalve uitstekend worden gebruikt om nader vast te stellen objecten te evalueren aan althans dit effect. Het voorstel van Ax en Creemers gooit te veel zaken op een hoop (verschillende soorten productie, gewenste en ongewenste effecten) om op een zinnige manier in de context van effectevaluatie bruikbaar te zijn.

Evaluatieresultaten worden niet gebruikt om consequenties te trekken voor units of measurement, maar om conclusies te trekken over de effectiviteit van programma's en dergelijke. Evaluatieresultaten zijn dientengevolge object van een wetenschappelijke discussie over hun houdbaarheid, net als de resultaten van ander onderzoek. Dientengevolge staat het een ieder vrij om een evaluatie-object te kiezen en onderzoek te doen met behulp van een set van effectcriteria. De resultaten zijn voorwerp van een voortgaande empirische en

eventueel theoretische discussie. Vanuit een evaluatieperspectief is het volstrekt legitiem als Van der Linden met zijn enkelvoudig criterium verschillende vormen van onderzoeksbeleid met elkaar vergelijkt en eventueel constateert dat sommige vormen van beleid meer effect sorteren in termen van zijn criterium. Dit is legitiem zolang hij maar geen consequenties trekt voor personen of groepen die minder effect sorteren in termen van zijn criterium: dan stapt hij over van het evaluatieperspectief naar het beoordelingsperspectief en dan moet zijn systeem aan andere eisen voldoen.

Slotopmerking

Het verschil tussen evalueren en beoordelen is in de discussie over de kwaliteitsbepaling van de wetenschappelijke productie in onderwijskunde onvoldoende doordacht. Te veel wordt gecapitaliseerd op het evaluatieperspectief, terwijl men in feite, gezien de beoogde consequenties, vanuit een beoordelingsperspectief zou moeten redeneren. Vergeleken met evaluatie stelt beoordeling extra eisen en voorwaarden aan het beoordelingssysteem. Aangezien de discussie in feite gaat over kwaliteitsbeoordeling, zou de discussie rekening moeten houden met legitieme accenten in taakinterpretatie, met het dekkingsprobleem van de beoordelingscategorieën, met wenselijke en onwenselijke effecten en met beroepsprocedures. Indien een dergelijke discussie, hetgeen waarschijnlijk lijkt, uitkomt op meervoudige beoordelingscriteria, dan zou een compensatorische combinatie van indicatoren serieuze aandacht verdienen boven de al te eenvoudige sommering.

REFERENTIES

- Ax, J. & Creemers, B. (1988). Kwaliteitsbepaling wetenschappelijke productie in de onderwijskunde. *VOR-bulletin*, 12, 118-121.
- Claessen, J.F.N. (1988). Outputbepaling onderwijswetenschappen op juiste spoor. *VOR-bulletin*, 12, 60-71.
- Hofstee, W.K.B. (1987). Voorwoord in W.Th.J.G. Hoebe. *Evaluatie van onderwijsverzorging, een samenhangend programma van evaluatieonderzoek. Een pre-advies ten behoeve van de Commissie Programmavoorstel Evaluatie Onderwijsverzorging*. 's-Gravenhage, 1987, p. 1-7 (publicatie van het Ministerie van Onderwijs en Wetenschappen).
- Linden, W.J. van der (1989). De meting van wetenschappelijke productiviteit. *VOR-bulletin*, 13, 56-59.
- Scriven, M. (1988). The new crisis in teacher evaluation: the improper use of 'research-based' indicators. *Professional Personnel Evaluation News*. January (1), 4-8.
- Sijde, P.C. van der (1988). Output: kwaliteit of kwantiteit. *VOR-bulletin*, 12, 294-295.

Ontvangen publikaties

- Brugman, D. (1988). Personaliseren van de leerstof. Een interventie-onderzoek naar het ontwikkelen van waardegebieden door leerlingen tijdens lessen sociale wereldoriëntatie. Proefschrift RU Leiden.
- CITO (1988). Balans van het rekenonderwijs in de basisschool.
- Corporaal, A.H. (1988). Bouwstenen voor een opleidingsdidactiek. De Lier: Academisch Boeken Centrum. f 40,-.
- Corte, E. de e.a. (1988). Groeien in onderwijzen nr. 1. (Leuven: Wolters).
- 'Espoir, N.E.J.C., e.a. (1988). Curriculumoverzicht Faculteit der Geneeskunde. Rijksuniversiteit Limburg.
- Groot, J. de (1988). Schoolboeken tussen uitgeverij, onderwijs en overheid. Groningen: Vakgroep Sociologie (doctoraal-scriptie).
- Hegger, W. (1988). Gezondheidseducatie. Een praktische handleiding voor programma-ontwikkeling. Assen/Maastricht: Van Gorcum. f 25,-.
- Hout, J.F.M.J. van (1988). Onderzoekers in opleiding. Een verklaringsmodel voor problemen van promotie-assistenten en assistenten in opleiding. Proefschrift KU Nijmegen. Nijmegen: IOWO.
- Houtkoop, W. & Kamp, M. van der (Red). (1988). Oriëntaties in de volwasseneneducatie. Amsterdam: Swets & Zeitlinger (Onderwijskundige Studies 7).
- Keizer, H. J. (1988). Cognitieve leerdoelen. Handleiding voor het identificeren en beschrijven van cognitieve leerdoelen en het classificeren van toetsvragen. Arnhem: CITO (algemene publikatie nr. 57).
- Kingma, R. (1988). Didactiek agrarisch praktijkschoolonderwijs in Nederland. Deelverslag 3: mas- en has-leerlingen over praktijkweken. Wageningen: Vakgroep Agrarische Onderwijskunde (Agrarisch onderwijskundige reeks nr. 9).
- Koetsier, C.P. en Brinke, J.S. ten (1989). Experiment individuele eindstage - 2 bij de vakgroep didactiek van de biologie. Utrecht: Centrum voor Didactiek van Wiskunde en Natuurwetenschappen.
- Koetsier, C.P., Driel, C. van en Janzen, R. W. (1989). Handleiding voor de beoordeling van de student in de cursus 'oriëntatie op het beroep van leraar'. Utrecht: Centrum voor Didactiek van Wiskunde en Natuurwetenschappen.
- Meijer, C.J.W. (1988). Verwijzing gewogen. Een studie naar de determinanten van verwijzing naar speciaal onderwijs. Proefschrift RU Leiden. Groningen: RION.
- Odink, J.G. e.a. (1988). Onderwijsbeleid: (contra-) productief? Nijmegen: Dekker & Van de Vegt (NVO-publikaties nr. 1).
- Oijen, P.M.M. van (1988). Beleid voor kennis. Bijdrage tot een methodologie van de vraag. Proefschrift RU Utrecht. Utrecht: Vakgroep Planning en Beleid.
- Pijl, S.J. (1988). Het gebruik van diagnostische informatie. Proefschrift RU Leiden. Groningen: RION.
- Pijning, H.F. (1988). Psychologie in het onderwijs van motorische vaardigheden. Groningen: Wolters-Noordhof, f 24,75.
- Sprangers, M. (1988). Response shift and the retrospective pretest. Proefschrift Uv Amsterdam. Den Haag: SVO (Selecta reeks).
- Teyken, C. (1988). Ontwerpen als proces: een exploratief onderzoek naar ontwerpprocessen van beginners vanuit onderwijskundig perspectief. Amsterdam/Lisse: Swets & Zeitlinger.
- Visser, J. de (1988). Acterschap en wetenschap. Constructieprincipes voor onderzoek als transportvorm van ervaringen in het maatschappelijk verkeer. Proefschrift VU Amsterdam. Amsterdam: Vrije Universiteit.
- Vries, B. de (1988). Het leven en de leer. Een studie naar de verbinding van leren en werken in de stage. Proefschrift KU Nijmegen. Nijmegen: ITS.
- Woerden, W. van, Bertels, F. & Blom, Chr. (1988). Onderwijsproject/Projectonderwijs: structurering van onderwijs in projectvorm. Delft: Delftse Universitaire Pers.
- Zuidema, J. (1988). Efficiënt spellingonderwijs. Een leer-en expertmodel voor het spellen. Proefschrift RU Utrecht. Leuven/Amersfoort: ACCO. f 40,75.

serie

Selecta-reeks

uitgegeven door

SVO, Instituut voor onderzoek van het onderwijs.

No. 88

K. de Glopper

Schrijven beschreven

Zorg over de inhoud en opbrengsten van het schrijfonderwijs heeft geleid tot een internationaal onderzoek, het *IEA Written Composition Study*, waaraan ook Nederland heeft deelgenomen. Dit is het verslag van het empirische gedeelte van het onderzoek in Nederland. Het boek beschrijft de feitelijke inhoud van het schrijfonderwijs in de eerste vier leerjaren van het voortgezet onderwijs; de onderzoeksresultaten d.m.v. een nationale peiling in het derde leerjaar voortgezet onderwijs; en de exploratieve bepaling van de relaties tussen onderwijsresultaten, de inhoud van het onderwijs en achtergrondvariabelen.

1989, 176 blz.
ISBN 90 647 2138 6
NUGI 724
f. 24,50

No. 89

P. Lem, S. Veerman
F. Nijssen, E. Roelofs

Omgaan met combinatieklassen

Verslag van een teamgerichte nascholing.

Werken in combinatieklassen blijkt een aantal problemen met zich mee te brengen, het onderzoek waarvan dit boek verslag doet, geeft aan hoe op deze problemen een antwoord kan worden gegeven. Daarvoor is een teamgerichte nascholingscursus opgezet. Op basis van effectief en succesvol gebleken onderwijsvaardigheden krijgen leraren richtlijnen aangeboden om hun onderwijs in combinatieklassen te verbeteren. Nascholing blijkt vooral effectief als deze niet gericht is op de individuele leraar, maar op het gehele schoolteam.

1989, 208 blz.
ISBN 90 647 2140 8
NUGI 724
f. 34,00

SWETS & ZEITLINGER B.V.

Heereweg 347, 2161 CA Lisse, tel. 02521-35111

Niet-lineaire monotone transformaties van de afhankelijke variabele: Wanneer geoorloofd en wanneer gewenst?¹

Charlotte J. de Wolff²

Faculteit der Psychologie, Universiteit van Amsterdam

ABSTRACT

Dependent variable measures may be nonlinearly transformed for use in analysis of variance for a number of reasons, such as to obtain normally distributed data and homogeneous variances, and to cancel out scale-induced interactions. In some cases, however, nonlinear transformations introduce severe problems in the interpretation of the experimental results. In this paper attention is drawn to the benefits and dangers of data transformations. Additionally, a study in which reading times were measured is discussed in order to illustrate the decision process concerning the question whether one should transform or not.

1. INLEIDING

Een transformatie is een vertaling van een verzameling numerieke gegevens in een verzameling andere getallen. De systematische relatie tussen de oorspronkelijke schaal en de getransformeerde schaal wordt de transformatie-functie genoemd. Deze functie is niet-lineair monotoon wanneer de interval-eigenschappen van de gegevens verloren gaan en wanneer de ordinale eigenschappen van de gegevens behouden blijven. Een voorbeeld van een niet-lineaire monotone transformatie is de wortelfunctie ($y = \sqrt{x}$).

Bij het analyseren van de gegevens kan op verschillende manieren gebruik gemaakt worden van niet-lineaire transformaties. Games (1983) maakt een onderscheid tussen drie vormen van transformatiegebruik: transformaties van beschrijvende statistische parameters, transformaties van predictoren, en transformaties van afhankelijke variabelen. In dit artikel beperk ik me tot de laatste categorie. In tegenstelling tot de eerste twee vormen van transformatiegebruik, lopen de meningen over de zin van niet-lineaire transformaties van de afhankelijke variabele uiteen. Dit zal in de loop van het artikel duidelijk worden.

De beslissing om de afhankelijke variabele te transformeren wordt meestal voorafgegaan door een complexe afwegingsprocedure. De onderzoeker behoort gegronde redenen te hebben om de data te transformeren en moet duidelijk beseffen welke gevolgen de transformatie heeft voor de interpreteerbaarheid van de gegevens. In Van den Brink en Koele (1984) wordt een voorbeeld gegeven van een onderzoek waarin een niet-lineaire transformatie van de afhankelijke variabele zinnig is. Dit onderzoek betrof het onthouden van geleerde tentamenstof op verschillende tijdstippen na de tentamen-datum. Er werd gezocht naar het verband tussen de aanwezige kennis R (retentie) en het tijdstip na de tentamen-datum (t). Voor de theoretische waarde van R werd een exponentieel model gepostuleerd: $R = ae^{bt}$ waarbij $b < 0$. Dit model werd op de gegevens gepast. Hiertoe werd de afhankelijke variabele 'retentie' logaritmisch getransformeerd. Dit levert: $\ln R = \ln a +$

¹ Met dank aan T.A.B. Snijders, W.P. van den Brink en G.J. Mellenbergh

² Adres: Faculteit der Psychologie, Universiteit van Amsterdam, Weesperplein 8, 1018 XA Amsterdam

bt. Met behulp van lineaire regressie kon nu naar een best passende functie worden gezocht. De logaritmische transformatie had geen nadelige consequenties voor de interpreteerbaarheid van de afhankelijke variabele 'retentie'.

In de volgende paragraaf zullen de gevolgen die een niet-lineaire transformatie heeft voor de interpreteerbaarheid van de gegevens aan de orde komen. Deze gevolgen zijn afhankelijk van meettheoretische factoren. Vervolgens zal een (geenszins uitputtend) aantal redenen voor niet-lineaire monotone transformaties van de afhankelijke variabele besproken worden. Ik beperk me daarbij voornamelijk tot het specifieke geval dat de gegevens met behulp van variantie-analyse geanalyseerd worden. Tot slot wordt aan de hand van een experiment waarin leestijden werden gemeten geïllustreerd hoe men zich een beeld kan vormen van het nut van niet-lineaire monotone transformaties met betrekking tot een specifiek data-bestand.

Wanneer men concludeert dat het zinvol is de data te transformeren, komt men voor de vraag te staan welke transformatie-functie het meest geschikt is. Er bestaan diverse methoden om te bepalen wat de meest adequate transformatie-functie is met betrekking tot een specifiek data-bestand (b.v. Scheffé, 1959; Smith, 1976; Tukey, 1977; Stoto en Emerson, 1984). Deze methoden zullen hier niet verder aan de orde komen. Geïnteresseerden worden verwezen naar de hierbovengenoemde auteurs.

2. MEETTHEORETISCHE ASPECTEN

In stricte zin zijn data-transformaties alleen dan toegestaan, wanneer de wezenlijke structuur van de gegevens behouden blijft en er geen informatie-verlies optreedt. Een voorbeeld van een toegestane data-transformatie is een lineaire functie van een numeriek systeem op interval-niveau; deze functie behoudt de betekenis van verschillen en de betekenis van verschillen tussen verschillen. Een ander voorbeeld is een monotone niet-lineaire functie van een numeriek systeem op ordinaal niveau; de niet-lineaire functie behoudt de betekenis van verschillen.

In de sociale wetenschappen is het vaak moeilijk vast te stellen of een variabele van ordinaal of interval-niveau is. De variabele heeft bijvoorbeeld wel een kwantitatief karakter, maar het is onduidelijk of verschillen van gelijke grootte dezelfde betekenis hebben. Dit is bijvoorbeeld het geval bij beoordelingsschalen ('rating-scales'). In dergelijke gevallen is de klasse toegelaten transformaties niet scherp gedefinieerd.

In de praktijk komt het nogal eens voor dat men besluit een transformatie op data toe te passen die in bovenstaande stricte zin niet toegestaan is, bijvoorbeeld wanneer een niet-lineaire transformatie wordt toegepast op een variabele van interval-niveau. Deze niet-lineaire transformatie verandert de betekenis van verschillen tussen verschillen. Volgens Henning (1978) is de representatie van de niet-lineair getransformeerde interval-variabele op één van de volgende manieren te interpreteren:

1. De representatie kan als een 'afgeleide' maat beschouwd worden. Wanneer er bijvoorbeeld een reciproke transformatie (d.i. $1/x$) wordt toegepast op reactietijden, is het resultaat te interpreteren als 'snelheid'. De betekenis van de variabele die als 'afgeleide' maat beschouwd kan worden staat op zichzelf; zij hoeft niet afgeleid te worden uit de betekenis van de oorspronkelijke variabele. Het meetniveau van de nieuwe variabele staat dan ook los van het meetniveau van de oorspronkelijke variabele. In dit voorbeeld is de variabele 'snelheid' van minstens interval-niveau. De getallen kunnen dan ook parametrisch getoetst worden.

2. De representatie kan beschouwd worden als 'alternatief' voor de oorspronkelijke afbeelding.

Stel dat er een logaritmische transformatie wordt uitgevoerd op een variabele van interval-niveau, bijvoorbeeld op de variabele 'reactietijd'. De getransformeerde variabele is dan te beschouwen als het resultaat van een isomorfe 'alternatieve' afbeelding waarbij de meetwaarden getransformeerd zijn in logaritmische waarden en de multiplicatieve voorschriften 'getransformeerd' zijn in optellingsvoorschriften. Indien aan de ongetransformeerde scores het volgende multiplicatieve model ten grondslag ligt:

$$X'_{ijk} = \mu' \cdot \alpha'_i \cdot \beta'_j \cdot \alpha\beta'_{ij} \cdot \epsilon'_{ijk}$$

ligt het volgende lineaire model ten grondslag aan de getransformeerde scores:

$$X''_{ijk} = \mu'' + \alpha''_i + \beta''_j + \alpha\beta''_{ij} + \epsilon''_{ijk}$$

waarbij $\mu' = \log \mu'$, enz.

De structuur van het oorspronkelijke numerieke systeem verschilt van die van het systeem waarop is afgebeeld, en beide afbeeldingen hebben verschillende kansverdelingen. De 'alternatieve' representatie heeft, in tegenstelling tot de representatie die als 'afgeleide' maat beschouwd kan worden, geen op zichzelf staande betekenis; de betekenis van de nieuwe afbeelding moet afgeleid worden uit de oorspronkelijke afbeelding. Hoewel de parameters van het multiplicatieve model van de ongetransformeerde scores eenvoudig af te leiden zijn uit die van het lineaire model, ontstaan er problemen wanneer de resultaten van de F-toets van het lineaire model gegeneraliseerd moeten worden naar de scores op de niet-getransformeerde schaal. Doordat beide modellen verschillende kansverdelingen hebben, zijn significante parameters in het lineaire model niet noodzakelijk significant in het multiplicatieve model en vice versa. Er kan derhalve geen betekenis worden gehecht aan de uitkomsten van variantie-analyses op niet-lineair getransformeerde scores die als 'alternatieve' afbeelding worden beschouwd.

3. De representatie kan beschouwd worden als een 'zwakkere' representatie waarvoor zwakkere assumpties gelden.

Na een niet-lineaire monotone transformatie van een interval-schaal zijn de ordinale relaties nog intact. De representatie is dus met behulp van niet-parametrische toetsen nog adequaat te analyseren.

In de praktijk zal men meestal geen genoegen nemen met het feit dat de getransformeerde gegevens niet meer met behulp van parametrische toetsen geanalyseerd kunnen worden, omdat de scores juist getransformeerd zijn om tegemoet te komen aan de assumpties die ten grondslag liggen aan variantie-analyse (zie paragraaf 3.3 en 3.4), of omdat men het noodzakelijk acht om de interval-eigenschappen van de data in de analyse 'mee te nemen' (zie paragraaf 3.5). Derhalve zal men meestal, wanneer de nieuwe afbeelding als 'alternatieve' afbeelding of als 'zwakkere' afbeelding geïnterpreteerd moet worden, liever de oorspronkelijke gegevens analyseren; de niet-lineaire transformatie is gepaard gegaan met te veel informatieverlies.

3. REDENEN OM TE TRANSFORMEREN

3.1 Het lineariseren van functies die niet lineair zijn in de coëfficiënten

In de inleiding werd een voorbeeld beschreven waarbij een exponentieel verband tussen variabelen werd gepostuleerd. Een dergelijk model is op vereenvoudigde manier te toetsen wanneer de afhankelijke variabele zodanig niet-lineair getransformeerd kan worden, dat er een lineair verband tussen de afhankelijke en de onafhankelijke variabele(n) ontstaat, terwijl na de transformatie de verstoringen ongeveer normaal verdeeld zijn en er sprake is van homogene variantie in de eventuele verschillende experimentele groepen (zie

paragraaf 3.3). Na de transformatie kan met behulp van lineaire regressie het model op de gegevens worden gepast, en kan o.a. uit de fractie verklaarde variantie worden afgeleid of het model in overeenstemming is met de gegevens.

3.2 De 'ware' aard van de variabele

Soms leidt een afbeelding van een oorspronkelijke responsmaat in een nieuw numeriek systeem tot een meer 'intrinsieke' weergave van het gedrag dat bestudeerd wordt. Dit is bijvoorbeeld het geval bij een onderzoek naar de snelheid waarmee ratten in een doolhof naar hun voedsel rennen. In plaats van de 'rentijd' kan een snelheids-score worden berekend die de afgelegde afstand per tijdseenheid weerspiegelt. Deze maat heeft als voordeel dat hij relatief onafhankelijk is van de experimentele opzet van het doolhof; daardoor is hij vergelijkbaar met andere snelheidsmaten uit experimenten die op andere wijze opgezet zijn.

Over de 'ware' aard van de variabele is niet altijd even veel bekend. Er wordt dikwijls verondersteld dat de variabele normaal verdeeld is en dat de variabele een homogene variantie bezit wanneer zij gemeten wordt in meerdere groepen. Om deze reden wordt de gemeten variabele in afwijkende gevallen zodanig getransformeerd dat zij deze eigenschappen gaat bezitten. Een plezierig bijverschijnsel van een dergelijke transformatie is dat er tevens wordt voldaan aan de assumpties die aan variantie-analyse ten grondslag liggen. In de praktijk kan dit er echter toe leiden, dat er meer om de geldigheid van statistische assumpties wordt getransformeerd dan om het verkrijgen van een meer waarheidsgetrouwe weergave van de variabele. Hier zal meer over worden gezegd in de volgende paragraaf.

3.3 Statistische voorwaarden

In de praktijk komt het veel voor dat men besluit data te transformeren omdat ze niet voldoen aan de voorwaarden voor het gebruik van een bepaalde statistische toets. Bij variantie-analyse gelden de volgende assumpties ten aanzien van de data:

- onafhankelijkheid van de waarnemingen
- homogene variantie over de groepen
- normale verdeling in de populatie waar de data uit afkomstig zijn.

Onafhankelijkheid van de waarnemingen kan alleen verkregen worden door middel van een correcte experimentele procedure. In designs met herhaalde metingen zijn de verschillende maten vaak in meer of mindere mate afhankelijk. In de 'mixed model'-aanpak van dergelijke designs wordt de assumptie van onafhankelijkheid verzwakt tot die van gelijke varianties en gelijke covarianties (zogenaamde 'compound symmetry'). Bij een dergelijke aanpak kan de heterogeniteit van de varianties gereduceerd worden met behulp van niet-lineaire transformaties. Transformaties voor herhaalde-metingen-designs zijn echter zeer gecompliceerd. Keppel (1973, p. 557) raadt de onderzoeker dan ook aan in deze gevallen een statisticus te raadplegen.

De geldigheid van de tweede en de derde assumptie is afhankelijk van de aard van de afhankelijke variabele. Bij reactietijden gaat dikwijls geen van beide op : vaak bezitten ze een heterogene variantie in verschillende experimentele groepen en zijn ze scheef verdeeld met een lange rechter staart. De heterogene varianties of de scheefheid van de verdeling zijn te verhelpen met behulp van niet-lineaire transformaties, maar het kan moeilijk of zelfs onmogelijk zijn beide euvels op te lossen met één en dezelfde transformatie.

Wanneer er echter een functionele relatie bestaat tussen de celgemiddelden en de varianties, is het mogelijk dat een 'symmetriserende' transformatie tevens een homogeniserende werking heeft op de variantie (zie b.v. Smith, 1976).

Het is de vraag in hoeverre het juist is data te transformeren enkel om de homogeniserende of normaliserende werking. Ten eerste moet men rekening houden met het in de vorige paragraaf besproken probleem van de interpreteerbaarheid van de getransformeerde variabele. Ten tweede zijn schendingen van de homogeniteits- en normaliteits-assumptie niet altijd zodanig rampzalig, dat maatregelen noodzakelijk zijn. Variantie-analyse wordt doorgaans behoorlijk robuust geacht tegen deze beide soorten schendingen, mits er sprake is van gelijke celfrequenties (Scheffé, 1959; Glass, Peckham & Sanders, 1972).

Ten aanzien van heterogene varianties wordt bij gelijke celfrequenties meestal de vuistregel gebruikt, dat de geldigheid van de F-toets weinig wordt aangetast door variantie-schommelingen die kleiner zijn dan een factor vier. Bij ongelijke celfrequenties is het onderscheidingsvermogen en de robuustheid nog steeds behouden in situaties waarin grote steekproeven zijn getrokken uit populaties met kleine varianties. In de overige situaties van ongelijke celfrequenties kunnen er afwijkingen in onderscheidingsvermogen optreden en kan het feitelijke significantie-niveau sterk afwijken van de nominale alfa; dan kan het zinvol zijn om een niet-lineaire transformatie uit te voeren (Budescu & Appelbaum, 1981).

Ten aanzien van schendingen van de normaliteits-assumptie is de situatie onduidelijker. Hoewel men er doorgaans vanuit gaat dat de F-toets conservatief is bij scheve verdelingen, bestaat er enige twijfel in hoeverre dit, met name bij kleine steekproeven, gepaard gaat met een verlies aan onderscheidingsvermogen. Naar aanleiding van een uitgebreid onderzoek naar het effect van schendingen van assumpties van variantie-analyse op fouten van de eerste en de tweede soort concluderen Glass et al. (1972) dat scheefheid een verwaarloosbaar klein effect heeft op zowel het significantie-niveau als het onderscheidingsvermogen van de F-toets.

Games en Lucas (1966) komen eveneens tot de conclusie dat een scheve verdeling op zich geen aanleiding hoeft te zijn om data te transformeren. In een simulatie-onderzoek vergeleken zij de 'power-curves' voor diverse scheve verdelingen met de 'power-curves' voor 'theoretische' normale verdelingen. Hierbij hielden ze de waarde van de niet-centraliteitsparameter Φ (d.i. een parameter die aangeeft in hoeverre experimentele groepen van elkaar verschillen) constant. Games en Lucas merkten op dat '... the use of the appropriate transformation increased the deviation from the theoretical power more often than it decreased it...'. Paradoxaal genoeg was het onderscheidingsvermogen voor scheve verdelingen hoger dan het onderscheidingsvermogen voor normale verdelingen.

Het artikel van Games en Lucas is aanleiding geweest voor een heftige discussie tussen Games enerzijds en Levine en Dunlap anderzijds (Levine & Dunlap, 1982; Games, 1983; Levine & Dunlap, 1983; Games, 1984).

Evenals Games en Lucas, onderzochten Levine en Dunlap in een simulatie-onderzoek in hoeverre het onderscheidingsvermogen voor scheve verdelingen verschilt van het onderscheidingsvermogen voor normale verdelingen. Zij kwamen echter tot een tegenovergestelde conclusie: '...if a transformation can be found that minimizes skew(...) power will be increased' (1982, p. 272). Levine en Dunlap konden tot deze bevinding komen doordat hun werkwijze essentieel verschilde van die van Games en Lucas. Terwijl Games en Lucas gegeven enkele vaste Φ -waarden het onderscheidingsvermogen voor scheve verdelingen vergeleken met het onderscheidingsvermogen voor 'theoretische' normale verdelingen, vergeleken Levine en Dunlap het onderscheidingsvermogen voor scheve verdelingen met het onderscheidingsvermogen voor de bijbehorende, tot een normale verdeling getransformeerde variant. Daardoor werd de niet-centraliteitsparameter Φ niet constant gehouden, maar 'meegetransformeerd'. Dit had tot gevolg dat de gemiddelden en varianties in de getransformeerde verdeling verschilden van de gemiddelden en varianties in de oorspron-

kelijke scheve verdeling. Het resulterende grotere onderscheidingsvermogen voor de getransformeerde data was dan ook niet enkel toe te schrijven aan de vorm van de verdeling, maar ook aan de nieuwe gemiddelden en varianties. Games en Lucas konden hun bevindingen daarentegen geheel toeschrijven aan de vorm van de verdeling, doordat zij vergelijkingen maakten tussen scheef verdeelde en normaal verdeelde data met dezelfde gemiddelden en varianties.

Veel praktische waarde heeft het onderzoek van Games en Lucas niet. Immers, bij het uitvoeren van een F-toets op getransformeerde data zijn de gemiddelden en varianties verschillend van de oorspronkelijke statistische parameters. Maar ook het onderzoek van Levine en Dunlap heeft een beperkte waarde. Zij onderzochten het effect van transformaties van slechts enkele theoretische verdelingen. Hun conclusie heeft dus alleen betrekking op deze klasse verdelingen. Generalisatie naar andere verdelingen is niet zonder meer mogelijk. Er zal veel onderzoek met een verscheidenheid aan verdelingen verricht moeten worden om tot een generalisatie te kunnen komen, daar er tot op heden geen algemeen bewijs gevonden is voor de stelling dat het onderscheidingsvermogen voor scheve verdelingen lager is dan voor normaal-getransformeerde verdelingen (Games, 1984).

Een vervelend verschijnsel in verdelingen van experimentele data dat in deze paragraaf nog niet aan de orde is gekomen is de aanwezigheid van extreem hoge of lage waarden. Een paar van deze uitbijters kunnen de resultaten van analyses onevenredig sterk beïnvloeden. Keppel (1973, p.577) raadt als remedie voor dit euvel aan de data niet-lineair te transformeren. Ik ben het echter met Henning eens wanneer deze stelt dat 'die theoretische Rechtfertigung dieser Empfehlung jedoch aussteht und in ihrer Konsequenz nur zu gewissen kosmetischen Effecten führt' (Henning, 1984, p.246). De invloed van uitbijters kan beter uitgeschakeld worden met behulp van methoden die afkomstig zijn uit de robuuste statistiek (zie Wu, 1985; Jackson, 1986), zoals 'trimming' of 'Winsorization' (zie b.v. Winer, 1971; Conover, 1971).

Samengevat verdient het dus meestal geen aanbeveling om zuiver en alleen uit 'statistische-assumptie-overwegingen' data te transformeren, met name wanneer er sprake is van gelijke celfrequenties of grote steekproeven. Variantie-analyse is bij een design met niet sterk verschillende aantallen waarnemingen in de verschillende cellen robuust ten aanzien van heterogene varianties en vermoedelijk ook ten aanzien van niet-normale verdelingen, hoewel de resultaten van Levine en Dunlap voorzichtigheid gebieden wat betreft deze laatste conclusie. Relevantere overwegingen met het oog op niet-lineaire monotone transformaties zijn respectievelijk de meettheoretische aspecten (zie paragraaf 2) en de overige statistische aspecten die hier beschreven worden (zie paragraaf 3.1 en 3.3 t/m 3.5). Wanneer bij gelijke celfrequenties een transformatie die een homogene variantie en normale verdeling tot gevolg heeft niet tevens een optimalisering van één of meer van deze factoren teweeg brengt, kan deze beter achterwege gelaten worden.

3.4 Heterogene varianties bij contrasten

In de vorige paragraaf kwam het probleem van de heterogene varianties aan de orde. Er werd gesteld dat de heterogene variantie van een steekproef bij gelijke celfrequenties doorgaans geen catastrofale gevolgen heeft voor de grootte van de fouten van de eerste en de tweede soort. Deze stelling behoeft echter enige nuancering.

Stel dat er bij een bepaalde data-verzameling met gebalanceerd design de celvarianties uiteenlopen. De F-toets is dan ongevoelig ten aanzien van de heterogene variantie in die zin dat het gemiddelde onderscheidingsvermogen en de gemiddelde significantie over alle mogelijke contrasten gelijk is aan die van gelijksoortige steekproeven met homogene varianties. Er ontstaan echter problemen wanneer men geïnteresseerd is in één specifiek contrast. Wanneer de analyse van de contrasten uit gaat van een 'pooled-error-term' (dat

wil zeggen het gemiddelde van de varianties binnen alle cellen) worden contrasten tussen cellen met kleine varianties te streng, en contrasten tussen cellen met grote varianties te mild geëvalueerd.

Dit probleem is op te lossen door de error-variantie te baseren op enkel de variantie binnen de in het contrast betrokken cellen. Normaal zou deze error-term minder precies zijn dan de 'pooled error'-term; hij is immers gebaseerd op minder vrijheidsgraden.

Wanneer er een functionele relatie bestaat tussen de celgemiddelden en de varianties binnen de cellen, is er een extra oplossing voorhanden: er kan een homogeniserende transformatie worden uitgevoerd, mits de data op de nieuwe schaal nog interpreteerbaar zijn (zie ook Smith, 1976).

3.5 Interactie-overwegingen

Het begrip 'interactie' is voor tweeërlei uitleg vatbaar: als theoretisch en als statistisch concept. In theoretische zin vormt interactie een formeel bestanddeel van een theorie, bijvoorbeeld de uit de ontwikkelingspsychologie afkomstige 'Aptitude-Treatment-Interaction'. Als statistisch concept komt interactie, in termen van het algemene statistische model, tot uiting in curvilineariteit (bij regressie-analyse), niet-additiviteit (bij variantie-analyse), of verschillende hellingen (bij covariantie-analyse). De statistische interactie is ordinaal of disordinaal, dat wil zeggen dat de trendcurven elkaar respectievelijk niet en wel kruisen (Lindquist, 1953; De Kroon & Van der Laan, 1981).

De grootte van een interactie is te manipuleren door de gegevens niet-lineair te transformeren. Ordinale interacties zijn zelfs geheel te voorschijn of weg te transformeren. Wanneer een theorie een interactie voorspelt en er in de data geen interactie gevonden wordt, kunnen de data vaak zodanig getransformeerd worden, dat er een ordinale interactie optreedt. Vice versa kunnen data vaak zodanig getransformeerd worden dat een gevonden ordinale interactie verdwijnt; dit zou gedaan kunnen worden wanneer de theorie juist geen interactie doet vermoeden.

Dit lijkt een gevaarlijke eigenschap van niet-lineaire transformaties. Immers, een sociale wetenschapper zou de verleiding wel eens moeilijk kunnen weerstaan zijn of haar data zodanig te transformeren dat ze overeenkomen met zijn of haar hypothese. Hoewel dit gevaar niet te elimineren is, zijn er wel enige richtlijnen te geven die, wanneer ze opgevolgd worden, de wetenschapper er voor kunnen behoeden misbruik te maken van transformaties.

In de eerste plaats is er voorzichtigheid geboden bij het niet-lineair transformeren van gegevens van interval-niveau, aangezien dit de betekenis van de interacties verandert. Dit heeft kwalijke gevolgen wanneer de nieuwe representatie als 'alternatieve' afbeelding moet worden beschouwd (zie paragraaf 2). De interacties zijn dan betekenisloos geworden.

Bij gegevens van ordinaal niveau is de klasse toegestane transformaties groter dan bij gegevens van interval-niveau. Het gevaar dat men hier bij het transformeren zodanig te werk gaat dat de kans dat de nul-hypothese verworpen wordt zo klein mogelijk wordt, is dan ook aanzienlijk groter.

Krantz (1974) stelt dat de theorie die aan de (ordinale) gegevens ten grondslag ligt een richtlijn moet zijn voor de transformatie-keuze. Wanneer er een additief statistisch model wordt gepostuleerd, dient er een transformatie te worden gekozen die de interacties minimaliseert. Wanneer de theorie daarentegen een interactie voorspelt, dient er een transformatie te worden gekozen die de interacties maximaliseert.

Dat het beslissingscriterium van Krantz niet altijd tot eenduidige interpretaties leidt, blijkt uit een door Henning besproken voorbeeld uit de leertheorie. In de leertheorie bestaan twee modellen ten aanzien van het verband tussen 'drive' (D) en 'incentive' (I) ten opzichte van 'habit-strength' (H), namelijk:

Hull: Response strength = $H * D * I$

Spence: Response strength = $H * (D + I)$

'Response strength' is een variabele van ordinaal niveau. Afhankelijk van de theorie waar vanuit wordt gegaan, wordt de $D * I$ - interactie als essentieel respectievelijk als artefact opgevat. Het model van Hull vereist een transformatie die een $D * I$ -interactie maximaliseert; het model van Spence daarentegen één die deze interactie minimaliseert. Aan de hand van de empirie is het onmogelijk uit te maken welk model het juiste is, omdat willekeurige monotone transformaties worden toegelaten.

In het bovenstaande werd er steeds vanuit gegaan dat het schaalniveau van de variabele bekend was en er een theorie voorhanden was die het gewenste statistische model dicteerde. Er bestaan echter ook situaties waarin er weinig theoretische of empirische voorkennis aanwezig is over de verbanden tussen de variabelen of het schaalniveau van de variabele. Men doet er dan verstandig aan te kiezen voor een eenvoudig statistisch model; wat dit betreft is een additief model te verkiezen boven een model met interactie (Winer, 1971, p. 397).

Er wordt eveneens voorkeur gegeven aan additieve statistische modellen wanneer er geen theoretische verwachtingen zijn over de interacties, en de interactie-parameter 'confounded' is met de andere parameters, iets wat optreedt bij onvolledige designs. Er is dan in principe geen onafhankelijke schatting van de interactie-parameters mogelijk.

De hier beschreven interactie-overwegingen zijn van groot belang voor de keuze van een juiste transformatie. Een uitgebreide bespreking hiervan is verder te vinden in Henning (1978).

3.6 De grootte van de 'error'-term in verhouding tot de experimentele effecten

Een transformatie kan een verkleinde 'error'-term tot gevolg hebben en zodoende leiden tot een significant hoofdeffect. Een voorbeeld wordt gegeven door Myers (1972, p.176). In zijn experiment met herhaalde metingen doen zich de volgende metingen van reactietijden voor:

	A1	A2	A3	
S1	1.7	1.9	2.0	
S2	4.4	4.5	5.7	$F = 2.87$ (n.s.)
S3	6.6	7.4	10.5	

Een reciproke transformatie geeft de volgende resultaten. De getransformeerde variabele kan als 'snelheid' opgevat worden. De nieuwe representatie kan worden beschouwd als 'afgeleide' maat.

	A1	A2	A3	
S1	.58	.52	.50	
S2	.22	.22	.17	$F = 13.56$ (s., $p = .01$)
S3	.15	.13	.09	

Het hoofdeffect is hier ineens significant, omdat de 'error-variantie' als gevolg van de transformatie een stuk kleiner is geworden.

In speciale gevallen kunnen niet-significante resultaten na een transformatie dus zeer significant worden. Hierin ligt wederom een gevaar voor ongefundeerde transformaties; er zal dan ook buiten de data-set om een rechtvaardiging voor de transformatie moeten bestaan.

4. ILLUSTRATIE

4.1 Het leestijd-experiment

In deze paragraaf worden de afwegingen die gemaakt dienen te worden wanneer men overweegt data niet-lineair te transformeren geïllustreerd aan de hand van een gedeelte van de gegevens uit een experiment door De Wolff (1988). In dit experiment werden op een videoscherm zin voor zin verhalende teksten aan proefpersonen aangeboden. De proefpersonen konden door een knop in te drukken de volgende zin oproepen en zodoende hun eigen leestempo bepalen. De teksten waren vanuit twee verschillende gezichtspunten te lezen. Proefpersonen werden aselekt ingedeeld in drie groepen. Aan twee groepen werd telkens voor het lezen van een tekst medegedeeld vanuit welk perspectief ze de tekst moesten lezen (P1 of P2). De derde groep kreeg een neutrale instructie. Door de leestijd te registreren werd onderzocht in hoeverre het perspectief, een manipulatie van het leesdoel, van invloed is op processen die plaats vinden tijdens het lezen. De teksten waren opgebouwd uit de volgende soorten zinnen:

- zinnen die inhoudelijke relevant zijn voor perspectief 1 (R1-zinnen)
- zinnen die inhoudelijke relevant zijn voor perspectief 2 (R2-zinnen)
- zinnen die inhoudelijk irrelevant zijn voor beide perspectieven, en volgen op R1-zinnen (I1-zinnen)
- deze zinnen waren onderverdeeld in
 - zinnen die een relevant concept voor perspectief 1 bevatten dat eveneens in de vorige zin werd genoemd (I1c+ - zinnen)
 - zinnen die voor geen van beide perspectieven een relevant concept bevatten (I1c- - zinnen)
- zinnen die inhoudelijk irrelevant zijn voor beide perspectieven, en volgen op R2-zinnen (I2-zinnen)
- deze zinnen waren onderverdeeld in
 - zinnen die een relevant concept voor perspectief 2 bevatten dat eveneens in de vorige zin werd genoemd (I2c+ - zinnen)
 - zinnen die voor geen van beide perspectieven een relevant concept bevatten (I2c- - zinnen)
- zinnen die in een neutrale context stonden, inhoudelijk voor geen van beide perspectieven relevant waren, en geen relevante concepten bevatten. (NM-zinnen)

De zinnen hadden ongeveer een gelijke lengte (38-42 letters). Tussen zinnen door stonden ten behoeve van de leesbaarheid af en toe zinnen van afwijkende lengte waarvan de leestijd niet werd geregistreerd.

De waarden die respectievelijk twee maal de standaardafwijking boven en beneden het gemiddelde van enerzijds de leestijden over proefpersonen uit een instructie-groep en over de items uit een materiaalcategorie lagen, werden beschouwd als uitbijters. Deze uitbijters werden vervangen door een waarde die geschat werd op grond van de afwijkingen van enerzijds de leestijden van proefpersonen uit de bijbehorende instructie-groep, en anderzijds de afwijkingen van de leestijden op alle items uit de bijbehorende materiaalcategorie van het algemene gemiddelde van de bijbehorende instructie-groep en materiaalcategorie. Verder werden de data van proefpersonen die in het algemeen zeer veel sneller of langzamer lazen dan de overige proefpersonen verwijderd.

De condities zijn hier geherformuleerd in 'Eigen perspectief' en 'Vreemd perspectief'. Hiertoe zijn de 'Perspectief-instructie-1-materiaal-1'- en 'Perspectief-instructie-2-materiaal-2'-combinaties enerzijds en 'Perspectief-instructie-1-materiaal-2'- en 'Perspectief-instructie-2-materiaal-1'-combinaties anderzijds samengenomen. Als gevolg hiervan zijn er enkel nog binnen-proefpersoon-factoren, te weten Re ('eigen'), Rv ('vreemd'), Ie met daarin genest Iec+ en Iec-, en Iv met daarin genest Ivc+ en Ivc-. Voor het gemak worden in

Tabel 1. Gemiddelde leestijden van één van de zes teksten in msec.

Re	1674
Rv	1574
Iec+	1710
Iec-	1715
Ivc+	1566
Ivc-	1533

de komende paragrafen de neutrale instructie en het neutrale materiaal buiten beschouwing gelaten. De gemiddelde leestijden op één van de teksten staan vermeld in tabel 1.

Uit statistische analyses bleek dat er langere leestijden waren op zinnen die voor het eigen perspectief relevant waren (Re) dan op zinnen die voor het vreemde perspectief relevant waren (Rv). Verder waren de leestijden op irrelevante zinnen die relevante zinnen voor het eigen perspectief opvolgden (Ie) langer dan de leestijden op irrelevante zinnen die relevante zinnen van het vreemde perspectief opvolgden (Iv). Er werd verwacht dat er langere leestijden op Iec+-zinnen (irrelevante zinnen die relevante zinnen voor het eigen perspectief opvolgen en die een relevant concept bevatten) dan op Iec-zinnen en Iv-zinnen zouden zijn. Deze verwachting kwam niet uit¹.

4.2. Niet-lineair transformeren?

4.2.1 Meettheoretische aspecten

De meetschaal van leestijden is minstens van interval-niveau. Verschillen in leestijden zijn belangrijk in de context van de theorie: de onderzoeker is geïnteresseerd in de verschillen in verwerkingstijd in 'real time'. Aan verschillen van gelijke grootte wordt dezelfde betekenis toegekend.

Een niet-lineaire transformatie van de leestijden zou de betekenis van verschillen tussen verschillen veranderen. Tenzij de leestijden reciprook getransformeerd worden, moet de afbeelding die verkregen wordt na een niet-lineaire transformatie van de leestijden als 'alternatieve' afbeelding worden beschouwd (zie paragraaf 2). Dit impliceert dat conclusies die gebaseerd zijn op variantie-analyses van bijvoorbeeld logaritmisch getransformeerde data niet te generaliseren zijn naar de oorspronkelijke schaal: '...effects that are significant in the multiplicative model need not be significant in the additive model and vice versa... Consequently, a logarithmic transformation can not be used when one is interested in the description of mental processes in real time' (Noordman-Vonk, 1977, p.14).

De parametrische analyse van niet-lineair getransformeerde gegevens uit dit onderzoek is dus, ongeacht om welke reden ze ook getransformeerd zijn, niet geoorloofd. Normaliter zou een onderzoeker na dit feit erkend te hebben, niet meer overwegen of er redenen zijn die een niet-lineaire transformatie wenselijk zouden kunnen maken. Desondanks zal, ten behoeve van de illustratie, in de rest van dit artikel besproken worden of er redenen zijn de data niet-lineair te transformeren.

4.2.2 Statistische voorwaarden

Hoewel de celfrequenties uit het experimentele design gelijk zijn, en we zodoende weinig te

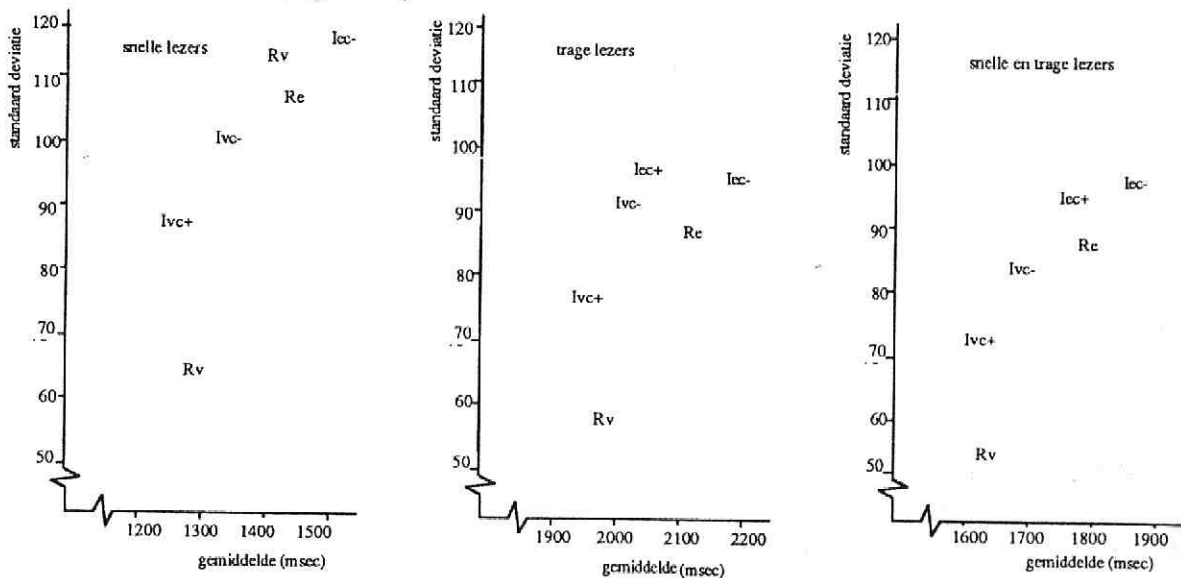
¹ Het zou te ver voeren hier in te gaan op de theoretische achtergrond van de hypothesen. Geïnteresseerden worden verwezen naar De Wolff (1988).

vrezen hebben voor schendingen van de assumpties van variantie-analyse in verband met de robuustheid daarvan, inspecteren we toch de scheefheids- en kurtosis-waarden van de verdelingen van de data om te zien hoeveel ze afwijken van de normale verdelingen. Het blijkt dat de data alle enigszins scheef naar rechts en 'platykurtic' verdeeld zijn. Uit de contouren die door Bowman en Shenton (1975) voor scheefheid en kurtosis zijn opgesteld trekken we echter de conclusie dat de verdelingen van de data niet afwijken van normale verdelingen.

Afwijking van normaliteit is in dit onderzoek dus geen reden om de data te transformeren. Maar hoe zit het met de homogene variantie?

Om een beeld te krijgen van de variantie-homogeniteit maken we een 'spread-versus-level-plot' (Tukey, 1977), waarin we voor het gemak in plaats van de mediaan het gemiddelde en in plaats van de 'fourth spread' de standaarddeviatie uitzetten (zie figuur 1).

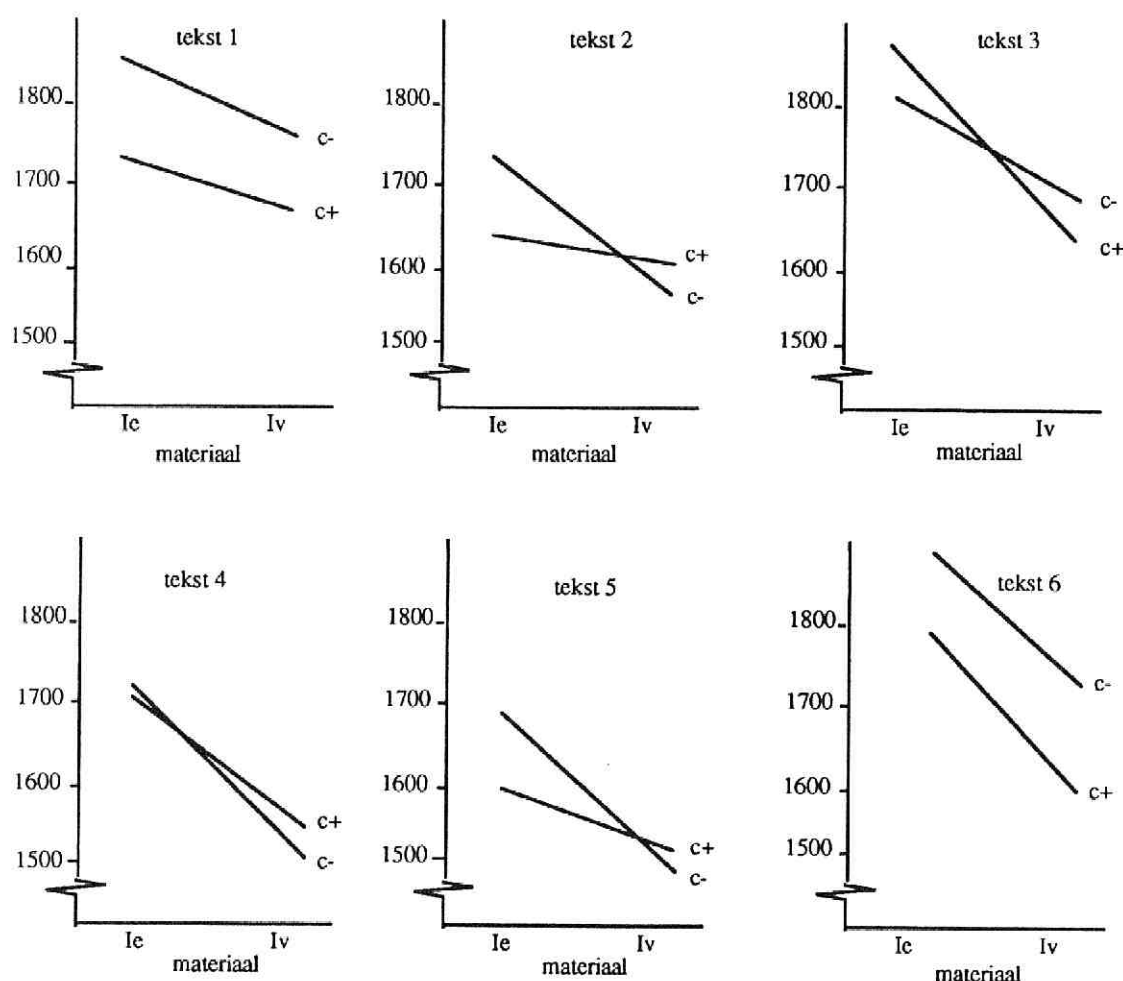
Deze grafieken worden eerst voor de vierentwintig langzamere en vierentwintig snellere lezers afzonderlijk gemaakt, aangezien het niet denkbeeldig is dat de aard van de samenhang voor deze beide groepen verschilt. Hieruit blijkt dat de samenhang voor de snellere lezers sterker is dan voor de langzamere lezers. De trend is echter voor beide groepen gelijk, wat ons doet besluiten om de twee groepen in het vervolg samen te nemen. Voor de gehele groep zien we weer een duidelijk positief verband tussen het gemiddelde en de standaarddeviatie (figuur 1).



Figuur 1. 'Spread-versus-level-plots' van leestijd op materiaalcategorieën

Zoals reeds eerder gesteld is, is heterogene variantie van deze orde van grootte (ongeveer een factor vier) geen reden om de data uit 'statistische-assumptie-overwegingen' te transformeren, gezien het behoud van onderscheidingsvermogen en de robuustheid van variantie-analyse bij gelijke celfrequenties. Het is echter mogelijk dat bij contrasten als gevolg van de heterogene variantie cellen met kleine varianties te streng en cellen met grote varianties te mild worden geëvalueerd (zie paragraaf 3.3). Aangezien alle gevonden effecten zeer significant zijn, zal er weinig veranderen wanneer we een betere schatter hebben van de varianties van de in de contrasten betrokken cellen. Een iets mildere of strengere beoordeling van de celvarianties zal zeer waarschijnlijk onze conclusies onveranderd laten.

Al met al zijn er geen redenen om vanuit statistische-assumptie-overwegingen de data te transformeren; de celfrequenties zijn gelijk en de data zijn nagenoeg normaal verdeeld. Er is wel sprake van heterogene variantie, maar dat is op zich geen aanleiding de data te transformeren.



Figuur 2. Interactie-effecten van de factoren Irrelevant materiaal en Concept.

4.2.3. Interactie-overwegingen

Vanuit de achterliggende theorie werd er verwacht dat de leestijden op Iec+ -zinnen (irrelevante zinnen (I) die een relevante zin voor het eigen perspectief opvolgen (e) en een relevant concept bevatten (c+)) relatief langer zouden zijn dan die op overige I -zinnen. Er werd dus een interactie verwacht tussen de factoren Irrelevant materiaal (niveau's: Ie en Iv) en Concept (niveau's: c+ en c-), namelijk in die zin dat de combinatie Iec+ samengaat met langere leestijden dan de combinaties Iec-, Ivc+, en Ivc-. Er bestonden geen verwachtingen ten aanzien van de hoofdeffecten van de factoren.

In figuur 2 zijn de resultaten van de zes teksten weergegeven.

De leestijden laten, behalve voor tekst 3, geen interacties in de door ons verwachte richting zien. Zoals reeds in paragraaf 3.4 gesteld is, is het mogelijk met behulp van transformaties bepaalde interacties te creëren of te laten verdwijnen, wat echter wel het gevaar met zich meebrengt dat de data naar het gewenste resultaat toe gemanipuleerd worden. In hoeverre is het in dit speciale geval nu mogelijk interacties te creëren?

Wanneer de data met behulp van een logaritmische functie getransformeerd worden, zal bij tekst 1 en 6 een interactie in de gewenste richting ontstaan. Immers, de verschillen tussen de hoge waarden komen na een dergelijke transformatie dicht bij elkaar te liggen.

Het is echter onmogelijk voor tekst 2, 4 en 5 interacties in de gewenste richting te verkrijgen met behulp van transformaties. Bij deze teksten is er namelijk sprake van een

disordinaire interactie in de omgekeerde richting. Deze interacties kunnen niet van 'richting' veranderen na een transformatie, aangezien een monotone transformatie de rangorde van de data intact laat.

Het is in dit onderzoek dus niet mogelijk een transformatie uit te voeren om de verwachte interacties te verkrijgen. Na een logaritmische of wortel-transformatie zou hoogstens in de helft van de gevallen een interactie in de gewenste richting ontstaan: een zeer pover resultaat. Zelfs wanneer de verwachte interacties wel verkrijgbaar zouden zijn door de data niet-lineair te transformeren, zou een transformatie uit louter interactie-overwegingen ongewenst zijn, omdat de leestijden dan niet meer eenduidig te interpreteren zouden zijn. Niet-lineair transformeren zou dan neerkomen op het ongerechtvaardigd omkneden van het data-bestand om de theoretische verwachtingen te doen uitkomen.

5. CONCLUSIE

Aan de hand van een experiment waarin leestijden werden gemeten is geïllustreerd hoe men zich een beeld kan vormen over het nut van niet-lineaire transformaties met betrekking tot variantie-analyse van een specifiek data-bestand. In dit voorbeeld was niet-lineaire transformatie niet geoorloofd: de resultaten van parametrische toetsen op niet-lineair getransformeerde leestijden zouden slecht te interpreteren zijn. Naast het feit dat niet-lineaire transformatie ongeoorloofd was, was het ook niet gewenst: de data voldeden aan de statistische eisen, de vanuit de achterliggende theorie verwachte interacties waren niet te verkrijgen met behulp van een transformatie, en de gevonden effecten dermate statistisch significant, dat een kleinere error-term in verhouding tot de experimentele effecten niet meer van belang was.

Dit voorbeeld kan echter niet representatief geacht worden voor de gang van zaken bij de meeste reactietijd-experimenten. Leestijden en reactietijden zijn bijna altijd scheef verdeeld met een lange rechter staart; vaak wenst men een eenvoudig additief statistisch model, en wil men daarom interacties minimaliseren; en meestal is een kleinere error-term in verhouding tot de experimentele effecten zeer gewenst. Niet-lineaire transformaties zijn in dergelijke gevallen het overwegen waard. Men dient er echter wel op te letten dat het data-bestand na de transformatie nog eenduidig te interpreteren is. Deze laatste vraag komt in de meeste statistische handboeken niet of nauwelijks aan de orde. Lord (1954) en Bennet (1954) geven aan dat de veronachtzaming van de meettheoretische eigenschappen van data door statistici voortkomt uit het feit dat de meettheoretische eigenschappen niet besloten liggen in de getallen, maar er daar buiten om aan zijn toegekend. In het algemeen zijn niet-lineaire data-transformaties slechts dan zinvol, wanneer de data op de getransformeerde schaal eenduidig te interpreteren zijn en de transformatie het meetniveau geen geweld aandoet.

REFERENCES

- Bennett, E.M.(1954). On the statistical mistreatment of index numbers. *The American Psychologist*, 9, 264.
- Brink, W.P. van der & Koele, P.(1985). *Statistiek, Dl. 1: datareductie*. Meppel:Boom.
- Budescu, D.V. & Appelbaum, M.I.(1981). Variance stabilizing transformations and the power of the F-test. *Journal of Educational Statistics*, 6, 1, 55-74.
- Conover, W.J.(1971). *Practical nonparametric statistics*. New York: Wiley.
- Games, P.A.(1983). Curvilinear transformations of the dependent variable. *Psychological Bulletin*, 93, 382-387.
- Games, P.A.(1984). Data transformations, power, and skew: a rebuttal to Levine and Dunlap. *Psychological Bulletin*, 95, 345-347.

- Games, P.A. & Lucas, P.A.(1966). Power of the analysis of variance of independent groups on non-normal and normally transformed data. *Educational and Psychological measurement*, 26, 311-327.
- Glass, G.V., Peckham, P.D. & Sanders, J.R.(1972).Consequences of failure to meet the assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237-288.
- Henning, H.J.(1978). Interaction, transformation, andempirical significance. *Archiv für Psychologie*, 130, 120-138, 236-264.
- Jackson, P.R.(1986). Robust methods in statistics. In: A.D. Lovie, *New developments in statistics for psychology and the social sciences*. New York: The British Psychological Society and Methuen.
- Keppel, G.(1973). *Design and analysis: A researchers handbook*. Englewood Cliffs: Prentice Hall.
- Krantz, D.H.(1974). Measurement theory and qualitative laws in psychofysics. In: D.H. Krantz, R.D.Luce, Atkinson & P.Suppes, *Measurement, psychofysics, and neural information processing*, Vol 2. San Francisco: Freeman & Co.
- Kroon, J. de & Laan, P. van der (1981). Distribution free procedures in two-way lay-outs; a concept of rank-interaction. *Statistica Neerlandica*, 35, 4, 189-211.
- Levine, P.W. & Dunlap, W.P (1982). Power of the F-test with skewed data: should one transform or not? *Psychological Bulletin*, 93, 3, 596-599.
- Levine, P.W. & Dunlap, W.P.(1983). Data transformation, power and skew: a rejoinder to Games. *Psychological Bulletin*, 93, 3, 596-599.
- Lord, F.M. (1954). Further comment on 'Football numbers'. *The American Psychologist*, 9, 264-265.
- Myers, J.L. (1972). *Fundamentals of experimental design*. Boston: Allyn & Bacon.
- Noordman-Vonk, W.(1977). *Retrieval from semantic memory*. Dissertatie. Groningen.
- Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.
- Smith, J.E.K.(1976). Data transformation in analysis of variance. *Journal of Verbal Language and Verbal Behavior*, 15, 339-346.
- Stoto, M.A. & Emerson, J.D.(1984). Powertransformations for data analysis. In: N.B.Tuma (Ed.), *Sociological Methodology*. Jossey-Bass, San Francisco.
- Tukey, J.W.(1977). *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.
- Winer, B.J.(1971). *Statistical principles in experimental design*. New York: Mc Graw-Hill.
- Wu, L.L.(1985). Robust M-estimation of location and regression In: N.B. Tuma (Ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.
- Wolff, Ch. J. de (1988). *Het perspectief van de lezer: selectieve verwerking van informatie uit tekst*. Stage-verslag. Nijmegen: Max Planck-Institut für Psycholinguistik/ Groningen: Instituut voor experimentele psychologie.

Manuscript ontvangen 30-9-1988

Definitieve versie ontvangen 27-1-1989

Functionele schrijfoopdrachten en stelopdrachten: een onderzoek naar de construct-validiteit van schrijfvaardigheidsmetingen¹

Huib van den Bergh²

Rijksuniversiteit te Utrecht, Vakgroep Taalbeheersing

ABSTRACT

In this study the construct validity of two types of writing assignment (functional writing tasks versus traditional essays) is investigated. Students (N = 619) from LBO (Lower vocational education) and MAVO (Lower secondary education) were assigned to five writing tasks (three short functional writing tasks and two traditional (long) essays) and 32 tests in order to measure 16 abilities from the semantic domain of the Structure-of-Intellect model. Results indicate that with both kind of writing tasks the same intellectual abilities are measured. That is to say, only if the writing products are rated in the same way. If an analytic scoring system is used other intellectual abilities are rated than is the case when the writing products are rated in a holistic way.

1. INLEIDING.

In het schooljaar 1985-1986 is het Centraal Schriftelijk Examen (CSE) Nederlands voor LBO- en MAVO-kandidaten gewijzigd. Eén van de veranderingen betrof het onderdeel schrijfvaardigheid. Voorheen moesten MAVO-eindexamenkandidaten een opstel schrijven, terwijl de schrijfvaardigheid van LBO-examinandi alleen in het schoolonderzoek getoetst werd. De schrijfvaardigheid van LBO-kandidaten wordt nu ook in het CSE getoetst; zij moeten drie à vier functionele schrijfoopdrachten maken. De MAVO-kandidaten kunnen (in principe) kiezen tussen een traditionele stelopdracht of dezelfde functionele opdrachten. Onder een functionele schrijfoopdracht verstaan we in dit verband een opdracht die gebaseerd is op een alledaagse taalgebruikssituatie, zoals het schrijven van een (sollicitatie)brief, het invullen van een formulier of het schrijven van een kattebelletje aan huisgenoten (Blok, 1987; Van den Bergh, 1988a).

Voor de functionele opdrachten moeten de leerlingen relatief korte teksten produceren op basis van een gevalsbeschrijving. De traditionele stelopdracht behoeft weinig toelichting; de leerlingen kunnen kiezen uit een aantal onderwerpen (titels) waarover zij één langere tekst moeten produceren.

Over het algemeen is in het onderwijs positief gereageerd op deze veranderingen in het type schrijfoopdracht (Van Schooten & Van den Bergh, 1987). Wel spreken sommige

¹Dit onderzoek is tot stand gekomen dankzij een subsidie van het Instituut voor Onderzoek van het Onderwijs (project: 4075)

²Hierbij wil de auteur Mindert Eiting, Jan Elshout, Gideon Mellenbergh en Gert Rijlaarsdam bedanken voor hun commentaar op een eerder versie van dit artikel.

docenten hun twijfel uit over de vergelijkbaarheid van beide type opdrachten; zo zou bij voorbeeld met de functionele opdracht geen creativiteit gemeten worden.

Ook op basis van onderzoeksresultaten kan men enige twijfel hebben over de vergelijkbaarheid van beide opdrachtvormen. In het algemeen bleek een geringe samenhang tussen schrijfprestaties, als zij werden uitgelokt door verschillende opdrachten (Traxler & Anderson, 1935; Coffman, 1966; Wesdorp, 1974; Lundsteen, 1976; Graves, 1981). Bovendien lijkt de samenhang tussen verschillende schrijfprestaties afhankelijk van de lengte van de te produceren tekst. Behoeft er slechts weinig tekst geproduceerd te worden, dan is de samenhang tussen de prestaties vaak bijzonder laag (Wesdorp, Van den Bergh, Bos, Hoeksma, Oostdam & Triesscheijn, 1985; Blok, 1987). De samenhang tussen prestaties bij 'lange' opdrachten is meestal beduidend hoger (Quellmalz & Capell, 1979; Quellmalz, Capell & Chou, 1982; De Glopper, 1986a; Van den Bergh, Glopper & Schoonen, 1987). Sommige onderzoekers verklaren dit onderzoeksgegeven door aan te nemen dat met 'korte' opdrachten een te kleine steekproef uit het vaardigheidsdomein van de schrijver getrokken wordt, hetgeen in een onbetrouwbare meting resulteert (Quellmalz & Capell, 1979; Van den Bergh e.a., 1987).

Andere verschillen tussen beide opdrachttypen die van belang lijken voor de meting van de schrijfvaardigheid zijn: het publiek is bij de functionele opdracht minder diffuus dan bij de traditionele stelopdracht, maar ook het verschaffen van een situatiebeschrijving kan van invloed zijn, daar de leerling minder informatie behoeft te genereren.

De vraag naar de validiteit van de nieuwe schrijfopdrachten is van belang, daar de doelstellingen voor dit examenonderdeel niet gewijzigd zijn. Immers, wanneer de functionele opdracht niet valide is ten opzichte van het traditionele opstel, dan is onbedoeld een verandering in de vereiste (schrijf)vaardigheden doorgevoerd. In dit onderzoek wordt getracht na te gaan of bij het maken van een functionele schrijfopdracht en het schrijven van een traditioneel opstel dezelfde intellectuele vaardigheden betrokken zijn.

In dit verband is nog een tweede wijziging betreffende het onderdeel schrijven in het CSE van belang. Parallel aan de geschetste veranderingen in type schrijfopdracht is namelijk ook de beoordelingsprocedure gewijzigd. Over de functionele schrijfprodukten wordt geen holistisch oordeel gegeven, zoals bij het opstel gebruikelijk is; functionele schrijfprodukten worden gescoord aan de hand van een analytisch scoringsvoorschrift. Welke intellectuele vaardigheden aan het schrijven toegeschreven kunnen worden is rechtstreeks afhankelijk van hetgeen beoordeeld is. Wordt bij de beoordeling alleen gelet op de inhoud van het schrijfprodukt, dan zullen andere vaardigheden van belang zijn, dan wanneer gelet wordt op de stijl, of het aantal spelfouten, of een combinatie van verschillende aspecten, zoals in een holistisch globaal oordeel.

Eventuele verschillen qua betrokken intellectuele vaardigheden tussen beide typen examenopdrachten kunnen twee oorzaken hebben. De verschillen kunnen inherent zijn aan het type opdracht (functionele versus stelopdrachten), maar ook kunnen zij een gevolg zijn van de veranderde beoordelingsprocedure (scoring versus een holistisch oordeel). Een docent kan bij het beoordelen van de globale kwaliteit van een opstel letten op andere aspecten dan in een scoringsvoorschrift tot uiting komen (vergelijk: Prater & Padia, 1983).

Het doel van het onderhavige onderzoek is ambitieuzer dan louter een uitspraak over de validiteit van de nieuwe schrijfopdracht ten opzichte van de oude schrijfopdracht. We willen ook aan kunnen geven welke intellectuele vaardigheden (eventueel) differentieel betrokken zijn bij het maken van een functionele schrijfopdracht en het schrijven van een traditioneel opstel. Vandaar dat aansluiting gezocht wordt bij een theorie of model over intellectuele vaardigheden. Daar een adequaat model expliciet gericht op schrijfvaardigheden niet voorhanden is (vergelijk: Braddock, Lloyd-Jones & Schoer, 1963; Britton,

Burges, Martin, McLeod & Rosen, 1975; Wesdorp, 1983; Hillocks, 1986), sluiten we aan bij een meer algemeen model voor intellectuele vaardigheden.

Een van de meest genuanceerde modellen voor het intellectuele functioneren is het 'Structure-of-Intellect' model (SI-model) van Guilford (1971, 1979; Guilford & Hoepfner, 1971). In dit vaardigheidsmodel worden 150 verschillende vaardigheden onderscheiden op drie dimensies: operaties, inhouden en produkten. Daar we in dit onderzoek geïnteresseerd zijn in verschillen in intellectuele vaardigheden die te maken hebben met de verwerking van betekenisvol materiaal, beperken we ons tot de semantische vaardigheden uit het SI-model. Hiermee wordt het aantal relevante vaardigheden gereduceerd tot dertig. Deze dertig vaardigheden kunnen we onderscheiden op twee dimensies, namelijk: een operatie-dimensie en een produkt-dimensie. Op de operatie-dimensie worden in het SI-model vijf verschillende operaties onderscheiden. Deze operaties weerspiegelen vijf elementaire intellectuele handelingen. Op de produkt-dimensie zijn zes verschillende produkten gedefinieerd, die aangegeven hoe de te verwerken informatie is georganiseerd (zie: Tabel 1).

Tabel 1. Een overzicht van de dertig semantische vaardigheden uit het SI-model.

Operaties		Produkten	
C	Cognitie	C	Klassen
D	Divergente produktie	I	Implicaties
E	Evaluatie	R	Relaties
M	Geheugen	S	Systemen
N	Convergente produktie	T	Transformaties
		U	Eenheden

Elke combinatie van een operatie, een inhoud (betekenisvol materiaal) en een produkt duidt een unieke vaardigheid aan. Deze vaardigheid kan aangegeven worden met een lettertrigram, waarvan de eerste letter de te verrichten operatie aangeeft, de tweede letter de inhoud (hier: de letter M, van 'seMantisch') en de derde letter het produkt (zie: Tabel 1). Het trigram EMI bij voorbeeld verwijst naar de evaluatie van betekenisvolle implicaties en het trigram CMU duidt de cognitie van betekenisvolle eenheden, of woordkennis aan.

Welke van de dertig vaardigheden uit het semantische deel van het SI-model zijn van belang voor de schrijfvaardigheid? Of meer precies: welke vaardigheden zijn differentieel betrokken bij het maken van een functionele schrijfofdracht en het schrijven van een traditioneel opstel? In de vele studies naar schrijfvaardigheid of -heden wordt zelden een verantwoording gegeven van onderscheiden vaardigheden. Meestal worden tekstuele kenmerken als vaardigheden beschouwd. Aspecten die regelmatig onderscheiden worden zijn: organisatie, inhoud, stijl etcetera. Een uitzondering is het onderzoek van Benton & Kiewra (1986). Zij toonden zowel een verband aan tussen de vaardigheid convergente produktie van semantische systemen (NMS) en zowel directe als indirecte metingen van de schrijfvaardigheid. Ook in onderzoek van Guilford (1971) wordt aannemelijk gemaakt dat in elk geval één van de SI-vaardigheden, namelijk de divergente produktie van betekenisvolle relaties (DMR), bij het schrijven betrokken is.

De relatie tussen te onderscheiden tekstuele kenmerken van schrijfprodukten en intellectuele vaardigheden is moeilijk te leggen. Aanknopingspunten biedt het procesonderzoek. Met name het proces-model van Hayes & Flower (1980; Flower & Hayes, 1980, 1981) kan in dit verband goede diensten bewijzen. In dit schrijfprocesmodel worden drie componenten onderscheiden: het lange termijn geheugen, de taakomgeving en het werkgeheugen. Het feitelijke schrijfproces wordt in het werkgeheugen uitgevoerd, waarbij

voortdurend informatie uit het lange duur geheugen opgehaald wordt en weer in het lange duur geheugen opgeslagen wordt (Shaughnessy, 1977; Daiute, 1981). De schrijver werkt hierbij onder de restricties van de taakomgeving (de schrijfopdracht en de tot dusver geproduceerde tekst).

In het werkgeheugen worden drie macroprocessen onderscheiden: plannen, formuleren en reviseren. In de planningsfase bepaalt de schrijver het doel van de tekst, genereert en organiseert de informatie. Deze processen bouwen voort op de uit het lange duur geheugen gegenereerde informatie. De informatie wordt in het werkgeheugen in een talige vorm gegoten gedurende het formuleerproces. Tijdens de revisiefase moet de schrijver fouten detecteren en corrigeren. Revisie betekent dat de schrijver woorden en zinnen moet controleren op hun grammaticale correctheid, de passendheid in de context, de ambiguïteit, de inhoudelijke adequaatheid etcetera. Gebleken fouten moeten hersteld worden, door bij voorbeeld een ander woord te kiezen, dan wel een geheel nieuwe formulering te bedenken.

Op basis van dit model is het plausibel te veronderstellen dat verschillende semantische SI-vaardigheden van belang zijn. Allereerst lijken de geheugen-vaardigheden cruciaal. Immers, het plan van de te produceren tekst moet op een efficiënte manier in het lange duur geheugen opgeslagen worden, opdat deze informatie makkelijk toegankelijk is bij de verder te plannen en te produceren tekst.

Ook divergente en convergente produktie-vaardigheden lijken van belang bij het schrijven. Deze vaardigheden zijn essentieel wanneer nieuwe informatie uit het lange duur geheugen opgehaald moet worden. Afhankelijk van de opdracht zal er in meer of mindere mate een beroep op divergente produktie-vaardigheden gedaan worden; bij een vrije opdracht meer dan bij een zeer nauw omschreven opdracht. De schrijver perkt door de reeds geproduceerde tekst zijn probleemruimte in, waardoor het mogelijke vervolg van de tekst aan steeds meer restricties onderhevig is. Vandaar dat bij een gevorderde tekst de convergente produktie-vaardigheden van groter belang zullen zijn. Bij de revisie van een tekst zullen naar verwachting evaluatievaardigheden van groot belang zijn. Immers, de schrijver moet de geproduceerde tekst beoordelen op zijn interne consistentie, maar ook moet de geschreven tekst beoordelend vergeleken worden met hetgeen hij tot uitdrukking wilde brengen. Ook een cognitie-vaardigheid als de cognitie van semantische transformaties (CMT) lijkt in deze fase onontbeerlijk. Immers, deze cognitievaardigheid verwijst naar de gevoeligheid voor ambiguïteiten. Voor het herstellen van fouten en onvolkomenheden lijken de convergente produktie van betekenisvolle systemen (NMS) en transformaties (NMT) van belang, als mede de divergente produktie van relaties.

Het spreekt haast vanzelf dat de schrijver over kennis moet beschikken eer hij iets kan schrijven. Dat wil zeggen: de schrijver moet beschikken over kennis van het onderwerp, woordkennis, kennis van de grammaticale structuren, kennis van het te schrijven teksttype, en kennis betreffende het publiek; de mate van efficiëntie waarmee een schrijver toegang heeft tot dergelijke coderingen of representaties bepaalt zijn plaats op het continuum van de verschillende cognitie-vaardigheden.

Bovenstaande houdt in dat alle vijf de operaties waarschijnlijk van belang zijn voor schrijven. Van de produkten lijken vooral de systemen, transformaties en implicaties van belang. Immers, de schrijver moet samenhangende gehelen (zinnen, alinea's en tekstfragmenten; systemen) produceren, onthouden en evalueren. Transformatie-produkten lijken in deze fase van belang, daar de schrijver reeds bestaande formuleringen moet herformuleren. Om een coherente tekst te produceren moet de schrijver rekening houden met de implicaties van de tekst tot dusver. Dit kan op twee momenten: tijdens de reguliere tekstproduktie, in welk geval de convergente produktie van betekenisvolle implicaties als probleemtype de overhand heeft, maar ook achteraf in de revisiefase van de tekst, in dat

geval zou de evaluatie van betekenisvolle implicaties (EMI) van belang kunnen zijn (in paragraaf 2.2.2 wordt een overzicht gegeven van de geoperationaliseerde vaardigheden).

Bovenstaande indeling van het schrijfproces is sterk vereenvoudigd. Er is geen rekening gehouden met het gegeven dat de schrijver de verschillende processen niet in een lineaire volgorde doorloopt (Emig, 1982). Een schrijver zal bij voorbeeld tijdens het formuleren verschillende bewoordingen beproeven en/of bijschaven. Het schrijfproces kan derhalve niet beschouwd worden successievelijke opeenvolging van verschillende processen of vaardigheden. Getracht is met de bovenstaande analyse van het schrijfproces in vaardigheden de belangrijkste probleemttypen per onderscheiden fase aan te geven. Echter, bovenstaande analyse impliceert niet dat andere, niet genoemde, vaardigheden tijdens een bepaalde schrijffase geen rol spelen. Zo zal bij voorbeeld de woordkennis (CMU) ook van belang zijn tijdens de revisiefase, wanneer de schrijver bestaande formuleringen herziet. Ook de divergente produktie van betekenisvolle relaties (DMR) kan tijdens fase van invloed zijn, indien de schrijver frekwent gebruikte woorden wil vervangen door synoniemen. Naar verwachting zullen evaluatie-vaardigheden ook van invloed zijn tijdens het genereren van de benodigde informatie. Immers, de schrijver moet de waarde van deze informatie bepalen ten opzichte van het schrijfdoel, de schrijfo opdracht of de tot dusver geproduceerde tekst (vergelijk het SIPS-model (Guilford, 1971), waarin gedemonstreerd wordt dat dezelfde vaardigheden op verschillende momenten een rol kunnen spelen).

Gezien de verschillen tussen de functionele en de traditionele stelopdracht, verwachten we dat er verschillen zullen zijn in intellectuele vaardigheden, die bij het maken van beide opdrachten een rol spelen. Met name verwachten we verschillen in de divergente produktie-vaardigheden en de cognitie-vaardigheden. Immers, een stelopdracht is een veel vrijere taak dan een functionele opdracht. Bij deze laatste moet de leerling op basis van een gevalsbeschrijving, waarin het leeuwedeel van de te vermelden informatie verschaft wordt, een tekst produceren. Het teksttype van de te produceren tekst is veel beter gedefinieerd bij een functionele opdracht dan bij het traditionele opstelopdracht. Kortom, de leerling moet met veel meer zaken rekening houden, waardoor er van divergente produktie weinig sprake zal zijn.

De leerling moet door de definiëring van de opdrachtsituatie en het teksttype op de hoogte zijn van vele conventies die in de te produceren tekst gewoonlijk gehanteerd worden. Een sollicitatiebrief bij voorbeeld mag qua toon niet familiaar zijn. Door deze verschillen in vereiste kennis tussen beide opdrachttypen wordt verwacht dat cognitie-vaardigheden, met name de cognitie van semantische systeem (CMS) en woordkennis (CMU), van groter belang zijn voor het maken van functionele opdrachten dan voor het schrijven van een traditioneel opstel.

We kunnen pogen ook de consequenties van de veranderde beoordelingsprocedure (scoring versus holistische oordelen) in SI-vaardigheden te vangen. Het zwakke punt van scoringsvoorschriften is dat er geen oordeel verkregen wordt over de coherentie van de gehele tekst. Ook relatief veel ambigue zinnen in een schrijfprodukt lijden niet automatisch tot een lagere score over dit produkt. We verwachten daarom dat de vaardigheden, die betrekking hebben op de coherentie en de ambiguïteit van een tekst (CMT, EMI, MMI, NMI) minder tot uiting komen in scores dan in oordelen.

In oordelen daarentegen zullen eerder zaken positief gewaardeerd worden die te maken hebben met een treffende woordkeus, iets wat zich niet eenvoudig in een scoringsvoorschrift laat vangen. Naar verwachting zullen divergente produktie-vaardigheden, met name DMR, wel blijken uit oordelen, maar niet uit scores over schrijfprodukten.

Rest de vraag wanneer te spreken van een substantieel verschil tussen functionele en stelopdrachten qua betrokken intellectuele vaardigheden? Niet elk verschil behoeft belangwekkend te zijn. En zeker in een studie als deze kunnen licht conclusies getrokken

worden op relatief arbitraire gronden. Reden te meer om een maat voor de grootte van een substantieel effect a priori te specificeren. Het lijkt verstandig de eventuele verschillen tussen opdrachttype te relateren aan de verschillen binnen de opdrachttypen; verschillen in intellectuele vaardigheden tussen de functionele schrijfopdracht en het traditionele opstel worden gerelateerd aan de verschillen in intellectuele vaardigheden tussen opdrachten van hetzelfde type. We zullen derhalve alleen van een substantieel of betekenisvol verschil spreken indien de verschillen in betrokken intellectuele vaardigheden tussen de opdrachttypen groter zijn dan de verschillen tussen opdrachten van hetzelfde type.

2. ONDERZOEKSONTWERP

2.1. Proefpersonen.

Aan het onderzoek werkten 619 leerlingen mee. De leerlingen zaten in negentien derde klassen van twaalf aslect gekozen scholen voor LBO en/of MAVO. De leerlingen hebben in totaal zeventien lesuren, verspreid over het gehele schooljaar 1987/1988, aan het onderzoek besteed; tien lesuren zijn besteed aan het maken van de SI-vaardigheidstoetsen en zeven lesuren aan de schrijfvaardigheidsoopdrachten.

2.2. Instrumentatie.

2.2.1. Schrijfvaardigheidsoopdrachten.

De leerlingen maakten vijf schrijfvaardigheidsoopdrachten, waarvan er vier in eerder onderzoek bij dezelfde leeftijdsgroep succesvol bleken. Drie van de vijf opdrachten (een 'Ooggetuigeverslag' (Blok, 1987), een 'Sollicitatiebriefje' (De Glopper, 1986b) en een 'Formulier invullen' (Van den Bergh e.a. 1986)) waren zogenaamde functionele opdrachten. Dat wil zeggen: deze opdrachten zijn gebaseerd op (schrijf-)activiteiten die in het dagelijkse leven een rol kunnen spelen. Twee opdrachten waren traditionele opstellen (een opstel over 'Het nut van huiswerk' (Rijlaarsdam, 1986) en een 'Vrij opstel').

Voor de drie korte functionele opdrachten moest de leerling op basis van gegeven informatie een (kort) briefje schrijven van ongeveer een halve pagina A4, en op basis van fictieve correspondentie een formulier invullen.

Bij de traditionele opstellen werden veel langere schrijfprodukten vereist (minimaal anderhalf à twee pagina's A4). Voor de eerste opdracht over het nut van huiswerk had de leerling de beschikking over zes pagina's documentatie-materiaal, waar hij gebruik van moest maken. De tweede opdracht was een examenopdracht; uit tien gegeven titels moest de leerling er één kiezen waarover een opstel van minimaal 500 woorden geschreven moest worden.

De functionele opdrachten zijn gescoord door twee personen. Honderd produkten zijn door beiden gescoord, om de kwaliteit van de scoring te controleren. De overeenstemming, geïndiceerd door de correlatie tussen per produkt gesommeerde scores, bleek redelijk (zie: Tabel 2). Naast de overeenstemming tussen beide scorers is ook de homogeniteit van de elementen van het scoringsvoorschrift berekend. Ook deze bleek voor alle drie de scoringsvoorschriften aanvaardbaar (zie: Tabel 2).

De kwaliteit van de langere stelprodukten is bepaald door ervaren docenten Nederlands. Elk stelprodukt is beoordeeld door twee docenten op een tienpuntsschaal, al waren het de produkten van hun leerlingen. De overeenstemming tussen de holistische oordelen van de docenten bleek alleszins redelijk (zie: Tabel 2).

Naast de in Tabel 2 gerapporteerde gegevens betreffende de scoring en beoordeling van respectievelijk de functionele en de traditionele stelprodukten, zijn voor de produkten

Tabel 2. De kwaliteit van de scoring en de holistische docentenoordelen, een betrouwbaarheidschatting (betr), het gemiddelde (gem) en de standaarddeviatie (sd) van het gesommeerde juryoordeel per opdracht en een homogeniteitschatting voor de elementen uit het scoringsvoorschrift (N: aantal beoordeelde schrijfprodukten).

Opdracht	Beoordelingswijze	betr	gem	sd	KR-20	N
Formulier	scoring	.94	11.02	6.43	.80	607
Ooggetuigeverslag	scoring	.89	8.42	4.23	.79	614
Sollicitatiebriefje	scoring	.96	10.95	3.62	.75	619
Het nut van huiswerk	holistisch oordeel	.68	10.60	2.98	-	586
Vrij opstel	holistisch oordeel	.67	10.90	2.79	-	581

behorende bij drie schrijfoopdrachten ('Ooggetuigeverslag', 'Sollicitatiebriefje', 'Het nut van huiswerk') ook schaaloordeelen verzameld (zie: Tabel 3). Dat wil zeggen de beoordelaars, twee per beoordeeld aspect, hadden de beschikking over 'opstelschalen' (zie: Blok & Hoeksma, 1984; Van den Bergh & Rijlaarsdam, 1987). Zij moesten gebruik maken van in kwaliteit oplopende voorbeeldopstellen waarvan de waardering gegeven was. De taak van de beoordelaars is in principe gereduceerd tot het plaatsen van het te beoordelen opstel op de opstelschaal, waarna de bijbehorende waardering als het ware afgelezen kan worden. Met behulp van de opstelschalen zijn drie aspecten van de schrijfprodukten beoordeeld: Globale kwaliteit, Taalgebruik, en Inhoud en organisatie. De overeenstemming tussen de verschillende beoordelaars bleek redelijk (zie: Tabel 3).

Tabel 3. Betrouwbaarheidschattingen (betr), gemiddelden en standaarddeviaties van het gesommeerde jury-oordeel (gem) voor de schaaloordeelen (N: aantal beoordeelde schrijfprodukten).

Opdracht	Beoordeeld aspect	betr	gem	sd	N
Ooggetuigeverslag	globale kwaliteit	.93	170.38	37.08	614
	taalgebruik	.92	179.81	32.09	614
	inhoud & organisatie	.93	179.11	34.25	614
Sollicitatiebriefje	globale kwaliteit	.93	204.07	30.78	619
	taalgebruik	.91	186.78	32.25	619
	inhoud & organisatie	.95	200.89	27.67	619
Het nut van Huiswerk	globale kwaliteit	.78	195.02	105.43	586
	taalgebruik	.77	187.66	98.93	586
	inhoud & organisatie	.80	195.27	110.27	586

Uit Tabel 2 en 3 blijkt dat de schrijfprodukten over het algemeen redelijk betrouwbaar beoordeeld en gescoord zijn, zij het dat de overeenstemming tussen de verschillende beoordelaars iets lager is voor de beide opstellen dan voor de korte functionele opdrachten.

2.2.2. SI-vaardigheidstoetsen

Ten behoeve van dit onderzoek zijn 32 intellectuele vaardigheidstoetsen geselecteerd uit een batterij van 47 vaardigheidstoetsen (Van den Bergh, 1989). Het belangrijkste criterium hierbij was de gemeenschappelijke variantie met de andere toetsen voor dezelfde vaardigheid. Tezamen appeleren deze 32 deelvaardigheidstoetsen aan zestien vaardigheden uit het semantische domein van het SI-model; elke vaardigheid wordt door twee toetsen geïndiceerd. In Tabel 4 wordt een overzicht en een abstracte omschrijving gegeven van de zestien intellectuele vaardigheden welke in dit onderzoek een rol spelen.

Bij de constructie van deze vaardigheidstoetsen is zoveel mogelijk aangesloten bij eerder onderzoek, waar vergelijkbare operationalisaties adequaat gebleken zijn (zie onder

Tabel 4. Een overzicht en een omschrijving van de geoperationliseerde intellectuele vaardigheden.

Vaardigheid	
Cognitie van semantische systemen (CMS)	Het begrijpen van relatief gecompliceerde ideeën, waarvan de constituerende elementen een soort samenhang, ordening of patroon vertonen.
Cognitie van semantische transformaties (CMT)	Het zien van mogelijke herinterpretaties van gegeven woorden, objecten e.d.
Cognitie van semantische eenheden (CMU) Divergente produktie van semantische relaties (DMR)	Woordkennis. Het produceren van vele passende relaties tussen een gegeven begrip en andere begrippen.
Divergente produktie van semantische eenheden (DMU)	Het produceren van verschillende ideeën die aan dezelfde (eenvoudige) eis voldoen.
Evaluatie van semantische implicaties (EMI)	Het vergelijken beoordelen van de implicaties van betekenisvolle stimuli.
Evaluatie van semantische systemen (EMS)	Het vergelijken beoordelen van complexere gehelen van objecten of ideeën.
Evaluatie van semantische eenheden (EMU)	Het vergelijken beoordelen van woorden of een eenvoudige zin(sdeel) ten opzichte van een standaard.
Geheugen van semantische implicaties (MMI)	Het onthouden van logische of pragmatische implicaties.
Geheugen van semantische systemen (MMS)	Het onthouden van gestructureerde verzamelingen concepten.
Geheugen van semantische transformaties (MMT)	Het onthouden van veranderde, herziene of geherdefinieerde informatie.
Geheugen van semantische eenheden (MMU)	Het onthouden van niet samenhangende betekenisvolle begrippen (woorden).
Convergente produktie van semantische implicaties (NMI)	Het deduceren van informatie die geïmpliceerd is in de aangeboden (begripsmatige) informatie.
Convergente produktie van semantische relaties (NMR)	Het produceren van de best passende relatie tussen gegeven begrippen.
Convergente produktie van semantische systemen (NMS)	Het koppelen van (componenten van) objecten, concepten of woorden gericht door een gegeven doelstelling tot een samenhangend geheel.
Convergente produktie van semantische transformaties (NMT)	Het produceren van een unieke oplossing door gegeven informatie te herdefiniëren of the herinterpreteren.

andere: Nihira, Guilford, Hoepfner & Merrifield, 1962; Guilford & Hoepfner, 1971; Kamstra, 1971; Elshout, 1976; Kunst, 1978; Hoeks, 1985; Schoonen & Wolf, 1984).

De correlatieve structuur van de toetsscores bleek beter aan te sluiten bij een SI-model dan bij alternatieve modellen voor de structuur van intellectuele vaardigheden (Van den Bergh, 1989). Dat wil zeggen: een SI-model met gecorreleerde vaardigheidsfactoren paste redelijk bij de geobserveerde correlaties ($gfi = .98$; $agfi = .96$; $RMR = .04$.) In Tabel 5 wordt een overzicht gegeven van deze 32 toetsen en de geschatte communaliteiten.

Uit Tabel 5 blijkt dat de geschatte communaliteiten van de vaardigheidstoetsen over het algemeen bevredigend is. Ook de geschatte correlaties tussen de vaardigheidsfactoren

Tabel 5. De communaliteitsschattingen van de 32 SI-vaardigheidstoetsen.

Toets	SI-vaardigheid	Communaliteit
Familie relaties	CMS	.632
Transitiviteiten	CMSS	.578
Andere betekenissen	CMT	.626
Woordplaatjes vertalen	CMT	.544
Nauwkeurigheid van betekenissen	CMU	.646
Woordkennis	CMU	.669
Gerichte associaties	DMR	.762
Getals-associaties	DMR	.412
Gebruiken genereren	DMU	.579
Problemen genereren	DMU	.619
Inferenties	EMI	.507
Syllogismen	EMI	.604
Matrijzen evaluatie	EMS	.630
Titels verzinnen	EMS	.582
Algemene uitspraken	EMU	.812
Zinnige zinnen	EMU	.823
Naam en dier	MMI	.845
Stad en produkt	MMI	.863
Gegevens onthouden	MMS	.616
Volgorde onthouden	MMS	.904
Dubbele betekenissen onthouden	MMT	.388
Gelijkklinkende woorden onthouden	MMT	1.000
Plaatjes onthouden	MMU	.716
Woorden onthouden	MMU	.741
Ontbrekende schakels	NMI	.440
Volgorde associaties	NMI	.878
Tegenstellingen	NMR	.661
Vocabulaire completie	NMR	.869
Tijdvolgorde	NMS	.368
Zinsvolgorde	NMS	.563
Object synoniemen	NMT	.663
Ongebruikelijke voorwerpen	NMT	.679

bleken niet extreem hoog, alleen de correlatie tussen NMS en NMT week in dit opzicht af ($r = .95$). De gemiddelde correlatie tussen de vaardigheidsfactoren werd geschat als .22.

Het is in het kader van dit artikel ondoenlijk om alle intellectuele vaardigheidstoetsen uitgebreid te beschrijven. We beperken ons daarom tot de twee toetsen voor de vaardigheid evaluatie van betekenisvolle implicaties (EMI). De vaardigheid EMI verwijst naar het vergelijkend beoordelen van implicaties of extrapolaties ten opzichte van een standaard. De twee (geselecteerde) toetsen ter meting van deze vaardigheid zijn: 'Inferenties' en 'Syllogismen'.

Voor de toets 'Inferenties' moet de leerling de geldigheid van drie conclusies evalueren gegeven een feitelijke uitspraak. De conclusies zijn zo geformuleerd, dat alle drie op het eerste gezicht min of meer plausibel lijken. De leerling moet derhalve de verschillende alternatieven steeds met de gegeven uitspraak vergelijken om tot het goede antwoord te komen. Eén van de tien items uit deze toets luidde:

Honderd jaar geleden was men gemiddeld gezien korter dan tegenwoordig.

1. De laatste honderd jaar zijn veel korte mensen geëmigreed.
2. Lange ouders hebben meer kinderen dan korte ouders.
3. Jan is waarschijnlijk langer dan zijn grootvader.

De tweede toets om de vaardigheid EMI te meten bestond uit acht syllogismen. Een syllogisme is een geldige redenering bestaande uit twee premissen en een conclusie. De leerling moest ter beantwoording van de vragen steeds kiezen uit drie sterk op elkaar gelijkende conclusies, waardoor evaluatie van de gegeven conclusies ten opzichte van de beide gegeven premissen noodzakelijk is. Een voorbeeld:

Sommige huizen zijn niet verwarmd.

Alle huizen zijn gebouwen.

1. Alle gebouwen zijn huizen.
2. Sommige gebouwen zijn niet verwarmd.
3. Verwarmde gebouwen zijn huizen.

Voor beide toetsen moet de leerling de alternatieven evalueren ten opzichte van de implicaties van de gegeven premissen. Vandaar dat beide toetsen de vaardigheid EMI vertegenwoordigen.

In het hierboven gegeven voorbeeld worden bij beide toetsen driekeuze-vragen gesteld. Dit is lang niet bij alle toetsen het geval. Integendeel, er is gestreefd naar een zo groot mogelijke variatie in toetsvorm; open vragen, twee-, drie-, vier-, vijf-, en zelfs zeskeuze-vragen. Meestal zijn verschillende vraagvormen gebruikt bij de toetsen die één vaardigheid beoogden te meten, alleen bij de evaluatie- en divergente produktie-vaardigheden is dit niet het geval. Dit is echter inherent aan de te meten vaardigheden. De vragen bij de zes toetsen die evaluatie-vaardigheden meten zijn alle meerkeuze-vragen, terwijl de vier divergente produktie-toetsen enkel open vragen bevatten.

3. VOORGENOMEN ANALYSES

De onderzoeksvraag, doen het maken van een functionele opdracht en het schrijven van een traditionele opstel een beroep op dezelfde intellectuele vaardigheden, verdient nog enige uitwerking. Deze vraag kan op verschillende manieren opgevat worden, zoals in de inleiding gesuggereerd is. Voor de generalisatie naar het CSE is het van belang dat er een vergelijking gemaakt wordt tussen de intellectuele vaardigheden zoals die tot uiting komen in de docentenoordeelen over de stelprodukten en de scores van de functionele opdrachten. Vinden we een verschil, dan kunnen we dit niet zondermeer aan verschillen in type opdracht toeschrijven, daar een gevonden verschil met evenveel recht aan de verschillende beoordelingsprocedures toegeschreven kan worden. Er worden daarom twee vragen onderscheiden:

1. Wordt met de drie functionele schrijfoopdrachten (in gelijke mate) een beroep gedaan op hetzelfde complex SI-vaardigheden als met de beide traditionele stelopdrachten, wanneer de eerste *gescoord* worden met behulp van een analytisch scoringsvoorschrift en de tweede door ervaren docenten *beoordeeld* worden? (In deze analyses worden dus alleen de analytische scores over de functionele schrijfprodukten en de oordelen van de docenten over de traditionele stelprodukten betrokken). De analyses die betrekking hebben op deze vraag noemen we de 'score-oordeel-analyses'.
2. Doen functionele schrijfoopdrachten en traditionele stelopdrachten (in gelijke mate) een beroep op hetzelfde complex van intellectuele deelvaardigheden, wanneer de produkten van beide opdrachttypen op *dezelfde* manier beoordeeld worden (in deze analyses worden derhalve alleen de schaaloordeelen betrokken voor de opdrachten

'Ooggetuigeverslag', 'Sollicitatiebrief' en 'Het nut van huiswerk'). De analyses die betrekking hebben op de beantwoording van deze onderzoeksvraag noemen we verder de 'oordeel-oordeel-analyses'.

In deze beide analyses kunnen de geschatte regressies van de schrijfvaardigheidsoordelen op de intellectuele vaardigheidsfactoren moeilijk geïnterpreteerd worden, daar deze vertroebeld worden door de gehanteerde beoordelingsprocedure. Immers, in de eerste analyse wordt de oppositie functionele scores - docentenoordelen geanalyseerd, en in de tweede analyse kunnen de resultaten vertekend worden doordat slechts één type oordelen (schaaloordelen) in de analyse betrokken is. Daarom wordt een extra analyse uitgevoerd, waarbij ook aan de schrijfvaardigheidsoordelen een structuur opgelegd wordt. Door ook aan de geobserveerde schrijfvaardigheidsvariabelen een structuur op te leggen kan met verschillende variantiebronnen rekening gehouden worden.

3. In de derde analyse stellen we ons de vraag: uit welke intellectuele componenten schrijfvaardigheid is opgebouwd. 'De schrijfvaardigheid' definiëren we operationeel als de gemeenschappelijke variantie van alle veertien schrijfvaardigheidsoordelen (zie: Tabel 2 en 3). In het vervolg zullen we de analyses die betrekking hebben op deze derde onderzoeksvraag de 'schrijfvaardigheidsanalyses' noemen.

De analyses worden uitgevoerd met LISREL (Jöreskog & Sörbom, 1986). We kunnen dan de relatie tussen de geobserveerde schrijfvaardigheidsoordelen en de schrijfvaardigheidsfactoren weergeven in een lineair model. In dit model wordt met y een kolomvector met geobserveerde schrijfvaardigheidsoordelen aangeduid, terwijl η een kolomvector met schrijfvaardigheidsfactoren aangeeft, met ongecorrleerde elementen η_i . De relatie tussen y en η kan dan gedefinieerd worden als:

$$y = \Lambda_y * \eta + \epsilon, \quad (1)$$

waarin met Λ_y een matrix is, met parameters op gekozen posities die corresponderen met de (veronderstelde) structuur van de schrijfvaardigheidsfactoren. Deze parameters weerspiegelen de regressies van de geobserveerde variabelen y op de latente factoren (η). De laatste term, ϵ is een kolomvector met residuen, met variantie Θ_ϵ .

De relatie tussen de geobserveerde SI-vaardigheidsscores en de SI-factoren komt in Vergelijking 2 aan de orde. In deze vergelijking wordt met x een kolomvector met geobserveerde intellectuele vaardigheidsscores aangegeven, terwijl ξ een kolomvector is met zestien factoren, corresponderend met de zestien onderscheiden SI-vaardigheden.

$$x = \Lambda_x * \xi + \delta. \quad (2)$$

In Vergelijking 2 is Λ_x een matrix met regressiegewichten van de geobserveerde intellectuele vaardigheidsscores op de latente factoren. δ is een kolomvector met residu-varianties op de diagonaal. Ook nemen we aan dat ϵ en δ niet covariëren. De correlaties tussen de SI-factoren ten slotte kunnen geschat worden in de matrix Φ . De regressie van de SI-factoren op de schrijfvaardigheidsfactoren tot slot kunnen we specificeren als:

$$\eta = \Gamma * \xi + \zeta. \quad (3)$$

Hierbij wordt met de matrix Γ een matrix aangeduid met regressiegewichten (γ_{ij}) van de schrijfvaardigheidsfactoren, gespecificeerd in de vector η , op de SI-vaardigheidsfactoren,

gespecificeerd in de vector ξ . Zeta (ζ) tot slot is een kolomvector met residuele-scores.

De drie onderzoeksvragen kunnen getoetst worden door restricties aan de matrix Γ op te leggen. De analyses voor de drie onderzoeksvragen zijn in principe hetzelfde. Vandaar dat we deze uitgebreid toelichten aan de hand van de eerste onderzoeksvraag, en voor de tweede en derde onderzoeksvraag slechts de verschillen met de eerste aanstippen.

Voor de eerste onderzoeksvraag, waar de regressie van de docentenoordelen over de beide stelprodukten op de SI-factoren vergeleken wordt met de regressie van de gescoorde prestaties op de drie functionele opdrachten, kunnen drie modellen getoetst worden. Voor de 'score-oordeel-analyses' is de matrix Γ een matrix van 5 X 16 regressiegewichten (γ_{ij}), daar vijf schrijfvaardigheidsfactoren (i ; één per opdracht) en zestien SI-vaardigheidsfactoren (j) onderscheiden worden. In het eerste model veronderstellen we dat er geen verschil is in regressiegewichten tussen de vijf opdrachten. We veronderstellen:

$$\gamma_{1j} = \gamma_{2j} = \gamma_{3j} = \gamma_{4j} = \gamma_{5j} \quad (j = 1, 2, \dots, 16) \quad (4)$$

Bovendien geldt, onder de restrictie dat de variantie van η_i gefixeerd is op één, dat wanneer alle regressiegewichten aan elkaar gelijk zijn ook de proporties verklaarde variantie van de vijf schrijfvaardigheidsfactoren aan elkaar gelijk zijn. De proporties onverklaarde variantie zijn derhalve ook gelijk voor de verschillende opdrachten. De tweede restrictie is dan:

$$\text{var } \zeta_i = \text{var } \zeta_{i'} \quad (i, i' = 1, 2, \dots, 5; i \neq i')$$

Dit model waarin noch verschillen binnen opdrachttype als tussen beide typen schrijfopdrachten toegestaan worden, kunnen we heel goed het 'geen-verschil-in-regressies-tussen-schrijfproduktenmodel', of kortweg 'gelijke-regressies-model' noemen.

In het tweede model versoepelen we de restricties enigszins. We veronderstellen dat de regressies van de SI-factoren wel gelijk zijn voor de beide traditionele stelopdrachten 'Het nut van huiswerk' en 'Het vrije opstel' ($i = 1, 2$) en gelijk voor de drie functionele opdrachten 'Formulier', 'Ooggetuigeverslag' en 'Sollicitatiebriefje' ($i = 3, 4, 5$). We kunnen deze restricties schrijven als:

$$\gamma_{1j} = \gamma_{2j}, \gamma_{3j} = \gamma_{4j} = \gamma_{5j}, \quad (j = 1, 2, \dots, 16) \quad (6)$$

en

$$\text{var } \zeta_1 = \text{var } \zeta_2, \text{var } \zeta_3 = \text{var } \zeta_4 = \text{var } \zeta_5. \quad (7)$$

In het tweede model worden dus geen verschillen in intellectuele vaardigheden toegestaan tussen produkten die uitgelokt zijn door hetzelfde type schrijfopdracht. Dit model noemen we daarom het 'geen-verschil-binnen-opdrachten-van-hetzelfde-type-model', of korter het 'opdrachttype-model'.

In het derde en laatste model leggen we geen restricties op aan de γ -parameters (en de varianties van de storingstermen ζ); alle regressiegewichten mogen vrij geschat worden. Dit model is duidelijk het minst restrictieve van de drie. Het wordt, omdat alle regressiegewichten vrij geschat mogen worden, het 'verschil-tussen-schrijfopdrachten-model' genoemd. In het vervolg zullen we aan dit model kortweg refereren als het 'verschil-model'.

Door de passing van deze drie (geneste) modellen te vergelijken kan het belang van de verschillende restricties gekwantificeerd en vergeleken worden. Het verschil in passing

tussen het 'opdrachttype-model' (met gelijke regressies per opdrachttype) en het 'verschil-model' (waarin alle regressiegewichten vrij geschat mogen worden) is onze maat voor de effectgrootte. Indien het verschil in passing tussen deze beide modellen kleiner is dan het verschil in passing tussen het 'gelijke-regressies-model' (gelijke regressies op alle schrijffactoren) en het 'opdrachttype-model' dan concluderen we dat er sprake is van een substantieel verschil in intellectuele vaardigheden betrokken bij het schrijven van een opstel en het maken van een functionele schrijfp opdracht.

Wel moet nog een opmerking gemaakt worden betreffende de schatting van de parameters uit beide meetmodellen (zie: Vergelijking 1 en 2). De parameters uit deze modellen worden namelijk niet geschat, doch gefixeerd op eerdere schattingen. De Λ_y -parameters worden gefixeerd op enerzijds de wortel uit de beoordelaarsbetrouwbaarheden en anderzijds de op de wortel uit de overeenstemming tussen de scoorders (zie: Tabel 2; de residuele varianties van de geobserveerde variabelen worden gefixeerd op: 1 - de betrouwbaarheid, waardoor de varianties van de η -variabelen gefixeerd zijn op één).

De parameters uit het tweede meetmodel worden eveneens gefixeerd. De Λ_x -parameters en de Θ_δ -parameters worden gefixeerd op de waarden geschat in een eerdere oplossing (zie: Tabel 5; Van den Bergh, 1989). Hierdoor hebben de verschillen tussen de beoordelaars geen storende invloed op de analyses; in feite hebben de analyses betrekking op de ware-oordelen (de oordelen gecorrigeerd voor de beoordelaarsbetrouwbaarheden).

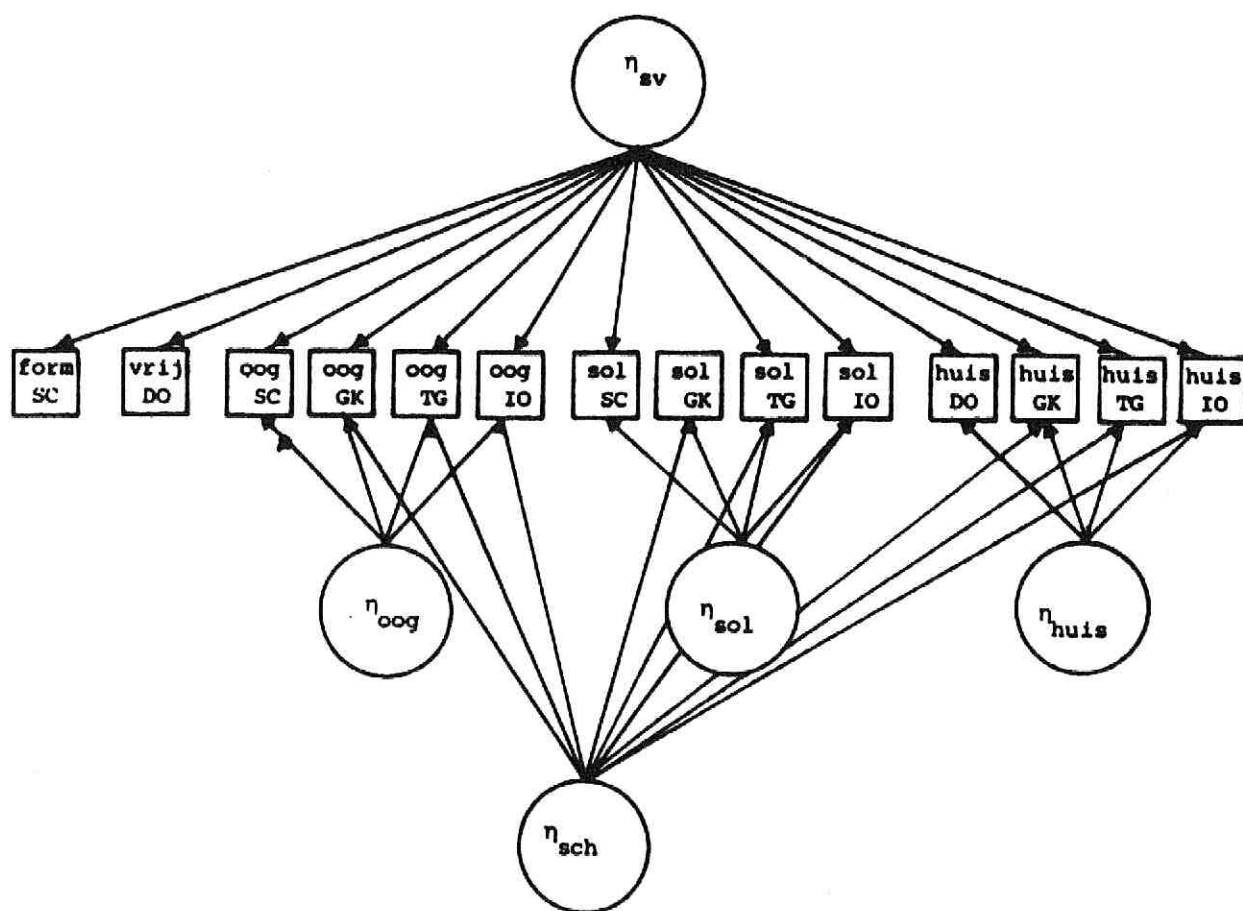
Modeltoetsingen worden vaak gekritiseerd wegens de grote aantallen modellen die gepast worden voor een acceptabele 'fit' verkregen wordt; er wordt dan te veel op kans gekapitaliseerd. Dit probleem is getracht te omzeilen door de te passen modellen van te voren te specificeren. Bovendien splitsen we bij de passing van de modellen de verzameling van intellectuele vaardigheidstoetsen in twee delen. Deze splitsing geschiedt zodanig dat elke SI-vaardigheid door één van de twee vaardigheidstoetsen vertegenwoordigd wordt. Elk model wordt dan tweemaal getoetst, waardoor kruisvalidatie op de variabelen verwezenlijkt wordt. De te trekken conclusies dienen te gelden voor beide variabelenverzamelingen. Alleen de te presenteren parameterschattingen zullen betrekking hebben op alle vaardigheidstoetsen, daar de standaardfout van de schattingen kleiner is naarmate de SI-factoren door meer toetsen geïndiceerd worden.

Hoewel de structuur van de opeenvolgende modellen voor de tweede onderzoeksvraag nagenoeg identiek is, zijn er enkele kleine verschillen die genoemd moeten worden. Bij de toetsing van de modellen voor de 'oordeel-oordeel-analyses' zijn drie schrijfvaardigheidsvariabelen betrokken en derhalve drie schrijfvaardigheidsfactoren. Het betreft de schaaloordeelen over de globale kwaliteit van het opstel over 'Het nut van huiswerk', het 'Ooggetuigeverlag', en het 'Sollicitatiebriefje' (zie: Tabel 3). De regressiegewichten van de geobserveerde schrijfvaardigheidsoordelen op de drie schrijfvaardigheidsfactoren worden gefixeerd op de wortel uit de interbeoordelaarsbetrouwbaarheid (zie: Tabel 3). De matrix Γ is bij deze analyses dus van de orde 3 maal 16.

Bij de toetsing van de modellen voor de derde onderzoeksvraag spelen alle veertien geobserveerde schrijfvaardigheidsvariabelen een rol (zie: Tabel 2 en 3). Het grote voordeel van deze 'schrijfvaardigheidsanalyses' is dat 'de' schrijfvaardigheid operationeel gedefinieerd kan worden als de gemeenschappelijke variantie van alle scores en oordelen. De schrijfvaardigheidsfactor aldus gedefinieerd bevat noch beoordelingsmethodespecifieke noch opdrachtspecifieke variantie, waardoor de regressies op 'de' schrijfvaardigheid bestudeerd kunnen worden.

Deze analyse bestaat uit twee delen. Eerst is getracht inzicht te verkrijgen in de samenhang tussen de veertien geobserveerde schrijfvaardigheidsvariabelen (Van den Bergh, 1988b) en vervolgens wordt de relatie tussen deze schrijfvaardigheidsfactoren en de SI-factoren onderzocht.

Om inzicht te verkrijgen in de samenhang tussen de schrijfvaardigheidsvariabelen, of beter de verschillende (gesommeerde) oordelen, zijn eerst enkele modellen getoetst waarin verschillende structuren aan de correlaties tussen de oordelen opgelegd werden (Van den Bergh, 1988b). Het meest wenselijke model met drie schrijfvaardigheidsfactoren, één voor elk onderscheiden aspect (globale kwaliteit, taalgebruik, en inhoud en organisatie), bleek zeer slecht bij de geobserveerde correlaties te passen ($X^2 = 2835.23$; $df = 73$; $agfi = .39$). Plausibel bleek een model waarin onderscheiden werden: één algemene schrijfvaardigheidsfactor, drie opdrachtspecifiekefactoren en één factor voor alle aspecten die met behulp van schaaloordeelen beoordeeld waren (zie: Figuur 1). Dit model bleek redelijk bij de geobserveerde correlaties tussen de jury-oordelen te passen ($X^2 = 204.23$; $df = 56$; $agfi = .92$). Uit de parameterschattingen bleek dat de opdrachtspecifiekefactoren van groot belang zijn, terwijl de schrijfvaardigheidsfactor slechts een gering deel van de variantie in scores en oordelen naar zich toe trekt (Van den Bergh, 1988b)



Figuur 1. Een grafische weergave van een model dat redelijk bij de geobserveerde correlaties tussen de veertien geobserveerde schrijfvaardigheidsvariabelen past η_{oog} , η_{sol} en η_{huis} zijn opdrachtspecifiekefactoren voor respectievelijk het 'Ooggetuigeverslag', het 'Sollicitatiebriefje' en het opstel over het nut van huiswerk; η_{sch} is een beoordelingsmethodespecifiekefactor voor alle aspecten die met behulp van opstelschalen beoordeeld zijn.

Bij de 'schrijfvaardigheidsanalyses' worden in principe weer dezelfde drie modellen getoetst (het 'gelijke-regressies-', het 'opdrachttype-' en het 'verschil-model'). Bij deze drie modellen worden de regressies van de SI-factoren op de algemene schrijfvaardigheidsfactor (η_{sv}) en de regressies op de schaalbeoordelingsfactor (η_{sch}) in alle modellen vrij geschat.

Tabel 6. De passing van de drie modellen per variabelenverzameling per toetsing (gfi: goodness of fit index; agfi: adjusted goodness of fit index; RMR: gemiddelde gekwadrateerde residuen som).

Analyses: Var. verzameling: Model	scores-oordelen			oordeel-oordeel			schrijfvaardigheids		
	I	II	II	I	II	II	I	II	II
	gfi	agfi	RMR	gfi	agfi	RMR	gfi	agfi	RMR
Gelijke-regr.	.949	.946	.057	.954	.952	.058	.993	.993	.020
Opdrachttype	.988	.986	.029	.994	.993	.022	.998	.997	.014
Verschil	.996	.994	.017	.999	.998	.009	1.000	.999	.005
							.995	.995	.021
							.919	.913	.069
							.936	.924	.065
							.941	.928	.063
							.946	.943	.062

De restricties die aan de regressiegewichten in de Γ -matrix (orde: 4 maal 16) opgelegd worden betreffen de regressies van de opdrachtspecifiekefactoren voor de opdrachten 'Het nut van huiswerk' (η_{huis}), 'Ooggetuigeverslag' (η_{oog}), en 'Sollicitatiebriefje' (η_{sol}) op de intellectuele vaardigheidsfactoren.

Samenvattend: we onderscheiden drie vraagstellingen, die beantwoord worden op basis van de resultaten van evenveel analyses (de 'score-oordeel-analyses', de 'oordeel-oordeelanalyses' en de 'schrijfvaardigheidsanalyses'). Elke vraagstelling wordt onderzocht met behulp van drie modellen (het 'gelijke-regressies-model', het 'opdrachttype-model' en het 'verschil-model'). Elk model wordt tweemaal getoetst, waarbij een andere verzameling intellectuele vaardigheden een rol speelt. In totaal worden dus achttien ($3 \times 3 \times 2$) toetsingen uitgevoerd.

4. RESULTATEN.

In de onderstaande Tabel 6 zijn de resultaten van de modelpassingen voor de relaties tussen de SI- en de verschillende schrijfvaardigheidsfactoren weergegeven. Bij de toetsing van deze modellen is gebruik gemaakt van de kleinste kwadraten methode ter schatting van de modelparameters, omdat de scores op de SI-vaardigheidstoetsen verre van multivariaat normaal verdeeld zijn (Van den Bergh, 1989).

Uit Tabel 6 blijkt dat de modelpassingen voor beide variabelenverzamelingen goed convergeren: conclusies die getrokken zouden worden voor één van beide variabelenverzamelingen gelden ook voor de andere variabelenverzameling.

Voor de 'score-oordeel-analyses' is het verschil in passing tussen het 'gelijke-regressies-' en het 'opdrachttype-model' groter dan het verschil tussen het 'opdrachttype-' en het 'verschil-model', ongeacht welke van de drie gepresenteerde passingsmaten (gfi, agfi, RMR) in ogenschouw genomen wordt. Conform de definiëring van de effectgrootte moeten we derhalve constateren dat er een substantieel verschil is in regressie van de functionele schrijffactoren en de stelfactoren op de SI-factoren, mits de eerste gescoord worden en de stelprodukten door ervaren docenten beoordeeld worden.

Blijkens de passing van de modellen voor de 'oordeel-analyses' is er geen substantieel verschil tussen functionele schrijfprodukten en stelprodukten aangetoond; het passingsverschillen tussen de drie getoetste modellen zijn telkens zeer klein. Zo klein zelfs dat een onderscheid tussen deze modellen niet zinvol lijkt. Uit overwegingen van spaarzaamheid geven we de voorkeur aan het meest restrictieve model: Het 'gelijke-regressies-model'. We kunnen derhalve concluderen dat het schrijven van een opstel en het maken van een functionele schrijfoopdracht een beroep doen op hetzelfde complex SI-vaardigheden, mits de produkten op dezelfde manier beoordeeld worden.

Vergelijken we de resultaten van de toetsingen voor de 'oordeel-oordeel-analyses' met de resultaten van de vorige analyse (de 'score-oordeel-analyses'), dan kunnen we concluderen dat er kennelijk geen verschil is tussen de beide type schrijfoopdrachten qua betrokken intellectuele vaardigheden ten gevolge van het type schrijfoopdracht, maar wel ten gevolge van de gehanteerde beoordelingsmethode (scoring versus beoordeling).

Ook voor de derde onderzoeksvraag, waarbij restricties aan de regressiegewichten van de opdrachtspecifiekefactoren op de SI-vaardigheidsfactoren opgelegd werden, bleek bij beide variabelenverzamelingen nauwelijks een verschil tussen schrijfprodukten gebaseerd op functionele opdrachten en schrijfprodukten ontlokt door traditionele stelopdrachten; wederom zijn de passingsverschillen tussen de drie modellen zo klein dat een onderscheid tussen deze modellen niet zinvol lijkt. Daarom wordt de voorkeur gegeven aan het restrictieve 'gelijke-regressies-model'.

Welke SI-vaardigheden van belang zijn voor de schrijfvaardigheid? In Tabel 7 zijn de schattingen van de regressiegewichten weergegeven. Deze regressiegewichten zijn geschat onder de restricties van het 'opdrachttype-model' (verschillen tussen functionele opdrachten en stelopdrachten qua betrokken intellectuele vaardigheden) voor de 'scores-oordelen-analyses', en onder de restricties van het 'gelijke-regressies-model' (geen verschil tussen opdrachttypen qua betrokken intellectuele vaardigheden) voor de 'schrijfvaardigheidsanalyses'. Bij deze laatste toetsing moet vermeld worden dat de ladingen van de SI-factoren op de opdrachtspecifiekefactoren bijzonder laag zijn ($\tilde{\gamma}=.087$). De passing van een model waarin alle regressies op de opdrachtspecifiekefactoren gefixeerd waren op nul verschilde dan ook nauwelijks qua passing ten opzichte van een model waarin deze ladingen wel geschat mochten worden (vergelijk: gfi: .937 versus .935; agfi: .934 versus .933; RMR: .062 versus .063). De drie opdrachtspecifiekefactoren kunnen dus niet met de gemeten SI-vaardigheden voorspeld worden. In Tabel 7 hebben we de opdrachtspecifiekefactoren daarom buiten beschouwing gelaten.

Tabel 7. De regressies van de verschillende schrijfvaardigheidsfactoren op de SI-factoren voor de 'scores-oordelen analyses' (een functionele-scores- en een docentenoordelen-factor) en de 'schrijfvaardigheidsanalyses' (een schrijfvaardigheids- en een schaalbeoordelingsfactor).

Model: SI-vaardigheid	Scores-oordelen analyses opdrachttype-model		Schrijfvaardigheidsanalyses gelijke-regressies-model	
	funct. scores	doc.oordelen	schrijfv.	schaaib.
CMS	.097	.210	.180	.034
CMT	.051	.380	.023	.680
CMU	.245	.238	.232	.463
DMR	.166	.656	.503	.124
DMU	.170	.520	.356	.038
EMI	.155	.755	.212	.175
EMS	.067	.169	.269	.140
EMU	.208	.187	.004	.064
MMI	.098	.013	.060	.064
MMS	.155	.145	.394	.010
MMT	.327	.078	.094	.217
MMU	.319	.068	.009	.314
NMI	.023	.208	.268	.194
NMR	.069	.128	.072	.013
NMS	1.000	1.000	1.000	.031
NMT	.914	.980	.752	.073
percentage verklaarde variantie	24.2	30.7	52.3	38.2

Het eerste dat in Tabel 7 opvalt is dat van de variantie in de 'latente scores' van de functionele schrijfprodukten en de variantie in de 'latente docentenoordelen' over de stelprodukten relatief weinig verklaard kan worden met behulp van de zestien SI-vaardigheden (respectievelijk: 24% en 31%). De reden hiervoor is dat een groot deel van de geobserveerde variantie in schrijfvaardigheidsoordelen en -scores onverklaarbare opdrachtspecifiekevariantie bleek (zie: Figuur 1).

Uit Tabel 7 blijkt dat de holistische docentenoordelen over de stelprodukten beduidend meer CMT, DMU, DMR en EMI weerspiegelen dan de scores van de functionele opdrachten, die op hun beurt weer een duidelijkere geheugen-component bevatten. Deze verschillen tussen de functionele-scores en de docentenoordelen moeten waarschijnlijk

toegeschreven worden aan de verschillen in beoordelingsprocedure in plaats van verschillen inherent aan het type schrijfo opdracht. Uit Tabel 7 blijkt voorts dat zowel divergente als convergente produktie-vaardigheden van groot belang zijn voor de schrijfvaardigheid, wanneer deze gedefinieerd is als de gemeenschappelijke variantie van veertien oordelen over de produkten ontlokt door vijf schrijfo opdrachten. Met name de vaardigheden DMR, DMU, NMS en NMT lijken essentieel voor de schrijfvaardigheid. Het belang van de andere vaardigheden lijkt beduidend minder. Bovendien kan met de zestien gemeten SI-vaardigheden een flink deel van de ware schrijfvaardigheidsvariantie verklaard worden (52%).

Tevens blijkt uit Tabel 7 dat de gemeenschappelijke variantie van de schaaloor delen niet eenvoudig als invaliditeitsfactor geïnterpreteerd kan worden. De schaaloor delen weerspiegelen wel intellectuele vaardigheden die voor het schrijven van belang zijn, maar in een ietwat andere verhouding.

5. DISCUSSIE

In dit onderzoek is getracht na te gaan of met verschillende functionele schrijfo opdrachten dezelfde intellectuele vaardigheden gemeten worden als met een traditionele stelopdracht. Op grond van de verrichte analyses moeten we concluderen dat er met beide type schrijfo opdrachten een beroep gedaan wordt op hetzelfde complex van intellectuele deelvaardigheden. Deze conclusie geldt alleen indien de schrijfprodukten op dezelfde manier beoordeeld worden. Dit is de examensituatie niet het geval: de functionele schrijfprodukten worden gescoord aan de hand van een analytisch scoringsvoorschrift, terwijl over de stelprodukten een holistisch oordeel gegeven wordt, aan de hand van (globale) voorschriften. In dit geval is er wel sprake van verschillen tussen het schrijven van een opstel of het maken van een functionele schrijfo opdracht. Deze verschillen lijken voornamelijk te maken te hebben met de coherentie (EMI, NMI), eenduidigheid (CMT) en gevarieerdheid van woordkeuze (DMR) van de geproduceerde tekst, zaken die lastig in een analytisch scoringsvoorschrift te vatten zijn. Vandaar dat de vraag klemmt, of deze verschillen inherent zijn aan de voor dit onderzoek ontwikkelde scoringsvoorschriften, of dat hier een algemeen manco van scoringsvoorschriften naar voren treedt. Misschien kunnen scoringsvoorschriften dusdanig geconstrueerd worden dat een dergelijk verschil in beoordeelde intellectuele vaardigheden niet op zal treden. Hoewel het zeer moeilijk is gebleken scoringsvoorschriften te ontwikkelen die vergelijkbaar zijn met holistische oordelen (vergelijk: Prater & Padia, 1983).

Opmerkelijk is dat de opdrachtspecifiekevariantie niet verklaard kan worden met behulp van de gemeten semantische SI-vaardigheden. Het lijkt onwaarschijnlijk, gezien de nul-ladingen van alle zestien SI-factoren op de opdrachtspecifiekefactoren én het gegeven dat de opdrachtspecifiekefactoren als orthogonaal gespecificeerd zijn, dat deze factoren een niet in de toetsbatterij opgenomen SI-vaardigheid weerspiegelen. We moeten welhaast concluderen dat opdrachtspecificiteit van schrijfvaardigheidsprodukten weinig of niets te maken heeft met het construct schrijfvaardigheid als een relatief constante dispositie van een persoon. Als de opdrachtspecifiekefactoren niet geïnterpreteerd kunnen worden als schrijfvaardigheidsfactoren, hoe deze dan te interpreteren? Er zijn verschillende mogelijkheden. Allereerst kunnen de opdrachtspecificiteiten een variantie in achtergrondkennis met betrekking tot het onderwerp weerspiegelen; de ene leerling weet bij voorbeeld meer van het ene onderwerp en heeft daar wel een passend tekstschema voor bij de hand, terwijl een andere leerling meer weet van het andere onderwerp (vergelijk: Boodoo & Garlinghouse, 1983). We kunnen dan veronderstellen dat variantie in schrijfvaardigheid pas tot

uiting komt als de leerlingen geen problemen ondervinden in het genereren van de benodigde informatie waarover geschreven moet worden, het genereren van een adequaat tekstschema, of het kennen van de voor een tekst specifieke conventies etcetera.

Een tweede mogelijkheid is dat verschillende opdrachten aanleiding zijn tot het beoordelen van verschillende aspecten van de schrijfprodukten. De aanwezigheid van bijvoorbeeld spelfouten of een slordig handschrift in een sollicitatiebriefje kan door de beoordelaars als zodanig storend ervaren worden dat het betreffende schrijfprodukt op geen van de beoordeelde aspecten meer hoog kan scoren. Er is dan sprake van een 'halo-effect'. Zou dit het geval zijn, dan is het des te belangrijker om in de te toetsen modellen beoordeelaarsspecifiekefactoren te onderscheiden. Dit aspect wordt overigens in veel onderzoek verwaarloosd; een hoge overeenstemming tussen beoordelaars is een noodzakelijke, maar niet voldoende voorwaarde voor valide oordelen.

Als derde verklaring van de onverklaarde opdrachtspecificiteiten kan proefpersoonvariabiliteit genoemd worden. Schrijven is een dermate complexe multidimensionele activiteit dat het niet verwonderlijk is dat een leerling het ene moment veel beter presteert dan het andere moment. We kunnen dan schrijfvaardigheid op één moment gemeten niet meer opvatten als een betrouwbare indicator van een relatief stabiel kenmerk van een leerling (vergelijk: Lundsteen, 1977; Chase, 1986). De consequenties van deze verklaring zijn natuurlijk hoogst onwenselijk: metingen van de schrijfvaardigheid in onderwijs en onderzoek zouden dan een hachelijke zaak zijn.

Deze studie kan ook opgevat worden als een (beperkte) validatie van het 'Structure-of-Intellect' model. Immers, een complexe vaardigheid als schrijfvaardigheid kan met (enig) succes ontleed worden in een beperkt aantal SI-vaardigheden. Misschien is validatie hier een enigszins zwaar begrip, desalniettemin is met een betrekkelijk gering aantal 'context-arme' intellectuele vaardigheden ruim de helft van de variantie van de 'ware schrijfvaardigheid' verklaard; een resultaat waarmee een deel van de kritiek op het 'Structure-of-Intellect' model weerlegd wordt (vergelijk bij voorbeeld: Lammers, 1984). Dit levert ons de mogelijkheid om te trachten het construct 'schrijfvaardigheid' psychologisch te interpreteren, alhoewel een dergelijke interpretatie bemoeilijkt wordt door de uitkomst dat verschillende manieren om de kwaliteit van een schrijfprodukt te bepalen een duidelijke invloed hebben op de betrokken intellectuele vaardigheden en het gegeven dat ongeveer de helft van de ware schrijfvaardigheidsvariantie verklaard kan worden.

Het belang van informatie generatie is duidelijk aangetoond, gezien de hoge regressies van de schrijfvaardigheid op de verschillende divergente en convergente produktievaardigheden. Het genereren van informatie is essentieel. Het lijkt logisch te veronderstellen dat naarmate een leerling meer tekst geproduceerd heeft het belang van convergente produktievaardigheden toeneemt. Echter, bij een dergelijke interpretatie moeten we zeer voorzichtig zijn, daar we nauwelijks informatie hebben over het niveau waarop een vaardigheid (op een bepaald moment werkzaam is). Zo kan met evenveel recht verondersteld worden dat convergente produktievaardigheden, als NMS, van groter belang zijn naarmate een zin vordert.

Voor de door Flower & Hayes (1981; Hayes & Flower, 1980) onderscheiden globale planningsfase is in dit onderzoek weinig steun gevonden. De globale planning of de constructie van een schrijfplan zou moeten blijken uit een hoge lading van de geheugenvaardigheden. Met name het geheugen van semantische systemen (MMS) zou in dit verband van belang kunnen zijn. Echter, de ladingen van de verschillende geheugenvaardigheden zijn lager dan verwacht mag worden indien deze planningsfase cruciaal is. Het lijkt veeleer alsof de schrijver tekstfragmenten genereert die op papier gezet worden, zonder dat de schrijver al een duidelijk vervolg in het hoofd heeft. Verondersteld wordt dat planningsactiviteiten zich veeleer op zinsniveau in plaats van tekstniveau afspelen.

Gedurende de tekstproductie lijkt een relatief goede schrijver verschillende bewoordingen te beproeven alvorens er een feitelijk op papier te zetten, gezien de hoge lading van de convergente productie van semantische transformaties (NMT). De betere schrijvers houden bij tekstproductie meer rekening met de implicaties van hetgeen zij schrijven (EMI) en de mate waarin een nieuw fragment aansluit bij reeds geproduceerde tekst (NMI).

De actieve woordenschat van de leerlingen blijkt nauwelijks een struikelblok voor de tekstproductie. Kennelijk is de actieve woordenschat van de meeste leerlingen voldoende. Wel blijkt de vaardigheid om in verschillende bewoordingen hetzelfde tot uitdrukking te brengen, bij voorbeeld het gebruik van synoniemen, van belang; leerlingen die hoog presteren op de divergente productie van semantische relaties (DMR) blijken beter schrijvers.

Samenvattend kan geconcludeerd worden dat de nieuwe examenopdracht valide is ten opzichte van de stelopdracht qua betrokken intellectuele vaardigheden, mits de kwaliteit van de verschillende produkten op vergelijkbare wijze gekwantificeerd wordt. Echter, wanneer we de validiteit van een schrijfvaardigheidsmeting in het algemeen bekijken, dan blijkt slechts een gering deel van de geobserveerde variantie als schrijfvaardigheidsvariantie geoormerkt te kunnen worden. Onderzoek naar deze opdrachtspecificiteit, die op grond de in dit artikel gepresenteerde resultaten als invaliditeitsfactor geïnterpreteerd is, lijkt noodzakelijk. Met name de relatie tussen de opdracht en de proportie opdrachtspecificitevariantie verdient de nodige aandacht.

LITERATUUR

- Bergh, H. van den (1988a). Onderzoek naar functionele doelstellingen voor het moedertaal onderwijs. *Pedagogische Studiën*, 65, 139-148.
- Bergh, H. van den (1988b). De relationele structuur van enkele schriftelijke taalvaardigheden. *Tijdschrift voor Taalbeheersing*, 1138-59.
- Bergh, H. van den (1988c). Schrijven en schrijven is twee: Een onderzoek naar de samenhang tussen prestaties op verschillende schrijftaken. *Tijdschrift voor Onderwijsresearch*, 13, 311-324.
- Bergh, H. van den, Glopper, K. de & Schoonen, R. (1987). *Directe metingen van schrijfvaardigheid: Validiteit en taakeffecten*. Paper gepresenteerd op het vierde VIOT-congres op 16, 17 en 18 december 1987.
- Benton S.L. & Kiewra K.A. (1986). Measuring the organizational effects of writing ability. *Journal of Educational Measurement*, 23, 377-386.
- Blok, H. (1987). *Taal voor alledag: Feiten en meningen over het taalgebruik van LBO- en MAVO-leerlingen in alledaagse situaties*. Den Haag: SVO (Selecta-reeks).
- Blok, H. & Hoeksma (1984). *Opstellen geschaald*. Amsterdam: SCO.
- Braddock, R., Lloyd-Jones, R. & Schoer, L. (1963). *Research in written composition*. Champaign, Ill.: National Council of the Teachers of English.
- Britton, J., Burges, T., Martin, N., McLeod, A. & Rosen, H. (1975). *The development of writing abilities*. School Council Research Studies: Macmillan Education.
- 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.
- Chase, C.I. (1986). Essay test scoring: Interaction of relevant variables. *Journal of Educational Measurement*, 23, 33-41.
- Coffman, W.E. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement*, 22, 41-45.
- Daiute, C.A. (1981). Psycholinguistic foundations of the writing process. *Research in the Teaching of English*, 15, 5-22.
- Elshout, J.J. (1976). *Karakteristieke moeilijkheden in het denken* (Academisch proefschrift). Amsterdam: Universiteit van Amsterdam.

- Flower, L.S. & Hayes, J.R. (1980). The dynamics of composing: Making plans and juggling constraints. In: L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.
- Flower, L.S. & Hayes J.R. (1981). A cognitive proces theory of writing. *College Composition and Communication*, 32, 365-387.
- Graves, D. (1981). A new look on research in writing. In: S.M. Haley-Jones (Ed.), *Perspectives on writing in grades 1-8*. Urbana: NCTE.
- Glopper, K. de (1986a). *Meettechnische aspecten van het onderzoek naar de stelvaardigheid van leerlingen in het derde leerjaar LTO, LHNO, MAVO, HAVO, en VWO*. Amsterdam: SCO.
- Glopper, K. de (1986b). *Schrijffprestaties in het voortgezet onderwijs: Een onderzoek naar de stelvaardigheid van leerlingen in het derde leerjaar LTO, LHNO, MAVO, HAVO, en VWO*. Den Haag: SVO.
- Guilford, J.P. (1971). *The nature of human intelligence*. London: McGraw-Hill.
- Guilford, J.P. (1979). *Cognitive psychology with a frame of reference*. San Diego: Edits.
- Guilford, J.P. & Hoepfner, R. (1971). *The analysis of intelligence*. New York: McGraw-Hill.
- Hayes, J.R. & Flower, L. (1980). *A cognitive model for the writing proces in adults*. Final Report. Pittsburg: Carnegie Mellon University.
- Hoeks, J. (1985). *Vaardigheden in begrijpend lezen* (Academisch Proefschrift). Amsterdam: Universiteit van Amsterdam.
- Hillocks, G. (1986). *Research on written composition*. Urbana, Ill.: ERIC Clearinghouse on Reading and Communication.
- Jöreskog, K.G. & Sörbom, D.G. (1986). *Lisrel: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods*. Uppsala: University of Uppsala.
- Kamstra, O.W.M. (1971). *De dimensionaliteit van het geheugen* (Academisch proefschrift). Amsterdam: Universiteit van Amsterdam.
- Kunst (1978). *Cognitie van semantische systemen: Strategieën en interpretaties bij een structure-of-intellect vaardigheid* (Academisch proefschrift). Amsterdam: Universiteit van Amsterdam.
- Lammers, D.H. (1984). Taalvaardigheid als complex begrip. *Tijdschrift voor Taalbeheersing*, 6, 101-117.
- Lundsteen, S. (Ed.) (1976). *Help for the teacher of written composition*. Urbana: ERIC/NCTE.
- Nihira, K., Guilford, J.P., Hoepfner, R. & Merrifield, P.R. (1962). *A factor analysis of semantic-evaluation abilities*. California: University of Southern California. Prater, D. & Padia, W. (1983). Develloping parallel holistic and analytic scoring guides for assessing elementary writing samples. *Journal of Research and Development in Education*, 17, 20-24.
- Quellmalz, E.S. & Cappell, F.J. (1979). *Defining writing effects of discourse and response mode*. L.A.: California University, Center for Study of Education. Quellmalz, E.S. Cappel, 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. (1985). *Effecten van leerlingrespons op aspecten van taalvaardigheid*. Amsterdam: SCO.
- Shaughnessy, M.P. (1977). *Errors and expectations: A guide for the teacher of basic writing*. New York: Oxford University Press.
- Schoonen, R. & Wolf, T. (1984). *Leesvaardigheid en leesprocessen: Een empirisch onderzoek naar lezen als taalvaardigheid en cognitief proces*. Amsterdam: Instituut voor Neerlandistiek van de Universiteit van Amsterdam.
- Schooten, E. van & Bergh, H. van den (1987). Effecten van examenveranderingen: Interviews met docenten op het LBO en MAVO. *Levende Talen*, 417, 37-43.
- Traxler A.E. & Anderson, H.A. (1935). The reliability of an essay examination English. *School Review*, 43, 534-539.
- Wesdorp, H. (1974). *Het meten van de produktief-schriftelijke taalvaardigheid*. Purmerend: Muusses.
- Wesdorp, H. (1983). *Schrijven in het voortgezet onderwijs*. Den Haag: SVO (SVO-reeks no 71).
- Wesdorp, H. Bergh, H. van den, Bos, D.J., Hoeksma, J.B., Oostdam, R. Scheerens, J. & Triesscheijn, B. (1985). *De haalbaarheid van periodiek peilingsonderzoek*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 19-10-1988

Definitieve versie ontvangen 20-2-1989

Schools over the Gangway: An Experiment on Response Improving-Procedures

Cor Suhre

Institute for Educational Research (RION) of the University of Groningen

ABSTRACT

The response rate to mailed questionnaires in educational research conducted in the Netherlands is generally low. There is only a limited number of response-improving procedures that researchers can undertake to improve the response rate. For four promising procedures an experiment was set up to find out in which form they are most effective. The procedures which were systematically manipulated were: advance notice of the research, the use of promised incentives, a follow-up and form aspects of the questionnaire. The results of the experiment reveal that advance notice and a follow-up are best conducted by telephone (compared to use of a letter) and that a promised incentive is also useful, no matter whether a fixed amount of money is concerned or entry in a lottery with attractive prizes. The format of the questionnaire shows to be of little influence with respect to response rate. With respect to the combined use of the procedures it turns out that the form in which the procedures are most effective is independent of the form in which the other procedures are administered. Further analysis with regard to background features of respondents and nonrespondents shows no difference in effect for the alternative forms of the advance notice and follow-up procedures with respect to selective response. However, for the use of incentives the study tells a different story.

1. INTRODUCTION

A lot of educational research in the Netherlands is conducted in the form of mail surveys. Typical of this kind of research is that in a short time span, at a low cost, data are collected mostly from teachers, in schools spread all over the country. Response to mailed questionnaires, however, has gradually declined to a rather low level during the last decennium. Response rates under 40% are not uncommon anymore. The most probable reasons for these low response rates are the frequency with which schools are contacted for cooperation in research issues and the perceived importance of the research questions. Whether teachers will cooperate need not be a consequence of personal motives only; it can also be dependent on reactions and opinions of colleagues in the schoolteam. This fact makes it extra important for researchers to specify in advance exactly which persons need to be contacted for what information.

Low response rates can seriously affect the external validity of research. As a consequence of nonresponse, results can be biased in several respects (see e.g. De Leeuw and Hox, 1987). Taking effective procedures against nonresponse is a step all evaluators have to take. The problem which most researchers face, however, is doing something about nonresponse within budget constraints. In order to select an optimal set of response-improving procedures researchers have to know two things. Firstly, which procedures or

The research project in which the response improving procedures were implemented, was sponsored by the Institute for Educational Research in the Netherlands (SVO project no. 6087)

Adres: Postbus 1286, 9701 BG Groningen

combinations of procedures are really effective in improving the response rate and secondly, what the effects of these procedures are on the selection of respondents.

In the experiment reported on, the effects on response rate were studied for four procedures which were taken in alternative forms. The procedures were implemented in an ongoing research project which studied the general state and progress of schools with respect to innovation in primary education. The first question addressed in this experiment was: In which form are the procedures most successful in improving response rate? Although a high response rate reduces the chance of serious misrepresentation, any nonperfect response rate still leaves room for questions regarding the representativeness of the group of respondents. It is conceivable, therefore, that the form in which procedures are administered has an effect on the degree of representativeness for the total population (Nederhof, 1981; Lockhart, 1984). The second question we sought to answer was whether there are differences between alternative forms of the procedures with respect to selective response. This second question is no less important than the first one. The two questions will be subsequently dealt with in this article.

2. PRACTICAL AND THEORETICAL CONTEXT

2.1 Research on response improving procedures

From quantitative analyses of the response rate to mailed questionnaires [see e.g. the comparison of results from studies by Heberlein and Baumgartner (1978), Goyder (1982), Eichner and Habermehl (1982)] the conclusion may be that the responsiveness in any research area can be influenced by a limited number of variables. These variables can be subsumed under four categories:

- (a) importance and salience of the research subject to the addressee;
- (b) incentives
- (c) follow-up procedures
- (d) form aspects of the questionnaire. On this last category there is, however, less equivocal evidence for relevant aspects than for the three others. Nevertheless, in his Total Design Method (TDM) approach, Dillman (1978, p. 121) draws special attention to visually observable characteristics of the questionnaire. For instance he strongly advocates the use of a booklet in (approximately) A-5 format.

From research in our own institute (Meijer, 1982) and also from De Leeuw and Hox (1987) there is evidence for one other category that can stimulate response rate:

- (e) advance notice: a short introduction of the research before sending the questionnaire.

2.2 The selected response improving procedures

In the experiment reported on here, four procedures were selected from each of the categories b-e. These procedures were selected for systematic manipulation because of their presumed effectiveness and general applicability for every researcher. Three of these procedures (from categories b, c and e) are in accordance with the philosophy behind Dillman's TDM (Dillman, 1978). The experiment was meant to find out in which form the procedures could best be applied to achieve high and nonselective response. A related issue is whether, in the combined use of alternative forms of the procedures, there is a consistent pattern for the degree of effectiveness for the most effective forms. In order to study these questions an ordinary looking questionnaire was used to ensure generalizability of the results of the experiment for most researchers. The printing procedures Dillman proposes have not been applied in the construction of the questionnaire. The questionnaire that was used resembles the most commonly used questionnaires in educational research.

The selected procedures are:

- R1: advance notice: introduction of the research in which schools would be asked to cooperate. The introduction was performed either by means of a letter or by telephone.
- R2: incentive: schools were promised either a gift coupon worth US \$10 or participation in a lottery with a small number of valuable prizes (4 soundmix machines and one keyboard (worth US \$700)). The prizes chosen for the lottery were thought to be especially attractive to schools since at the time the experiment was carried out a television programme for children was running that made use of these apparatuses. These procedures were set in contrast to no incentive.
- R3: format of the questionnaire: either in A-4 format or in a handy A-5 format.
- R4: follow-up: a reminder four weeks after date of sending. The reminder was carried out by mail or by telephone.

In the experiment only one follow-up contact was used because in most educational survey research projects time and financial resources do not permit more than one follow-up.

3. METHOD

The questionnaire

The questionnaire contained questions regarding general background features, innovation aspects of education and questions regarding school effectiveness characteristics. Most of the questions regarding innovation and school effectiveness were part of Likert-type scales. The questionnaire consisted of 15 pages. The expected filling time was 45 minutes to an hour.

Sample

The units of interest were school principals. The sample was a random sample of 995 primary schools out of all primary schools in the Netherlands. Seven schools could not be contacted since they had ceased to exist at the time the study was carried out. Therefore the effective sample consisted of 988 schools.

Design

The schools to be contacted were assigned randomly to any of all possible combinations of conditions in which the procedures were to be administered. The random assignment was organized in such a way that approximately equal numbers of schools received the different combinations of conditions.

Procedure

Depending on the conditions which were assigned to a school beforehand, the principals received information about the research subject and incentives. The information school principals received about the research was the same in the two announcement conditions. In the letter accompanying the questionnaire schools were informed of the kind of reward (if any) they would get if they returned the questionnaire. For returning the questionnaire the school principals could make use of an accompanying return envelope of the same size (A4 or A5) as the questionnaire. The follow-up, either by mail or by telephone, was planned four weeks after mailing the questionnaire. The phone calls in the experiment were conducted by one and the same student assistant. The deadline for returning the questionnaire was set a seven weeks after receipt.

Analysis

The differential effects of alternative forms of response improving procedures on response rate were analyzed by means of logit analysis (a special form of loglinear analysis, see Goodman, 1972). To this end a number of loglinear models were formulated which differed in the specification of main effects (differences in response rate between alternative forms of the procedures) and interaction effects between alternative forms of the response improving procedures on response rate. The effects of differences in response rate for the incentive conditions were incorporated in the models as Helmert contrasts. The first Helmert contrast (R2(1)) refers to the difference between no incentive vs both incentive conditions. The second Helmert contrast (R2(2)) refers to the difference between the fixed amount of money condition and the lottery condition. The plausibility of interaction effects and main effects was investigated by a sequential testing procedure. This amounted to evaluating the general fit of the formulated models and the decrease in fit between complex and simple models in terms of the Likelihood Chi-square ratio test statistic. The sequential testing procedure makes it necessary to compare several plausible models.

4. RESULTS ON RESPONSE RATE

The mean response rate in this study was 52.4%. Seventy four percent of the response was received within three weeks after mailing the questionnaire.

The technical results of the analysis are presented in the following tables and are summarized below:

- there is no evidence for interaction effects between alternative forms of the response improving procedures. Model 3 containing only main effects of the alternative forms of the procedures did not produce a weaker goodness of fit, as evidenced by the conditional Likelihood ratio chi square.
- there is evidence for main effects. From Tables 1 and 2 it is clear that the discrepancy in response rate for the advance notice conditions cannot be left out of any model and therefore shows a real effect. With respect to the incentive conditions and the follow-up conditions there is some friction when leaving out the discrepancy in response rate from model 3. Discrepancy in response rate for school principals, who received the questionnaire in a different format, is not statistically significant.

Tabel 1: Chi square statistics for some explanatory models for response rate

Manipulation of response-improving procedures:

R1: advance notice (written or by telephone)

R2: incentive (none/US \$ 10 / participation in a lottery)

R3: format of the questionnaire booklet (A4 vs A5)

R4: follow-up (by letter or by telephone)

Model:	LIKELIHOOD RATIO CHI SQUARE	DF	P
1) full model - 3rd order interaction	2.9	2	.23
2) full model - 2nd and 3rd order interaction	10.1	9	.35
3) R1+R2+R3+R4	18.7	18	.41
4) R1+R2+R4	20.5	19	.37
5) R1+R2+R3	22.1	19	.28
6) R1+R3+R4	24.3	20	.23
7) R2+R3+R4	38.8	19	.01
8) R1+R2+(1)R4	20.7	20	.41

Note: R2(1) refers to the contrast effect of no reward vs both incentive conditions.

Tabel 2: Chi square statistics for conditional tests on the discrepancy between models

Model:	CHANGE IN LIKELIHOOD RATIO CHI SQUARE	DF	P
(2/1)	7.1	7	.41
(3/2)	8.6	9	.47
(4/3)	1.8	1	.18
(5/3)	3.4	1	.07
(6/3)	5.6	2	.06
(7/3)	20.1	1	.00
(8/4)	0.2	1	.63

– further analysis on the differences in response rate for the incentive conditions yields no difference in effect between the lottery condition and the fixed amount of money condition. Removal of the second Helmert contrast for the incentive effect from model 4 shows to be not detrimental with regard to the goodness of fit (see Table 1).

The implications of these technical results are the following. Manipulation with respect to the format of the questionnaire cannot be considered effective. Manipulation of the other three procedures raises enough evidence as to which form is expected to be most effective. For procedures R1 and R4 the telephone conditions produced significantly better response rates. With regard to the use of the incentives (taking part in a lottery and promising a fixed amount of money), both incentives are about equally successful in raising response rate. The amount of differential effectiveness for the alternative forms of the procedures seems to hold regardless of the use of the form of any other response improving procedure. The expected differences in effect for the three procedures are portrayed in Table 3. These differences are based on the log odds parameters for model 8 (see Table 1).

Tabel 3: Expected difference between alternative forms of the response improving procedures.

Measures	Expected dif. response rate
Advance notice (by letter or by phone)	14%
Incentive (no/ fixed amount of money/lottery)	8%
Follow-up (by mail/phone)	6%
Total for combination of most effective forms	25%

5. REPRESENTATIVENESS

The question whether a less than optimal response rate from a random sample is detrimental to the representativeness for the population is a difficult one. One can never be certain that respondents do not differ from the population in certain respects. Not all population characteristics are relevant, however. Which characteristics bear relevance is highly influenced by the research questions under study. In this study we consider the response rate of the sampled units with respect to two background features, denomination of the school and composition of the social cultural background of the pupils. Schools differ in the principle upon which they are based (public vs religious) and also in the diversity of social cultural background of the pupils. These features are usually of relevance in two broad subjects in educational evaluation research: educational innovation and differences between schools in test results for pupils in basic skills. Evidence on

the relevance of the background features to these research subjects can be found in recent research in these two areas within our research institute [see Harskamp, Pijl and Prins (1984, p.70), Harskamp and Rijken (1988) and van der Werf (1988, p. 123)].

The variation between schools in relative number of children from different social and cultural minorities is expressed in a social cultural weight which ranges from 1.00 (meaning that a school does not have children with working-class parents or children from cultural minorities) to 1.9 (meaning that all children come from specific cultural minorities).

A key question in this study was whether there are differences in effectiveness for alternative forms of the response-improving procedures for schools differing in the two background features. This question was tackled by means of a logit analysis. The procedure for testing was nearly the same as in the evaluation of differences in effect for the alternative forms of the response-improving procedures. As a starting model a logit model was proposed which incorporates besides main effects of the two background features and the first order interaction between these characteristics, simple main effects of the three procedures for which there were differences for the alternative forms and first order interactions of the procedures with background features. We evaluated this model against four other models. The first alternative model leaves out interactions between the background features and the different forms of the advance notice and follow-up procedures (R1 and R4). This model tests for differences between specific schools in response behavior to alternative forms of these procedures. The next three models explore the effects of reward on different groups of schools. It is often suggested (see e.g. Nederhof, 1981) that reward attracts different kinds of volunteers. These models explicitly seek to test that suggestion. The effects of the incentive conditions in these models are specified as simple contrasts to the no incentive condition.

The alternative models are evaluated by testing the goodness of fit of the models and the reduction in fit as compared to the more complex model.

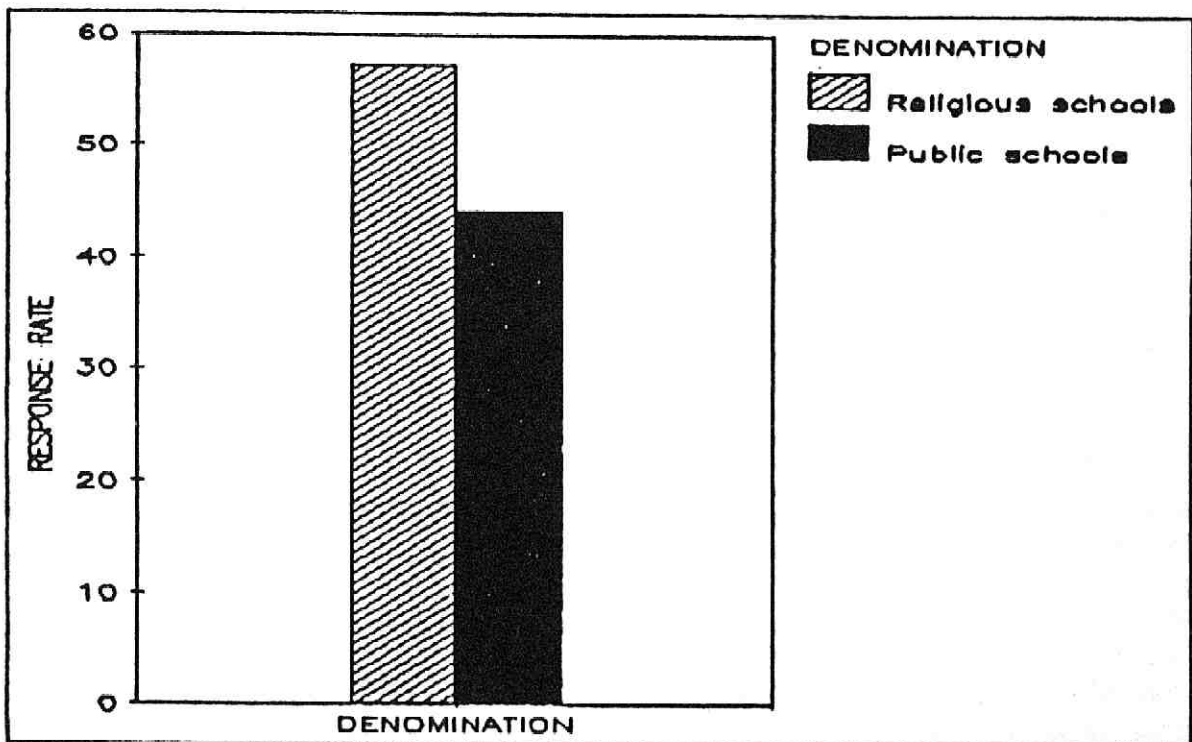


Fig. 1. Response rate for religious and for public schools.

Tabel 4: Description of background features and models:

Background features:			
denomination: foundation principle, religious/public social cultural weight score: based on the number of children from various social cultural background			
Manipulation of response-improving procedures:			
R1: announcement (written or by telephone)			
R2: incentive (none vs a fixed amount of money and entry in the lottery)			
R4: follow-up (by letter or by telephone)			
Plausible Models:			
(1) simple main effects background features + first order interaction of background features + simple main effects response improving procedures + first order interactions of response improving procedures with background features			
(2) same as under (1) except for the omission of interactions between R1 and R4 with the background features;			
(3) same as under (2) and removal of interactions for the contrast between no incentive vs entry in the lottery with the background features;			
(4) same as under (2) and removal of interactions for the contrast of no incentive vs the fixed amount of money condition with the background features;			
(5) simple main effects background features + first order interaction of background features + simple main effects response improving procedures			
Model:	LIKELIHOOD RATIO CHI SQUARE	DF	P
1)	9.5	8	.30
2)	13.8	12	.31
3)	21.0	14	.10
4)	15.0	14	.38
5)	21.2	16	.17
Conditional tests:			
(2/1)	4.3	4	.36
(3/2)	7.2	2	.03
(4/2)	1.2	2	.55
(5/4)	6.2	2	.05

6. RESULTS ON REPRESENTATIVENESS

a. Differences in background features.

For the two background features there are significant differences in response rate in the present study.

The difference in denomination (respective Chi-square 15.6, $df=1$, $p=.00$) is depicted in Figure 1. There is no difference in median on the social cultural weight score for respondents and nonrespondents. There is, however, a difference in social cultural weight for schools which had a score higher than the median value. For schools with a weight higher than the median value, the weight is a predictor for nonresponse (the Mann-Whitney U test for testing the difference between respondents and nonrespondents in social cultural weight being significant at the .001 level). Above a certain level schools with higher social cultural weight scores are less likely to respond.

The differences in response rate for schools differing on the two background features lead to some bias in response. The differences with respect to denomination are quite large. There are about 8% more religious schools and therefore about 8% less public schools in the group of respondent schools than in the population. With respect to the cultural weight

score the differences are less marked. Combining the information about the bias in response, we get an indication which schools are selectively missing. These are public schools from the bigger cities in the country. This finding fits in with experience in other research areas (see Nederhof, p. 37).

b. Difference in effects of alternative forms of the procedures on response from schools.

The results of the test procedures do not present evidence on differential effectiveness for alternative forms of the advance notice and follow-up procedures for different subgroups within the sample (Table 4). The same holds true for the financial incentive used in the experiment. For the incentive condition of entry in the lottery, however, there is evidence for a differential effect. Model 4, the model that incorporates the interactions of the contrast between the lottery condition and the no incentive condition with the background features, provides a better fit than the model that does not contain these interactions (model 5). In exploring the cause for the better fit of model 4 relative to model 5, we find that this is almost entirely attributable to the differential effectiveness of the lottery condition for religious and public schools. The lottery incentive especially attracted school principals from public schools, whereas it didn't attract the principals from religious schools. Concluding this section it can be stated that the use of advance notice and follow-up by telephone as well as an incentive in the form of a fixed amount of money represent unbiased means of raising response. The use of a lottery introduces bias in response by attracting volunteers from public schools to a greater extent.

7. DISCUSSION

Although there is evidence for differential effectiveness between the alternative forms of the advance notice and follow-up procedures with respect to response rate, there appear to be no differences between alternative forms of the procedures with respect to the response pattern of schools who differ in the two background features. The alternative forms we studied apparently do not have a different impact on the representativeness of the respondents (or in this study, on the selective response). With respect to the incentive conditions the study tells a different story. An incentive in the form of a fixed amount of money is equally successful in attracting volunteers from religious or public schools. An incentive in the form of a lottery does not attract volunteers from religious schools. The reason why this is so might be that the lottery concept is not in accordance with the philosophy of religious schools. Although the incentive of a lottery is not attractive to volunteers from religious schools it is to volunteers from public schools. Since public schools do not respond well (see Figure 1) the use of a lottery with large prizes might be an effective procedure in gaining more response from public schools and therefore in reducing selective nonresponse. Another (although less preferable) possibility in ensuring representativeness lies in taking advantage of known response behavior for schools differing in relevant background features. With this knowledge in mind one could design a sampling scheme that would enable one to get a more representative group of respondents. This would be preferable to increasing sample size so as to raise the absolute number of respondents. Further research on relevant characteristics for broad educational issues, as well as filing of these characteristics of schools could therefore be of use.

In the experiment we carried out only one follow-up was used. It could be argued that more follow-ups might reduce the extent of differential effectiveness between the alternative forms of the response-improving procedures. In our view this would mean that more follow-ups would successfully recruit more volunteers who had not been confronted

with the apparently more effective alternative forms of the response-improving procedures, i.e. there should be evidence for interactions between later follow-ups and the alternative forms in which the procedures were carried out. Whether this assumption holds true is a question open for research. At this point we can only conclude that there is evidence for differences in effectiveness between alternative forms in which the studied response-improving procedures were carried out and that for each procedure these differences are independent of the form in which the other procedures are carried out. We also know that, except for the lottery procedure, there is no evidence that the procedures lead to selective response. Whether this is also true for more follow-ups is a question open for research too.

A few words on the costs of the procedures. Costs for the two telephone conditions of the advance notice and the follow-up procedures were about equal per unit. These procedures are less efficient with regard to allocation of financial resources because they are administered to both respondents and nonrespondents. The promised fixed amount of money condition is the most cost efficient in that respect. In comparing total expected costs for implementation of the most effective forms (as used in this experiment) of the response-improving procedures, however, the balance is in favor of the telephone conditions of the advance notice and follow-up procedure.

REFERENCES

- De Leeuw, D.E., & Hox, J.J. (1987). Artifacts in Mail Surveys: The influence of Dillman's Total Design Method on the quality of the responses. In W.E. Saris & I.N. Gallhofer, *Sociometric Research, Vol. I*. London: Macmillan.
- De Leeuw, D.E., & Hox, J.J. (1987). The Effect of an Advance Letter, a Reminder by Telephone, and the Enclosure of a Second Questionnaire on the Response to Mail Surveys. *Kwantitatieve methoden*, 25, 127-139.
- Dillman, D.A. (1978). *Mail and Telephone Surveys: the Total Design Method*. New York: Wiley.
- Eichner, K., & Habermehl, W. (1981). Predicting Response Rates to Mailed Questionnaires. *American Sociological Review*, 46, 361-363.
- Goodman, L.A. (1972). A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables. *American Sociological Review*, 37, 28-46.
- Goyder, J.C. (1982). Further Evidence on Factors Affecting Response rates to Mailed Questionnaires. *American Sociological Review*, 47, 550-553.
- Harskamp, E.G., Pijl, Y., & Prins, J. (1984). *Werken aan de Nieuwe Basisschool. Eindrapport*. RION, Groningen.
- Harskamp, E.G., & Rijken, V. (1988). *Periodieke Peiling Stand van Zaken Basisscholen. Eindrapport*. RION, Groningen.
- Heberlein, T.A., & Baumgartner, R.M. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. *American Sociological Review*, 43, 447-462.
- Lockhart, D.C. (1984). The Stages of Mailed Questionnaire Returning Behavior. In D.C. Lockhart (ed.), *Making Effective Use of Mailed Questionnaires*. New Directions for Program Evaluation, no. 21. San Francisco: Jossey-Bass.
- Meijer, C.J.W. (1982). L.O.M.-Onderwijs. *Verwijzing en toelating. Eindrapport*. Rion, Haren.
- Nederhof, A.J. (1982). *Some sources of artifact in social science research: nonresponse, volunteering, and research experience of subjects. Dissertation*. University of Leiden, Leiden.
- Werf, van der, M.P.C. (1988). *Het schoolwerkplan in het basisonderwijs. Dissertation*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 11-8-1988

Definitieve versie ontvangen 17-1-1988

De oordelen van kinderen over zelfverzekerd gedrag van volwassenen

H.J.G. Soppe en M. van Heiningen-Scali
Katholieke Universiteit Brabant

ABSTRACT

Videorecordings of seven adults, presented as teachers, were shown to 8- to 13-year old children. Each adult was shown in a dominant and in a submissive role. Each child saw only one role of every adult. The children's task was to judge the targets on 19 five-point scales, each measuring a psychological attribute or teacher quality. The results showed that (1) the judgements of the subjects had a clear factorial structure; (2) the role played by the targets influenced the children's judgements about them: in their dominant role more teacher qualities and more positive characteristics were ascribed to them; (3) the subjects saw many differences between the targets on the measured characteristics. It was concluded that short observations of nonverbal behaviour may be sufficient for children to ascribe psychological characteristics to unknown people and that their judgements are more organized than hitherto assumed.

Wanneer op scholen sollicitatieprocedures lopen voor het aantrekken van nieuwe leerkrachten gebeurt het vaak dat ook de leerlingen, gevraagd of ongevraagd, een oordeel over de sollicitanten geven. Dit oordeel kan gebaseerd zijn op waarnemingen tijdens een proefles maar ook op een toevallige observatie van de sollicitant. Aan deze oordelen wordt door de leerkrachten dikwijls veel waarde gehecht. Verondersteld wordt blijkbaar dat kinderen, zelfs op basis van betrekkelijk kortdurende waarnemingen, kunnen vaststellen wat zij aan iemand hebben.

Deze opvatting wijkt af van wat ontwikkelingspsychologisch onderzoek heeft opgeleverd ten aanzien van persoonsoordelen van kinderen. Volgens dit onderzoek zouden kinderen beneden de leeftijd van acht jaar nauwelijks psychologische eigenschappen aan anderen toeschrijven (Shantz, 1983, voor een overzicht) en, voor zover zij dit wel doen, zouden de toegekende eigenschappen in ieder geval niet stabiel of dispositioneel zijn (Rholes & Ruble, 1984). Dit laatste betekent dat de betreffende eigenschappen niet de basis zouden vormen voor verwachtingen over het gedrag van de beoordeelde personen. Kinderen tussen acht en twaalf jaar zouden wel in behoorlijke mate psychologische disposities aan anderen toedichten, maar tussen de toegeschreven eigenschappen zou weinig samenhang bestaan (Flavell, 1985). Met andere woorden, het zou hen ontbreken aan een impliciete persoonlijkheidstheorie (Bruner & Tagiuri, 1954). Bovendien zouden kinderen in dit leeftijdsinterval nauwelijks tussen andere personen differentiëren op de gebruikte dimensies (Barenboim, 1981).

Er zijn diverse redenen om aan te nemen dat het gememoreerde onderzoek de vaardigheden van kinderen onderschat. Berndt & Heller (1985) zijn bijvoorbeeld van mening dat onderzoekers vaak de door jongere kinderen toegeschreven psychologische eigenschappen ten onrechte als niet-dispositioneel hebben gecodeerd. Kritiek is ook mogelijk op het feit dat de gegevens op basis waarvan de kinderen moeten oordelen altijd

bestaan uit verhaaltjes. Deze testsituatie is weinig representatief voor de wijze waarop in het dagelijks leven informatie over anderen wordt verworven. Daarin zijn persoonsoordelen meestal gebaseerd op de directe waarneming van gedrag en van de contexten waarin het plaats vindt. Misschien representeert de 'weigering' van jonge kinderen om op basis van louter verbale informatie psychologische disposities toe te schrijven niet meer dan de afwezigheid van een 'person attribution bias' (sommigen spreken zelfs van 'error') zoals volwassenen die hebben (Kassin & Pryor, 1985). Wat betreft het veronderstelde gebrek aan differentiatie en organisatie kan worden opgemerkt dat kinderen hiertoe op andere gebieden wel in staat zijn. Zo zijn zij vanaf zes jaar in staat tot serieren en dus tot het differentiëren tussen fysieke objecten (Piaget, Grize, Szeminska & Vinh Bang, 1977). Ontwikkelingspsychologisch geheugenonderzoek heeft voorts laten zien dat schoolkinderen in behoorlijke mate in staat zijn tot het integreren en elaboreren van en het maken van afleidingen uit aangeboden informatie (Flavell, 1985). Ook is gevonden dat kinderen al vanaf vier jaar structuren in emotionele expressies onderkennen (Bullock & Russell, 1984). Tenslotte is de vaststelling van Rogers (1978) relevant dat de interpersoonlijke wereld van kleuterschoolkinderen niet chaotisch is. Dit suggereert dat zij weten wat zij aan elkaar hebben en wat ze van elkaar kunnen verwachten.

Met het voorgaande zijn diverse argumenten gegeven voor nader onderzoek naar de persoonsoordelen van kinderen van twaalf jaar en jonger. In dit artikel wordt over een dergelijk onderzoek gerapporteerd. Het doel hiervan was na te gaan in hoeverre kinderen op basis van de waarneming van gedrag psychologische eigenschappen aan anderen toeschrijven en in hoeverre samenhang tussen die eigenschappen bestaat. In het onderzoek werden de kinderen geconfronteerd met videobeelden van zeven jong volwassenen, namelijk leerlingen van een Pedagogische Academie (verder studenten genoemd), die werden gepresenteerd als leerkrachten. Hun taak was om deze personen te beoordelen op een aantal psychologische dimensies. De te beoordelen personen werden als leerkrachten gepresenteerd omdat dit de kinderen een zinvolle en relevante context voor hun oordelen zou verschaffen. Teneinde de invloed van *gedragsverschillen* op de psychologische oordelen te kunnen onderscheiden van de invloed van verschillen in fysieke kenmerken, speelde ieder der zeven studenten twee rollen, die varieerden op de dimensie zelfverzekerdheid. Voor deze dimensie werd gekozen omdat zelfverzekerdheid voor kinderen een belangrijke eigenschap van goede leerkrachten is (Docking, 1980). Aangenomen mag daarom worden dat deze dimensie een belangrijke rol speelt bij de oordelen van kinderen over personen die als leerkrachten worden gepresenteerd.

Drie vragen stonden dus in het onderzoek centraal: 1. Schrijven de kinderen op basis van de waarneming van gedrag psychologische eigenschappen aan anderen toe?; 2. Is er samenhang tussen de toegeschreven eigenschappen?; 3. Zijn de toegeschreven eigenschappen een functie van de mate van zelfverzekerdheid in het waargenomen gedrag?

METHODE

Proefpersonen.

Drie groepen kinderen, ieder bestaand uit ongeveer de helft jongens en de helft meisjes, deden aan het onderzoek mee. De leeftijdsranges en de aantallen kinderen in deze groepen waren respectievelijk: groep I, 97 - 112 maanden, 72 kinderen; groep II, 118 - 132 maanden, 68 kinderen; groep III, 138 - 153 maanden, 47 kinderen. De ppn waren gerecruteerd uit diverse scholen in Tilburg.

Stimulusmateriaal.

Van ieder van zeven studenten (vier mannen, drie vrouwen) van de Pedagogische Academie te Tilburg werden twee video-opnamen gemaakt. Opgenomen werden beide keren het fietsen (ongeveer 30 sec.) en het daaropvolgende lopen (ongeveer 70 sec.) naar de Academie en het betreden van een leslokaal, inclusief het plaatsnemen aan een tafel en het openslaan van een boek (ongeveer 20 sec.). Ten behoeve van de eerste opname werden de studenten geïnstrueerd zich zelfverzekerd te gedragen, ten behoeve van de tweede opname onzeker. Daartoe werden zij geïnformeerd over de typische verschillen in gedragskenmerken van deze houdingen, zoals rechte vs gebogen lichaamshouding, veel en ruime vs weinig en beperkte bewegingen, hoofd rechtop vs gebogen, de blik ronddwalend vs teneergeslagen, geen vs wel een glimlach, wenkbrauwen gefronst vs opgetrokken (Harper, 1985). Van de aldus verkregen opnamen werden twee videobanden samengesteld. Op elke band kwam elke student eenmaal voor, eerst de drie vrouwen, dan de vier mannen. De volgorde van de rollen op band I was: onzeker, onzeker, zeker, zeker, onzeker, onzeker, zeker, en die op band II: zeker, zeker, onzeker, onzeker, zeker, zeker, onzeker.

Om na te gaan of de studenten hun rol volgens instructie hadden gespeeld, werden beide banden aan 25 volwassen beoordelaars (psychologie-studenten) voorgelegd. Deze beoordelaars werd gevraagd iedere student te beoordelen op een zeven-puntzelfverzekerschaal, lopend van -3 = 'zeer onzeker' naar +3 = 'zeer zelfverzekerd'. Voor alle beoordeelde personen waren de verschillen tussen de gemiddelde zelfverzekerscores voor de twee versies significant met $p < .001$. De beoogde verschillen waren dus inderdaad aanwezig. Gesommeerd over de zeven studenten bedroeg de gemiddelde score voor de onzekere rol -1.2, voor de zelfverzekerde rol +1.3.

Beoordelingsdimensies.

Besloten werd de stimuluspersonen te laten beoordelen op een aantal psychologische eigenschappen en vaardigheden die leraren volgens kinderen wel of juist niet dienen te bezitten. Docking (1980) geeft een overzicht van eigenschappen en vaardigheden die leerlingen uit het voorgezet onderwijs belangrijk vinden. Om te verifiëren of basisschoolleerlingen dezelfde opvattingen hebben, werd een vooronderzoek gedaan. Hierin werd aan 29 tweede klassers, 29 derde klassers, 21 vierde klassers, 25 vijfde klassers en 24 zesde klassers een vragenlijst voorgelegd waarin een honderdtal mede aan Docking (1980) ontleende psychologische eigenschappen en vaardigheden was opgenomen. De proefpersonen dienden voor elk item aan te geven of zij de betreffende eigenschap/vaardigheid al of niet gewenst achtten voor leraren. Gekozen kon worden uit de antwoordmogelijkheden 'ja', 'nee', 'soms wel - soms niet' en 'het maakt me niet uit'. Items die niet werden begrepen, mochten worden overslagen.

Voor elk item werd per klas vastgesteld hoeveel procent van de kinderen ieder der antwoordmogelijkheden had gekozen. Vervolgens werd nagegaan welke items aan de volgende voorwaarden voldeden: a) in minstens drie van de vijf groepen kiest minimaal 75% van de kinderen het ja (nee) antwoord en b) in de eventuele overige groepen is het aantal ja (nee) antwoorden groter dan 50% en is de som van de ja (nee) antwoorden en de 'soms wel - soms niet'- antwoorden $\geq 75\%$. In totaal 40 items voldeden aan dit criterium. Hieruit werden 19 items geselecteerd voor opname in de beoordelingslijst, te weten: slim, eerlijk, rustig, aardig, aansteller, verlegen, saai, zeker van zichzelf, streng, lief voor kinderen, sloom, geduldig, kan zorgen dat kinderen hard werken, kan goed les geven, kan goed grapjes maken, kan een klas rustig houden, kan kinderen begrijpen, kan goed tegen een grapje, kan spannend vertellen.

Procedure.

Het onderzoek werd klassikaal verricht. De helft van de klassen zag band I, de andere helft band II. Steeds nadat een stimulus-persoon volledig was getoond, werd de volledige lijst voor die persoon ingevuld, door achter elk item één van zes antwoordmogelijkheden te omcirkelen. Deze waren respectievelijk: 'heel erg', 'erg', 'tamelijk', 'weinig', 'heel weinig' en 'ik weet het niet'. Van de laatste mogelijkheid werd slechts sporadisch gebruik gemaakt: in alle leeftijdsgroepen in minder dan één promille van de gevallen.

RESULTATEN

Als ruwe data waren 7 (studenten) $\times$ 19 (schalen) = 133 scores per proefpersoon beschikbaar. De eerste stap in de analyse was een reductie van dit aantal. Op het gereduceerde aantal scores werden vervolgens factoranalyses uitgevoerd, waarna de daaruit resulterende factorscores aan een MANOVA werden onderworpen.

Om een analyse van het effect van de verschillen in zelfverzekerdheid op de oordelen mogelijk te maken, werden voor elke pp per schaal twee scores berekend, te weten de gemiddelde score van de studenten die een zelfverzekerde rol hadden gespeeld en de gemiddelde score van de studenten die een onzekere rol hadden gespeeld. In de eerste factoranalyse (A) werden de verbanden geanalyseerd tussen de 19 gemiddelde scores behorend bij de vier studenten die op band I de zelfverzekerde en op band II de onzekere rol hadden gespeeld, in de tweede factoranalyse (B) werden de 19 scores betrokken behorende bij de drie studenten die op band I de onzekere en op band II de zelfverzekerde rol speelden.

Factoranalyse A resulteerde na varimax rotatie en bij een eigenwaardecriterium ≥ 1.00 in drie factoren, die tezamen 60% van de variantie verklaarden. De eerste factor kan het best worden gekarakteriseerd als 'Leraarscapaciteiten'. De schalen die $\geq .40$ laadden op deze factor zijn: goed les geven (.75), kinderen hard laten werken (.74), een klas rustig houden (.72), kinderen begrijpen (.67), geduldig (.57), eerlijk (.54), lief (.53), slim (.44), goed kunnen vertellen (.42), aardig (.43), rustig (.42). Op de tweede factor laadden ($\geq .40$) de schalen sloom (.83), saai (.75), aansteller (.74), verlegen (.66) en streng (.59). Deze factor laat zich goed karakteriseren door de term 'Ontoegankelijkheid': saai, sloom en verlegen gedrag getuigt van weinig toegankelijkheid voor anderen. Strengheid zet grenzen op de gedragsmogelijkheden van anderen en maakt daardoor de grenzenstellende persoon als persoon onbereikbaar. Iemand tenslotte die zich aanstelt wordt waarschijnlijk ervaren als met zijn aandacht bij zichzelf. Centraal bij factor III staat 'Onderhoudendheid', zowel in actieve als in passieve zin. Op deze factor laadden de schalen 'grapjes kunnen maken' (.84), 'grapjes kunnen verdragen' (.77), 'lief' (.55), 'aardig' (.50) en 'goed kunnen vertellen' (.49).

Op de factorscores resulterend uit factoranalyse A werd een 3 (Groepen) $\times$ 2 (Geslacht ppn) $\times$ 2 (Rol) MANOVA uitgevoerd. Significant ($p \leq .05$) waren de effecten van Rol en Groep. Univariante analyses wezen uit dat het Roleffect wordt verklaard uit het feit dat de studenten in hun zelfverzekerde rol als minder 'ontoegankelijk' (factor II) werden gezien. Wat betreft het Groepseffect wezen univariante analyses uit dat met de toename van de leeftijd van de ppn minder leraarscapaciteiten (factor I) en een grotere ontoegankelijkheid (factor II) aan de studenten worden toegeschreven.

Factoranalyse B leverde voor een belangrijk deel dezelfde resultaten op als factoranalyse A, zij het dat nu vier factoren werden gevonden. Deze verklaarden gezamenlijk 58% van de variantie. De eerste factor betrof wederom leraarscapaciteiten. Op deze factor laadden: grapjes kunnen maken (.78), eerlijk (.74), klas rustig kunnen houden (.72), goed kunnen vertellen (.72), grapjes kunnen verdragen (.70), aardig (.68), goed les geven (.67), kinderen

begrijpen (.64), kinderen hard laten werken (.64), slim (.58), lief (.49). In deze analyse vormen dus de factoren I en III uit analyse A één factor. Factor II kan wederom als 'Ontoegankelijkheid' worden gekarakteriseerd. Ook nu laadden daarop de schalen saai (.80), verlegen (.68), streng (.68), aansteller (.66) en sloom (.51). Factor III is moeilijk interpreteerbaar. Op deze factor laadden geduldig (.70), zeker van zichzelf (.69) en lief (.51). Op factor IV tenslotte laadden 'rustig' (.80) en 'sloom' (.58). 'Kalmte' lijkt een goede karakterisering van deze factor.

Een MANOVA op de factorscores resulterend uit factoranalyse B leverde wederom een significant effect op van Rol en Groep. Univariante vervolganalyses wezen uit dat het Roleffect wordt verklaard uit het feit dat de studenten in hun zelfverzekerde rol hoger scoren op de factor 'Leraarscapaciteiten' en lager op de factor 'Ontoegankelijkheid' en op de factor 'Kalmte'. Wat betreft het Groepseffect wezen de univariate analyses uit dat met de toename van de leeftijd van de ppn minder leraarscapaciteiten aan de studenten worden toegeschreven.

Behalve voor de verschillen tussen de rollen waren de ppn ook gevoelig voor de natuurlijke verschillen tussen de studenten. Dit blijkt uit de resultaten van negentien 7 (Studenten) x 3 (Groepen) x 2 (Geslacht ppn) MANOVA's waarin de studenten met elkaar werden vergeleken op elk der 19 variabelen. Bij alle variabelen behalve 'zeker van zichzelf' en 'kinderen begrijpen' was het hoofdeffect Student significant ($p \leq .05$). Met andere woorden, op nagenoeg alle dimensies werden significante verschillen tussen de zeven studenten gezien. Slechts tweemaal was er een interactie Student x Groep, zodat geconcludeerd kan worden dat de leeftijdsgroepen i.h.a. op dezelfde wijze de verschillen tussen de studenten waarnamen.

DISCUSSIE

Alle drie de onderzoeksvragen blijken bevestigend beantwoord te kunnen worden. Ten aanzien van de eerste vraag 'Schrijven de kinderen op basis van de waarneming van gedrag psychologische eigenschappen aan anderen toe?' bleek dat slechts in een zeer beperkt aantal gevallen van de 'ik weet het niet'-antwoordmogelijkheid gebruik werd gemaakt. De proefpersonen hadden dus blijkbaar geen moeite met het verzoek de getoonde personen te beoordelen. Men zou nog kunnen tegenwerpen dat de proefpersonen om redenen van meegaandheid met de proefleider afzagen van het gebruik van het 'ik weet het niet'-antwoord. Een dergelijke houding zou echter hebben moeten leiden tot willekeurig invullen van de schalen en dientengevolge tot een gebrek aan samenhang tussen de beoordelingsdimensies en tot het ontbreken van verschillen in oordelen over de verschillende studenten en over de beide rollen van elke student. Uit de resultaten bleek echter duidelijk dat dit niet het geval is. De tegenwerping kan dus terzijde worden geschoven.

Volgens de literatuur zou bij 8-12 jarige kinderen nog niet of nauwelijks samenhang bestaan tussen hun psychologische oordelen over anderen (Flavell, 1985; Shantz, 1983). Dit onderzoek laat zien dat deze opvatting dient te worden bijgesteld, in ieder geval wat betreft eigenschappen waarop kinderen hun leerkrachten beoordelen. Uit beide factoranalyses kwamen immers enkele merendeels goed interpreteerbare factoren naar voren. Hierbij is van belang dat de oordelen van de ppn gebaseerd waren op een zeer beperkt 'contact' met de te beoordelen personen waarin bovendien niet of nauwelijks specifieke informatie verschaft werd over de status van de studenten op de beoordelingsdimensies. De ppn hadden dus blijkbaar een aantal ideeën in hun hoofd over het samengaan van de onderzochte kenmerken. Met andere woorden, kinderen van deze leeftijd houden er

kennelijk al een impliciete persoonlijkheidstheorie op na die hen 'in staat stelt op grond van een geringe hoeveelheid informatie over een persoon of de sociale groep waartoe die persoon behoort allerlei veronderstellingen te maken over de verdere eigenschappen en gedrag van die persoon' (Willemsen & v.d. Kloot, 1987).

De toegeschreven eigenschappen en capaciteiten varieerden met de mate van zelfverzekerdheid die de te beoordelen personen uitstraalden. De kinderen merkten dus de verschillen in zelfverzekerdheid van gedrag op en dit had gevolgen voor hun beoordelingen van de studenten. In hun zelfverzekerde rol werd over de studenten gunstiger geoordeeld dan wanneer zij de onzekere rol speelden. Hieruit blijkt opnieuw de werkzaamheid van een impliciete persoonlijkheidstheorie bij kinderen tussen acht en twaalf jaar.

De oordelen van de kinderen variëren ook met de te beoordelen persoon. Immers op bijna alle dimensies wordt verschillend geoordeeld over de zeven studenten. Mogelijk hangen die verschillen in oordelen samen met natuurlijke verschillen tussen de studenten in de mate van zelfverzekerdheid die ze ten toon spreiden. Ook andere factoren (fysionomie, kleding) kunnen echter een rol gespeeld hebben. Nader onderzoek zou dat moeten uitwijzen. Ongeacht de uitkomst daarvan is het interessant dat slechts een korte waarneming van personen over wie zeer beperkte informatie wordt verschaft voldoende is voor het vormen van een oordeel over deze personen.

Tenslotte blijkt ook de leeftijd van de ppn effect te hebben op de oordelen: de jongste ppn oordeelden een aantal malen gunstiger over de studenten dan de oudere kinderen. Dit resultaat komt overeen met de bevindingen in eerder onderzoek (Rholes & Ruble, 1984, voor ook een overzicht). Jongere kinderen blijken er positievere verwachtingen op na te houden over het gedrag van anderen en in het algemeen een grotere voorkeur te hebben voor positieve informatie. Verschillende factoren zouden voor de afnemende positiviteit in oordelen over anderen verantwoordelijk kunnen zijn. Rond de leeftijd van tien jaar beginnen kinderen aanzienlijk kritischer over anderen te worden (Lievegoed, 1946). Wellicht hebben ze daarom meer informatie over anderen nodig alvorens tot eenzelfde niveau van positiviteit in oordelen te kunnen komen als kinderen van jongere leeftijd en zijn hun oordelen bij gebrek aan die informatie gematigder. Ook neemt met toenemende leeftijd het inzicht toe dat gedrag voorgewend kan zijn toe (Gnepp & Hess, 1986). Dit zou eveneens kunnen leiden tot een grotere voorzichtigheid in het toeschrijven van positieve eigenschappen.

In dit onderzoek werden slechts eerste indrukken gemeten. Een mogelijke vraag is of deze eerste indrukken, conform de heersende populaire opvattingen hierover, ook juist waren. Ook zou men zich kunnen afvragen of en hoe deze eerste indrukken bij toenemende bekendheid met de beoordeelde personen veranderen. Het onderzoek was er niet op gericht deze vragen te beantwoorden. Ook zonder antwoorden op deze vragen zijn de uitkomsten echter van belang. Praktisch gezien is dit het geval omdat aangenomen mag worden dat de wijze waarop kinderen een nieuwe leerkracht tegemoet treden sterk wordt bepaald door de eerste indrukken die zij van hem/haar hebben. In die zin is het van belang te weten dat (ook) kinderen zich zeer snel een eerste indruk van anderen kunnen vormen op psychologische dimensies. Theoretisch zijn de uitkomsten van belang omdat zij, indien bevestigd in nader onderzoek, nopen tot bijstelling van de in de literatuur aan te treffen opvattingen terzake.

LITERATUUR

- Barenboim, C. (1981). The development of person perception in childhood and adolescence: from behavioral comparisons to psychological constructs to psychological comparisons. *Child Development*, 52, 129-144.

- Berndt, Th.J. & Heller, K.A. (1985). Measuring children's personality attributions: responses to open-ended questions versus trait ratings and prediction of future behavior. In: S.R. Yussen (Ed.) *The growth of reflection in children*. New York: Academic Press.
- Bruner, J.S. & Tagiuri, R. (1954). The perception of people. In: G. Lindzey (Ed.) *Handbook of social psychology*, Vol. 2. Cambridge, Mass.: Addison-Wesley.
- Bullock, M. & Russell, J.A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioral Development*, 7, 193-214.
- Docking, J.W. (1980). *Control and discipline in schools*. London: Harper & Row.
- Flavell, J.H. (1985). *Cognitive Development*. Englewood Cliffs, N.J.: Prentice Hall.
- Gnepp, J. & Hess, D.L.R. (1986). Children's understanding of verbal and facial display rules. *Developmental Psychology*, 22, 103-108.
- Harper, R.G. (1985). Power, dominance and nonverbal behavior. An overview. In: S.L. Ellyson & J.F. Dovidio (Eds.) *Power, dominance and nonverbal behavior*. New York: Springer.
- Kassin, S.M. & Pryor, J.B. (1985). The development of attribution processes. In: J.B. Pryor & J.D. Day (Eds.) *The development of social cognition*. New York: Springer Verlag.
- Lievegoed, B.C.J. (1946). *Ontwikkelingsfasen van het kind*. Zeist. Uitgeverij Vrij Geestesleven.
- Piaget, J., Grize, J.B., Szeminska, A.D., & Vinh Bang (1977). *Epistemology and psychology of functions*. Dordrecht: D. Reidel.
- Rholes, W.S. & Ruble, D.N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, 55, 550-560.
- Rogers, C. (1978). The child's perception of other people. In: H. McGurk (Ed.) *Issues in childhood social development*. London: Methuen.
- Shantz, C.U. (1983). Social Cognition. In: P.H. Mussen (Ed.) *Handbook of Child Psychology*, Vol. III.
- Willemsen, T.M. & Kloot, W.A. van der (1987). Een ruimtelijk model voor attributie processen. In: A. van Knippenberg, M. Poppe, J. Extra, G.J. Kok, E. Seydel (Eds.) *Fundamentele Sociale Psychologie*, deel 2. Tilburg: Tilburg University Press.

Manuscript ontvangen 25-5-1988

Definitieve versie ontvangen 15-3-1989

Notities en Commentaren

VOORRANG AAN ACHTERSTAND?

EEN COMMENTAAR OP HET GELIJKNAMIGE ARBO-ADVIES

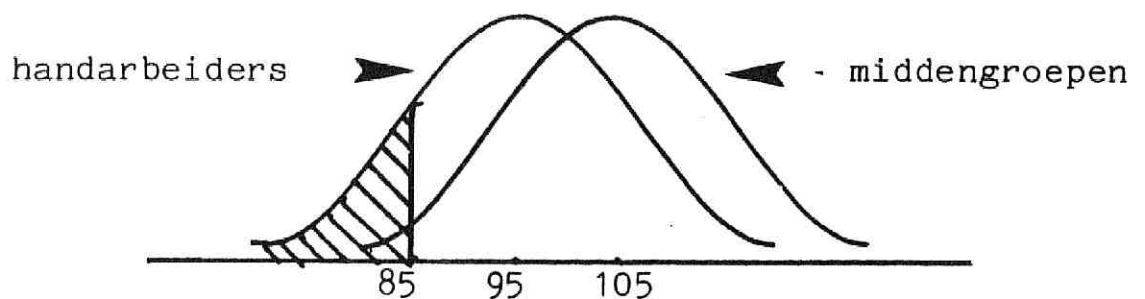
G. W. Meijnen

Universiteit van Amsterdam, Vakgroep Onderwijskunde

Op 20 januari 1987 verwierp de Eerste Kamer het wetsvoorstel voor het onderwijsvoor-rangsbeleid. Hoewel de discussie zich toespitste op problemen inzake de uitvoering, speelden verschillen in ideologische benadering van de problematiek waarschijnlijk de belangrijkste rol. Het onderwijsvoor-rangsbeleid is gestart als initiatief ter politieke linkerzijde tijdens het kabinet Den Uyl, zijnde het antwoord op de milieuspecifieke selectie in het onderwijs. Later werd de problematiek van etnische minderheden en – in mindere mate – meisjes ook onder de werking van dit beleid gebracht. Extra faciliteiten werden, en worden, aan scholen toegekend op grond van de sociale en/of etnische herkomst van leerlingen. Een facilitering op basis van **groepskenmerken** dus. Collectieve emancipatie van de genoemde groepen is daarbij het oogmerk. Empirisch manifesteert de achterstand van zowel arbeiderskinderen als leerlingen uit etnische minderheden zich door lagere gemiddelde scores met betrekking tot zowel de leervakken als de gemeten intelligentie. Dat lagere gemiddelde wordt veroorzaakt door òn een **onder**vertegenwoordiging in de hogere prestatiecategorieën òn een **over**vertegenwoordiging in de lagere.

In termen van spreiding en gemiddelden laat de problematiek zich als overlappende normaalverdelingen beschrijven. (zie figuur 1)

Bestrijding van onderwijsachterstanden als groepskenmerk is vanuit deze filosofie gericht op het naar rechts doen opschuiven van de normaalverdeling van de prestaties van de doelgroepen. Een actueel redelijk presterend arbeiderskind is wellicht een potentiële hoogpresteerder!



Figuur 1. Spreiding en gemiddelden van prestatie-scores van handarbeiderskinderen en kinderen uit de middengroepen.
(De gemiddelden zijn enigszins arbitrair, afhankelijk als ze zijn van wat onder de beide statusgroepen wordt begrepen.)

Ter rechterzijde in de politiek wordt een meer individualistisch georiënteerde visie aangehangen en wordt positieve discriminatie gereserveerd voor leerlingen die, om welke reden dan ook, aanmerkelijk achterblijven. Herhaaldelijk is vanuit deze hoek dan ook gepleit voor een toewijzing van extra middelen aan scholen op basis van werkelijk geconstateerde achterstanden, te bepalen volgens een daartoe geëigende test- en/of toetsprocedure. In termen van de bovengeschetste figuur zou het dan de leerlingen betreffen die bijvoorbeeld lager dan 85 scoren. Daaronder bevinden zich dan, gelet op het gearceerde vlak, wel disproportioneel veel kinderen uit de doelgroepen van het huidige onderwijsvoorrangsbeleid.

Ook met betrekking tot de voorgestane oplossingsstrategieën zijn er verschillen te constateren. De groepsgerichte benadering legt de nadruk op de interactie tussen buiten- en binnenschoolse factoren (o.a. milieubias in beoordelingen van leerkrachten) en op cultuurverschillen (omgangsvormen, cognitieve stijlen, taalgebruik, waardenoriëntaties) en propageert vervolgens o.a. maatregelen ter overbrugging van de verschillen tussen school- en thuiscultuur, onderwijs in eigen taal en cultuur, roldoorbrekend onderwijs en het mobiliseren van buitenschoolse hulpbronnen. De individugerichte visie benadrukt in veel mindere mate de sociale achtergronden van de onderwijsachterstanden en richt de inspanning rechtstreeks op de te verbeteren instructie aan deze leerlingen binnen de muren van de school.

Het ARBO-advies nu, kiest het politieke midden: de toewijzing van de extra middelen moet gebaseerd blijven op de huidige groepskenmerken, maar de gepropageerde oplossingsstrategieën zijn sterk onderwijstechnologisch van aard en niet groepsspecifiek. Achterstand is achterstand en door aanwending van maatregelen die gebaseerd zijn op het onderzoek naar de effectieve school (differentiatie, leerlingvolgsystemen, accentuering instrumentele vaardigheden) dient deze te worden bestreden.

Milieu- en etnospecifieke ongelijke onderwijskansen worden qua stijl van analyse en remediëring weer aangepakt als in het begin van de zestiger jaren. Een maatschappelijke visie op ongelijke onderwijskansen ontbreekt geheel. Of we zouden deze moeten zoeken achter de aanbevelingen om scholen contracten te laten afsluiten met welzijnsinstellingen als bibliotheken, peuterspeelzalen en buurt- en clubhuizen over de inzet van de middelen die de overheid steekt in het welzijnswerk ten behoeve van het onderwijsvoorrangsbeleid.

Het voorgaande impliceert niet dat al de aanbevelingen derhalve geen betekenis hebben voor het bestrijden van de ongelijke onderwijskansen van arbeiderskinderen en allochtone leerlingen. Ook zij zijn ongetwijfeld gebaat bij effectieve instructie, adequate leerlingvolgsystemen en goede rapportagesystemen. Geredeneerd vanuit de driehoeksrelatie school-gezin-buurt en in het bijzonder vanuit de overtuiging dat het gezinsmilieu een dominerende rol speelt in de schoolcarrière van kinderen is de aanpak van de ARBO echter nogal eenzijdig. Waarom worden voor de inzet van de bibliotheek of voor het betrekken van ouders bij de school niet even uitgewerkte maatregelen gepresenteerd als voor wat binnen de school moet gebeuren?

Materiaal waarop de Raad zich ook in dit opzicht kan baseren is er in overvloed (GEON, OSM).

Tenslotte iets over de door de Raad voorgestane wettelijke inkadering van het onderwijsvoorrangsbeleid. Zij pleit voor het onderbrengen binnen de huidige sectorwetgeving en niet voor een aparte onderwijsvoorrangswet. Vanuit de visie van de Raad dient de school centraal te staan en dienen de welzijnsactiviteiten daarop te worden afgestemd. Dat betekent een verschuiving in de machtsverhoudingen: verschillen in visies op te voeren beleid tussen scholen en welzijnsinstellingen en daaruit voortvloeiende conflicten zullen afnemen. Een efficiency-verhogend effect is daarvan het gevolg. Er dreigt echter ook een gevaar: als de toegewezen middelen niet worden geormerkt voor de

samenwerking met welzijnsinstellingen noch voor het overbruggen van de verschillen tussen school- en thuiscultuur, dan is de kans groot dat het onderwijsvoorrangsbeleid inderdaad verwordt tot uitsluitend zorgverbredingsbeleid.

Ontvangen 10-10-1988

Boekbesprekingen

J. B. J. Koelman, *Kosten van de verzuiling; een studie over het lager onderwijs*. Den Haag: VUGA, 1987.

Hoewel deze dissertatie bij zijn verschijnen veel stof deed opwaaien, is het geen ééndagsvlieg, die vergeten zal worden zodra het moment van de emotie vervlogen is. Het is een zeer informatieve studie over de kosten van ons verzuild onderwijsstelsel, die nog lange tijd bruikbaar en actueel zal blijken. Dat is allereerst te danken aan de hardnekkigheid van de verzuiling binnen het onderwijs. Maar dat is ook te wijten aan de grote boog die de meeste onderzoekers van het onderwijs om dit onderwerp heen maken, laat staan dat zij zich met de financiële zijde ervan zouden willen bezighouden.

De dubbele probleemstelling van dit boek is helder:

1. Leidt de vormgeving van de aan het verzuilde stelsel gebonden bekostigingssystematiek voor de materiële exploitatie-uitgaven tot additionele uitgaven?
2. Leiden de bekostigingsvoorwaarden, doordat schaalvoordelen te weinig kunnen worden benut tot additionele uitgaven?

Voordat Koelman aan de eigenlijke beantwoording van deze vragen toekomt, schetst hij in een aantal inleidende hoofdstukken een historisch perspectief van het openbaar en bijzonder lager onderwijs, de bekostigingsvoorwaarden en -systematiek van de lager onderwijswet, en de relatie tussen pluriformiteit en verzuiling in het lager onderwijs. Dit gebeurt helder en zonder al te moeilijk jargon. Het zou het een bruikbaar boek in universitair onderwijs maken. Toch bevatten deze inleidende hoofdstukken een ernstig tekort, vooral vanuit het gezichtspunt van de huidige stand van kennis in de onderwijs-research. Koelman meent, in navolging van Boef-van der Meulen en Van Kemenade (1986), dat de belangrijkste verklaring van de blijvende verzuiling in het onderwijs de verankering van deze verzuiling in wetten en politieke machtsverhoudingen is. Deze verklaring was echter reeds verouderd toen Boef-Van der Meulen en Van Kemenade hem opschreven. Een andere verklaring is dat de openbare en bijzondere scholen verschillen in hun effectiviteit. Leerlingen op bijzondere scholen hebben betere schoolloopbanen dan leerlingen op openbare scholen. Wat ook de oorzaken van dit effectiviteitsverschil zijn, het kan alvast verklaren waarom ook onkerkelijke ouders een voorkeur vertonen voor confessioneel onderwijs, ondanks de secularisering van de samenleving en los van de wettelijke verankering van de verzuiling. Koelman kent deze verschillen in effectiviteit tussen openbare en bijzondere scholen niet, ondanks het feit dat de eerste onderzoeksresultaten daarover op dezelfde Onderwijsresearch Dagen gepresenteerd werden als waarop hij zijn onderzoekopzet naar voren bracht (o.a. Van Laarhoven, e.d.). Bovendien waren daarvoor reeds buitenlandse onderzoeksresultaten bekend, die lieten zien dat er effectiviteitsverschillen bestaan tussen openbare en private scholen (o.a. Coleman).

Dit leidt tot twee tekorten in zijn analyses. Allereerst onderzoekt hij niet of naast de additionele kosten die voortvloeien uit de verzuiling ook additionele opbrengsten van de verzuiling bestaan. In feite aanvaardt hij ten onrechte stilzwijgend de juridische fictie dat alle Nederlandse scholen hetzelfde eindproduct opleveren. Het verwaarlozen van deze additionele opbrengst van verzuiling brengt tegelijkertijd met zich mee dat zijn schatting van de additionele uitgaven, voortvloeiend uit de verzuiling van het onderwijs (1,5 à 2 miljard gulden) onjuist kan zijn. Ten tweede betekent deze negatie van effectiviteitsverschillen tussen scholen dat Koelman in zijn analyses van hoofdstuk VIII het aantal leerlingen ten onrechte hanteert om de additionele kosten van verzuiling te schatten, in

plaats van een echte produkt-meting (bijv. het aantal leerlingen dat op een bepaald niveau de basisschool verlaat).

Een sterk punt van deze dissertatie is dat Koelman niet alleen de bestaande samenhangen tussen verzuilingsgraad, schoolgrootte en onderwijskosten analyseert (hoofdstuk VIII), maar ook een simulatie-model opstelt, waarmee hij de financiële gevolgen van verschillende aspecten van verzuiling (met name het aantal kleine scholen) kan berekenen. Met behulp van een aantal beslissingsstappen wordt een besluitvormingsproces geconstrueerd (of gesimuleerd), dat tot een bepaald eindprodukt leidt. In het onderhavige geval gaat het besluitvormingsproces over het aantal scholen dat in een gemeente gesticht of gehandhaafd wordt; de beslissingsstappen gaan over de voorwaarden om die scholen te mogen stichten of handhaven en het eindprodukt betreft de personele kosten nodig voor die scholen. Simulatie-modellen zijn in de onderwijsresearch helaas zeldzaam, ondanks hun grote potentiële mogelijkheden. Vanuit dat gezichtspunt is het prijzenswaardig dat Koelman een simulatie-model opstelt en gebruikt. Toch vind ik de uitkomsten van het simulatie-model niet geheel overtuigend. Het model blijkt vooral gevoelig voor de gebruikte normen voor het leerlingenaantal per school, vereist bij de stichting van een school in een gemeente. Het model blijkt minder gevoelig voor het aantal verzuilde scholen binnen de eerder genoemde normen voor het leerlingenaantal per school. Koelman leidt daaruit af dat vooral de te soepele toepassing van de stichtings- en opheffingsnormen voor scholen binnen het huidige verzuilde onderwijs verantwoordelijk zijn voor de additionele kosten van de verzuiling. Hoewel zijn analyse van de bestaande samenhangen tussen verzuilingsgraad, schoolgrootte en onderwijskosten, dit belang van schoolgrootte als schakel tussen verzuilingsgraad en onderwijskosten per gemeente ondersteunen, ben ik niet overtuigd door de uitkomsten van het simulatie-model. Te veel rusten deze uitkomsten op de opdeling van Nederland in denkbeeldige voedingsgebieden voor de scholen. Het zou derhalve goed zijn geweest indien Koelman zijn analyse van de bestaande samenhangen tussen verzuilingsgraad, schoolgrootte en onderwijskosten (hoofdstuk VIII) gebruikt had, voor een tweede schatting van effecten van verschillende normen voor de stichting van scholen bij uiteenlopende verzuilingsgraden. De daarvoor vereiste methode is binnen de sociologie reeds lang bekend en wordt geregeld toegepast in Nederland.

Al deze bezwaren nemen niet weg dat ik deze dissertatie alleszins de moeite waard vindt en hem zeer bruikbaar acht voor de onderwijsresearch. Dat laatste is echter alleen mogelijk indien het ook door Koelman gesignaleerde 'taboe' op verzuilingsvraagstukken verdwijnt.

J. Dronkers
Katholieke Universiteit Brabant
Postbus 90153
5000 LE Tilburg.

W. H. N. Jansweijer, *PDP: Een benadering vanuit de kunstmatige intelligentie van probleemoplossen en leren door doen in een semantisch rijk domein*. Academisch proefschrift: Universiteit van Amsterdam, 1988.

Eén van de belangrijkste manieren waarop wij mensen ons vaardigheden eigen maken en perfektioneren is door veel te oefenen. Vaak is een mondelinge of schriftelijke instructie onvoldoende om te leren. In Jansweijers proefschrift 'PDP: Een benadering vanuit de kunstmatige intelligentie van probleemoplossen en leren door doen in een semantisch rijk domein' wordt een poging beschreven om het mechanisme te modelleren dat ten grondslag ligt aan dit soort prestatieverbeteringen. Hij koos daarvoor het domein van het oplossen van opgaven in de warmteleer.

Jansweijer ging daarbij op de volgende manier te werk. Hij deed een formele analyse van de taak die hij aan zijn proefpersonen voorlegde, met name van de kennis en vaardigheden waarop de taak een beroep doet. Verder observeerde hij het feitelijke aanpakgedrag van beginnende en gevorderde probleemoplossers door ze een serie opgaven hardop denkend te laten uitvoeren. Op grond van deze twee gegevensbronnen werden twee computermodellen geconstrueerd: PDP-1, een model dat het probleemoplossen van de gevorderde probleemoplosser simuleert en PDP-0, een model dat het probleemoplossen van de beginner simuleert. Vergelijking van deze twee modellen zou dan het gewenste inzicht in 'leren door doen' moeten opleveren. Jansweijer richt zich in zijn proefschrift op theorievorming en modellering. Hij komt er helaas maar zeer ten dele aan toe hypothesen die voortkomen uit deze theorieën te toetsen aan nieuw empirisch materiaal.

PDP-1, het model van de gevorderde probleemoplosser, bevat twee soorten kennis, domeinkennis en kennis over de manier waarop een probleem in de warmteleer wordt aangepakt en opgelost. Verder bevat het model een **probleemoplosser** die met behulp van deze kennis de eigenlijke probleemoplossingsactiviteiten mogelijk maakt. Door het probleemoplossingsgedrag van de menselijke probleemoplosser te vergelijken met het gedrag van het model, is het mogelijk na te gaan of het model een goede simulatie oplevert. De resultaten van zo'n vergelijking suggereren dat PDP-1 een goede simulatie vormt van probleemoplossen door gevorderden in de warmteleer.

In tegenstelling tot de oplossingsweg van gevorderden en van het model, leken de beginnersprotocollen gekenmerkt te worden door een ongestructureerd verloop. Hun probleemoplossen leek een aaneenrijging van mislukkingen, vastlopen en oplossen van impasses. Toch waren de beginners in staat om 78% van de vraagstukken op te lossen, maar bij slechts 9% van de protocollen trad geen impasse op. Deze impasses waren het gevolg van deficiënties in de kennis, die zowel van feitelijke aard waren als de wijze van probleem aanpakken betroffen.

Wanneer in de kennis van PDP-1 vergelijkbare hiaten werden aangebracht, dan bleek het model vast te lopen bij het oplossen van de gestelde opgaven. Het is niet in staat om, zoals de menselijke probleemoplosser, deze impasses te repareren. Een beginner is dus *meer* dan iemand die het nog niet weet. Dat *meer* is gevonden in de toevoeging van een **meta-probleemoplosser** die impasses signaleert, classificeert en met behulp van kennis over typen impasses en bijpassende reparatiemogelijkheden het probleemoplossingsplan bijstelt. Daarna wordt de **probleemoplosser** weer geactiveerd die het bijgestelde plan verder uitvoert. Jansweijer geeft helaas geen kwantitatieve gegevens over de overeenstemming tussen de oplossingswegen van dit beginnersmodel, PDP-0, en beginnende menselijke probleemoplossers in dit domein. Hij laat wel aan de hand van één vergelijking tussen mens en model zien, dat PDP-0 in staat is om impasses die ontstaan door gebrekkige kennis te overwinnen.

Wat zeggen deze PDP-modellen nu over het leerproces dat door het oplossen van series

opgaven plaats zou vinden? Noch PDP-1, noch PDP-0 bevatten een leercomponent. Jansweijer moet dus op andere bronnen terugvallen wanneer hij in hoofdstuk 8 van zijn proefschrift dit onderwerp aansnijdt. Hij baseert zijn theorie over leren door doen op een aantal pijlers: op psychologische en computationele modellen van het menselijk leerproces, op observaties van het leerproces dat zijn proefpersonen leken door te maken en op de observatie van Anzai en Simon (1979) dat in hun onderzoek de belangrijkste leerwinst optrad ten gevolge van een tussentijdse reflectie op de taak en de gehanteerde oplossingsmethoden.

Jansweijer concludeert dat de impasses en het repareren ervan de momenten bij uitstek zijn waarop geleerd kan worden, bijvoorbeeld doordat de leerling op zo'n moment het leerboek zal raadplegen om alles nog eens precies na te kijken. Verder wordt door het optreden en repareren van impasses geleerd welke methoden onder welke omstandigheden tot moeilijkheden kunnen leiden en daarom met extra zorg moeten worden uitgevoerd. De laatste en verreweg interessantste vorm van leren door impasses die Jansweijer beschrijft, bestaat eruit dat de leerling achteraf een rationale construeert voor de gehanteerde oplossingsmethode. De leerling beredeneert dan bijvoorbeeld waarom een bepaalde aanname toegestaan is of waarom de ene geprobeerde methode tot succes leidde waar de andere faalde.

Jansweijers theorie over impasse-gestuurd leren is zeer aantrekkelijk. Ze doet vermoeden dat de instelling waarmee een leerling opgaven maakt van invloed is op het leereffect dat uitgaat van oefenen. Met name de laatste wijze van leren zal sterk afhangen van de doelstelling die de leerling zich stelt. Wie de sommen zo snel mogelijk af wil hebben, zal zich er geen zorgen over maken waarom de gekozen methode toepasbaar is. Ook wat dat betreft doet de theorie van Jansweijer realistisch aan.

Helaas moet de geformuleerde theorie het verder van deze geloofwaardigheid hebben. Jansweijer toetst haar niet formeel, noch rapporteert hij illustratieve protocolfragmenten. Het in de appendix afgedrukte hardop-denken protocol van een beginnende probleemoplosser, laat zien dat deze proefpersoon bij deze opgave inderdaad informatie opzoekt in de cursustekst; voor de andere gepostuleerde processen geeft het geen aanwijzingen.

Sinds de opkomst van de cognitivistische benadering van begrijpen en leren heeft leren door doen lange tijd buiten de aandacht van psychologen gelegen. Jansweijers proefschrift levert alleen al daarom een belangrijke bijdrage aan de psychologie. Daarnaast heeft het proefschrift ook onderwijskundig gezien een aantal interessante aanknopingspunten. De eerste is, dat we er niet als vanzelfsprekend vanuit moeten gaan dat het gewenste leerresultaat vanzelf zal optreden wanneer er maar genoeg wordt geoefend. De andere is, dat het zaak is erop aan te sturen dat leerlingen die niet uit zichzelf een rationale achteraf voor de gekozen oplossingsweg construeren, daartoe op één of andere wijze moeten worden aangezet. Dit betekent een grote aandacht voor een kwalitatieve evaluatie van het antwoord, zowel door de docent als door de leerling. Beide aspecten zijn belangwekkend genoeg om verder empirisch onderzoek te rechtvaardigen.

Literatuur

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, 86, 124-140.

H. P. A. Boshuizen
 Vakgroep Onderwijsontwikkeling en Onderwijsresearch
 Rijksuniversiteit Limburg
 Postbus 616
 6200 MD Maastricht

C. J. W. Meijer, *Verwijzingen Gewogen. Een studie naar determinanten van verwijzing naar speciaal onderwijs*. Proefschrift Leiden, Monografieën onderwijsonderzoek 3, RION, Groningen, 1988. (132 p. - met lit. opg. - met samenvatting in het Engels).

Deze studie beperkt zich tot een beknopte analyse van de uitstroom uit het basisonderwijs naar het speciaal onderwijs. De nadruk valt op de twee grootste schooltypen voor speciaal onderwijs: het LPOM- en het MLK-onderwijs. Tussen 1970 en 1985 is bij deze twee schooltypen samen een duidelijke procentuele stijging te constateren van 0,58% naar 0,74%. Dit betekent een toename van 28%. Deze ontwikkeling roept veel vragen op.

Op basis van literatuurstudie en een vooronderzoek zijn een aantal verwachtingen geformuleerd ten aanzien van factoren die van invloed zouden zijn op de verwijzing naar speciaal onderwijs. Die verwachtingen liggen op vier niveaus: van de leerkracht, van de klas/groep, van de school en van de interactie tussen leerkracht en leerling.

Wat de leerkracht betreft hebben die verwachtingen betrekking op de opleiding en deskundigheid, ervaring in het speciaal onderwijs, het aantal dienstjaren, de attitude ten aanzien van probleemleerlingen, de perceptie van eigen kunnen, het geslacht en het zelf hebben van kinderen.

Bij de klas of groep houden die verwachtingen verband met de klassegrootte, het al of niet combinatieklas zijn en de mate van differentiatie.

Schoolkenmerken die in de verwachtingen werden betrokken zijn het al of niet deelnemen aan innovatieprojecten, de bereikbaarheid van speciaal onderwijs, aanwezige deskundigheid in het team, de sociale status van de leerlingpopulatie, de attitude van hoofd en leerkrachten ten aanzien van probleemleerlingen, het teamfunctioneren, de signatuur en de grootte van de school.

Van een drietal leerlingkenmerken, met name geslacht, milieu en het soort probleem wordt de statistische interactie nagegaan met de genoemde leerkrachtkenmerken.

Het onderzoek bestaat uit twee fasen: een vooronderzoek van 1980-1982 en een hoofdonderzoek van 1983-1986. Op grond van in het vooronderzoek gegenereerde hypothesen is een opzet gemaakt voor het hoofdonderzoek. De rapportage heeft in hoofdzaak betrekking op het hoofdonderzoek.

Als meetinstrumenten zijn vragenlijsten gebruikt, waarmee bij 400 leerkrachten uit groep 4 informatie is verzameld over de relevante leerkrachtkenmerken en de afhankelijke variabele 'verwijzingsgeneigdheid'. Daarnaast is met behulp van het aanbieden van gevalsbeschrijving de interactie tussen leerkracht en leerling geanalyseerd. De wijze waarop de verschillende constructen in het onderzoek zijn geoperationaliseerd wordt vrij duidelijk aangegeven.

De gerapporteerde resultaten vallen uiteen in twee groepen. De eerste groep heeft betrekking op school-, onderwijs- en leerkrachtkenmerken, die gerelateerd worden aan verwijzingscijfers naar het speciaal onderwijs. De tweede groep betreft de relatie tussen leerkrachtfactoren en verwijzingsgeneigdheid.

Wat de eerste groep gegevens betreft worden de voorafgestelde hypothesen met betrekking tot de relatie tussen schoolfactoren en de verwijzing naar het speciaal onderwijs voor het merendeel *niet* bevestigd. Oppervlakkige schoolkenmerken, waaronder worden gerekend de signatuur, de grootte en de afstand tot speciale onderwijsvoorzieningen, noch meer onderwijsnabije kenmerken zoals deskundigheid, attitude ten aanzien van probleemleerlingen, innovatie-status en teamfunctioneren, blijken op een systematische wijze in te werken op de uitstroom uit het basis- naar het speciaal onderwijs. Opmerkelijk is dat een aantal resultaten als contra-intuïtief beschouwd kunnen worden. Zowel bij klassegrootte als bij differentiatie wordt een significant verband gevonden met verwijzingscijfers, maar

die verbanden zijn tegengesteld aan de verwachtingen. Zo worden uit kleine klassen meer leerlingen verwezen dan uit grote klassen. Als verklaring hiervoor geeft de auteur de zogenaamde sensitiviteits-these: naarmate leerkrachten meer sensitief zijn voor leerlingen met problemen en daarover meer informatie inwinnen en verkrijgen, neemt het aantal verwijzingen toe. Dit is een plausibele verklaring, maar niet de enig mogelijke. Het is bekend dat bij differentiatie de kans op het zogenaamde Mattheus-effect toeneemt. In kleinere klassen zal men eerder overgaan tot differentiatie. Dit kan leiden tot minder aandacht voor de zwakke leerlingen, zoals bijvoorbeeld blijkt in combinatieklassen. Juist minder intelligente leerlingen hebben behoefte aan interactie met de leerkracht en zijn minder goed in staat zelfstandig, gemotiveerd verder te werken. Daardoor kan de effectieve leertijd dalen. Mogelijk zijn beide verklaringen van toepassing. Nader onderzoek is hier gewenst.

Bij de tweede groep resultaten staat het construct 'verwijzingsgeneigdheid' als afhankelijke variabele centraal in plaats van het verwijzingsgedrag. De reden is dat de verwijzingsfrequentie laag is en slechts weinig variantie in de verwijzingsaantallen wordt gevonden. Meer dan de helft van de leerkrachten heeft geen leerlingen in de klas bij wie verwijzing is overwogen of geëffectueerd. Dit heeft zijn consequenties voor de interpretatie van de resultaten, vooral doordat de stap van verwijzingsgeneigdheid naar empirische verwijzing ontbreekt in dit onderzoek. Daar komt nog bij dat gegevens over de betrouwbaarheid van de beoordelingen bij de gevalsbeschrijvingen, waaruit die verwijzingsgeneigdheid wordt afgeleid, ontbreken. Enige voorzichtigheid lijkt daarom op zijn plaats.

Er werden acht factoren die als determinant worden gezien voor verwijzingsgeneigdheid, onderzocht: de deskundigheid van de leerkrachten, ervaring in het speciaal onderwijs, het aantal dienstjaren, het geslacht van de leerkracht, het zelf kinderen hebben, vertrouwen in eigen kunnen, de grootte en samenstelling van de klas, interactie tussen bepaalde leerling- en leerkrachtkenmerken. Vrijwel alle verwachtingen met betrekking tot de relatie tussen leerkrachtkenmerken (en de interactie daarvan met leerlingkenmerken) en verwijzingsgeneigdheid dienen te worden verworpen. Alleen het al dan niet 'vertrouwen hebben in eigen kunnen' van de betreffende leerkracht is van invloed op de verwijzing.

De gevonden resultaten geven veel stof tot discussie. Aannemelijk lijkt de stelling, dat verwijzing naar het speciaal onderwijs niet de resultante is van een interactie tussen leerkracht en leerling, maar dient te worden opgevat als de resultante van een als onhoudbaar gepercipieerde classesituatie. Deze stelling is echter nog niet bewezen. Onderzoek dat zich richt op de 'analyse van onhoudbare situaties' zou wel eens vruchtbaar kunnen zijn. Immers, situaties zijn veranderbaar en vormen daarmee een effectieve ingang voor een onderzoeksprogramma, zoals de auteur terecht opmerkt.

Niet iedereen zal even gelukkig zijn met de resultaten van dit onderzoek. Bepaalde gegevens kunnen misbruikt worden om bezuinigingen te 'rechtvaardigen'. Maar ze kunnen ook worden 'gebruikt' om nog eens kritisch te kijken naar meningen die wat al te gemakkelijk verkondigd worden, zonder dat ze voldoende aan de realiteit zijn getoetst. Dit proefschrift verdient alle waardering en vormt hopelijk een aanzet voor verder onderzoek in de richting van 'classroom ecology' waarvan zorgverbreding een onderdeel dient uit te maken.

M. J. C. Mommers
K.U. Nijmegen
Vakgroep Interdisciplinaire Onderwijskunde
Erasmusplein 1-20
6525 GG Nijmegen

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

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

Jrg. 13, nr. 3

Juni 1989

Inhoudsopgave

Verslag ledenvergadering 25 mei 1989	198
Ontvangen publikaties	199

Verslag ledenvergadering 25 mei 1989

Aanwezig: de leden Creemers en Stijnen

de bestuursleden Beintema, Claessen, Roede en Van der Werf

1. Opening

Er is bericht van verhindering van Gerritsen van der Hoop, van Hout, Kanselaar en Scheerens.

2. Mededelingen/ingekomen stukken

Er zijn geen mededelingen en ingekomen stukken.

3. Verslag vorige ledenvergadering

Het verslag wordt vastgesteld.

4. Verslag verenigingsjaar 1988

N.a.v. punt 2 Onderwijs Research Dagen: De Cie themaplaning is van mening dat de door de VOR vastgestelde thema's te specifiek worden ingevuld. Die Cie-themaplaning toetst de invulling marginaal en heeft daarom weinig mogelijkheden specifieke invulling te verhinderen. Het VOR-bestuur wordt verzocht zich te bezinnen op een structuur, bijvoorbeeld in de vorm van divisies, waarin de thema's duidelijker zijn te herkennen. Het bestuur zegt toe zich hierover te zullen buigen.

N.a.v. punt 5 Gedragsregels voor onderwijsonderzoekers: gevraagd wordt wat het standpunt is van het bestuur ten aanzien van verbreding van groepen die de regels kunnen onderschrijven. Het bestuur heeft momenteel de volgende opstelling: de gedragsregels zijn in principe bedoeld voor VOR-leden. In het bestuur wordt wel gesproken over mogelijkheden voor een bredere vereniging of federatie van onderwijskundigen, waarvan de VOR deel uitmaakt.

Er blijkt namelijk behoefte aan een brede organisatie van onderwijskundigen te zijn. De eerste opstelling van het bestuur is echter dat deze organisatie niet direct moet worden opgehangen aan de VOR. Het bestuur heeft echter nog geen definitief standpunt bepaald. In dit verband wordt er op gewezen dat de ORD wel breder en algemener is.

5. Jaarrekening 1988

De kascontrolecommissie is accoord gegaan met de jaarrekening en spreekt haar waardering uit voor de inzichtelijke wijze waarop de penningmeester financieel verslag heeft uitgebracht. De aanbevelingen zullen door het bestuur worden besproken. In aansluiting hierop beveelt de ledenvergadering aan het International Journal actiever te verspreiden, bijvoorbeeld op ORD en andere congressen/symposia. Verder wordt geadviseerd meer aan ledenwerving te doen, bijvoorbeeld door aio's korting te geven. Het bestuur zal deze aanbevelingen bespreken.

6. W.v.t.t.k.

Er worden geen punten meer ingebracht.

7. Sluiting

De voorzitter sluit de vergadering met dank aan de belangstellende leden.

Ontvangen publikaties

- ARBO (1988). *Progressie in professie: advies over de ontwikkeling van de PABO*. Zeist: ARBO.
- Bakker, B.F.M., Dronkers, J. & Meijnen, G.W. (1989). *Educational opportunities in the welfare state*. Longitudinal studies in educational and occupational attainment in the Netherlands. Nijmegen: ITS.
- Bergh, H. van den (1988). *Examens geëxamineerd*. Den Haag: SVO (Selecta reeks).
- Bijlmakers, L.A. & Meijers, J.L. (1989). *Agrarisch cursusonderwijs: achtergronden en deelname*. Wageningen: Landbouwuniversiteit, Vakgroep Agrarische Onderwijskunde.
- Boshuizen, H.P.A. (1989). *De ontwikkeling van medische expertise: Een cognitief-psychologische benadering*. (Proefschrift RU Limburg) Haarlem: Thesis.
- Caluwé, L. de, Marx, E.C.H. & Petri, M.W. (1988). *Schooldevelopment: models and change*. Leuven/Amersfoort: ACCO (ISIP-technical report). f 38,50.
- Dyck-Lovinck, M.E. van (1988). *Mammoetwet en arbeidsmarkt: Het geval LHNO*. Proefschrift RU Leiden.
- Eberg, J.W., Elskamp, Th.C.M. van den, Stokhuyzen, Y.L.M. & Stokking, K.M. (1989). *Aansluiting tussen vraag naar en aanbod van automatiseringsopleidingen*. Utrecht: ISOR.
- Esch, W. van (1988). *Schoolbesturen en de vrijheid van onderwijs*. (Proefschrift Universiteit van Amsterdam) Nijmegen: ITS.
- Geer, J.P. van de (1988). *Analyse van kategorische gegevens*. Deventer: Van Loghum Slaterus. f 75,00.
- Gennip J. van (1989). *Werken met invoeringsprogramma's door basisscholen*. Nijmegen: ITS.
- Gennip, J. van & Genemans, L. (1988). *De vormgeving van de beleidsmaatregel invoeringsprogramma's basisonderwijs door onderwijsbegeleidingsdiensten*. Nijmegen: ITS.
- Gennip, J. van & Genemans, L. (1989). *Begeleiden van basisscholen met invoeringsprogramma's*. Nijmegen: ITS.
- Glopper, K. de (1988). *Schrijven beschreven: Inhoud, opbrengsten en achtergronden van het schrijfonderwijs in de eerste vier leerjaren van het voortgezet onderwijs*. Den Haag: SVO (Selecta reeks).
- Hopkins, D. (1988). *Doing school based review. Instruments and Guidelines*. Leuven/Amersfoort: ACCO (ISIP-technical report). f 38,50.
- Koetsier, C.P. & Brinke, J.S. ten (1989). *Experiment individuele eindstage 2 bij de vakgroep Didactiek van de biologie. Analyse van de interviews met schoolpracticum docenten*. Utrecht: Centrum voor Didactiek van Wiskunde en Natuurwetenschappen.
- Koetsier, C.P. & Brinke, J.S. ten (1989). *Experiment individuele eindstage 2 bij de vakgroep Didactiek van de biologie. Analyse van het rondetafelgesprek met de instituutsbegeleiders*. Utrecht Centrum voor Didactiek van Wiskunde en Natuurwetenschappen.
- Kok, F. (1988). *Vraagpartijdigheid. Methodologische verkenningen*. (Proefschrift Universiteit van Amsterdam) Amsterdam: SCO.
- Kral, M., Visser, H. & Gennip, J. van (198..). *Afhaken; afstel of uitstel?* Nijmegen: ITS.
- Perrenet, J.C., Herfs, P.G.P. & Terwel, J. (1988). *Met AGO op weg. Adaptief Groeps Onderwijs; Een werkwijze voor (wiskunde)onderwijs in heterogene groepen*. Utrecht: ISOR.

- Pieters, J.M. (Red.) (1988). *Onderzoek naar leren en instructie in arbeidsorganisaties*. (Onderzoeksthema-groep Onderwijsleerprocessen) Enschede: Universiteit Twente.
- Raemakers, J. & Gennip, J. van (1989). *Analyse van invoeringsprogramma's*. Nijmegen: ITS.
- SER (1989). *Milieu en economie. Advies inzake milieumaatregelen door overheid en bedrijfsleven*. 's-Gravenhage: SER.
- Smit, F. (Red.) (1989). *Praktijkbezoek medezeggenschap voortgezet onderwijs*. Deventer: Kluwer Onderwijsuitgaven.
- Spit, P. (1988). *Het leren van technische handvaardigheden*. Proefschrift KU Brabant.
- Steen, A.M.J. van der (1988). *Agrarische opleidingscentra in Nederland. Deelverslag 1: Uitgangspunten voor de bevordering van de onderwijskwaliteit*. Wageningen: Landbouwwuniversiteit, Vakgroep Agrarische Onderwijskunde.
- Streun, A. van (1988). *Heuristisch wiskunde-onderwijs*. Proefschrift RU Groningen.
- SVO (1988). *Jaarverslag 1987*.
- Terwel, J., Herfs, P., Perrenet, J. & Ploeg, D. van der (1988). *Ontwerpen van adaptief onderwijs*. Utrecht: ISOR.
- Verijdt, H.J.P.M. & Diederens, J.H.I.W. (1989). *KMBO'ers in een administratief beroep*. Nijmegen: ITS.
- Voogt, J.C. & Bollen, R. (1988). *SBR In Nederland. Een studie over schooldiagnose*. Leuven/Amersfoort: ACCO. f 21,50.
- Wiegersma, S. (1989). *Innovatie van het hoger onderwijs*. Groningen: Wolters-Noordhoff (Hoger onderwijs reeks). f 49,50.

Computergeleid Enquêteren; een overzicht van technologische vernieuwingen*

E. D. De Leeuw^{1,2}

Vrije Universiteit Amsterdam

ABSTRACT

Computer assisted surveys: a review of the technological changes in data collection procedures

Until recently computers were mainly used for data analysis and statistics. However, the advancing technology in electronics in general and in personal computers in particular has started a new development in data collection procedures. In this paper an overview is given of the state of computer assisted data collection procedures. Computer assisted equivalents of the three main data collection modalities are described: computer assisted telephone interviewing (CATI), computer assisted personal interviewing (CAPI), and computer assisted self-administered questionnaires (CASAQ), of which the "tele-panel" is a special case. Special attention is given to the possible application of these methods in educational research. Advantages and disadvantages of computer assisted data collection are discussed, and a selected bibliography with suggested readings is offered.

1. INLEIDING

De vragenlijst is een van de meest gebruikte instrumenten voor dataverzameling in het sociaal wetenschappelijk onderzoek. Binnen vragenlijstonderzoek zijn drie hoofdvormen te onderscheiden: het mondelinge interview, het telefonische interview, en de schriftelijke enquête. Met name het telefonisch interview, en in mindere mate de schriftelijke enquête, hebben in de laatste jaren het mondelinge interview steeds meer verdrongen (Thornberry, Nichols & Kulpinski, 1982; Turner & Martin, 1984). Nieuwe ontwikkelingen op het gebied van telefonisch en schriftelijk enquêteren hebben geleid tot responspercentages die vergelijkbaar zijn met het mondelinge interview (Dillman, 1978; Frey, 1983), tegen lagere kosten (Groves en Kahn, 1979). Bovendien resulteren deze alternatieven voor het mondelinge interview in gegevens van goede kwaliteit (Groves en Mathiowetz, 1984; De Leeuw en Hox, 1985, 1988; De Leeuw en Van der Zouwen, 1988).

Op dit moment voltrekt zich een stille elektronische revolutie binnen de dataverzameling. Al tientallen jaren wordt de computer gebruikt bij de verwerking en analyse van gegevens, en ook bij de rapportage van onderzoek is de tekstverwerker haast niet meer weg te denken. De informatisering voltrekt zich nu op een ander terrein, dat van de dataverzameling. In eerste instantie heeft de computer vooral veel invloed gehad op de werkwijze bij het telefonisch interviewen. Het traditionele telefonisch interview met papier

Adres: Koningslaan 22-24, 1075 AD Amsterdam

* NOOT

Deze bijdrage is bedoeld als (voortgezette) inleiding in nieuwe technische en methodische ontwikkelingen op het terrein van onderwijsonderzoek. De Redactie heeft het voornemen ook in de toekomst dergelijke 'tutorials' op te nemen.

en pen heeft hoe langer hoe meer plaatsgemaakt voor CATI: Computer Assisted Telephone Interviewing. Maar ook bij het mondelinge interviewen en zelfs bij de schriftelijke enquête heeft de computer haar intrede gedaan. Analooq aan CATI noemt men deze nieuwe wijze van werken respectievelijk CAPI (Computer Assisted Personal Interviewing) en CASAQ (Computer Assisted Self Administered Questionnaire).³

In dit artikel wordt een overzicht gegeven van de belangrijkste ontwikkelingen op het gebied van computergeleid enquêteren. Hierbij ligt de nadruk op de invloed van deze technologische vernieuwingen op de methoden van dataverzameling voor sociaal wetenschappelijk onderzoek en in het bijzonder onderwijskundig onderzoek. Allereerst wordt in paragraaf 2 een beschrijving gegeven van de gang van zaken bij CATI, waarbij een aantal belangrijke begrippen binnen het computergeleid enquêteren worden behandeld. In paragraaf 3 wordt het computergeleid persoonlijk interviewen (CAPI) behandeld; paragraaf 4 gaat over het gebruik van computers bij schriftelijk enquêteren (CASAQ); en in paragraaf 5 wordt een speciale vorm van CASAQ behandeld, namelijk het tele-interview. Tenslotte wordt in paragraaf 6 expliciet ingegaan op de voor- en nadelen van computergeleid enquêteren.

2. COMPUTERGELEID TELEFONISCH INTERVIEWEN (CATI)

2.1. Voorgeschiedenis

De eerste, nog vrij primitieve, CATI-systemen zijn rond 1970 ontwikkeld ten behoeve van het Amerikaanse bedrijfsleven (Nicholls en Groves, 1986). CATI werd in korte tijd populair. Het is kosten-effectief en snel, en naar men algemeen aanneemt ook zeer accuraat (Shanks, 1983). In het begin van de jaren tachtig zijn CATI-systemen gemeengoed geworden in het Amerikaanse marktonderzoek en doen ze hun intrede in Europa (Fink, 1983). In 1975 werd CATI geïntroduceerd in de Amerikaanse universiteiten. Binnen de universitaire onderzoeksprogramma's werden de mogelijkheden van CATI uitgebreid om aan de specifieke eisen van universitair enquête-onderzoek te kunnen voldoen. Op dit moment maken in Amerika een groot aantal universiteiten en aan universiteiten gelieerde onderzoeksinstituten gebruik van CATI. In Europa begint CATI net in de universitaire wereld door te dringen. Voor een overzicht zie Spaeth (1987).

De eerste CATI-systemen werkten op een centrale computer ofwel mainframe, die verbonden is met meerdere terminals waaraan de interviewers zitten. Alle informatie die tijdens het interview nodig is, wordt van de centrale computer naar het beeldscherm van de betreffende interview(st)er gezonden. Alle informatie die bij een respondent(e) verzameld is, wordt via de terminal van de interview(st)er ingevoerd in de centrale computer.

Tegenwoordig worden hoe langer hoe meer microcomputers of personal computers gebruikt, en zijn ook interviewprogramma's voor deze machines commercieel verkrijgbaar (voor een overzicht zie De Bie et al., 1989). Alle informatie die tijdens het interview nodig is, is opgeslagen in de microcomputer van de interview(st)er. De antwoorden van de respondenten worden gedurende het interview in de micro opgeslagen. Aan het einde van een interviewperiode kunnen de gegevens van de voltooide interviews getransporteerd worden naar een centrale (mini) computer via een speciale operatie. Deze opstelling is veel minder storingsgevoelig. Legde een storing in het mainframe nog alle interviews stil, een sporadisch optredende storing in een micro zal slechts één interview onderbreken (zie ook Palit en Sharp, 1983). Ook bij kleinschaliger onderzoek kan CATI nu vrij eenvoudig gebruikt worden. Zo geeft Dillman (1978) een voorbeeld van een universitaire ruimte met microcomputers, die gedeeltelijk overdag door studenten gebruikt wordt, en gedeeltelijk, met name 's avonds en in het weekend, voor telefonische interviews ten behoeve van onderzoek.

De allerlaatste ontwikkelingen gaan uit van microcomputers die met elkaar verbonden zijn via een 'netwerk'; hierbij zijn de voordelen van het oude mainframe systeem, zoals meer centrale regeling en controle, gecombineerd met de voordelen van het gebruiken van personal computers (onder andere risico spreiding bij technische storingen, en kortere wachttijd tussen de vragen).

In Nederland wordt CATI nog voornamelijk gebruikt binnen het marktonderzoek (zie De Bock, 1987). Het Centraal Bureau voor de Statistiek is in 1982 begonnen met CATI-experimenten. Hiervoor werd een door een marktonderzoekorganisatie ontwikkeld CATI-systeem gebruikt. Sinds 1984 wordt CATI gebruikt voor een aantal periodiek terugkomende onderzoeken, zoals het Consumenten Conjunctuur Onderzoek (Kerssemakers et al., 1987). Binnen het CBS wordt gewerkt aan een geïntegreerd interviewsysteem voor MS-DOS computers: BLAISE. Met behulp van het BLAISE-systeem kunnen vragenlijsten worden geconstrueerd, computergeleide interviews worden afgenomen, en gegevens worden ingevoerd en gecontroleerd (Bethlehem, 1989; Denteneer et al., 1987). Het Geografisch Instituut van de Rijksuniversiteit Utrecht voert sinds 1984 CATI-interviews uit met een binnen dit instituut zelf ontwikkeld programma voor microcomputers.

2.2. Wat is CATI

Er bestaan vele verschillende systemen voor computergeleid telefonisch interviewen. Een algemene definitie van CATI is dan ook moeilijk te geven. Globaal gezien is CATI een geautomatiseerde vorm van telefonisch interviewen. Het ene systeem kent echter veel meer mogelijkheden dan het andere. Om een overzicht te geven zijn de verschillende mogelijkheden hieronder puntsgewijs samengevat. Dit overzicht is voor een deel gebaseerd op Nicholls en Groves (1986).

OVERZICHT VAN MOGELIJKHEDEN BIJ VERSCHILLENDE CATI-SYSTEMEN

1. Steekproeftrekken:

Het systeem trekt een steekproef bijvoorbeeld uit een bekende lijst met telefoonnummers of via 'random digit dialing'. Een veel gebruikte methode voor 'random digit dialing' is een steekproef van bekende nummers uit het telefoonboek te trekken en dan de laatste twee of drie cijfers te vervangen door toevalscijfers. Deze werkwijze heeft als voordeel dat ook geheime nummers, net aangesloten abonnees, en net verhuisde abonnees in principe bereikt kunnen worden, zonder dat veel niet-bestaande nummers nodeloos gedraaid worden. Dit laatste is het grote bezwaar van volstrekt a-select bellen. Bovendien houdt het CATI-systeem de status van ieder steekprofelement bij. Zo wordt bij gestratificeerde en quota steekproeven bijgehouden hoeveel steekprofeenheden er van iedere specifieke groep gebeld zijn en hoeveel er nog gebeld moeten worden.

2. Bijhouden van een opbelschema:

Wanneer een interview(st)er klaar is voor een volgend interview, geeft het systeem het nummer van een getrokken steekprofelement door, of draait zelf het nummer. Wanneer een nummer geen gehoor geeft, zorgt het systeem ervoor dat dit nummer op een ander tijdstip opnieuw gebeld wordt.

3. Administratie:

Het systeem houdt bij hoeveel nummers er gedraaid zijn, in hoeveel telefoongesprekken dit

resulteerde, en wat het resultaat daarvan was. Verder wordt de response ratio en de produktiviteit bijgehouden. Deze gegevens zijn op ieder moment beschikbaar, zowel per interview(st)er als in totaal.

4. Interactief interviewen:

Het interview zelf wordt gekenmerkt door de volgende gang van zaken.

- De tekst van de te stellen vraag, de toegestane antwoordmogelijkheden met code en eventuele instructie aan de interview(st)er verschijnen op het beeldscherm.
- Het systeem past de tekst van de vragen aan op basis van de antwoorden op voorgaande vragen of informatie uit voorgaand onderzoek. Bijvoorbeeld: 'Hoe oud is Uw oudste dochter?', als al bekend is dat er een zoon en twee dochters in het gezin zijn.
- Bij 'gesloten' vragen wordt de code voor het gegeven antwoord direct ingetikt. Het systeem controleert of dit een bestaande en toegestane code is, en of het antwoord consistent is met voorgaande antwoorden. Wanneer er een fout ontdekt wordt dan resulteert dit ofwel in een verzoek om de code nogmaals te controleren, ofwel in het verschijnen van additionele controle vragen op het scherm.
- Bij 'open' vragen kan een geruime hoeveelheid tekst direct worden ingetikt. Deze kan dan later gecodeerd en verder verwerkt worden.
- Het overslaan van vragen en het stellen van specifieke vragen aan specifieke subgroepen (skipping en branching) wordt automatisch door het systeem verzorgd op basis van eerdere antwoorden. Op het juiste moment biedt de computer de juiste controle- en vervolgvragen aan.
- Interviewers kunnen indien nodig een interview halverwege stoppen en later op een afgesproken tijdstip op het juiste punt weer verdergaan. Het systeem staat ook toe dat teruggegaan wordt naar een eerdere vraag en dat antwoorden gecontroleerd en eventueel veranderd worden. Commentaar van interviewers kan worden toegevoegd.

5. Kwaliteitsbewaking (monitoring):

Bij grotere systemen is het mogelijk om van iedere interview(st)er het interview te volgen op een centrale terminal. Niet alleen kan men zo fragmenten van interviews beluisteren, het is ook mogelijk om te zien wat er op dat moment door de interview(st)er wordt ingetikt, en wat er op de vorige vragen is ingevuld.

6. Maken van een bestand:

Het systeem maakt een databestand in een zodanig formaat dat dit bestand zonder problemen direct verwerkt en geanalyseerd kan worden.

Opgemerkt moet nog worden dat bovenstaande een generisch overzicht is en dat lang niet alle CATI-systemen alle genoemde mogelijkheden bezitten. Zo kan bij een klein CATI-systeem bijvoorbeeld het automatisch toewijzen van nieuwe telefoonnummers aan een interview(st)er (punt 2) ontbreken, en wordt met gedrukte lijsten met telefoonnummers gewerkt. Ook bij gebruik van CATI op micro-computers is dit het geval, bovendien wordt de kwaliteitsbewaking (punt 5) dan vaak uitgevoerd door een supervisor die tijdens het interviewen rondloopt. In de toekomst zullen echter steeds vaker microcomputers via een lokaal netwerk (LAN) aan elkaar gekoppeld zijn. Hierdoor wordt het mogelijk de voordelen van een groot centraal systeem te combineren met de voordelen van microcomputers.

2.3. Voordelen voor onderwijskundig onderzoek

Het telefonische interview wordt in het algemeen aangeraden voor onderzoek bij geografisch sterk gespreide organisaties en instellingen. Een getrainde interview(st)er kan

telefonisch snel langs secretaresses, conciërges en andere 'gatekeepers' tot de deskundige doordringen, en dan blijkt dat er voor het beantwoorden van enkele vragen altijd wel even tijd is (Dillman, 1978; Tull en Hawkins, 1984). Dit maakt deze methode met name geschikt voor onderwijskundig beleidsonderzoek. Een voorbeeld is het onderzoek naar de behoefte aan de 'DOP-regeling' onder leerkrachten, zoals door het ITS onlangs is uitgevoerd.

Tot voor kort was CATI-onderzoek alleen mogelijk op grote commerciële instituten. De dalende prijzen van microcomputers en het beschikbaar komen van adequate programmatuur, maken echter dat CATI ook binnen het bereik van universitaire instituten komt. Voor veel onderzoek is één toegewijde onderzoek(st)er, die de beschikking heeft over een 'buitenlijn' en een personal computer al voldoende om in relatief korte tijd een groot aantal interviews in het land af te nemen (zie ook Baker en Lefes, 1988). Een voorbeeld hiervan is een onderzoek, dat momenteel op de Universiteit van Amsterdam wordt uitgevoerd onder hoofden van scholen over de bestuurlijke heroriëntatie in het lager beroepsonderwijs.

Het gebruik van een CATI-systeem heeft het bijkomende voordeel dat geen vragenlijsten en antwoordformulieren gedrukt hoeven worden, er geen extra data-invoer fase meer nodig is, en dat de taak van de interviewers verlicht wordt. Hierdoor zijn in het algemeen de resultaten sneller bekend, en kunnen deze een grotere actualiteitswaarde hebben.

Wanneer een longitudinaal onderzoek gepland wordt heeft computergeleid enquêteren nog het bijkomende voordeel dat informatie uit vorige interviews tijdens het interview gebruikt kan worden bij het stellen van vragen: een vorm van 'tailored' enquêteren. Men denke hierbij aan onderzoek naar de studievoortgang van Open Universiteit-studenten in het land, of aan een onderzoek naar de ervaringen van de eerste generatie Assistenten en Onderzoekers in Opleiding aan de verschillende universiteiten in Nederland.

3. COMPUTERGELEID PERSOONLIJK INTERVIEWEN (CAPI)

Bij het computergeleid persoonlijk interviewen gaat de interview(st)er met een draagbare micro op pad. Dankzij nieuwe ontwikkelingen op hardware-gebied worden micro's steeds kleiner, lichter en handzamer. Mede hierdoor is het computergeleid persoonlijke interviewen, waar al vaak over gespeculeerd werd, nu ook werkelijkheid geworden.

Net als bij CATI heeft men bij CAPI het meeste profijt van de computer bij complexe vragenlijsten met veel 'skipping' en 'branching'. Net als bij CATI worden bij CAPI ook standaardcontroles uitgevoerd op de ingetikte antwoorden. In vergelijking met het CATI-systeem is de controle op de interviewers echter geringer. Het interview vindt immers bij de respondenten thuis plaats, en niet op een centrale plaats waar de supervisor mee kan kijken of luisteren. Bij het gebruik van CAPI kan in tegenstelling tot bij CATI gebruik worden gemaakt van visueel materiaal. Dit kan bestaan uit de klassieke kaarten met antwoordmogelijkheden die in het 'gewone' persoonlijke interview aan de respondenten worden voorgelegd, maar men kan ook voor een visuele presentatie de grafische mogelijkheden van het beeldscherm uitbuiten. Ook kunnen respondenten rustig iets nazoeken, wat bij telefonisch interviewen beduidend moeilijker is.

Een belangrijk verschil met computergeleid telefonisch interviewen is, dat bij het persoonlijke interview de respondenten zien dat er een computer gebruikt wordt. Op dit moment wordt CAPI nog slechts op beperkte schaal gebruikt, en uitgebreid vergelijkend onderzoek op dit gebied is nog niet gedaan. De invloed van de computer op de respondenten lijkt echter mee te vallen. Volgens interviewers die met CAPI gewerkt hebben, hebben respondenten meer bezwaren tegen het (vaak standaard) gebruik van een bandrecorder bij het interview dan tegen een computer. Er zijn geen aanwijzingen dat het gebruik van een computer het aantal weigeringen bij interviews beïnvloedt (vgl. Hindricks, 1986; Sikkels, 1987).

CAPi combineert een groot aantal voordelen van het klassieke persoonlijke interview met de voordelen die van CATI bekend zijn; en het zal ongetwijfeld steeds vaker gebruikt gaan worden. Sinds begin 1987 voert het C.B.S. maandelijks zo'n 10000 persoonlijke interviews uit met behulp van een draagbare microcomputer. Het betreft hier de Enquête Beroeps Bevolking, de geautomatiseerde opvolger van de ArbeidsKrachtenTelling (Van Basterlaer et al, 1987). Ook binnen het marktonderzoek wordt gebruik gemaakt van computergestuurd persoonlijk interviewen. Zo werken bijvoorbeeld zowel Inter/View als het NIPO met CAPi-systemen (De Bock, 1987).

In de nu volgende paragraaf zal een algemene beschrijving van computergeleid persoonlijk interviewen gegeven worden. In de praktijk kan de werkwijze per CAPi-systeem enigszins verschillen. De interviewers krijgen naast de gebruikelijke interviewtraining, een algemene introductie in het gebruik van een microcomputer. Doorgaans hebben interviewers thuis een modem, zodat ze via de telefoon verbinding kunnen maken tussen hun microcomputer en de centrale hoofdcomputer. Via het modem kunnen voltooide interviews overgezonden worden naar de centrale computer. Ook kunnen zo nieuwe vragenlijsten met een korte instructie en een nieuw adressenbestand naar de interviewers toegezonden worden. De interviewers kunnen dan thuis eerst rustig de vragenlijst op het scherm bekijken en eventueel een proefinterview doen en vervolgens op pad gaan.

Nadat de interview(st)er door de respondent(e) is binnengelaten wordt de microcomputer te voorschijn gehaald en wordt een korte toelichting op het gebruik van de computer gegeven. De interview(st)er start daarna het programma en begint met het eigenlijke interview.

Bij speciale vragen, bijvoorbeeld over gevoelige onderwerpen, bestaat de mogelijkheid dat de respondent(e) zelf achter de computer gaat zitten en de antwoorden in alle rust intikt. De interview(st)er blijft op een afstand aanwezig om eventuele toelichting of technische assistentie te kunnen geven. Deze werkwijze komt overeen met de tijdens 'gewone' interviews gebruikelijke werkwijze om voor gevoelige vragen een schriftelijke vragenlijst aan de respondent te geven, die dan tijdens het interview door de respondent zelf wordt ingevuld, in een envelop gedaan en dichtgeplakt. Uit experimenten met onder andere vragen naar alcoholmisbruik (Waterton, 1984) blijkt dat het op deze wijze gebruiken van een computer bij gevoelige onderwerpen goede resultaten geeft.

Aan het einde van het interview zorgt de interview(st)er ervoor dat het systeem correct wordt afgesloten.

Directe toepassingen van CAPi binnen onderwijskundig onderzoek zijn nog niet bekend. Een speciale variant voor observatieonderzoek is ontwikkeld bij het ITS en toegepast in een onderzoek waarbij steeds twee observatoren, ieder met een draagbare microcomputer, dezelfde leerlingen in een schoolklas observeerden. Het observatieprogramma is zo ontworpen dat na het invoeren van een code voor de betreffende gedragscategorie het systeem automatisch een klok activeert. Wanneer het gedrag beëindigd wordt is één commando voldoende om de klok te deactiveren. Zo wordt per leerling zowel het type gedrag als de tijdsduur nauwkeurig bijgehouden. Ook bij het Amsterdams Pedologisch Centrum is dergelijk onderzoek gedaan.

4. COMPUTERGELEID ZELF INVULLEN VAN VRAGENLIJSTEN (CASAQ)

Inmiddels heeft ook bij 'schriftelijk' onderzoek, waarbij respondenten zelf de vragenlijsten invullen, de microcomputer zijn intrede gedaan. Bij schriftelijk onderzoek kunnen de gegevens op verschillende manieren worden verzameld: 1) De interview(st)er overhandigt tijdens het interview een vragenlijst, of laat bij vertrek een vragenlijst achter. 2) Een groep,

die in een ruimte bijeen is, vult tegelijkertijd een vragenlijst of test in. 3) Een vragenlijst wordt per post naar potentiële respondenten opgestuurd (postenquête).

In de vorige paragraaf werd al een voorbeeld gegeven hoe interviewers tijdens een CAPI-interview respondenten zelf achter de computer kunnen laten plaatsnemen, zodat deze in alle privacy de vragen kunnen beantwoorden. Bij groepsgewijze afname van vragenlijsten wordt de computer steeds vaker gebruikt. Zo worden bij computer ondersteund onderwijs de onderdelen regelmatig afgesloten met een test per computer, waardoor de leerlingen onmiddellijk feedback krijgen op hun prestaties. En bij een automatiserings- of een informaticacursus ligt een onderwijsevaluatie per computer voor de hand.

Zelfs voor de postenquête zijn computer ondersteunde alternatieven ontwikkeld. Hierbij wordt gebruik gemaakt van bestaande computernetwerken en elektronische post- en mededelingen systemen (bulletin boards) voor computergebruikersgroepen. Onderzoekers kunnen aan de systeembeheerders vragen of zij een mededeling en vragenlijst in het bulletin board willen opnemen. Wanneer dat gebeurt is krijgt iedereen, die via telefoon en modem contact maakt met het bulletin board, een verzoek op het scherm om een aantal vragen te beantwoorden. Na toestemming verschijnen de vragen op het scherm, en de ingetikte antwoorden worden (anoniem) in een bestand opgeslagen. Deze laatste ontwikkelingen staan nog in de kinderschoenen, en slechts beperkte groepen kunnen momenteel zo ondervraagd worden. Ervaringen met dit type onderzoek in Amerika zijn echter veelbelovend (Kiesler en Sproull, 1986).

Een combinatie van dataverzamelen per computer en de traditionele postenquête is ook goed mogelijk. Zo heeft het ITS voor een evaluatieonderzoek leerkrachten een geautomatiseerde vragenlijst op 'floppy' per gewone post toegestuurd. Het betrof hier een onderzoek naar een nascholingscursus voor leerkrachten, die verbonden waren aan scholen die net computers toegewezen hadden gekregen. De leerkrachten vulden de vragen via hun schoolcomputer in, en stuurden de beantwoorde vragenlijst op een 'floppy' weer terug.

5. HET TELE-INTERVIEW

Een speciale variant op de postenquête per computer is het Teleinterview (zie ook Saris et al., 1987). Het betreft hier panelonderzoek waarbij respondenten regelmatig een aantal vragenlijsten met een home-computer invullen. Hiertoe wordt bij een groot aantal huishoudens een kleine home-computer en een modem geplaatst. Ieder weekend wordt er via dit modem automatisch contact gemaakt met een centrale computer, en worden een aantal vragenlijsten naar de home-computer gestuurd en op schijf opgeslagen. De respondenten kunnen zelf bepalen wanneer zij de vragenlijsten gaan invullen. Iedere keer wanneer de computer wordt opgestart, verschijnt er op het scherm een overzicht van de in te vullen vragenlijsten en van de namen van leden van het huishouden waarvoor speciale vragenlijsten bestemd zijn. De respondent identificeert zich, en krijgt de vragen die voor hem of haar bestemd zijn. Het interview-programma registreert welke leden van het huishouden de vragen al beantwoord hebben. De volgende keer dat de computer gebruikt wordt, neemt het programma de draad weer op en stelt vragen aan de resterende leden van het huishouden. Wanneer alle vragen door iedereen beantwoord zijn worden de antwoorden automatisch weer teruggestuurd naar de centrale computer. Als een bepaald huishouden een week niet geantwoord heeft, wordt per telefoon persoonlijk contact opgenomen, en eventuele assistentie aangeboden. De hele procedure is zo eenvoudig mogelijk gemaakt en de respondenten hebben geen speciale kennis van computers nodig om de vragen te kunnen beantwoorden. Wanneer er vragen zijn of in het geval dat er onverhoopt technische storingen optreden, kan altijd een lid van het onderzoeksteam

telefonisch bereikt worden. Het NIPO werkt al enige tijd succesvol op deze wijze met een 'telepanel'.

Het werken met een telepanel heeft naast de al eerder genoemde voordelen van computergeleid interviewen, zoals controle op onmogelijke en foute antwoorden, en het op eenvoudige wijze presenteren van complexe vragenlijsten met veel skipping en branching, twee duidelijke voordelen.

In de eerste plaats kunnen met behulp van een goed computersysteem de mogelijkheden van panelonderzoek uitgebuit worden. Informatie over panelleden kan in het systeem worden opgeslagen en in het verdere onderzoek gebruikt worden. Zo kunnen aansluitend op de installatie van de computer bij de respondenten thuis de verschillende leden van het huishouden ondervraagd worden. Daarna zijn bijvoorbeeld namen, leeftijden, en hobbies van de leden van het huishouden bekend. Op grond van deze informatie kunnen vragenlijsten voor speciale sub- of doelgroepen nu direct aan de vertegenwoordigers van deze groepen worden voorgelegd.

In de tweede plaats is deze werkwijze zeer effectief om een van de grootste problemen van panelonderzoek, de paneluitval, tegen te gaan. Zolang een huishouden meedoet aan het panel heeft het immers gratis de beschikking over een home-computer. Het belonende aspect van het hebben van een home-computer kan nog versterkt worden door iedere week naast nieuwe vragenlijsten ook een nieuw computerspelletje of een volgende aflevering van een computercursus elektronisch toe te sturen.

Een systeem voor computergeleid schriftelijk ondervragen, zoals hier boven geschetst, vergt een hoge initiële investering in apparatuur en deskundigen. Zo'n systeem is dan ook alleen maar kosten-effectief wanneer zeer regelmatig een groot aantal vragenlijsten naar een vast panel worden gezonden en dit panel ook voor lange tijd gebruikt wordt. Voor incidenteel onderzoek en kortlopende panels zijn de traditionele papier-en-pen methoden zoals het toezenden van postenquêtes en door panelleden bij te houden dagboekjes op dit moment financieel beter haalbaar. Bij een langlopend informatiseringsproject op scholen zou deze werkwijze echter goed toegepast kunnen worden. De investering in apparatuur is dan immers al gedaan.

6. VOOR EN NADELEN VAN COMPUTERGELEID ENQUÊTEREN

6.1. Voordelen

Wanneer men het optimale computergeleide interview vergelijkt met het optimale papier-en-pen interview, dan heeft het eerste drie duidelijke voordelen: een snelle en adequate administratie, meer controlemogelijkheden en meer flexibiliteit.

Op het punt van administratie van interviews, zoals het bijhouden van opbelschema's, responspercentages en dergelijke, kan het gebruik van één centraal systeem de taak van de interviewers en de projectleiders enorm verlichten. Bovendien hoeven vragenlijsten en antwoordformulieren niet meer gedrukt te worden, de antwoorden hoeven niet meer op speciale codeervellen te worden overgezet en er is geen aparte pons- of datainvoerfase nodig.

Het gebruik van computergeleid enquêteren brengt op twee punten meer controlemogelijkheden met zich mee. Allereerst is er de interne controle van het systeem op de antwoorden van de respondenten. Hierbij wordt al tijdens het interviewen door het systeem nagegaan of een gegeven antwoord consistent is met al eerder gegeven antwoorden. In principe kunnen deze controles ook uitgevoerd worden bij een papier-en-pen interview. Zo kan men controleren of de eerder opgegeven leeftijd overeenkomt met antwoorden over de schoolcarrière en belangrijke tijdstippen daarin. Gewoonlijk gebeuren

dit soort controles dan achteraf, nadat het interview beeindigd is. Bij een computergeleid interview kan deze controle razendsnel tijdens het interview gedaan worden. De interview(st)er krijgt dan bijna onmiddellijk een melding op het beeldscherm en de tekst van een te stellen controlevraag. Bij inconsistente antwoorden kan er dus direct worden doorgevraagd. Het praktische probleem voor vragenlijstconstructeurs en interviewers is dan wel hoe een controlevraag op natuurlijke wijze aan de respondenten voor te leggen, zodanig dat de verstandhouding niet wordt aangetast.

In de tweede plaats biedt het gebruik van een computergeleid interview-systeem meer mogelijkheden voor het controleren van de interviewers zelf. Met name de grote systemen voor computergeleid interviewen hebben uitgebreide mogelijkheden om interviewers te volgen. Deze vorm van kwaliteitscontrole is vooral gericht op het tegengaan van bekende interviewereffecten. Uit de literatuur over interviewgedrag en interviewereffecten (zie ook Billiet, Loosveldt en Waterplas, 1984) zijn veel gevallen bekend hoe interviewers (onbewust) respondenten kunnen beïnvloeden en een antwoord kunnen suggereren. Om dit tegen te gaan zijn bij veel CATI-systemen voorzieningen getroffen waardoor supervisors tijdens een interview mee kunnen luisteren en tevens de door de interviewers ingetikte antwoorden op het eigen beeldscherm kunnen volgen. Indien nodig kan dan onmiddellijk worden ingegrepen en bijgestuurd. Gemaakte opnamen met voorbeelden van moeilijke momenten kunnen bij de nabespreking of bij (tussentijdse) interviewtrainingen gebruikt worden.

Bij geïndividualiseerde toepassingen van computergeleid enquêteren zoals CAPI zijn deze controle en bijsturingsmechanismen natuurlijk niet mogelijk. Wel kent ieder systeem voor computergeleid enquêteren eenvoudige controlemogelijkheden. Zo zijn van gesloten vragen de antwoordcategorieën bekend. Wanneer interviewers of respondenten een antwoordcode invoeren die buiten het bereik van bestaande codes valt volgt automatisch een foutmelding en krijgt men de mogelijkheid deze inconsistenties te verbeteren.

Een derde voordeel is de flexibiliteit van het systeem. Dit voordeel wordt groter naarmate de te gebruiken vragenlijst of het interviewschema gecompliceerder is. Bij een complexe vragenlijst kunnen interviewers op verschillende wijzen door een vragenlijst heengeleid worden. Dit 'skipping' of overslaan leidt vaak tot langdurig geblader in vragenlijsten. Ook is het mogelijk dat afhankelijk van een eerder antwoord een hele nieuwe serie vragen gesteld gaat worden ('branching'). Wanneer een vragenlijst veel skipping en branching-punten bevat, bijvoorbeeld wanneer men gedetailleerd de schoolcarrière van een respondent in kaart wil brengen, veroorzaakt dit veel opzoekwerk en bladeren tijdens het interview, en vaak lange pauzes tussen de vragen. De op papier gedrukte vragenlijsten zelf zijn dan ook lang en gecompliceerd. Bij een computergeleid interview verzorgt het systeem de 'routing' door de vragenlijst en verschijnt direct de juiste vervolgvraag op het scherm. Routingfouten worden zo vermeden en de interviewers kunnen zich nu beter concentreren op hun eigenlijke taak het op goede wijze communiceren met de respondenten.

De flexibiliteit van een CATI-systeem biedt onderzoekers méér mogelijkheden. Zo is het mogelijk om een aantal bij elkaar horende vragen in een willekeurige volgorde te laten stellen of de volgorde op een systematische wijze te variëren in een complex factorieel design. Dit is van speciaal belang wanneer men vreest voor volgorde effecten. Ook kunnen veelgebruikte vragen en achtergrondvragen als elektronische blokken in een bibliotheek bewaard worden. Deze modulen kunnen aangevuld worden met kwantitatieve gegevens uit eerder empirisch onderzoek en kwalitatief commentaar. Met name grotere instituten voor onderwijskundig onderzoek zullen hiervan profiteren.

Opvallend is dat er tot nu toe weinig empirisch onderzoek gedaan is naar de betrouwbaarheid van computergeleid interviewen in vergelijking tot de meer reguliere

methoden van dataverzamelen. De meeste referenties betreffen slechts de potentiële voordelen van deze nieuwe methoden. Sikkel (1987) verklaart dit o.a. uit de geringe motivatie om na het opzetten van een goedlopend computergeleid interviewsysteem nog eens een uitgebreid vergelijkend evaluatieonderzoek op te zetten.

Uit het tot nu toe beschikbare spaarzame empirische onderzoek blijkt zeker niet dat computergeleid enquêteren slechter zou zijn. Groves en Nicholls (1986) geven een samenvatting van de empirische kennis over de datakwaliteit bij computergeleid telefonisch interviewen. Zij concluderen dat er bij CATI beduidend minder fouten optreden bij zgn. doorverwijzingsvragen (skipping en branching) en dat er geen duidelijke verschillen zijn m.b.t. item-nonrespons. Ook bij vergelijking van CAPI met de traditionele interviewmethoden blijkt er geen verschil in itemnonrespons op te treden, zelfs niet wanneer de respondent 'gevoelige' vragen gesteld wordt (Sikkel, 1988). In het algemeen lijken zowel de interviewers als de respondenten deze nieuwe technieken met enig enthousiasme te ontvangen en te accepteren; van duidelijke weerstanden tegen de computer is geen sprake (Bastelaer, Kerssemakers en Sikkel, 1987). Tot nu toe zijn de resultaten veelbelovend. Er is echter nog veel systematisch vergelijkend onderzoek nodig voor een gefundeerd oordeel over de datakwaliteit bij computergeleid dataverzamelen gegeven kan worden.

6.2. Nadelen

Iedere medaille heeft een keerzijde, en ook computergeleid enquêteren kent zijn problemen. Deze zijn samen te vatten in de woorden investering, computerangst, en verandering van werkwijze en organisatiestructuur.

Bij de implementatie van een systeem voor computergeleid enquêteren is een investering in apparatuur en deskundigheid noodzakelijk. Bovendien dienen alle leden van het onderzoeksteam een zekere mate van 'computervriendelijkheid' aan de dag te leggen. Wanneer men werkt met een vast corps van ervaren interviewers zal een bijscholing op het gebied van computervaardigheden zeker aan te bevelen zijn. De voornaamste taak van interviewers is echter het interviewen zelf. Veel computergeleide enquête systemen zijn juist ontworpen om de technisch-administratieve taak van de interviewer te verlichten, zodat deze zich kan concentreren op een ontspannen gespreksfeer en een goede communicatie met de respondenten. Interviewers moeten dan ook kunnen vertrouwen op een eenvoudig te bedienen en goedwerkend systeem; in die zin dient het systeem dan ook 'interviewervriendelijk' te zijn.

De taak van de vragenlijstontwerpers wordt bij computergeleid enquêteren duidelijk verzaamd. Zij zullen deskundigheid op het gebied van gecomplexeerde vragenlijstconstructie moeten kunnen combineren met enige kennis van computersystemen. Zeker wanneer men de voordelen van computergeleid enquêteren (consistentiecontrole, automatische routing en dergelijke) ten volle wil uitbuiten dan vergt dat veel inventiviteit en vakmanschap bij de vragenlijstconstructie. Daarbij moet ervoor worden zorg gedragen dat een van de sterkste kanten van computergeleid enquêteren, de flexibiliteit, door alle controlemogelijkheden niet omslaat in een inflexibele en rigide opzet (vgl. House en Nicholls, 1988). Om tussen deze scylla en charibdis door te kunnen laveren is een gecontroleerd flexibele aanpak nodig. Dit vereist van de vragenlijstontwerpers naast kennis over het systeem zelf, het onderwerp en de te stellen vragen, ook uitgebreide kennis over de gang van zaken tijdens interviews en daarbij veel voorkomende problemen en kennis van de in het onderzoek gebruikte speciale interviewerinstructies en training.

De overgang van papier-en-pen vragenlijsten naar computergeleide enquêtes kan met name in grotere organisaties met een gesegmenteerde arbeidsverdeling tot problemen leiden. Voor het opzetten en uitvoeren van een computergeleide enquête is veel

samenwerking en interactie tussen de verschillende afdelingen een vereiste. De wijze van steekproeftrekken, dataverzameling, codering en gegevensinvoer vinden nu binnen één computersysteem plaats. Dit heeft tot gevolg dat de desbetreffende specialisten intensief met elkaar, met de vragenlijstontwerpers, en met de informatiseringsdeskundigen zullen gaan samenwerken, waarbij bestaande werkverdelingen doorbroken worden. Om deze vrij diepgaande veranderingen soepel te laten verlopen kan een reorganisatie van de managementstructuur en de managementoriëntatie nodig zijn. Een goed overzicht van de veranderingen in een onderzoeksorganisatie, die de overgang naar CATI met zich mee kan brengen, geven Berry en O'Rourke (1988).

6.3. Conclusie

Computergeleid enquêteren brengt naast vele voordelen ook nadelen met zich mee. Op dit moment is de noodzakelijke inspanning in tijd en moeite slechts in twee situaties echt lonend. In de eerste plaats natuurlijk wanneer in een grootscheeps onderzoek een vrij gecompliceerde vragenlijst gedurende langere tijd en/of bij een grote steekproef wordt afgenomen. Maar ook bij het andere uiterste is computergeleid enquêteren aan te raden. In een 'micro'-survey waarbij in essentie al het werk door één onderzoek(st)er wordt gedaan biedt het gebruik van een personal computer bij de dataverzameling grote voordelen. In deze situatie is er van bovengeschetste managementproblemen geen sprake, terwijl de vereenvoudiging van de soms lastige interviewtaak, de foutenreductie en de tijdsbesparing bij interviewen en datainvoer hier nu net het verschil uit kunnen maken tussen het wel of niet uitvoerbaar zijn van een -wat budget en personeel betreft- beperkt onderzoek. Voor de vele middelgrote onderzoeken, zeker als het onderzoeken met eenvoudige vragenlijsten betreft, biedt computergeleid enquêteren nu nog weinig voordelen.

Een eindoordeel over computergeleid enquêteren is nu echter nog niet te geven. Denkend aan de veranderingen die in de afgelopen twintig jaar hebben plaatsgevonden binnen de gegevensverwerking en aan de nieuwe mogelijkheden die de computertechnologie binnen het bereik van statistici heeft gebracht, lijkt een optimistische toekomstverwachting voor de gang van zaken bij de elektronische gegevensverzameling op zijn plaats.

Noten

¹ Vakgroep Methoden en Technieken, Faculteit SCW, Vrije Universiteit en Interuniversitair Onderzoeksinstituut voor Psychometrie en Sociometrie (IOPS).
Correspondentieadres: Koningslaan 22-24, 1075 AD Amsterdam.
Earn/Bitnet: V70UEDIT at HASARA11.

² Met dank aan: H. de Bock (Inter/View), F. Dekker (Geografisch Instituut Rijksuniversiteit Utrecht), J. Hox (Faculteit PAOW, Universiteit van Amsterdam), F. Kerssemakers (C.B.S.), S. Pols (NIPO), en J. Hutjes, C. de Grouw en H. van der Hoeven (ITS).

³ In dit artikel wordt gebruik gemaakt van de in de U.S.A. gebruikelijke benamingen voor computergeleid enquêteren, zoals o.a. beschreven staan in Nicholls en Groves (1986).

7. AANBEVOLEN LITERATUUR

- J. de Jong-Gierveld en J. van der Zouwen (1987). *De vragenlijst in sociaal onderzoek*. Deventer: Van Loghum Slaterus. (Overzicht van ontwikkelingen binnen vragenlijstonderzoek).
J.J. Hox en G. de Zeeuw (1988). *De Micro-computer in Sociaal Wetenschappelijk Onderzoek*. Lisse: Swets en Zeitlinger. (Algemene informatie over informatisering van onderzoek).
D.A. Dillman. (1978). *Mail and Telephone Surveys; the Total Design Method*. New York: Wiley.

(Praktische inleiding in telefonisch enquêteren en CATI)

- R.M. Groves et al. (1988). *Telephone Survey Methodology*. New York: Wiley. (State of the art overzicht op het terrein van telefonisch interviewen).
- CBS-select 4: Automation in Survey Processing* (1987). Den Haag: Staatsuitgeverij. (Recente ontwikkelingen bij verschillende vormen van geautomatiseerde gegevensverzameling).

8. LITERATUUR

- Baker, R.P. & Lefes, W.L. (1988). The design of CATI-systems: a review of current practice. In R.M. Groves et al. *Telephone Survey Methodology*. New York: Wiley.
- Bastelaer, A.M.L. van, Kerssemakers, F.A.M. & Sikkel, D. (1987). A test of the Netherlands continuous labor force survey with hand held computers: interviewer behaviour and data quality. In *CBS-select 4: Automation in Survey Processing*. Den Haag: Staatsuitgeverij.
- Berry, S.H. & O'Rourke, D. (1988). Administrative designs for centralized telephone centers: implications of the transition to CATI. In R.M. Groves et al. *Telephone Survey Methodology*. New York: Wiley.
- Bethlehem, J. (1989). *Blaise, een eerste kennismaking*. Voorburg: Centraal Bureau voor de statistiek.
- Billiet, J., Loosveldt, G. & Waterplas, L. (1984). *Het surveyinterview onderzocht*. Leuven: Sociologisch onderzoeksinstituut, K.U. Leuven.
- Bie, S.E. de, Stoop, I.A.L. & Vries, K.L.M. de (1989). *CAI Software: An evaluation of software for computer assisted interviewing*. Uitgave van de Vereniging Onderzoeks Instituten (VOI), p/a SISWO, Amsterdam
- Bock, H. de (1987). Technologische innovatie in sociaal wetenschappelijk onderzoek: maatschappelijke randvoorwaarden. In J. de Jong-Gierveld & J. van der Zouwen. *De vragenlijst in sociaal onderzoek*. Deventer: Van Loghum Slaterus.
- Denteneer, D., Bethlehem, J.G., Hundepool, A.J. & Keller W.J. (1987). The BLAISE-system for computer assisted survey processing. In *CBS-select 4: Automation in Survey Processing*. Den Haag: Staatsuitgeverij.
- Dillman, D.A. (1978). *Mail and Telephone Surveys; the Total Design Method*. New York: Wiley
- Fink, J.C. (1983). CATI's first decade; the Chilton experience. *Sociological Methods and Research*, 2, 153-168.
- Frey, J.H. (1983). *Survey Research by Telephone*. Beverly Hills: Sage.
- Groves, R.M. & Kahn, R.L. (1979). *Surveys by Telephone*. New York: Academic Press.
- Groves, R.M. & Mathiowitz, N.A. (1984). Computer assisted telephone interviewing: effects on interviewers and respondents. *Public Opinion Quarterly*, 48, 356-369.
- Groves R.M. & Nicholls, W.L. II. (1986). The status of Computer Assisted Telephone Interviewing: part 2- data quality issues. *Journal of Official Statistics*, 2, 117-134.
- Hindriks, M.J.H. (1986). *Computer Assisted Telephone Interviewing*, paper ter gelegenheid van het NSAV jubileum congres, Amsterdam, Vrije Universiteit.
- House, C.C. & Nicholls, W.L. II. (1988). Questionnaire redesign for CATI: design objectives and methods. In R.M. Groves et al. *Telephone Survey Methodology*. New York: Wiley.
- Kerssemakers, F.A.M., Mast, F.A.C. & Remmerswaal, P.W.M. (1987) Computer assisted telephone interviewing, some response findings. In *CBS-select 4: Automation in Survey Processing*. Den Haag: Staatsuitgeverij.
- Kiesler, S. & Sproull, L.S. (1986). Response effects in electronic surveys. *Public Opinion Quarterly*, 50, 402-413.
- Leeuw, E.D. de & Hox, J.J. (1985). Recente ontwikkelingen in schriftelijk onderzoek. In *Jaarboek van de Nederlandse Vereniging voor marktonderzoekers 1985*. Haarlem: De Vrieseborch.
- Leeuw, E.D. de & Hox, J.J. (1988). The effects of response stimulating factors on response rates and data quality in mail surveys: a test of Dillman's Total Design Method. *Journal of Official Statistics*, 4, 241-249.
- Leeuw, E.D. de & Zouwen, J van der (1988). Data quality in telephone and face to face surveys; a comparative meta analysis. In R.M. Groves et al. *Telephone Survey Methodology*. New York: Wiley.

- Nicholls, W.L. II. & Groves, R.M. (1986). The status of Computer Assisted Telephone Interviewing: part 1- introduction and impact on cost and timeliness of survey data. *Journal of Official Statistics*, 2, 93-115.
- Palit, C. & Sharp, H. (1983). Microcomputer Assisted Telephone Interviewing. *Sociological Methods and Research*, 2, 169-189.
- Saris, W.E., Doorn, L. & Pijper, W.M. (1987). A new approach to panel research: teleinterviewing. Amsterdam: Stichting voor Sociometrisch Onderzoek.
- Shanks, J.M. (1983). The current status of Computer Assisted Telephone Interviewing; recent progress and future prospects. *Sociological Methods and Research*, 2, 119-142.
- Sikkel, D. (1987). Computergestuurde gegevensverzameling. *Mens en Maatschappij*, 3, 289-301.
- Sikkel, D. (1988). *Quality Aspects of Statistical Data Collection*. Amsterdam: Sociometric Research Foundation.
- Spaeth, M.A. (1987). CATI-facilities at survey research organizations. *Survey Research*, 18, 18-22.
- Thornberry, O., Nicholls, W.L. & Kulpinsky, S. (1982). *Data collection methods in federal statistical surveys*. American Statistical Association, Proceedings of the 1982 conference, section on research methods.
- Tull, D.S. & Hawkins, D.J. (1984). *Marketing Research; Measurement and Method*. New York: McMillan.
- Turner, C.F. & Martin, E. (1984). *Surveying subjective phenomena*, Volume I. New York: Russell Sage Foundation.
- Waterton, J.J. (1984) *Reporting alcohol consumption: the problem of response validity*. American Statistical Association, Proceedings of the 1984 conference, section on survey research methods.

Manuscript ontvangen 9-3-1988

Definitieve versie ontvangen 22-6-1989

De effecten van een opleiding die tot doel heeft a.s. leraren te leren reflecteren¹

F.A.J Korthagen en Th. Wubbels

Instituut voor de Lerarenopleiding, Universiteit van Amsterdam
Vakgroep Natuurkundedidaktiek/PDI, Rijksuniversiteit Utrecht

ABSTRACT

The aim of this study was to establish whether a program designed primarily to promote reflective teaching among prospective teachers does indeed produce more reflective teachers, and whether graduates of such a program are also better teachers. In addition, the study examined whether these graduates showed a greater inclination towards innovation and a higher level of job satisfaction. To this end the graduates were compared with a control group, using a number of established criteria taken from the literature. According to the perceptions of their students with regard to teacher behaviour, the graduates of the program designed to promote reflective teaching performed better than the teachers in the control group. However, for a number of other criteria no differences were found between the two groups. The results of a further analysis indicate that it is primarily in the long term that the reflective teaching program may be expected to have a favourable effect on teacher performance.

INLEIDING EN PROBLEEMSTELLING

In het afgelopen decennium zijn nieuwe visies op de opleiding van onderwijsgevenenden ontstaan, waarvan het inquiry-oriented paradigma de meest opvallende representant is. Het wordt door Zeichner (1983) gezien als alternatief voor het behavioristische, het personalistische en het traditioneel-ambachtelijke paradigma. Een centraal begrip in dit nieuwe paradigma is (leren) reflecteren. Dit begrip is, zoals Gore (1987) opmerkt, gaan behoren tot de taal van de lerarenopleiding. In Nederland speelt het een belangrijke rol in de ontwikkeling van de universitaire lerarenopleiding. Zo noemt de ARCULO-werkgroep Toelating tot de Tweedefase Lerarenopleiding (1985) 'vaardigheid in het analyseren van onderwijs(achtige) activiteiten' als mogelijk selectie criterium. In de ARCULO-nota 'Bouwstenen voor het curriculum van de universitaire lerarenopleiding' (ARCULO-werkgroep Tweede Fase, 1985,p.7) wordt gesteld dat een eerstegraads leraar over reflectieve bekwaamheden moet beschikken. We citeren: 'Reflectieve bekwaamheden betekenen dat een leraar zijn of haar handelen zelf onderwerp kan maken van beschouwing met het oog op verbeteren van dat handelen'.

Er zijn verschillende opleidingsmodellen en -strategieën ontwikkeld die gebaseerd zijn op de doelstelling van het leren reflecteren. Er is evenwel in de literatuur een schaarste aan duidelijke beschrijvingen van die modellen en strategieën. Uitzonderingen zijn Feiman (1979), Korthagen (1982), Vedder (1984), Zeichner en Liston (1987) en Zeichner en Noffke

Adres: F.A.J. Korthagen, Herengracht 256, 1016 BV Amsterdam

(1987). Nog schaarser zijn effectstudies (Zeichner, 1987). Er is (nog) nauwelijks empirische ondersteuning van de stelling dat programma's die gericht zijn op leren reflecteren, betere leraren of leraren met andere kenmerken afleveren. Korthagen en Verkuyl (1986) toonden zelfs aan dat dergelijke programma's wel eens van weinig nut kunnen zijn voor studenten met een zgn. externe leeroriëntatie, d.w.z. een voorkeur voor leren door sturing van buitenaf, bijv. door een docent of boek.

De bovengenoemde aandacht voor (leren) reflecteren in combinatie met het ontbreken van een empirische basis voor dit opleidingsdoel was de aanleiding tot het hier beschreven onderzoek. Daarin stonden de volgende vragen centraal: Zijn leraren die opgeleid zijn volgens een op leren reflecteren gebaseerd opleidingsmodel in vergelijking met anders opgeleide leraren:

1. reflectiever ingesteld?
2. betere leraren?
3. meer bereid tot vernieuwingen in het onderwijs?
4. tevredener met hun beroep?

THEORETISCH KADER

Reflectie door leerkrachten is in het afgelopen decennium meer in de belangstelling komen te staan onder invloed van de toenemende aandacht voor 'teacher thinking'. Zoals Clark (1986) aangeeft werd de leraar vóór 1975 voornamelijk opgevat als 'decision-maker' (vergelijkbaar met een klinisch arts) en werd in het onderzoek naar het functioneren van leraren vrijwel uitsluitend het gedrag beschouwd. Rond 1975 werd ingezien dat meer onderzoek naar het denken van leraren nodig was: 'It is obvious that what teachers do is directed in no small measure by what they think' (National Institute of Education, 1975, p. 1). Dat leverde een verschuiving op naar een opvatting van de leerkracht als 'reflective professional who constructs meaning' (Clark, 1986; zie ook Schön, 1983).

Met deze verandering in het gehanteerde paradigma ontstond ook een verschuiving in de gehanteerde onderzoeksmethoden. Teneinde zicht te krijgen op de denkprocessen van leraren werd gebruik gemaakt van 'verbal report methods that attempt to obtain data on the intellectual processes used by subjects as they render judgments and make decisions or solve problems' (Shavelson, Webb & Burstein, 1986, p. 79). Hardop-denkmethoden, retrospectieve interviews en stimulated recall-technieken werden meer en meer gebruikt. Ze leverden nieuwe inzichten op, bijvoorbeeld inzicht in het belang van 'teachers' conceptions' (Duffy, 1977; Larsson, 1984), 'teacher perspectives' (Janesick, 1982), 'teacher constructs' (Olson, 1981), 'teacher beliefs' (Munby, 1983), 'teacher thinking criteria' (Halkes & Deijkers, 1984) en 'personal knowledge' (Lampert, 1985). Ook ontstond aandacht voor metacognitieve kennis en regulatievaardigheden (Simons & Vermunt, 1986). De discussie over de validiteit en betrouwbaarheid van op intro- of retrospectie gebaseerde onderzoeksmethoden is echter nog niet verstomd. Het is moeilijk, zo niet onmogelijk het denken van leraren rechtstreeks te analyseren, met name tijdens het lesgeven. Meer in het bijzonder geldt dat voor reflectie: men kan nu eenmaal niet in het hoofd van de reflecterende persoon kijken.

Mede omdat wij in ons onderzoek een groot aantal afgestudeerden wilden betrekken, die verspreid over geheel Nederland werkzaam zijn, hebben wij gekozen voor een andere benadering, die meer aansluit bij de in de sociale psychologie gangbare methoden van attitude-meting. Wij richten ons niet op de analyse van de feitelijke reflectie, maar trachten te bepalen in hoeverre de persoon een reflectieve instelling heeft. Daarnaast maken wij gebruik van gedragsbeschrijvende vragenlijsten van leerlingen om verbanden te kunnen

leggen tussen die reflectieve instelling en de kwaliteit van de onderzochte leraren.

We zullen hieronder eerst ingaan op de begrippen 'reflectie' en 'reflectieve instelling'. Daarna bespreken we de vraag wat wij in ons onderzoek verstaan onder de kwaliteit van een leraar. Tenslotte gaan we in op het verband tussen reflecteren en de leraarskenmerken uit de vragen 3 en 4.

Er bestaat in de literatuur geen overeenstemming over de definitie van het reflectiebegrip. Wij kiezen voor een definitie die berust op het principe dat reflectie betrekking heeft op mentale structuren. Deze structuren zijn volgens het epistemologisch subjectmodel van Groeben & Scheele (1977) persoonlijk en dynamisch. Immers, als iemand reflecteert worden die structuren niet als statisch gegeven aanvaard, maar gaat het er die persoon nu juist om, ze uit te breiden, te verfijnen of te veranderen. Het gevolg van reflectie is derhalve dat er (nieuwe) relaties ontstaan tussen de elementen van een mentale structuur. Dat leidt tot de volgende definitie, die wij overnemen van De Jong & Korthagen (1988): een persoon reflecteert als hij bezig is structuur aan te brengen in zijn perceptie van een situatie, zijn wijze van handelen of zijn wijze van leren, of als hij bezig is zo'n structuur te wijzigen. Een persoon heeft een reflectieve instelling als er sprake is van een neiging tot het aanbrengen of wijzigen van mentale structuren en daarmee van een gerichtheid op het eigen ontwikkelingsproces.

Wil men bij iemand vaststellen in hoeverre er sprake is van een reflectieve instelling, dan dient rekening gehouden te worden met het domeinspecifieke karakter van reflectie (Korthagen, 1987a; Nelissen, 1987). Korthagen & Verkuyl (1986) onderscheiden op basis van kwalitatief onderzoek bij a.s. wiskundeleraren vier domeinen, n.l. zichzelf, de leerlingen, de leerstof en de school.

Tot zover het theoretisch kader dat ten grondslag ligt aan de eerste onderzoeksvraag. Met betrekking tot de tweede vraag moet geconstateerd worden dat hoewel er in de literatuur wat algemeen onderwijskundige vaardigheden betreft, langzamerhand wel een beeld is ontstaan van effectief leraarsgedrag (zie Brophy & Good, 1986) er toch geen eenstemmigheid bestaat over de vraag wat een goede leraar is. In het hier beschreven onderzoek hanteerden wij twee empirisch onderbouwde criteria voor de kwaliteit van een leraar.

In de eerste plaats maakten wij gebruik van leerlingenpercepties van het leraarsgedrag. Voor de vaststelling van die percepties ontwikkelden Wubbels e.a. (1987a) de Vragenlijst voor Interactioneel Leraarsgedrag, de VIL. De VIL is gebaseerd op een model van Leary (1957) voor de analyse van interpersoonlijk gedrag. In dit model worden twee dimensies van gedrag onderscheiden, n.l. de machtsdimensie (boven-onder) en de nabijheidsdimensie (samen-tegen). Deze dimensies worden gerepresenteerd m.b.v. onderling loodrechte assen. In het zo ontstane assenstelsel worden vervolgens door nog eens twee lijnen aan te brengen, acht sectoren onderscheiden, n.l. Boven-Samen (BS, 'leidend'), Samen-Boven (SB, 'helpend/vriendelijk'), Samen-Onder (SO, 'begrijpend'), Onder-Samen (OS, 'ruimte gevend/latend'), Onder-Tegen (OT, 'onzeker'), Tegen-Onder (TO, 'ontevreden'), Tegen-Boven (TB, 'corrigerend') en Boven-Tegen (BT, 'streng'). Met elk van deze sectoren correspondeert een schaal van de VIL bestaande uit ca. 10 items die op een vijfpuntsschaal beantwoord worden.

Teneinde leerlingenscores op de VIL te kunnen interpreteren in het licht van de tweede vraagstelling maken we gebruik van de resultaten van een onderzoek van Brekelmans e.a. (1986) naar verbanden tussen leerlingenpercepties van interactioneel leraarsgedrag en cognitieve en affectieve leeropbrengsten. Uit dat onderzoek bleek dat de gemiddelde scores die leerlingen toekennen aan een bepaalde leraar op de schalen BS, SB en SO positief samenhangen met zowel cognitieve als affectieve opbrengsten: naarmate leerlingen vinden dat een leraar meer leidend, vriendelijk of begrijpend gedrag vertoont, behalen deze

leerlingen hogere prestaties op een cognitieve toets en waarderen ze de lessen en het vak meer. Met name voor BS zijn deze verbanden vrij sterk. Verder toonden Brekelmans e.a. (1986) aan dat de scores op OT, TO en TB negatief samenhangen met cognitieve en affectieve opbrengsten. Voor de schalen BT en OS zijn de verbanden met opbrengsten niet eenduidig. Uit deze resultaten, die in grote lijnen sporen met resultaten van proces-product-onderzoek (Brophy & Good, 1986) kan een criterium voor de kwaliteit van een leraar afgeleid worden. We noemen dit het leerlingenperceptiecriterium; volgens dit criterium is een leraar beter naarmate hij of zij een hogere score heeft op de schalen die positief met leerwinst correleren en lagere scores op de negatief correlerende schalen.

In de tweede plaats gebruiken we de adequaatheid van het zelfbeeld van leraren als criterium voor leraarskwaliteit. De adequaatheid van het zelfbeeld van een leraar kan worden afgemeten aan de overeenstemming tussen dat zelfbeeld en het beeld dat leerlingen van die leraar hebben, beide vastgesteld m.b.v. de VIL. Wij hebben de volgende redenen om dit zelfbeeldcriterium voor leraarskwaliteit te hanteren. In de eerste plaats betekent een geringe discrepantie tussen het zelfbeeld van een leraar en het leerlingenbeeld dat die leraar een betrekkelijk objectief beeld van het eigen gedrag heeft. Immers, volgens Hook & Rosenshine (1979) stemmen de leerlingenpercepties van het leraarsgedrag meer overeen met de percepties van het leraarsgedrag door getrainde observatoren dan de zelfpercepties van leraren. Er zijn aanwijzingen in de persoonlijkheidspsychologie dat mensen adequater functioneren naarmate ze een objectiever beeld van het eigen gedrag hebben (Leary, 1957; Van de Griend, 1975). Bovendien hebben Wubbels e.a. (1987b) aanwijzingen gevonden voor verbanden tussen de kwaliteit van een leraar (afgemeten aan de leeropbrengsten die hij of zij bereikt) en de mate van overeenstemming tussen het leerlingenbeeld en het leraarsbeeld van het leraarsgedrag. Naarmate leraren meer gedrag vertonen dat gunstig is voor leerwinst, verschilt hun zelfbeeld minder van het beeld dat hun leerlingen van het leraarsgedrag hebben.

De derde en vierde onderzoeksvraag zijn gebaseerd op verbanden die gelegd kunnen worden tussen reflectie enerzijds en innovatiebereidheid en arbeidssatisfactie anderzijds.

Het verband tussen reflectie en innovatiebereidheid komt bijv. naar voren bij Zeichner (1983, p.6) die over inquiry-oriented teacher education zegt: 'this orientation views the prospective teacher as an active agent in his or her own preparation for teaching and assumes that the more a teacher is aware of the origins and consequences of his or her actions and of the realities that constrain these actions, the greater is the likelihood that he or she can control and change both the actions and the constraints'. (Zie ook Zeichner & Liston, 1987, p. 26.) Vedder en Bannink (1987, p. 3) zeggen over de leraar: 'He must possess the potential for change, and also here, reflection is useful and necessary'.

Het verband tussen reflectie en arbeidssatisfactie wordt in de literatuur minder rechtstreeks gelegd, maar kan daaruit wel afgeleid worden. Wanneer het handelen van beroepsbeoefenaars meer en meer een routinematig karakter krijgt heeft dat als positief effect dat de complexe werkelijkheid beter beheersbaar wordt. Maar een negatief effect is dat, zoals Schön (1983, p.61) schrijft, het risico dreigt dat die beroepsbeoefenaars weinig meer nadenken over wat zij doen en dat zij in patronen vervallen die zij niet meer kunnen corrigeren. Bannink e.a. (1985, p. 8-9) verwijzen naar deze opvatting van Schön en voegen er aan toe: 'En hij [de beroepsbeoefenaar] leert selectief verschijnselen te veronachtzamen die niet overeenkomen met zijn 'handelingsoriëntaties'. Dat heeft verveling en 'burn-out' tot gevolg'. Net als Bannink e.a. veronderstellen wij dat routinematig handelen gemakkelijk tot sleur en verveling kan leiden, hetgeen met een geringe arbeidssatisfactie samen zal gaan.

ONDERZOEKSOPZET

Onderzoek naar de effectiviteit van opleidingsprogramma's is volgens Koehler (1985) vaak methodologisch zwak omdat er geen controlegroepen en voormetingen gebruikt worden en men geen lange-termijneffecten meet. In ons onderzoek hebben we wel een controlegroep gebruikt en ook naar lange-termijn effecten gekeken. Er werd gebruik gemaakt van een quasi-experimentele opzet waarbij gegevens werden verzameld over twee groepen leraren die tussen 1977 en 1986 waren afgestudeerd.

De eerste groep, die we verder de experimentele groep zullen noemen, bestond uit wiskundeleraren die opgeleid waren in de door Korthagen (1982) beschreven (Nieuwe Leraren-)opleiding. Blijkens het onderzoek van Korthagen (1982) en van Korthagen en Verkuyl (1986) was die opleiding in de periode dat de onderzochte leraren daar studeerden, uitdrukkelijk gericht op leren reflecteren. Dat kwam zowel tot uitdrukking in de curriculumelementen in de wiskunde- en onderwijskundecomponent van de opleiding als in de wijze van begeleiden van de studenten.

Bij de keuze van een opleiding die in aanmerking kwam om daarvan leraren voor een controlegroep te betrekken, baseerden we ons in eerste instantie op een analyse van de studieprogramma's zoals die in de studiegidsen van de Nieuwe Lerarenopleidingen voor het vak wiskunde worden beschreven. Gekozen werd voor de afgestudeerden van een opleidingsprogramma dat gezien de beschrijving in de studiegids gedurende de periode dat de controlegroep daar studeerde, veel nadruk legde op kennis van het schoolvak.

Alle oud-studenten van de twee opleidingen die afgestudeerd waren tussen 1977 en 1986, werden benaderd met de vraag of ze aan het onderzoek wilden meewerken. Van de aan het onderzoek deelnemende leraren werden achtergrondgegevens verzameld zoals het geslacht, de leeftijd, het jaar van afstuderen en de leerjaren en het schooltype waarin les gegeven werd. De variabele 'jaar van afstuderen' werd gebruikt om via een cross-sectionele benadering indicaties voor korte- en lange-termijneffecten te verkrijgen.

Er werden 248 afgestudeerden bereikt, van wie er slechts 143 in het onderwijs werkzaam bleken te zijn. Van hen verklaarden 81 leraren zich bereid mee te werken. Uiteindelijk is van 73 leraren een complete set gegevens verkregen. De experimentele groep (37 leraren) en de controlegroep (36 leraren) bleken niet significant te verschillen op de achtergrondvariabelen geslacht, leeftijd en leerjaren waarin les gegeven werd, maar wel op de variabele schooltype. De leraren uit de controlegroep werkten relatief meer in het lager beroepsonderwijs dan de leraren uit de experimentele groep ($\chi^2 = 14,2$, $df = 6$, $p = 0,01$).

Een controle op de juistheid van de bovengenoemde veronderstelling dat de opleidingen wezenlijk verschillend waren, werd uitgevoerd door de afgestudeerden (schriftelijk) de open vraag te stellen wat volgens hen de belangrijkste kenmerken van hun opleidingsprogramma waren geweest. De antwoorden werden geanalyseerd door een onafhankelijke beoordelaar die per respondent aangaf hoeveel uitspraken betrekking hadden op leren reflecteren en/of het leren sturen van de eigen ontwikkeling. Daarbij wist de beoordelaar niet van welke opleiding een respondent afkomstig was. De afgestudeerden van het op leren reflecteren gebaseerde opleidingsprogramma noemden gemiddeld 4,6 opleidingsprincipes. Dit is ongeveer twee keer zoveel als de 2,3 principes die in de controlegroep werden genoemd. Van de afgestudeerden in de controlegroep noemde slechts 6% een principe dat met het leren reflecteren te maken had. In de andere groep was dit 47%. De afgestudeerden in de controlegroep noemden het meest de nadruk op het belang van wiskundekennis (namelijk 33% van hen). Dit principe werd door geen van de leraren uit de experimentele groep genoemd. De afgestudeerden van het opleidingsprogramma met de doelstelling te leren reflecteren herkennen deze doelstelling dus inderdaad meer in de

opleiding die ze hebben gevolgd dan de afgestudeerden van de controlegroep.

We hanteerden verder de volgende methoden van gegevensverzameling.

1. In verband met de eerste onderzoeksvraag werd gebruik gemaakt van de reflectieschalen van de zgn. I.E.S.-vragenlijst. Van deze vragenlijst bestond een variant voor a.s. leraren zonder leservaring (Korthagen, 1987b) en een nog niet gepubliceerde, maar wel in eerder onderzoek gebruikte variant voor leraren met leservaring. We gebruikten een bewerking van deze laatstgenoemde variant die bestaat uit vier schalen die meten in hoeverre leraren reflectief zijn ingesteld op de vier domeinen zichzelf, de leerlingen, de leerstof en de mesositatie in de school. De respondenten beantwoordden de items op een vijfpuntsschaal van 'erg oneens' (1) tot 'erg eens' (5).
2. Gegevens over de variabelen innovatiebereidheid en arbeidssatisfactie (corresponderend met de derde en vierde onderzoeksvraag) werden verzameld met een bewerking van een leraarsvragenlijst die eerder was gebruikt in het kader van een Nederlandse optie van de Second International Science Study (Brekelmans e.a., 1986).

Tabel 1: Aantal items, karakteristiek item en interne consistenties (Cronbach's α) van de leraarschalen.

schaal	aantal items	karakteristiek item	coëfficiënt alpha
Reflectieve instelling			
- zichzelf	6	ik denk over mezelf na	0,81
- leerlingen	3	ik verdiep mij erin hoe de leerlingen met elkaar omgaan	0,71
- leerstof (wiskunde)	5	als ik wiskundeles heb gegeven, denk ik erover na welke wiskundige problemen er in de les naar voren kwamen	0,80
- schoolsituatie	6	ik denk na over het functioneren van mijn school als geheel	0,71
Innovatiebereidheid	7	doelen, inhoud en vorm van het onderwijs moeten voortdurend ter discussie worden gesteld	0,55
Arbeidssatisfactie	9	ik vind het leuk om leraar te zijn.	0,86

In tabel 1 wordt als maat voor de betrouwbaarheid de interne consistentie weergegeven van de schalen betreffende de reflectieve instelling, innovatiebereidheid en arbeidssatisfactie. Verder wordt een representatief item voor deze schalen gepresenteerd op basis van de item-totaalcorrelatie. Alle schalen blijken bruikbaar. De interne consistenties liggen behalve voor innovatiebereidheid (0,55) boven de door Nunnally (1967) als acceptabel genoemde norm van 0,70.

3. I.v.m. het leerlingenperceptie criterium en het zelfbeeldcriterium werden gegevens verzameld bij de leraren zelf en bij de leerlingen van twee klassen van elke leraar. Daartoe werd de leraar gevraagd de klas te selecteren waar de lessen het best gingen en de klas waar de lessen het slechtst gingen. Vervolgens werd voor beide klassen het zelfbeeld van de leraar en het gemiddelde leerlingenbeeld van die leraar bepaald met behulp van de VIL. De VIL is reeds vaak in onderzoek gebruikt en blijkt steeds betrouwbaar te zijn (zie Wubbels e.a. 1987a). Ook in dit onderzoek zijn de interne

consistenties van alle VIL-schalen hoger dan 0,80.

Aangezien de groepen in dit onderzoek niet erg groot zijn hebben we met het oog op de power van de gebruikte toetsen een significantieniveau van 5% gehanteerd. Teneinde kanskapitalisatie te voorkomen is het aantal uit te voeren toetsingen beperkt door scores tot één score samen te voegen waar dit inhoudelijk verantwoord was, dan wel door multivariate toetsen te gebruiken.

RESULTATEN

reflectieve instelling

In het kader van de eerste onderzoeksvraag zijn de experimentele en de controlegroep met een multivariate variantieanalyse vergeleken op de schalen betreffende de reflectieve instelling². Tabel 2 geeft de gemiddelde scores op de verschillende schalen en de resultaten van de toetsing. De groepen blijken ten aanzien van de meest voor de hand liggende effectmaat, hun reflectieve instelling, niet te verschillen. Om na te gaan of er verschillen zijn op kortere en wat langere termijn werden ook multivariate variantie-analyses uitgevoerd voor de gemiddelde schaalscores in de twee opleidingsgroepen voor respectievelijk de leraren die minder dan 3 jaar geleden waren afgestudeerd en leraren die 3 jaar of langer geleden waren afgestudeerd. Noch voor de leraren die kort geleden zijn afgestudeerd, noch voor de langer geleden afgestudeerden blijken er ten aanzien van de reflectieve instelling significante verschillen tussen de opleidingsgroepen te zijn.

Tabel 2. Gemiddelde scores in de twee groepen leraren op de leraarsschalen en de resultaten van multivariate variantie-analyses resp. t-toetsen. Dezelfde gegevens voor leraren 1 en 2 jaar resp. 3 jaar of meer na hun afstuderen.

Schaal ¹	alle leraren			leraren 1 en 2 jaar na afstuderen			leraren 3 jaar of meer na afstuderen		
	exp. groep n = 37	con- trole- groep n = 36	F/t- waarde	exp. groep n = 7	con- trole- groep n = 9	F/t- waarde	exp. groep n = 30	con- trole- groep n = 27	F/t- waarde
Reflectieve instelling									
- zichzelf	4,0	3,9		3,8	3,9		4,1	3,9	
- leerlingen	3,6	3,6		3,9	3,8		3,6	3,5	
- leerstof (wiskunde)	3,8	3,6	f=0,4 df=4/68	3,9	3,9	f=0,1 df=4/11	3,7	3,4	f=0,4 df=4/52
- schoolsituatie	3,4	3,5		3,5	3,4		3,4	3,6	
Innovatiebereid- heid	3,0	3,0	t=0,2	3,2	3,0	t=0,7	3,0	3,0	t=0,2
Arbeidssatisfactie	3,9	3,4	t=2,5**	3,9	3,7	t=0,8	3,9	3,3	t=2,4**

* $p < 0,05$; ** $p < 0,01$

¹ Alle schaalscores zijn getransformeerd zodanig dat ze een bereik hebben van 1,0 tot 5,0, overeenkomend met de vijf puntsschaal die bij elk item was opgenomen: 1: erg oneens; 5: erg eens.

Leerlingenperceptiecriterium

Voor elke klas werd de gemiddelde leerlingenscore op de acht schalen van de vragenlijst voor interactioneel leraarsgedrag bepaald. Daarna werden de scores van de twee klassen van één leraar gemiddeld. Aldus werden per leraar acht gemiddelde schaalscores verkregen als maat voor de leerlingenpercepties van het leraarsgedrag.

Tabel 3 toont de gemiddelde leerlingenscores op de acht schalen van de VIL in de twee groepen leraren. Uit de tabel blijkt, dat de leerlingenscores voor de experimentele groep hoger zijn dan die voor de controlegroep op alle schalen die positief met cognitieve en affectieve opbrengsten verbonden zijn (BS, leidend, SB, helpend/vriendelijk en SO, begrijpend). Voor alle negatief met opbrengsten verbonden schalen (OT, onzeker, TO, ontevreden en TB, corrigerend) zijn de scores van de experimentele groep lager dan die van de controlegroep.

Tabel 3: Leerlingenpercepties van het interactioneel leraarsgedrag in de twee groepen. Dezelfde gegevens voor leraren 1 en 2 jaar na afstuderen resp. 3 jaar of meer na afstuderen.

Schaal ¹	Alle leraren		leraren 1 en 2 jaar na afstud.		leraren 3 jaar of meer na afstud.		L.B.O.-klassen	
	exp. groep n = 37	contr. groep n = 36	exp. groep n = 7	contr. groep n = 9	exp. groep n = 30	contr. groep n = 27	exp. groep n = 19	contr. groep n = 38
BS-leidend	3,6	3,4	3,4	3,3	3,6	3,4	3,8	3,3
SB-helpend/ vriendelijk	3,9	3,6	3,6	3,7	3,9	3,6	4,0	3,5
SO-begrijpend	4,0	3,8	3,7	3,9	4,1	3,8	4,1	3,8
OS-ruimtegevend/latend	2,8	2,7	2,5	2,8	2,9	2,7	2,8	2,7
OT-onzeker	2,0	2,1	2,2	2,2	1,9	2,0	1,9	2,2
TO-ontevreden	2,3	2,4	2,5	2,5	2,2	2,4	2,2	2,4
TB-corrigerend	2,7	2,8	3,1	2,8	2,6	2,8	2,7	2,9
BT-streng	2,7	2,7	2,8	2,8	2,7	2,7	2,8	2,7

¹Bereik van schaalscores als bij tabel 2

Teneinde het leerlingenperceptiecriterium met één toets te kunnen onderzoeken is een globale indicatie voor de kwaliteit van een leraar berekend, namelijk de som van de BS-, SB- en SO-scores verminderd met de TB-, TO- en OT-scores. Met behulp van deze maat kunnen de twee groepen met elkaar worden vergeleken. Tabel 4 toont de globale kwaliteitsscore en de resultaten van een t-toets op de gemiddelde kwaliteitsscores in de twee groepen.

De gemiddelde kwaliteitsscore voor de experimentele groep blijkt significant hoger dan in de controlegroep. De effectgrootte is 0,6 standaarddeviatie hetgeen een effect van middelmatige grootte genoemd kan worden (Cohen, 1977).

Analyses voor de onlangs (minder dan 3 jaar geleden) afgestudeerden laten zien, dat er in de eerste twee beroepsjaren geen significant verschil in de kwaliteitsscore is (tabel 4). Voor de langer geleden afgestudeerden is er echter wel een significant verschil en uit tabel 3 blijkt dat de richting van de verschillen dezelfde is als voor de totale groep. Het beter functioneren van de afgestudeerden van de op leren reflecteren gebaseerde opleiding lijkt dus eerder en lange- dan een korte-termijn effect.

Tabel 4. Globale kwaliteitsmaat in de twee groepen en resultaten van een t-toets op de groepsgemiddelden.

	Exp. groep		contr. groep		t-waarde
	n	kw.	n	kw.	
alle leraren	37	4,5	36	3,5	1,9*
leraren 1 en 2 jaar na afstuderen	7	3,0	9	3,4	-0,3
leraren 3 jaar of meer na afstuderen	30	4,9	27	3,6	2,2*
LBO-klassen ¹⁾	19	5,1	38	3,0	3,5**

* $p < 0,05$; ** $p < 0,01$

¹⁾Deze analyse kon alleen op klasniveau worden uitgevoerd, terwijl de overige analyses zijn uitgevoerd nadat over de klassen van een leraar was gemiddeld.

Het beter functioneren van de leraren in de experimentele groep zou kunnen samenhangen met het eerder geconstateerde verschil in aantal LBO- en AVO-klassen in de twee groepen. Wanneer we veronderstellen dat LBO-leerlingen moeilijker zijn dan AVO-leerlingen, zou dat een oorzaak kunnen zijn voor de lagere kwaliteitsscores van de leraren in de controlegroep. Daarom werden de leerlingenpercepties voor de LBO-klassen in de twee groepen apart vergeleken. De t-toets (tabel 4) op de gemiddelde kwaliteitsscores laat zien dat de leraren van de experimentele groep ook in de LBO-klassen beter functioneren dan de leraren van de controlegroep. Dit resultaat kan nog veroorzaakt zijn doordat in de controlegroep onder de LBO-klassen meer slechtste klassen aanwezig kunnen zijn dan in de experimentele groep. In de experimentele groep blijkt echter van de LBO-klassen 47% als slechtste klas gekozen te zijn, terwijl dat in de controlegroep 34% is. In de experimentele groep zijn onder de LBO-klassen dus relatief zelfs meer slechtste klassen vertegenwoordigd.

We concluderen uit deze analyses van de leerlingenpercepties van het leraarsgedrag dat de afgestudeerden van de opleiding die als doel heeft a.s. leraren te leren reflecteren, meer gedragingen vertonen die gunstig zijn voor leeropbrengsten dan de afgestudeerden van een opleiding die deze doelstelling niet heeft. De leraren uit de experimentele groep functioneren wat het leerlingenperceptie criterium betreft beter dan de leraren uit de controlegroep.

Zelfbeeldcriterium

Teneinde de adequaatheid van de lerarenpercepties van het eigen gedrag te onderzoeken hebben we de leerlingenpercepties van het leraarsgedrag en de leraarspercepties van het eigen gedrag vergeleken. De vergelijking is uitgevoerd door voor elke klas van een leraar de som (over de acht schalen) van het absolute verschil tussen de gemiddelde leerlingenscore in die klas en de leraarsscore te berekenen. Daarna werden de gemiddelden van deze absolute verschilsscores in de twee groepen vergeleken met een t-toets². De absolute verschilscore is in de experimentele groep 3,4 en in de controlegroep 3,8. Deze scores blijken niet significant te verschillen: ($t = 1,5$ en $p = 0,13$). Ook voor de leraren die 1 of 2 jaren geleden zijn afgestudeerd is er geen significant verschil tussen de twee opleidingsgroepen. Voor de leraren die 3 jaar geleden of langer zijn afgestudeerd is er echter wel een significant verschil, waarbij de discrepantie tussen de leerlingenperceptie van het leraarsgedrag en de perceptie van de leraar zelf kleiner is in de experimentele groep (scores respectievelijk 3,1 en 3,8, $t = 2,2$ en $p = 0,02$, de effectgrootte is ongeveer 0,5 standaarddeviatie). We concluderen dat de leraren die meer dan 2 jaar geleden zijn afgestudeerd aan de op leren reflecteren gebaseerde opleiding een adequater zelfbeeld hebben dan de leraren in de controlegroep met meer dan twee jaar ervaring.

Innovatiebereidheid en arbeidssatisfactie

Tenslotte zijn de leraren uit de twee groepen in het kader van de derde en vierde onderzoeksvraag met behulp van t-toetsen (enkelzijdige toetsing) vergeleken op hun innovatiebereidheid en arbeidssatisfactie (zie tabel 2)². Voor innovatiebereidheid wordt geen significant verschil tussen de groepen gevonden, maar voor arbeidssatisfactie wel. Bij een analyse op de gemiddelden van de groepen leraren die minder dan 3 jaar geleden zijn afgestudeerd blijken er voor beide variabelen geen significante verschillen te zijn tussen de twee opleidingsgroepen. Echter, voor de leraren die drie jaar of langer geleden zijn afgestudeerd blijkt de arbeidssatisfactie van de experimentele groep significant hoger (effectgrootte 0,6 standaarddeviatie). Een effect van de op leren reflecteren gebaseerde opleiding kan dus zijn dat de door haar opgeleide leraren op de langere termijn tevredener zijn met hun beroep.

DISCUSSIE

De conclusies van dit onderzoek met betrekking tot de effecten van het opleidingsprogramma dat zich richt op het leren reflecteren van a.s. leraren zijn voor voorstanders van leren reflecteren als doelstelling van de lerarenopleiding deels bemoedigend en deels teleurstellend. Bemoedigend is dat de leraren meer gedrag vertonen dan anders opgeleide leraren dat gunstig is voor leerwinst van leerlingen. Verder is de hogere arbeidssatisfactie en de grotere adequaatheid van hun zelfpercepties op langere termijn verheugend. Teleurstellend is dat de leraren uit de experimentele groep niet geneigd zijn meer te reflecteren en dat ze geen grotere bereidheid tot innovatie hebben.

Opvallend is dat de verschillen tussen de twee groepen leraren niet direct na de opleiding maar pas later tot uiting komen in de hier onderzochte effectmaten. De op leren reflecteren gebaseerde opleiding draagt er dus niet zozeer toe bij dat de leraren als beginnend leraar in de klas beter functioneren, maar geeft hun wellicht een grotere doorgroeicompententie. Dit stemt overeen met de conclusies van Korthagen (1982, blz. 262). Een vergelijkbare suggestie wordt door Créton en Wubbels (1984, blz. 366) geformuleerd ten aanzien van de korte- en lange-termijneffecten van het door hen onderzochte begeleidingsprogramma voor beginnende leraren. Wellicht moet geconcludeerd worden dat opleidingsprogramma's vooral effect (kunnen) hebben voor de periode nadat de leraren beginnersproblemen te boven zijn gekomen. We willen hierbij aantekenen dat de verschillen tussen de leraren in de twee opleidingsgroepen die samenhangen met het tijdstip waarop ze zijn afgestudeerd ook een gevolg zouden kunnen zijn van verschillen in de kenmerken van de opleiding zoals die de laatste jaren is gegeven vergeleken met de jaren daarvoor.

Ten aanzien van de oorzaken van het gunstiger leraarsgedrag in de experimentele groep moeten we constateren dat dit niet kan worden toegeschreven aan hun reflectieve instelling, althans aan de instelling zoals deze in dit onderzoek is gemeten. Desondanks zouden de twee groepen in reflectieve instelling kunnen verschillen. We kunnen ons n.l. voorstellen dat leraren zich bij de beantwoording van de vragen laten leiden door een vergelijking van hun reflectieve instelling met een norm over hoe reflectief een leraar moet zijn. In dat geval kunnen de scores voor de twee groepen leraren gelijk zijn terwijl de norm in de twee groepen wel zou kunnen verschillen en daarmee dus ook de werkelijke reflectieve instelling. Er zijn redenen om te veronderstellen dat de afgestudeerden van de op leren reflecteren gebaseerde opleiding ten aanzien van een reflectieve instelling hogere normen hebben dan de leraren uit de controlegroep. In de eerste plaats kan dit veroorzaakt zijn door de nadruk die de opleiding heeft gelegd op de noodzaak om te reflecteren. Die nadruk kan de norm van de studenten hebben doen verschuiven. Verder kan door de

nadruk op reflecteren het kritisch vermogen van de leraren worden gestimuleerd, waardoor zij wellicht eerder vinden dat ze te weinig reflecteren.

Een mogelijke verklaring voor het beter functioneren van de reflectief opgeleide leraren is natuurlijk dat dit wordt veroorzaakt door andere kenmerken van het opleidingsprogramma dan de gerichtheid op het leren reflecteren. Zo'n kenmerk kan bijvoorbeeld zijn dat er een sterkere selectie plaats vindt, waardoor studenten met minder capaciteiten de opleiding voortijdig verlaten. Aanleiding voor een dergelijke interpretatie zijn de resultaten die Korthagen en Verkuyl (1986) vonden. Zij toonden aan dat er een impliciete selectie plaats vindt in de op leren reflecteren gebaseerde opleiding.

Ook kunnen verschillen in kenmerken van de studenten bij het begin van hun opleiding een rol spelen bij de verklaring van het beter functioneren van de afgestudeerden van de op leren reflecteren gebaseerde opleiding.

Tenslotte kunnen nog andere variabelen een rol spelen die niet in ons onderzoek betrokken zijn, zoals contextvariabelen die invloed hebben op de professionele ontwikkeling van de leraren. Met name kan verondersteld worden, dat leraren op scholengemeenschappen zich beter kunnen ontwikkelen tot goede leraren dan leraren die op categorale LBO-scholen werken.

Het hier beschreven onderzoek moet gezien worden als een eerste bijdrage aan een empirische kennisbasis over de resultaten van een lerarenopleiding gebaseerd op de doelstelling studenten te leren reflecteren. De gegevens geven aanleiding een dergelijke opleiding positief te waarderen, maar roepen twijfel op over de vraag of het leren reflecteren wel het werkzame bestanddeel is dat tot deze positieve waardering leidt.

We zien verschillende lijnen voor onderzoek in de toekomst. De eerste lijn betreft verder onderzoek naar verbanden tussen reflectie en de kwaliteit van leraren, waarbij ook methoden gebruikt kunnen worden, zoals ontwikkeld door Nelissen (1987) en Broeckmans (1987). In dit verband roept ons onderzoek de interessante vraag op of er een verband bestaat tussen het bezitten van een reflectieve instelling en het hebben van een adequaat zelfbeeld. Een tweede lijn betreft onderzoek naar de effecten van opleidingsprogramma's. Daarbij is het belangrijk om ook het onderwijsleerproces tijdens de opleiding in de beschouwing te betrekken en om gebruik te maken van voor- en nametingen. Tenslotte zou onderzoek gedaan moeten worden naar de invloed van contextfactoren, zoals de schoolorganisatie, op de ontwikkeling van leraren. Een aanzet tot een kennisbasis op dit gebied vinden we onder meer bij Wildman en Niles (1987).

NOTEN

1. We bedanken E. Ellinger en R. van den Dool voor hun assistentie bij de verwerking van de gegevens.
2. Er blijkt geen schooltype-effect te zijn op de scores op de reflectieschalen, de schalen voor innovatiebereidheid en arbeidssatisfactie en ook niet voor de absolute verschilscore van leerlingenperceptie en leraarsperceptie van het leraarsgedrag.

REFERENTIES

- Arculo-werkgroep toelating tot de tweedefase lerarenopleiding (1985). Advies over toelating en selectie i.v.m. de tweede-fase lerarenopleiding. Utrecht.
- Arculo-werkgroep tweede fase (1985). Bouwstenen voor het curriculum van de universitaire lerarenopleiding. Utrecht.
- Brophy, J.E. & Good, T.L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (ed.), *Handbook of research on teaching* (3rd ed.). New York: McMillan, 328-375.

- Bannink, P., et al. (1985). Verslag onderzoeksfase 4 schoolstageproject. Utrecht.
- Brekelmans, M. et al. (1986). Interactioneel leraarsgedrag, curriculum en onderwijsopbrengsten. In W.J. Nijhof & E. Warries (red.). *De opbrengst van onderwijs en onderzoek* (p. 41-56). Lisse.
- Broeckmans, J. (1987). Invloeden van begeleidersgedrag tijdens nabesprekingen op reflecties van aanstaande leraren. In J.J. Peters & H.H. Tillema (red.). *Scholing in onderwijs en bedrijf* (p. 35-46). Lisse.
- Clark, C.M. (1986). Ten years of conceptual development in research on teacher thinking. In M. Ben-Peretz, R. Bromme & R. Halkes (ed.). *Advances of research on teacher thinking*. Lisse.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York.
- Creton, H.A. & Wubbels, Th. (1984). *Ordeproblemen bij beginnende leraren*. Utrecht.
- Dewey, J. (1933). *How we think: a restatement of the relation of reflective thinking to the educative process*. Chicago.
- Duffy, G. (1977). A study of teacher conceptions of reading. Paper, gepresenteerd op de National Reading Conference, New Orleans.
- Feiman, S. (1979). Technique and inquiry in teacher education: A curricular case study. *Curriculum Inquiry*, 9 (1), p. 63-79.
- Gore, J.M. (1987). Reflecting on reflective teaching. *Journal of Teacher Education*, 38 (2), p. 33-39.
- Griend, P.C. van de (1975). In het perspectief van de angst. Zutphen.
- Groeben, N. & Scheele, B. (1977). *Argumente für eine Psychologie des reflexiven Subjekts*. Darmstadt.
- Halkes, R. & Deijkers, R. (1984). Teachers' thinking criteria. In R. Halkes & J.K. Olson (eds.). *Teacher thinking: a new perspective on persisting problems in education* (p. 149-162). Lisse.
- Hook, C. & Rosenshine, B. (1979). Accuracy of teachers reports of their classroom behavior. *Review of Educational Research*, 49, p. 1-12.
- Janesick, V. (1982). Of snakes and circles: making sense of classroom group processes through a case study. *Curriculum Inquiry*, 12 (2), p. 161-189.
- Jong, J.A. de (1988). *Je wordt aan het denken gezet, ontwikkeling en evaluatie van een model voor de schoolstage van de pabo*. Utrecht.
- Jong, J.A. de, & Korthagen, F. (1988). *Handelen, reflecteren, opleiden*. Paper gepresenteerd op het VULON-congres 1988 te Beekbergen.
- Kemmis, S., & McTaggart, R. (1982). *The action research planner*. Victoria, Australië.
- Koehler, V. (1985). Research on preservice teacher education. *Journal of Teacher Education*, 36 (1), p. 23-30.
- Kolb, D.A., & Fry, R. (1975). Towards an applied theory of experiential learning. In C.L. Cooper (ed.). *Theories of group processes* (p. 33-58). London, etc.
- Korthagen, F.A.J. (1982). *Leren reflecteren als basis van de lerarenopleiding*. 's-Gravenhage.
- Korthagen, F.A.J. (1987a). Leren reflecteren anno 1986: een overzicht van de literatuur en de stand van zaken in het onderzoek naar het leren reflecteren. In L.G.M. Prick et al. (red.). *Jaarboek '86 VULON, Lopend onderzoek in de lerarenopleiding in artikelen en kronieken* (p. 11-27). De Lier.
- Korthagen, F.A.J. (1987b). Intern en extern gestuurd leren in de opleiding van onderwijsgevenden. In J.J. Peters & H.H. Tillema (red.). *Scholing in onderwijs en bedrijf* (p. 23-34). Lisse.
- Korthagen, F.A.J., & Verkuyl, H.S. (1986). Vraag en aanbod, naar een gedifferentieerde opleidingsdidactiek voor onderwijsgevenden, gebaseerd op verschillen in leerconcepties. In N.A.J. Lagerweij & Th. Wubbels (red.). *Onderwijsverbetering als opdracht* (p. 149-160). Lisse.
- Lampert, M. (1985). How do teachers manage to teach: perspectives on problems in practice. *Harvard Educational Review* 55 (2), p. 178-194.
- Larsson, S. (1984). Describing teachers' conceptions of their professional world. In R. Halkes & J.K. Olson (eds.). *Teacher thinking: a new perspective on persisting problems in education* (p. 123-133). Lisse.
- Leary, T. (1957). *Interpersonal diagnosis of personality*. New York.
- Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the structure of behavior*. New York.
- Munby, H. (1983). A qualitative study of teachers' beliefs and principles. Paper, gepresenteerd op de AERA-conferentie, Montreal, Canada.
- National Institute of Education (1975). *Theory development*. (report of panel 10, National Conference of Studies in Teaching). Washington D.C.

- Nelissen, J.M.C. (1987). Kinderen leren over wiskunde, een studie over constructie en reflectie in het basisonderwijs. Gorinchem.
- Nunnally, J. (1967). Psychometric theory, New York.
- Olson, J. (1981). Teacher influence in the classroom. *Instructional Science*, 10, p. 259-275.
- Piaget, J. (1977). Recherches sur l'abstraction réfléchissante, deel 1. Parijs.
- Schön, D.A. (1983). The reflective practitioner, how professionals think in action. New York.
- Shavelson, R.J., Webb, N.M. & Burstein, L. (1986). Measurement of teaching. In M.C. Wittrock (ed.). *Handbook of research on teaching*, third edition (p. 50-91). New York, 1986.
- Simons, P.R.J. & Vermunt, J.D.H.M. (1986). Self-regulation in knowledge acquisition: a selection of Dutch research. In G. Beukhof & P.R.J. Simons (eds.). *German and Dutch Research on learning and instruction* (p. 101-135). Den Haag.
- Skemp, R.R. (1979). Intelligence, learning and action. Chichester, etc.
- Vedder, J. (1984). Oriëntatie op het beroep van leraar. Lisse.
- Vedder, J. & Bannink, P. (1987). The development of practical skills and reflection at the beginning of teacher training. Paper ATEE-conferentie 1987 te Berlijn.
- Wildman, T.M. & Niles, J.A. (1987). Essentials of professional growth. *Educational Leadership* 44 (5), 4-10.
- Wubbels, Th., e.a. (1987a). De perceptie van de leraar-leerlingrelatie; constructie en kenmerken van een instrument. *Tijdschrift voor onderwijsresearch*, 12, p. 3-16.
- Wubbels, Th., e.a. (1987b). Teacher behavior, an important aspect of the learning environment? In B.J. Fraser (ed.). *The study of learning environments* 3 (p. 10-25). Perth: Curtin University.
- Zeichner, K.M. (1983). Alternative paradigms of teacher education. *Journal of teacher education* 34 (3), p. 3-9.
- Zeichner, K.M. (1987). Preparing reflective teachers: an overview of instructional strategies which have been employed in preservice teacher education. *International Journal of Educational Research* 11, p. 565-575.
- Zeichner, K.M. & Liston, D.P. (1987). Teaching student teachers to reflect. *Harvard Educational Review* 57 (1), p. 23-48.
- Zeichner, K.M. & Noffke, S. (1987). Action research and teacher thinking. Paper AERA-conferentie 1987, Washington D.C.

Manuscript ontvangen 7-7-1988

Definitieve versie ontvangen 1-5-1989

Effecten van tweetalig onderwijs aan Marokkaanse kleuters

E. Wagenaar en E.M. Scholte

Vakgroep Orthopedagogiek Universiteit van Amsterdam

SUMMARY

Effects of bilingual education of Moroccan preschool children.

An experimental group of 15 Moroccan preschool children in The Netherlands were taught using the children's mothertongue, the standard Moroccan-Arabic, as medium of instruction. During a small part of each day the Dutch language was also used in classroom. At the beginning of the preschool period, the children showed considerable delays in cognitive and linguistic development, both in their first and in their second language (Arabic and Dutch, respectively). Socioeconomic status of the children was low in many relevant aspects (language and educational level of parents, unemployment, housing, etc.). After two years of bilingual education, the Moroccan-Arabic language level of the children had increased up to age level while a matched sample of 15 control children remained in arrears. The cognitive development had improved, both of the treated and the control children, up to the age level. In both groups the Dutch language performance had reached the same level as other Moroccan children in The Netherlands. Implications of the findings are discussed.

1 INLEIDING

Onderwijsinstructie in de moedertaal ten behoeve van allochtone kinderen is in de laatste jaren van een nauwelijks bekend tot een hoogst actueel, en hoogst omstreden onderwerp geworden. De discussie wordt echter nogal eens vertroebeld door oneigenlijke motieven en begripsverwarring. Zo wordt onderwijsinstructie in de moedertaal vaak opgevat als een vorm van 'Onderwijs in Eigen Taal en Cultuur' (OETC), en gepresenteerd als een alternatief voor de meer gangbare, en weinig bevredigende OETC-werkwijzen. Dit is onjuist. OETC is een vak, zoals rekenen of aardrijkskunde, dat aan bepaalde groepen kinderen wordt gegeven, en aan andere groepen niet; in tegenstelling echter tot het onderwijs in andere vakken kan het OETC nog niet gebruik maken van uitgewerkte leerplannen en werkwijzen. Mede daarom is het nog steeds onderhevig aan grote onderlinge kwaliteitsverschillen en onzekerheden, en vormt het bovendien een kwetsbaar onderdeel van het curriculum in tijden van financiële krapte. Verschillende recente publicaties gaan hier nader op in (o.a. De Haan en Nalbantoglu, 1986; Haijer en Meestringa, 1988). Onderwijsinstructie in de moedertaal daarentegen is geen apart vak, maar heeft betrekking op het medium van instructie en van communicatie in de klas. De rationale voor deze vorm van onderwijs ligt in de ervaring dat allochtone kinderen in de Nederlandstalige klas vaak onvoldoende kunnen meekomen, zowel op prestatiegebied als ten aanzien van sociaalemotionele en motivationele aspecten; als één van de redenen voor dit tekortschieten van het onderwijs wordt de onvoldoende beheersing door het allochtone

kind van de Nederlandse taal gezien. Deze zou het voor het kind zowel onmogelijk maken om zich in de Nederlandse klas 'thuis' te voelen als om de Nederlandse lessen met vrucht te volgen. Daarom kiest men voor de eigen taal als instructietaal. Het kind zou hierdoor niet alleen beter Nederlands leren, maar ook zou, doordat zijn sterke kanten - zijn moedertaal - naar voren worden gehaald, zijn algehele achterstandssituatie worden gecompenseerd. Verondersteld wordt dus dat onderwijsinstructie in de moedertaal niet alleen het aanleren van Nederlandse taalvaardigheid ten goede komt, maar ook de cognitieve, linguïstische en sociaal emotionele ontwikkeling in het algemeen. Wie OETC overbodig of zelfs schadelijk acht voor het allochtone kind in Nederland heeft daarmee nog niets gezegd over het nut van onderwijsinstructie in de moedertaal, en vice versa. Over het rendement van een onderwijsmodel met de moedertaal als instructietaal is echter nog weinig bekend. Het onderzoek waarover hier wordt gerapporteerd beoogt een aanzet te geven dit nader in kaart te brengen. Hiertoe is een onderwijsexperiment met moedertaalinstructie aan allochtone kinderen opgezet. De kleuterklas vormt een geschikte plaats om met dit experiment aan te vangen. Op dat moment immers is de taalvaardigheid van de kinderen in het Nederlands het laagst, om de eenvoudige reden dat deze kinderen nog maar zo kort de tijd en de gelegenheid hebben gehad die taal te leren. Ook vanuit een ontwikkelingspsychologisch standpunt komt men tot dezelfde keuze: kinderen van vier en vijf jaar staan nog zo dicht bij het ouderlijk milieu, en nog zo weinig in de buitenwereld, dat het tot gevoelens van onzekerheid en onbegrip kan leiden wanneer men hen aanspreekt in de taal van 'de wereld' in plaats van in die van thuis (vergelijk de oudere, sociolinguïstische onderzoeken van bijv. Labov, 1970; Bereiter en Engelman, 1966; Van Calcar, 1975). Daarnaast ligt specifiek bij jonge kinderen een belangrijk argument voor het werken met onderwijsinstructie in de moedertaal in het feit dat ook de moedertaal op die leeftijd veelal nog onvoldoende ontwikkeld is, met name bij kinderen in achterstandssituaties. Wanneer niet alleen de tweede, maar ook de eerste taal voor het kind een ontoereikend medium is om in te communiceren en te leren, worden zijn kansen op succes in het onderwijs uiteraard nog ernstiger bedreigd. In dit artikel wordt verslag gedaan van enkele effecten van het hiervoor genoemde experiment. Meer specifiek zal worden ingegaan op de vraag in hoeverre onderwijsinstructie in de moedertaal een positieve invloed heeft op de cognitieve, linguïstische en sociaal emotionele ontwikkeling van Marokkaanse kleuters.

2 METHODE

2.1 Vraagstelling

De vraagstelling van het onderzoek luidt: 'in hoeverre heeft instructie in de Arabisch-Marokkaanse moedertaal een gunstige uitwerking op de cognitieve, linguïstische en sociaal emotionele ontwikkeling van (jonge) Marokkaanse kinderen?'. In het navolgende zal de aandacht met name worden gericht op de cognitieve en linguïstische ontwikkeling. Over de sociaal emotionele ontwikkeling zal elders worden gerapporteerd (Wagenaar en Scholte, 1989b).

2.2 Het tweetalige onderwijsmodel

Het moedertaalexperiment is gebaseerd op een tweetalig onderwijsmodel: in een kleuterklas met uitsluitend Marokkaanse kinderen werd gedurende de gehele kleuterperiode in de ochtend les gegeven door een Marokkaanse leerkracht die het Marokkaans-Arabisch als instructietaal hanteerde, in de middag werd les gegeven door een Nederlandse leerkracht die het Nederlands als instructietaal hanteerde. De lesvakken, de doelstellingen en de inhoud van de lessen werden niet gewijzigd. Deze volgden het reguliere curriculum

zoals voor het kleuteronderwijs in Nederland door wet en regelgeving is vastgelegd.

2.3 Onderzoeksopzet

Het effect van onderwijsinstructie in de moedertaal kan worden vastgesteld door de ontwikkeling van kinderen met moedertaalinstructie gedurende enige tijd te volgen en te vergelijken met vergelijkbare kinderen die geen onderwijsinstructie in de moedertaal hebben gehad. Bij een positieve invloed mag verwacht worden dat eerstgenoemde kinderen een gunstiger ontwikkeling door zullen maken (Borg en Gall, 1979). Beide onderzoeksgroepen kunnen worden gevormd door Marokkaanse kleuters at random te verdelen over een groep die wel (proefgroep) en een groep die geen (controlegroep) onderwijsinstructie in de moedertaal krijgt. In de veldsituatie van dit onderzoek was at random verdeling van de Marokkaanse kinderen echter niet mogelijk. Om deze reden is het onderzoek opgezet als een intact veldexperiment met nonequivalente groepen (Cook en Campbell, 1979). Dit wil zeggen dat de controlegroep werd gevormd door in het veld een aantal Marokkaanse kinderen te zoeken die maximaal vergelijkbaar zijn met de kinderen in de proefgroep. Het verschil is alleen dat de controlekinderen geen moedertaalinstructie hebben gehad. Zonder de veldsituatie geweld aan te doen kunnen aldus valide conclusies betreffende het effect van moedertaalinstructie worden getrokken. Als controlegroep werd in dit onderzoek een groep Marokkaanse kinderen genomen, die op een met de proefschool vergelijkbare basisschool in dezelfde buurt zaten. Het betreft kinderen die het regulier onderwijs voor allochtone kinderen volgen, met andere woorden Nederlands als instructietaal hebben en twee uur per week Onderwijs in Eigen Taal en Cultuur krijgen. Ter optimalisering van de vergelijkbaarheid werd de controlegroep op de demografische kenmerken sekse, leeftijd en etniciteit met de proefgroep gematched. Het effect van onderwijsinstructie in de moedertaal kan nu worden bepaald door na te gaan in hoeverre de kinderen in de proefgroep zich gunstiger ontwikkelen dan de kinderen in de controlegroep. Variantieanalyse vormt hiervoor een geschikte techniek (Cook en Campbell, 1979). Indien verschillen statistisch significant zijn ($P < .05$), wordt ervan uitgegaan dat de gevonden vooruitgang een effect is van onderwijsinstructie in de moedertaal. Bij vergelijking van het aanvangsniveau kwam naar voren dat de proefgroep een iets lager algemeen ontwikkelingsniveau had dan de Marokkaanse controlegroep, één en ander gemeten volgens de UKKI (Baarda, 1977). Met deze verschillen dient bij de effectbepaling rekening te worden gehouden. Om deze reden is op de aan de UKKI gerelateerde ontwikkelingscriteria een covariantieanalyse toegepast. De invloed van eventuele aanvangsverschillen wordt in dit geval via regressieanalyse gecompenseerd (Cook en Campbell, 1979). Variantie- en covariantieanalyse kunnen echter slechts worden toegepast wanneer criteriummaten normaal verdeeld zijn, varianties van criteriummaten homogeen zijn en meetfouten in criteriummaten en covariaten at random verdeeld zijn (Hays, 1988). Via analyse van residuen (Norusis, 1988) werd vastgesteld dat het hier gerapporteerde onderzoek in voldoende mate aan deze voorwaarden voldeed (Hays, 1988). Hiernaast werden in de covariantieanalyses gecorrigeerd voor eventuele afwijkingen in de homogeniteit van de regressiecorrecties (Cook en Campbell, 1979).

2.4 Onderzoeksgroepen

In dit artikel wordt gerapporteerd op basis van 15 Marokkaanse moedertaalproefkinderen, en 15 Marokkaanse controlekinderen. Opgemerkt moet nog worden dat de beide 'Marokkaanse' groepen elk zijn samengesteld uit 11 kinderen van Marokkaanse afkomst en 4 kinderen van Berber afkomst. Dit laatste lijkt onlogisch. De instructietaal van de proefgroep is het Arabisch-Marokkaans en deze taal is voor Berbers in principe even vreemd als het Nederlands. Dat in de proefgroep toch aan Berber sprekende kinderen

onderwezen wordt komt door het feit dat het een intact veldexperiment betreft. De kleuters waren simpelweg door de ouders voor de proefklas aangemeld, omdat deze graag willen dat hun kinderen met het Arabisch-Marokkaans vertrouwd raken. Ze vormen daarmee als het ware een knikker die in het onderhavige veldonderzoek in het spel is. Het is voor het experiment echter van belang er enig zicht op te hebben of beide groepen Marokkaanse kleuters zich anders ontwikkelen onder invloed van de moedertaalinstructie. Dit zal in de (co)variantieanalyses via interactieanalyse worden nagegaan (Hays, 1988).

2.5 Onderzoeksinstrumenten

In het onderzoek wordt aandacht besteed aan de cognitieve, de linguïstische en de socialemotionele ontwikkeling van de kinderen. Het cognitieve ontwikkelingsniveau van de kinderen werd bij het begin van het experiment gemeten met behulp van de UKKI (Baarda, 1977) en aan het eind van de kleuterperiode met behulp van de Kleuter SON (SnijdersOomen, 1975). Het betreft hier een nonverbale intelligentietest, waardoor eventuele culturele invloeden op de testsituatie worden omzeild. Beide testen geven een indicatie van het algemene cognitieve ontwikkelingsniveau in de vorm van een Intelligentie Quotient (IQ).

Het talige ontwikkelingsniveau is zowel voor het Arabisch als voor het Nederlands vastgesteld. Voor de bepaling van het Arabische taalniveau is de Reynell taalontwikkelingstest gebruikt (Reynell, 1977). Deze test brengt in beeld in hoeverre het kind de taal begrijpt en hoe het niveau van spreken is. Het Arabische taalniveau werd bepaald door de Reynelltest in het Arabisch af te nemen door een moedertaal sprekende proefleider. Bij de bepaling van het Nederlandse taalniveau werd gebruik gemaakt van de Taaltest voor Allochtone Kinderen. Deze test geeft o.a. een indruk van de passieve en actieve woordenschat die het kind ter beschikking heeft. Daarnaast bepaalt de test of het kind op adequate wijze woorden vormt en of het de functies van zinnen begrijpt. Verder wordt getoetst in hoeverre het kind de betekenis van woorden kent en de relaties tussen woorden onderling begrijpt. Deze test is zowel voor autochtone als allochtone kinderen genormeerd (Verhoeven en Vermeer, 1986). De gedragsontwikkeling van de kinderen wordt in kaart gebracht aan de hand van de Child Behavior Checklist (Achenbach en Edelbrock, 1978; Achenbach 1979) in de Nederlandse bewerking van Verhulst (1985). Over de bevindingen ten aanzien van de sociaal emotionele gedragsontwikkeling wordt elders gerapporteerd (Wagenaar en Scholte, 1989b).

Tabel 1. Vergelijking schoolleerniveau proefgroep en controle-groep bij de aanvang van het veldexperiment (methode: variantieanalyse).

Criterium	testmethode	Moedertaal Controletest		F-waarde	P-waarde
		(N=15) Gemid.	(N=15) Gemid.		
Gemiddelde testleeftijd	Burg. staat	4.5*	4.5*	0.0	.85
Algemene ontwikkeling	UKKI-IQ	82.3	89.9	3.1	.15
Taalbegrip Nederlands	Reynell-Ned.	3.0*	3.0*	0.0	.89
Taalproductie Nederlands	Reynell-Ned.	3.0	2.8	0.6	.46
Taalbegrip Arabisch	Reynell-Arab.	3.8	4.0	0.1	.71
Taalproductie Arabisch	Reynell-Arab.	3.3	3.0	0.6	.43

* uitgedrukt in leeftijd

3 RESULTATEN

3.1 De beginsituatie van de kinderen

In Tabel 1 worden de belangrijkste kenmerken van de beginsituatie van de Marokkaanse kinderen kort samengevat; voor een uitvoeriger overzicht, waarbij tevens de demografische kenmerken zijn opgenomen, wordt verwezen naar Wagenaar en Scholte (1989a).

Aan het begin van het experiment is de leeftijd van de kinderen in beide onderzoeksgroepen gemiddeld vier en een half jaar. Het cognitieve niveau van de Marokkaanse groepen verschilt onderling niet significant. Wel lijkt de relatief hoge F-waarde bij het UKKI-IQ er naar te verwijzen dat de Marokkaanse proefgroep gemiddeld een iets lager algemeen ontwikkelingsniveau heeft dan de Marokkaanse controlegroep. Het cognitieve ontwikkelingsniveau ligt bij de aanvang van het experiment zowel in de proef als in de controlegroep aanzienlijk onder de Nederlandse leeftijdsnorm. Ook wat taalontwikkeling betreft vertonen beide groepen bij het begin van het onderzoek in vergelijking met de Nederlandse leeftijdsnorm grote achterstanden. Dit betreft zowel de moedertaal als het Nederlands: een achterstand van anderhalf jaar in het begrijpen en het spreken van het Nederlands, alsmede een achterstand van een half tot anderhalf jaar in het Arabisch (resp. voor het begrijpen en het spreken). De verschillen met de leeftijdsnorm zijn (zeer) significant. Uitzondering hierop vormt het taalbegrip in het Arabisch, waar de achterstand relatief klein is.

3.2 De cognitieve ontwikkeling

In Tabel 2 staan de gegevens betreffende de cognitieve ontwikkeling van de Marokkaanse kinderen aan het einde van de kleuterperiode. Deze zijn voor de volledigheid uitgesplitst naar de Arabische en de Berberkinderen. Weergegeven zijn de leeftijdsnorm en de gemiddelden in beide onderzoeksgroepen. In de laatste twee kolommen staat per criteriummaat telkens het resultaat van (co)variantieanalyse tussen de moedertaalgroep en de controlegroep vermeld.

Tabel 2. Vergelijking algemene intelligentie ontwikkeling proef-groep en controle-groep na anderhalf jaar Moedertaal instructie (methode: covariantieanalyse).

Criterium	test-methode	groep ¹	Norm Gemid.	Moedert. Gemid.	Contr. Gemid.	F-waarden covariantieanalyse	
Gemiddelde testleeftijd		Tot.	6.0	5.9	6.0	moedertaal:	0.4
		Mar.	6.0	5.9	6.0	etniciteit:	0.1
		Ber.	6.0	6.0	5.9	interactie:	2.3
Algemene ontwikkeling	KSON Mental Age	Tot.	6.0	5.7	5.7	moedertaal:	0.1
		Mar.	6.0	5.7	5.6	etniciteit:	0.1
		Ber.	6.0	5.4	5.7	interactie:	0.4
Intelligentie niveau	KSON IQ	Tot.	100.0	97.0	97.0	moedertaal:	0.2
		Mar.	100.0	99.0	95.0	etniciteit:	0.1
		Ber.	100.0	90.0	100.0	interactie:	0.5

¹ Tot. = totale groep, N-Moedertaal=15, N-Contr.=15

Mar. = Marokkaanse subgroep, N-Moedertaal=11, N-Contr.=11

Ber. = Berber subgroep, N-Moedertaal=4, N-Contr.=4

* $p < 0.05$

Uit de tabel blijkt dat het cognitieve niveau in beide groepen na de kleuterschoolperiode niet significant meer verschilt van de norm voor de leeftijdsnorm, al ligt het niveau nog wel

iets onder het gemiddelde. Dit geldt zowel ten aanzien van de KSON-mental age als ten aanzien van het KSON-IQ. De gegevens suggereren dat de cognitieve ontwikkeling van de subgroep Berber kinderen in de moedertaalgroep iets minder gunstig is verlopen dan die van de Arabische kinderen in deze groep. Dit verschil is echter niet statistisch significant.

3.3 De taalontwikkeling in het Arabisch

In Tabel 3 worden de gegevens weergegeven met betrekking tot de ontwikkeling van de taalvaardigheid in het Arabisch.

Tabel 3. Vergelijking Arabische Taalontwikkeling van de proef-groep en de controle-groep na anderhalf jaar Moedertaal instructie (methode: covariantieanalyse).

criterium	test- methode	groep ¹	Norm Gemid.	Moedert. Gemid.	Contr. Gemid.	F-waarden (co)variantieanalyse
Taalbegrip Arabisch	Reynell	Tot.	6.0	6.0	4.9	moedertaal: 10.3*
		Mar.	6.0	6.5	5.6	etniciteit: 2.3
		Ber.	6.0*	4.6*	2.9*	interactie: 0.4
Taalproductie Arabisch	Reynell	Tot.	6.0*	4.4*	3.4*	moedertaal: 6.4*
		Mar.	6.0*	4.7*	3.9*	etniciteit: 2.9
		Ber.	6.0*	3.3*	2.2*	interactie: 0.4

¹ Tot. = totale groep, N-Moedertaal=15, N-Contr.=15

Mar. = Marokkaanse subgroep, N-Moedertaal=11, N-Contr.=11

Ber. = Berber subgroep, N-Moedertaal=4, N-Contr.=4

* $p < 0.05$

De proefgroep die moedertaalinstructie kreeg blijkt na de kleuterperiode op leeftijdsniveau te functioneren wat het begrijpen van de Arabische taal betreft. De achterstand van een half jaar die bij de aanvang van het experiment bestond is nu ingelopen. De controlegroep is nu echter meer dan een jaar achtergeraakt. (Co)variantieanalyse wijst uit dat de gunstiger ontwikkeling in de moedertaalgroep met grote waarschijnlijkheid toegeschreven kan worden aan het gehanteerde onderwijsmodel. Wanneer we de resultaten uitsplitsen naar respectievelijk de Arabische en de Berberkinderen, dan blijken er tussen deze groepen verschillen te bestaan. In de moedertaalgroep functioneren de kinderen met een Arabische achtergrond boven het leeftijdsniveau, de kinderen met een Berberachtergrond functioneren echter anderhalf jaar daaronder. In de controlegroep blijken beide subgroepen onder de leeftijdsnorm te functioneren, echter ook hier doen de Berberkinderen het minder goed dan Arabische kinderen. Eerstgenoemde groep vertoont een half jaar achterstand in het Arabisch, de laatstgenoemde groep heeft een Arabische taalachterstand van drie jaar. Met andere woorden, beide subgroepen in de proefgroep lijken te hebben geprofiteerd van onderwijsinstructie in de moedertaal, maar de Arabische kinderen aanzienlijk meer dan de Berber kinderen. Dit laatste is overigens weinig verrassend, daar het Arabisch voor de Berber kinderen geen moedertaal is.

Verder blijkt de proefgroep wat de ontwikkeling van de spreekvaardigheid in het Arabisch betreft na de onderzochte periode van onderwijsinstructie in de moedertaal ruim een jaar onder leeftijdsniveau te functioneren. De kinderen in de controlegroep hebben wat Arabische spreekvaardigheid betreft een achterstand van twee en een half jaar. Uit de (co)variantieanalyse blijkt dat het verschil tussen beide groepen significant is. De gunstiger situatie in de proefgroep kan derhalve ook hier weer met grote waarschijnlijkheid worden toegeschreven aan de onderwijsinstructie in de moedertaal. Echter, beide groepen bevinden zich wat Arabische spreekvaardigheid betreft nog steeds in een aanzienlijke

achterstandssituatie. Beide groepen functioneren op een niveau dat significant van de leeftijdsnorm verschilt. Ook hier lijkt de situatie voor de Berber kinderen minder gunstig te zijn, en ook hier weer in de proefgroep in mindere mate dan in de controlegroep.

Het voorafgaande kan als volgt worden samengevat. Zesjarige Marokkaanse kinderen die tijdens de kleuterperiode Arabisch-Marokkaanse onderwijsinstructie in de moedertaal hebben genoten begrijpen het Arabisch goed wanneer dit thuis ook gesproken wordt. Wanneer van huis uit een Berbertaal wordt gesproken zijn de kinderen in het Arabisch vooruitgegaan. Het spreken van het Arabisch daarentegen vertoont in beide (sub)groepen nog steeds een achterstand. Marokkaanse kinderen met Nederlandstalig onderwijs begrijpen na anderhalf jaar het Arabisch in redelijke mate wanneer dit thuis ook gesproken wordt. Wanneer thuis een Berbertaal wordt gesproken begrijpen zij het Arabisch praktisch niet. Zowel de Arabisch als Berbersprekende Marokkaanse kinderen met Nederlandstalig onderwijs spreken aan het einde van de kleuterperiode zeer slecht Arabisch. Deze bevindingen wijzen uit dat instructie in de moedertaal de Arabische taalontwikkeling van Marokkaanse kinderen in gunstige zin beïnvloedt en dat deze beïnvloeding groter is wanneer thuis eveneens Arabisch gesproken wordt. De vraag die vervolgens rijst is hoe het met de Nederlandse taalontwikkeling van de kinderen is gesteld.

3.4 De ontwikkeling van de taalvaardigheid in het Nederlands

Tabel 4 geeft de gegevens met betrekking tot de Nederlandse taalvaardigheid. De weergegeven scores zijn ruwe scores. In de tabel staan de volgende gegevens vermeld. De eerste twee kolommen geven de normgegevens van de Taaltest voor Allochtone kinderen

Tabel 4. Vergelijking Nederlandse Taalontwikkeling van de proefgroep en de controlegroep na anderhalf jaar Moedertaal instructie (methode: covariantieanalyse).

criterium	groep ¹	Norm Ned.	Norm Mar.	Marok. Moedert.	Marok. Contr.	F-waarden (co)variantie	
Passieve woordschat Nederlands	Tot.	51.0*	36.0	27.5	34.8	moedertaal:	2.6
	Mar.			25.2	33.9	etniciteit:	0.6
	Ber.			29.0	35.4	interactie:	0.2
Actieve woordschat Nederlands	Tot.	28.0*	16.5	14.9	13.9	moedertaal:	0.2
	Mar.			14.8	14.5	etniciteit:	0.5
	Ber.			15.2	12.5	interactie:	0.3
Woordvorming Nederlands	Tot.	22.0*	12.0	12.0	13.0	moedertaal:	0.3
	Mar.			13.0	14.0	etniciteit:	2.6
	Ber.			9.0	11.0	interactie:	0.1
Zinsbegrip I Nederlands woordfuncties	Tot.	21.5	20.0	20.1	21.7	moedertaal:	2.7
	Mar.			20.0	21.0	etniciteit:	0.5
	Ber.			18.0	22.0	interactie:	1.3
Zinsbegrip II Nederlands binnen	Tot.	23.5	21.0	21.0	20.1	moedertaal:	1.3
	Mar.			20.6	20.8	etniciteit:	0.5
	Ber.			21.4	17.0	interactie:	1.7
Zinsbegrip III Nederlands buiten	Tot.	22.5	21.5	22.1	20.6	moedertaal:	2.5
	Mar.			21.3	20.7	etniciteit:	0.2
	Ber.			22.8	20.3	interactie:	0.7

¹ Tot. = totale groep, N-Moedertaal=15, N-Contr.=15

Mar. = Marokkaanse subgroep, N-Moedertaal=11, N-Contr.=11

Ber. = Berber subgroep, N-Moedertaal=4, N-Contr.=4

* $p < 0.05$

weer, de eerste kolom die voor Nederlandse kinderen en de tweede die voor Marokkaanse kinderen. De volgende twee kolommen geven de testuitslagen in de onderzoeksgroepen weer; de derde kolom geeft de gemiddelde scores van de proefgroep en de vierde die van de controlegroep.

Uit de (co)variantieanalyses blijkt dat er geen statistisch significante verschillen bestaan tussen de proefgroep en de controlegroep wat betreft de zeven gebruikte onderdelen van de Taaltoets voor Allochtone kinderen. Beide groepen blijken aan het einde van de kleuterperiode op een overeenkomstig niveau te functioneren wat Nederlandse taal betreft. Dit niveau komt overeen met de leeftijdsnorm die voor Marokkaanse kinderen in Nederland geldt. De uitkomsten van (co)variantieanalyses naar ethnische achtergrond wijzen uit dat de Berber kinderen en de Arabische kinderen zich ongeveer gelijkelijk hebben ontwikkeld. Een uitzondering hierop vormt de subtest Woordvorming. Deze wordt zowel in de moedertaalgroep als in de controlegroep door de Berber kinderen minder goed gemaakt dan door de Arabische kinderen.

Wanneer we de Nederlandse taalprestaties van de Marokkaanse kinderen echter beoordelen volgens Nederlandse normen dan blijken zowel de proefgroep, de controlegroep als de Marokkaanse normgroep zich in een aanzienlijke achterstandpositie te bevinden. De achterstand betreft met name de actieve en passieve woordenschat in het Nederlands, alsmede de Nederlandse woordvorming. Uit de uitslagen op de subtests Zinsbegrip I, II en III blijkt dat de Marokkaanse wel in min of meer gelijke mate Nederlandse zinnen begrijpen als Nederlandse kinderen.

Samenvattend kan derhalve worden gesteld dat zowel Marokkaanse kinderen in de kleuterklassen met onderwijsinstructie in hun Arabische moedertaal als Marokkaanse kinderen in de kleuterklassen met anderhalf jaar onderwijsinstructie in het Nederlands, de Nederlandse taal in gelijke mate beheersen. In beide groepen ligt het niveau op dat van de Marokkaanse leeftijdsnorm. Naar Nederlandse maatstaven gemeten presteren beide groepen Marokkaanse kinderen echter significant onder de landelijke norm die voor kinderen van hun leeftijd is vastgesteld.

4 CONCLUSIES EN AANBEVELINGEN

Op grond van het voorafgaande kan worden geconcludeerd dat het gebruik van het Arabisch-Marokkaans als instructietaal in de kleuterklas voor Marokkaanse kinderen met een Arabische taalachtergrond in vergelijking met Marokkaanse kleuters zonder dit type onderwijs leidt tot een toename in taalvaardigheid in de moedertaal. De ontwikkeling van de Nederlandse taalvaardigheid van deze kinderen blijft op het niveau van Marokkaanse leeftijdgenoten met Nederlands als instructietaal. Met betrekking tot de cognitieve ontwikkeling blijkt dat Marokkaanse kinderen met onderwijsinstructie in de moedertaal het niet slechter, echter ook niet beter doen dan hun Marokkaanse leeftijdgenoten zonder onderwijsinstructie in de moedertaal. Zoals in de inleiding reeds werd gesteld bestaan er aanwijzingen dat goede moedertaalontwikkeling een belangrijke voorwaarde kan zijn voor de algehele ontwikkeling van het kind, zowel op cognitief en linguïstisch als op sociaal emotioneel gebied (o.a. Appel, 1984; Teunissen, 1987). Of dit inderdaad zo is, en of onderwijsinstructie in de moedertaal daarbij een zinvolle en onderwijskundig verantwoorde weg kan zijn, zal uit followup onderzoek van de Marokkaanse kinderen van dit onderzoek nog moeten blijken. Vooralsnog verwijzen de hier gepresenteerde bevindingen ernaar dat het hanteren van de moedertaal als instructietaal voor Marokkaanse kleuters niet tot achterstanden leidt, noch op cognitief gebied, noch op het gebied van de Nederlandse taalvaardigheid, terwijl wel vooruitgang wordt geboekt met betrekking tot de

taalvaardigheid in de moedertaal. Indien op termijn geen positieve effecten aantoonbaar zouden blijken te zijn op de cognitieve en sociaalemotionele ontwikkeling kan men zich evenwel afvragen of de toegenomen Arabische taalvaardigheid wel in het belang is van de allochtone kinderen en de Nederlandse samenleving.

Het lijkt verder van belang de aandacht te vestigen op een tweede bevinding die in dit onderzoek wordt vastgesteld, namelijk dat de hier onderzochte Marokkaanse kinderen bij het begin van de kleuterperiode een ontwikkelingspeil vertoonden dat overeenkomt met het niveau van een gemiddeld kind in een peuterspeelzaal. De consequenties van die bevinding kunnen van betekenis zijn voor de inrichting van het onderwijs. Immers, juist in de eerste levensjaren van een kind betekent anderhalf jaar achterstand in ontwikkeling een levensgroot verschil. Voor deze kinderen zal het onderwijs, zoals dat standaard voor de betreffende leeftijdsgroep wordt geboden, ontoereikend en inadequaaf zijn. Leerkrachten van kleuterklassen met (grote aantallen van) dergelijke kinderen zullen daar in hun onderwijs rekening mee moeten houden. Zij zullen hun onderwijs terdege moeten aanpassen en behoeven daarbij van overheidswege meer ondersteuning dan tot op heden gegeven wordt.

Tot slot zijn *een aantal waarschuwingen* bij het interpreteren van de hier gepresenteerde onderzoeksbevindingen op hun plaats. In de eerste plaats zijn de onderzochte groepen kinderen beperkt van omvang. Hierdoor dienen de conclusies onder voorbehoud te worden gezien en in verder onderzoek nog te worden bevestigd. Om deze reden hebben wij replicaties van het quasiexperiment gepland onder andere jaargroepen Marokkaanse kleuters. Over de ontwikkeling van deze kinderen zal te zijner tijd worden gerapporteerd. Verder werd het beeld hier enigszins eenzijdig op de talige en cognitieve ontwikkeling van de kinderen gericht. Over de sociaalemotionele ontwikkeling van de kinderen wordt elders gepubliceerd (Wagenaar en Scholte, 1989b). In de tweede plaats kunnen de bevindingen niet zonder meer worden gegeneraliseerd naar andere taalgroepen: een algemene ontwikkelingsachterstand van een omvang als bij de hier onderzochte kinderen behoeft zich niet in gelijke mate ook in andere ethnische groepen voor te doen. Onnodig wellicht te zeggen, dat deze achterstanden een reden tot grote zorg vormen, en specifieke aandacht voor juist de Marokkaanse groeperingen in Nederlandse scholen noodzakelijk maken. In de derde plaats dient men te bedenken, dat de onderhavige conclusies uitsluitend betrekking hebben op jonge kinderen, in de leeftijd voordat het 'formele' onderwijs begint, en niet over de hele basisschoolperiode gegeneraliseerd kunnen worden. Waar opvang in de moedertaal wellicht zinvol blijkt voor kleuters, hoeft dat voor bovenbouwleerlingen niet zonder meer ook het geval te zijn: op die leeftijd immers is de moedertaalontwikkeling ook bij kinderen in achterstandssituaties in het algemeen voldoende om het onderwijs te kunnen volgen, terwijl ook de vaardigheid in het Nederlands zodanig is toegenomen dat de schoolparticipatie goed kan zijn. Om dan nog met onderwijsinstructie in de moedertaal te gaan werken, lijkt weinig rendabel. Voor jonge, beginnende leerlingen echter ligt de situatie geheel anders. Op deze leeftijd is de binding aan het ouderlijk milieu nog veel sterker; stoornissen in de communicatie kunnen lang doorwerken op de attitude en de motivatie van het kind en zijn ouder ten aanzien van school; het aanleren van cognitieve zaken als de lees en rekenvoorwaarden tenslotte, zal nauwelijks zinvol kunnen gebeuren, zolang de elementaire taalvaardigheid niet is ontwikkeld. Het is voor allochtone kleuters dan ook niet ondenkbaar dat onderwijsinstructie in de moedertaal een de algehele ontwikkeling begunstigende factor kan zijn.

5 LITERATUUR

- Achenbach, Th.M. (1978). The child behavior profile: I. Boys aged 6-11. *Journal of Consulting and Clinical Psychology*, 46(3): 478-488.
- Achenbach, Th.M., & Edelbrock C.S. (1978). The classification of child psychopathology. *Psychological Bulletins*, 85(6): 1275-1301.
- Appel, R. (1984). Immigrant children learning Dutch, sociolinguistic and psycholinguistic aspects of secondlanguage acquisition. Dordrecht: Foris Publications.
- Baarda, D.B. (1977). Utrechtse Korte Kleuter Intelligentie test. Lisse: Swets en Zeitlinger.
- Bereiter, C., & Engelman S. (1966). Teaching disadvantaged children in the preschool. Englewood Cliffs: Prentice Hall.
- Borg, W.R., & M.D. Gall (1979). Educational Research. New York: Longman.
- Calcar, C. van, e.a. (1975). Onderwijs aan (arbeiders)kinderen: school, buurt en maatschappij. Groningen: Wolters.
- Cook, T.D., & Campbell D.I. (1979). Quasiexperimentation, Design and analysis: issues for field setting. Chicago: Rand McNally.
- Cummins, J. (1981). Biliteracy, language proficiency and educational programs. In J. Edwards (Ed.), *The social psychology of reading*. Silverspring.
- Gemeente Amsterdam (1988). Discussienota Gemeentelijk Minderhedenbeleid Onderwijs.
- Gemeente Amsterdam (1988). Onderwijs en Migranten, Verslag van een conferentie op 12 en 13 april.
- Haan, D. de, & Nalbantoglu P. (1986). Onderwijs in eigen taal en cultuur op de basisschool: ontwikkeling en perspectieven. *Moer*, 1/2: 9-30.
- Haijer, M. & Meestringa Th. (1986). Tien jaar OETC aan kleuters in Zweden. *Samenwijs*, nr. 9 en 10
- Hays, W.L. (1988) *Statistics*. New York: Holt, Rinehart & Winston.
- Labov, W. (1970). The logic of nonstandard English. J.E. Alatis (ed.), *Linguistics and the teaching of standard English to speakers of other languages and dialects*. Washington.
- Norusis, M.J. (1988), *SPSSx*. New York: McGraw Hill.
- Reynell, J.K. (1977). Reynell Developmental Language Scales, manual. Windsor: NFER.
- Snijders, J.Th., & Snijders Oomen N. (1976). Niet Verbaal Intelligentie Onderzoek. Groningen: Wolters.
- Stichting Leerplan Ontwikkeling (1987). Voorstel voor een leerplan voor onderwijs in eigen taal en cultuur in het basisonderwijs. Enschede.
- Stichting Leerplan ontwikkeling (1988). Onderwijs aan kleuters in eigen taal en cultuur. Enschede.
- Teunissen, J.M.F.B.G. (1986). Een school twee talen, een onderzoek naar de effecten van een tweetalig bicultureel onderwijsprogramma voor Marokkaanse en Turkse leerlingen op basisscholen te Enschede. Utrecht: vakgroep Onderwijskunde.
- Verhoeven, L.T., & Vermeer A. (1986). Taaltoets Allochtone Kinderen. Tilburg: Zwijsen.
- Verhulst, F.C. (1985). Mental health in Dutch children. Meppel: Krips Repro.
- Wagenaar, E. (1989). Moedertaalopvang en onderwijs voor jonge Marokkaanse kinderen, in voorbereiding.
- Wagenaar, E., & Scholte E.M. (1989a). Tweetalig onderwijs aan Marokkaanse kleuters. *Pedagogische Studiën*, augustus 1989.
- Wagenaar, E., & Scholte E.M. (1989b). Bilingual education of young Moroccan children, in voorbereiding.
- Winter, M. de (1986). Het voorspelbare kind. Vroegtijdige onderkenning van ontwikkelingsstoornissen (VTO) in wetenschappelijk en sociaalhistorisch perspectief. Lisse: Swets en Zeitlinger.
- Wolf, J.C. van der (1984). Schooluitval, een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het reguliere onderwijs. Lisse: Swets en Zeitlinger.

Manuscript ontvangen 2-11-1988

Definitieve versie ontvangen 26-5-1989

Onderwijs en Maatschappelijke Ongelijkheid, een Overzicht van 25 jaar theorievorming en onderzoek in Nederland*

Jan C.C. Rupp en Anton A. Wesselingh
Universiteit van Utrecht en Nijmegen

ABSTRACT

Education and Social Inequality. A Review of 25 Years of Dutch Research.

This paper argues that the developments of the last two and a half decades in Dutch research on education and social inequality are characterized by a loss of importance of the debates on the various dimensions of the (essentially contested) concept of inequality, which were dominant in the first decade, and by a growing interest in the development of more specific theories and in doing historical and other empirical investigations. The three conceptual debates are resp. on deficit versus difference, on intelligence: heredity or environment, and on reproduction versus emancipation. These debates are important insofar as they urge the social scientist to define clearly what the problem to be investigated is and what his or her standards are. Fruitful they became when the research was directed at cultural differences between families, differences between and within schools, bilingual education and negotiations in the classroom, related with social class, gender and ethnic background. The theories of Bourdieu at the one hand and of Boudon at the other became very influential in the course of the eighties. A strong tradition was build up in research on schoolcareers following Coleman a.o. with implications mainly in the sphere of the political-arithmetic tradition. A characteristic feature of the Dutch research climate is a total disagreement on the question if and how a form of integration of secondary education must be realized. The reviewers plea for more comparative research in the way Ringer did.

1. INLEIDING EN KORTE SAMENVATTING

Wie gevraagd wordt een balans op te maken van vijftientwintig jaar theorievorming en onderzoek in Nederland op het gebied van onderwijs en maatschappelijke ongelijkheid, is genooddaakt de Nederlandse activiteiten in een internationaal kader te plaatsen. Buitenlandse ontwikkelingen oefenen invloed uit op de wetenschappelijke activiteiten in ons land en - zij het in mindere mate - omgekeerd. Een vergelijkend overzicht maakt tevens duidelijk wat typisch Nederlands is aan het overzicht, hetzij doordat er plekken wit zijn gebleven op de Nederlandse onderzoekskaart, hetzij omdat de betreffende wetenschappelijke activiteit voornamelijk in ons land wordt aangetroffen. De Nederlandse situatie behoeft op zich weer een verklaring. De discussie ligt in ons land bij een onderwerp als ongelijkheid bijvoorbeeld, zo gevoelig dat het niet zonder meer legitiem wordt gevonden er onderzoek naar te doen; organisatie van onderzoek en van onderzoekers in verbanden als

de Vereniging voor de bevordering van het Onderzoek naar de relatie tussen Onderwijs en Maatschappelijke Ongelijkheid (OOMO) en werkgroepen van SISWO, zijn noodzakelijk om legitimatie te verkrijgen en te behouden. De verticale verdeling in 'zuilen' en de verhouding tussen de zuilen is een vraagstuk dat veel belangrijker en eigener voor de samenleving wordt gevonden dan de horizontale verdeling in rangen en standen, in seksen en in etnische groeperingen. Het begrip eenheid wordt eerder geassocieerd met de pacificatie tussen de zuilen dan die tussen de sociale lagen (cf Vos e.a., 1985).

Een review heeft altijd een beoordelend karakter. Sommige ontwikkelingen worden als positief beoordeeld, andere minder. Richtinggevend bij zo'n beoordeling is de wetenschapsopvatting van de beoordelaar. De onze sluit aan bij die van Lakatos (Chalmers, 1981). Wij gaan er van uit dat er tegelijk meerdere onderzoeksprogramma's lopen, ieder met eigen domeinvooronderstellingen (mens- en maatschappijopvattingen) en harde kern en met eigen positieve heuristieken, die globaal richting geven aan het onderzoek, onder meer op het hier besproken terrein. Een voorbeeld is de positief heuristische waarde van de metatheorie van Olson (*The logic of collective action*) voor het onderzoek van Boudon (1973) naar de klassespecifieke afweging van de kosten en baten van eventuele sociale stijging van de kinderen. Een voorbeeld uit een ander programma is de positieve heuristiek van de kapitaaltheorie van Bourdieu (1979) voor de theorie over stratificatie en sociale klassen, waarin nu niet langer alleen economische hulpbronnen, maar ook culturele worden verdisconteerd (Sixma en Ultee, 1983).

In Nederland heeft het Leidse Talentenproject (Van Heek e.a., 1986) een belangrijke rol gespeeld. Het 'Talentenproject' is de verzamelnaam voor een serie onderzoeken naar de maatschappelijke factoren die de doorstroming van geschikte leerlingen ('talenten') naar het voortgezet onderwijs beïnvloeden. Uitgangspunt was de idee dat zich in de qua status lagere sociale lagen in Nederland een 'reservoir' aan niet-gebruikt talent zou bevinden, een reservoir dat Nederland hard nodig had voor zijn industrieel-technologische ontwikkeling. Ook in andere Westerse landen werden dergelijke projecten uitgevoerd.

De discussie binnen en rondom dit project maakte duidelijk hoezeer het begrip onderwijsongelijkheid een 'essentially contested concept' (Connolly, 1983) is, dat wil zeggen een begrip waaraan meerdere, soms onderlinge tegenstrijdige dimensies te onderscheiden zijn. Die dimensies hangen samen met brede theoretische orientaties, met onderzoeksprogramma's derhalve. Ook in ons land bestaan meerdere onderzoeksprogramma's (vgl. Harbers, 1986).¹

Lange tijd stonden in het wetenschappelijk debat over onderwijsongelijkheid drie kwesties in het brandpunt van de belangstelling: 'difference-deficit'; intelligentie, erfelijkheid en omgeving; en reproductie versus reductie van ongelijkheid. Deze debatten speelden zich in de jaren zeventig grotendeels af op het vlak van de domeinvooronderstellingen van de diverse onderzoeksprogramma's en niet op theoretisch niveau. Het ging daarbij om de vraag van welke aard het probleem van onderwijs en maatschappelijke ongelijkheid nu precies is, of het eigenlijk wel een probleem is en zo ja, of het onderwijs wel iets aan de maatschappelijke ongelijkheid kan veranderen. Het is belangrijk de discussie over domeinvooronderstellingen, normen en waarden te onderscheiden van die over theorieën die er op gericht zijn een verschijnsel, een geproblematiseerd deel van de werkelijkheid te

1. Ten onrechte stelt Harbers overigens dat de leiders van het Talentenproject zich bij hun onderzoek hebben laten leiden door hun meritocratische voorkeuren. De uitkomsten van het onderzoek waren, juist tegen de verwachtingen van de onderzoekers in, immers dat er van een reserve aan talenten onder de arbeidersbevolking geen sprake was. Toen andere onderzoekers aantoonde dat er wel een reserve was, gaven de leiders van het project dit niet zonder moeite toe.

verklaren. Een discussie over mens- en maatschappijopvattingen kan ook met argumenten, op rationele wijze worden gevoerd (Habermas, 1983). Maar een discussie op het morele niveau heeft geen verklarende kracht. Een voorbeeld uit het debat over onderwijsongelijkheid kan dat verduidelijken. De meeste onderzoekers - maar niet alle - vatten gelijkheid op in termen van gelijke schoolloopbanen bij gelijke prestaties. Deze meritocratische definitie fungeert als de ideale norm waartegen de feitelijke ontwikkeling kan worden afgezet. Die norm is arbitrair, maar niet willekeurig want één uit een beperkt aantal mogelijkheden. Een andere opvatting van gelijkheid is een gelijke verdeling van schoolloopbanen over scholen van diverse typen tussen kinderen van verschillende sociale herkomst, sekse en etnische herkomst. Een derde definitie, ideale norm is één schoolloopbaan voor alle kinderen. Een dergelijke norm kan gebruikt worden als maat voor het democratisch gehalte van een land. Bij internationale vergelijkingen gaat concurrentie tussen landen een rol spelen. Voorzover het internationale gezelschap van onderzoekers er al in slaagt consensus te bereiken over een norm voor (on)gelijkheid (bijvoorbeeld in termen van sociale mobiliteit), worden de resultaten van die vergelijkingen altijd sterk aangevochten (Ganzeboom, 1983). Van dergelijke definities zijn er meerdere in omloop (Matthijssen, 1971, 1975; Meijers, 1983). Er zijn ook onderzoekers die alle tot nu toe genoemde definities liberaal vinden omdat het probleem gedefinieerd wordt binnen het bestaande schoolsysteem terwijl dit systeem zelf bekritiseerd zou moeten worden in het belang van met name kinderen uit de arbeidersklasse (Van Calcar e.a., 1968; Brands e.a., 1977), van vrouwen (Veneberg, 1982) en van etnische groepen (Mullard, 1986). Hoe men hierover ook denkt, dergelijke debatten gaan over normen en verklaren niets, omdat zij nog geen heuristische waarde hebben. Het is belangrijk waar mogelijk resultaten van onderzoek bij de discussie over normen en waarden te betrekken. Het kan veel misverstanden voorkomen. Zo hoeft gelijkheid in het onderwijs nog niet gepaard te gaan met mobiliteit in de maatschappij: in bepaalde historische perioden blijkt een grotere mobiliteit in het onderwijs zelfs ten koste te gaan van de sociale mobiliteit (Ringer, 1979, p. 29; Dronkers, 1986, p. 59). Een ander voorbeeld is de bevinding dat in het onderwijsstelsel meerdere hoofdwegen en zijwegen bestaan (op weg naar de universiteit of één van de in een land voorkomende soorten universiteiten; op weg naar een leidinggevende positie in de samenleving) en dat de meningen over wat onder kwaliteit van het onderwijs te verstaan, ook onder de diverse elites, sterk uiteenlopen. Dit pleit tegen de aanname dat in het onderwijs sprake zou zijn van één ideologie, of van een dominante ideologie (Abercrombie, e.a. 1981; *British Journal of Sociology of Education*, 1986, 2; *Oxford Review of Education* 13, 1987; Jungbluth, 1985).

Gaandeweg is het accent verschoven van discussies over mens- en maatschappijopvattingen (de domeinvooronderstellingen van de onderzoeksprogramma's) naar discussies over meer specifieke uit de heuristiek van de programma's afgeleide theorieën als mogelijke verklaringen van het ongelijkheidsprobleem. Een rol daarbij hebben ook gespeeld de teleurstellende ervaringen opgedaan met de compensatie- en activeringsprogramma's. Deze vormden voor enige van de bij de stimuleringsprojecten betrokken sociale wetenschappers aanleiding zich hernieuwd en op een meer fundamentele wijze dan voorheen bezig te gaan houden met de vraag naar de achtergronden van onderwijsongelijkheid. Het debat over 'difference-deficit' is theoretisch productief geworden enerzijds in het onderzoek naar verschillen tussen en binnen scholen, anderzijds in het onderzoek naar tweetalige opvoeding (Jungbluth, 1985; Rupp, 1988). Het debat over de intelligentie is door Vroon (1980) voorlopig afgesloten met de conclusie dat het vooralsnog niet mogelijk is de normen van de onderzoeker en de gevonden feiten uit elkaar te halen. Het reproductiedebat is theoretisch produktief geworden toen men, geïnspireerd door de interpretatieve sociologie, verklaringen voor de ongelijkheid in en door het onderwijs mede is gaan zoeken in de richting van onderhandelingen en strijd in de klas (Wesselingh, 1985b, 96/110).

Een balans over Nederlandse theorievorming en onderzoek opgemaakt, geeft dan het volgende beeld te zien. Positief is de gestage groei van het *historisch onderzoek* op het probleemgebied. Kennis van de geschiedenis van een probleem en van de pogingen dat probleem op te lossen, is onontbeerlijk voor een goed begrip ervan. Historisch onderzoek wordt in Nederland niet verricht bij gebrek aan theorie, integendeel het onderzoek rust op een theoretische basis: Van Tijn (1977) plaatst het onderwijs in een sociaaleconomische kontekst, Frijhoff (1981) in een sociaal-culturele. Positief is het historisch onderzoek ook omdat er comparatief te werk wordt gegaan, een werkwijze waarvoor Idenburg (1961, 1962, 1964, 1967 en 1971) al sterk pleitte en die hij zelf met succes heeft toegepast. De Belgisch-Nederlandse Vereniging voor de Geschiedenis van Opvoeding en Onderwijs speelt de laatste jaren een stimulerende rol.

Een tweede positieve ontwikkeling situeren wij binnen het onderzoek naar *schoolloopbanen*. Ontwikkelingen in de sfeer van de statistische methoden en van modellen, vooral in de USA, worden hier produktief gebruikt. Vooral de jongste discussie over de diverse niveaus van analyse (individueel niveau, schoolniveau, regionaal niveau, landelijk niveau) is van grote importantie. Wil dergelijk onderzoek echter niet in empiristische en ad-hoc verklaringen blijven steken, dan is het wel nodig theorieën te ontwikkelen over de aard en werking van het sociale zowel als over de aard en werking van het individuele. Produktief is het schoolloopbaanonderzoek tot nu toe vooral geweest in het pogen ontwikkelingen in de tijd aan te geven rond het probleem van onderwijsongelijkheid en mogelijke invloeden van beleidsmaatregelen te traceren.

Een derde positieve ontwikkeling valt aan te wijzen op het vlak van de *theorievorming*. De rationele keuzetheorie van Boudon en de kapitaal-theorie van Bourdieu (cultureel-, sociaal- en economisch kapitaal) zijn daarvan de belangrijkste voorbeelden. Deze theorieën hebben vele Nederlandse onderzoekers geïnspireerd. Daarnaast heeft een organisatietheorie als die van Marx (1975) vruchtbaar gewerkt op het onderzoek naar verschillen in effectiviteit tussen scholen.

Minder positieve ontwikkelingen zijn er ook te melden. Vergeleken met de gang van zaken in andere landen dient de *universiteit* in Nederland vrijwel niet tot voorwerp van onderzoek. In andere landen is de universiteit van oudsher het meest belangrijke onderwerp van onderzoek en theorievorming, ook als het gaat om vraagstukken van ongelijkheid. In Nederland domineert de opvatting dat het basisonderwijs het belangrijkste is omdat dat het onderwijs voor allen is. Men kan met evenveel recht stellen dat de universiteit in de samenleving als het belangrijkste onderwijs geldt en dat daarom - als ideaal - iedereen een universitaire studie moet kunnen volgen. Het weinige onderzoek dat tot nu toe wel is verricht naar de universiteit en naar nieuwe vormen van intellectualiteit (o.a. Frijhoff, 1981, 1983; Wendrich, 1981), verkeerde lange tijd in een marginale positie in ons land. Hopelijk kan de in 1982 opgerichte wetenschappelijke vereniging Batavia Academica daarin verandering aanbrengen, en zal deze vereniging ook het ongelijkheidsvraagstuk in zijn onderzoeksprojecten meenemen.

Een tweede minder positieve ontwikkeling is de impasse in de discussie over de vormgeving van het *onderwijsvoorrangsbeleid* en van de integratie van (de eerste fase van) het *voortgezet onderwijs*. De overmatige belangstelling voor evaluatie-onderzoek duidt op een angstig gebrek aan inhoudelijke denkbeelden. Het is een te onderzoeken probleem op zich waarom in Nederland, in tegenstelling tot bijna alle andere Westerse landen, nog steeds nauwelijks geïntegreerd voortgezet onderwijs voorkomt en de bestaande middens scholen de bestrijding van ongelijkheid niet in hun doelstellingen hebben opgenomen². De

2. Zie Rupp (1983). Tijdens een internationale conferentie over de middenschool besteedde Lagerweij (1986) geen enkele aandacht aan dit probleem, in tegenstelling tot de sprekers uit de andere landen. In de plannen van de W.R.R. voor de basisvorming was in eerste instantie bestrijding van ongelijkheid niet als doelstelling opgenomen.

V.S. begonnen met de high school reeds aan het einde van de vorige eeuw; in Engeland richtte men direct na de tweede wereldoorlog comprehensive schools op; in de Bondsrepubliek bestaan in sommige deelstaten wel Integrierte Gesamtschulen en in andere niet.

Een derde negatief punt op de balans is de stagnatie in de *theorievorming ten aanzien van de ongelijkheid naar sekse en naar land van herkomst*. De snelle, positieve ontwikkelingen in de sekse-ongelijkheid kunnen niet verklaard worden met de beschikbare theorieën (cf Jungbluth, 1982), die weliswaar er op gericht zijn de ongelijkheid te verklaren, maar niet de condities aangeven waaronder ongelijkheid zou kunnen worden verminderd en historisch verminderd is. Op het gebied van ongelijkheid en etniciteit vindt de door geen sociale wetenschapper voorziene ontwikkeling plaats dat ouders Islamitische scholen gaan oprichten. Wellicht is een veelbelovende wetenschappelijke ontwikkeling het zoeken naar per etnische groep verschillende emancipatiewegen (Koot en Rath, 1987). Voor het overige worden er voornamelijk discussies op het niveau van domeinvooronderstellingen gevoerd over de betekenis van het behoud van taal en cultuur van het land van herkomst, over de ernst van discriminatie en dergelijke.

Moet een review-artikel over onderzoek op het gebied van onderwijsongelijkheid strikt genomen alleen gaan over de voortgang in de kennis over de rol die het onderwijs speelt ter verklaring van verschillen in schoolloopbanen die samenhangen met sociaal milieu, sekse en etnische herkomst? Wij vinden van niet. Ook factoren die buiten het onderwijs liggen zijn van groot belang. Zo is dan ook inzicht nodig in de vraag wat onder 'sociaal milieu' verstaan moet worden. Anders gezegd: welke verschillen in opvoedingspraktijken hangen samen met verschillen in schoolloopbanen?³ Hetzelfde geldt voor de invloed van de regio en van het land van herkomst. Maar de rol van het onderwijs zelf blijft natuurlijk een belangrijk studieobject. Na een langdurige impasse als gevolg van de resultaten van het onderzoek van Coleman (1966) en van Jencks (1972) is de kennis over de rol van het onderwijs goed op gang gekomen. Coleman vond indertijd geen specifieke bijdrage van (verschillen tussen) scholen aan het probleem van de ongelijkheid. Hij onderzocht vooral organisatorische kenmerken van scholen. Fijnere statistische analysetechnieken maken het intussen ook mogelijk de invloeden van het sociaal milieu medierende (reproducerende en versterkende) rol van scholen in kaart te brengen. Schoolloopbanen, ongelijkheid in stand houdende of veranderende kenmerken van schoolstelsels als geheel, van ontwikkelingen in die stelsels, van verschillen tussen scholen en verschillen in scholen gemaakt tussen leerlingen van verschillende herkomst, worden langzamerhand ontdaan van het stof dat hen onzichtbaar maakte. Dat geldt ook en al eerder voor de rol van de gezinsopvoeding, van de taalontwikkeling in relatie tot de sociale omgeving en van het onderzoek naar levenskansen in onderwijs, arbeid en maatschappij in het algemeen. We zullen van de stand van zaken op deze terreinen melding maken in dit overzicht. Voor verschillen tussen onderwijsgevend en hun activiteiten en tussen leerlingen van dezelfde herkomst bestaat nog maar weinig belangstelling, althans vanuit het gezichtspunt van onderwijsongelijkheid.

2. OVER HET SCHOOLLOOPBAANONDERZOEK

Het schoolloopbaanonderzoek kent een bredere doelstelling dan het leveren van een bijdrage aan de analyse en verklaring van de ongelijkheidsrelaties in het onderwijs (al is het met dat thema wel sterk begaan). Het loopbaanonderzoek opereert over een breder front

3. Ringer (1979) wijst op de sterk uiteenlopende opvattingen over het verschijnsel 'stratificatie' in de diverse landen.

van een algemene thematiek: het verklaren van empirische verschillen in schoolloopbanen in het algemeen en niet speciaal de sekse- of milieuspecifieke verschillen tussen groepen leerlingen. Een goed voorbeeld van deze brede aanpak biedt het werk van Blok en Saris (1980). Ook 'levenskansen' hoort dus bij dit terrein van onderzoek gelet op het belang van onderwijs voor de positietoewijzing in de maatschappij. Het is dan ook niet onze bedoeling de opkomst en de bloei van het loopbaanonderzoek uitputtend te beschrijven; daarvoor kan men elders terecht⁴.

Ons interesseert de bijdrage die het schoolloopbaanonderzoek heeft geleverd aan de verheldering van de ongelijkheidsproblematiek.

In Nederland loopt de ontwikkeling van het schoolloopbanenonderzoek chronologisch gezien als volgt: na een aantal jaren van betrekkelijke stilte na het Talentenproject komt een nieuwe stroom onderzoekingen op gang. Dat onderzoek bouwde voort op de nogal aangevochten resultaten van de eerste reeks onderzoekingen uit de jaren zestig (o.a. Van Heek e.a., 1968) en op nieuwe resultaten van buitenlands onderzoek (o.a. Jencks, 1972, Boudon, 1973 en Bowles en Gintis, 1976). Van Heek dacht, zoals later bleek ten onrechte (Boon van Ostade, 1972; Van Kemenade en Kropman, 1972; Van Heek e.a., 1973), dat er van een reserve aan onbenut talent niet echt sprake was. Jencks c.s. meenden dat de rol van de school betrekkelijk secundair is als het gaat om reproductie van maatschappelijke ongelijkheid. Bowles en Gintis stellen dat de school er wel degelijk toe doet en juist bijdraagt aan het handhaven van ongelijkheid. Men blijft ook nu vooral geconcentreerd op de overgang van het lager naar het voortgezet onderwijs en de opvoedingsprocessen in de voorschoolse periode.

In 1975 verschijnt de dissertatie van Peschar, waarin school- en beroepsloopbanen van leerlingen van uiteenlopende sociale herkomst worden vergeleken, onder controle van de gemeten intelligentie o.a. De discussie die op dit interessante project volgt, wordt voor een deel in dit tijdschrift gevoerd (TOR 1975). Dronkers en De Jong (1978) introduceren de padanalyse in het loopbaanonderzoek. Hun replicatie van Jencks (1972) met Nederlandse data werkt inspirerend maar blijkt uiteindelijk toch niet te voldoen: de auteurs stappen over naar het zogenaamde Wisconsin-model (Sewell and Hauser, 1980). De hausse van padanalyse en andere regressievergelijkingen die daarna volgt wordt door Tesser (1986) beschreven, evenals de opkomst van de Lisrel-techniek en de toepassing van enkele niet-metrische technieken, waaronder log-lineaire analyse. Tessers conclusie 'dat de ontwikkeling van het schoolloopbaanonderzoek goed beschreven kan worden als een ontwikkeling van data-analysetechnieken' (1986, 11) is juist. Inderdaad is de methodische geavanceerdheid van het schoolloopbaanonderzoek geen punt van discussie, die is algemeen erkend en het loopbaanonderzoek dankt zijn erkenning en reputatie vooral daaraan. Niet in de laatste plaats doordat dank zij het invoeren van nieuwe variabelen, het uitputtend analyseren van bestaande data en het invoeren van multi-level analyses het empirisch fundament onder het ongelijkheidsonderzoek is verstevigd.

Methodisch geavanceerd en ook beleidsgericht? Voor een deel zeker. Niet voor niets plaatst Tesser (1986, p.16) het Nederlandse schoolloopbanenonderzoek in de Engelse traditie van het zogeheten 'political-arithmetic' onderzoek. Dat impliceert "een gerichtheid op het beschrijven en dokumenteren van de toestand van de samenleving in verband met centrale sociale en politieke kwesties. Een waardebepaalde keuze van onderwerpen gaat daarbij gepaard met relatief objectieve en geavanceerde methoden van data-analyse."⁵

4. Bij Dronkers (1986) bijvoorbeeld. Overigens is de huidige bloei van het schoolloopbaan-onderzoek ook niet vanzelf tot stand gekomen. Vooral dank zij het het onvermoeibaar activeren, publiceren en congresseren van enkelen is veel bereikt.

5. Tesser verwijst hierbij naar Halsey, Heath and Ridge (1980).

Een dergelijke karakterisering lijkt hoogstens voor een deel van toepassing op het Nederlandse onderzoek. We komen daar nog op terug. Dronkers, sprekend over het complexe karakter van de materie en over de moeite die het kost de ingewikkelde verbanden te analyseren: 'Deze ontrafeling van effecten is van groot praktisch belang' (1986, p.62), daarmee expliciet wijzend op de ongefundeerdheid van talloze ideeën en politieke uitspraken. Maar, Tessers omschrijving van hiervoor van het loopbaanonderzoek en de 'sociale en politieke kwesties' waarmee dat onderzoek zich bezig houdt, is slechts ten dele correct en wijkt enigszins af van Dronkers' uitspraak die meer wijst op koele distantie, het belang van 'getting the facts', het leveren van een solide empirische basis op grond waarvan (door anderen) politieke uitspraken kunnen worden gedaan en beleidsbeslissingen genomen. De boeiende analyse van Dronkers bijvoorbeeld, over veranderingen in de samenleving (met name in de economie) en de groei van de onderwijsdeelname, laat wel iets vermoeden over de maatschappelijke consequenties van een en ander maar biedt geen voorspellingen (Dronkers, 1989a). We wijzen hier ook op het onderzoek van Faasse e.a. (1986). Veel schoolloopbaanonderzoek lijkt juist sterk naar binnen gekeerd, geobsedeerd door de complexe relaties, het ontrafelen daarvan en de discussie daarover. De behoefte om nieuwe grenzen te verkennen op het terrein van de methoden en technieken, leidt dan ook tot steeds andere secundaire analyses op bestaande (gerepareerde) data van soms lang geleden afgesloten cohorten. Nieuwe dataverzameling vindt zelden plaats.

Over de stand van zaken met betrekking tot de theorievorming moeten we enkele kanttekeningen maken. Op dit terrein worden aan de ene kant enige vorderingen gemaakt vooral rond het thema sociale stratificatie (Sixma & Ultee, 1983; Luykx e.a., 1986) en ook met betrekking tot een theorie van de bijdrage van cultuur en culturele hulpbronnen in het onderwijs (Ganzeboom & De Graaf, 1983; Ganzeboom e.a., 1987; De Graaf, 1987). Ook de verworvenheden op methodisch terrein die hebben geleid tot inzichten in de problemen van multi-level analyse, betekenen een grote vooruitgang in het sociaalwetenschappelijk onderzoek op dit terrein (vgl. Kreft, 1987).

Aan de andere kant is van een theorie over het proces van de schoolloopbaan en haar determinanten toch niet echt sprake en is de oogst bepaald mager. Het pionierswerk van Boudon over de keuzemomenten die mensen in hun school- en beroepsloopbaan tegenkomen, heeft de Nederlandse onderwijssociologen (en onder hen de loopbaanonderzoekers) wel geïnspireerd maar heeft hen niet kunnen helpen een 'loopbaantheorie' te ontwikkelen. Tesser is heel duidelijk op dit punt: 'Discussie over de relatie tussen analysetechnieken en toelaatbare empirische uitspraken op basis van de onderzoeksresultaten die met die technieken worden gegenereerd, heeft inhoudelijk geleid tot een zekere bescheidenheid.' (ibid.). Zo kun je het ook zeggen. Een voorbeeld: in het schoolloopbaanonderzoek is de aandacht terecht uitgegaan naar enkele variabelen die al na het Talentenproject leken 'afgeschreven', variabelen zoals regio en richting. Maar ondanks voorwerk van bijvoorbeeld Meester en De Leeuw (1984), zijn wij er nog steeds niet in geslaagd systematische vergelijkingen te maken tussen het gewicht van de variabelen SES, sekse en regio. Is er een relatie gelegd tussen regio en 'cultureel klimaat' in de zin van het arsenaal aan middelen waarover inwoners van een bepaalde regio kunnen beschikken om hun bestaan vorm te geven (vgl. Vliegen en De Jong, 1981)? En is het loopbaanonderzoek theoretisch in staat geweest te verklaren waaraan de forse inhaal-manoeuvre van vrouwen in het onderwijs (gelet op de deelnamecijfers aan v.o. en w.o.) moet worden toegeschreven?

Ook de volwasseneneducatie, van groot belang voor de levenskansen van mensen, blijft tot nu toe buiten beeld, zoals trouwens ook enkele andere deeltijdse vormen van onderwijs. Schoolloopbaanonderzoek beperkt zich grotendeels tot het reguliere dag-onderwijs en daarbinnen vooral tot het algemeen voortgezet onderwijs. Zo is tot nu toe ook geen aandacht besteed aan bedrijfsopleidingen, om een voorbeeld te noemen van deze beperking.

De vraag is ook of het totaal ontbreken van de variabele etniciteit in de theorievorming binnen het loopbaanonderzoek tot nu toe, ook moet worden gezien als een gevolg van de deficieten van de traditionele bestanden. Voorlopig houden we het maar bij deze pragmatische veronderstelling, immers de brede en algemene doelstelling van het schoolloopbaanonderzoek wordt nogal eens doorkruist door de beperkingen van het materiaal: gedateerde en onvolledige gegevens, niet te herstellen lacunes, en - niet te vergeten - overwegingen van bescherming en privacy van individuen, zoals die door de overheid (bijv. het CBS) de laatste jaren worden gehanteerd.

Ontwikkelingen op iets langere termijn zijn moeilijk in het onderzoek te vatten, ook in het schoolloopbaanonderzoek. Dronkers wijst daar terecht op (1986, p. 120 ev.). Dat impliceert dat school- en stelselfactoren in het lopende loopbaanonderzoek maar ten dele kunnen worden betrokken. Te denken valt daarbij aan vernieuwingen in het recente verleden als de invoering van de Mammoetwet, het dalende leerlingenaantal, de verlenging van de leerplicht en voor de nabije toekomst zaken als de integratie kleuter- en lager onderwijs en de forse groei van het speciaal onderwijs. Wat de invloed van beleidsmaatregelen voorstelt in hun effecten op de schoolloopbaan, is moeilijk aan te geven. Vermoedelijk wreekt zich hier het gebrek aan internationaal vergelijkend onderzoek waarbij Nederland betrokken is en waarbij stelselfactoren in vergelijkend opzicht beschreven kunnen worden en hun specifieke betekenis voor de Nederlandse situatie duidelijk kan worden. Uitspraken voor Nederland als geheel krijgen dan meer reliëf en zijn verantwoord (vgl. Meijnen, 1988). Ook past hier tenslotte de kanttekening dat het onderzoek een enigszins 'provinciaals' karakter heeft: op enkele uitzonderingen na (Peschar, 1987; Ganzeboom e.a., 1987) houdt het loopbaanonderzoek zich vooral bezig met het analyseren van data van Nederlandse kohorten van plaatselijke of regionale oorsprong.

3. KENMERKEN VAN SCHOOLSTELSELS, SCHOLEN, ONDERWIJSGEVENDEN EN LEERLINGEN

Schoolstelsels

In de jaren zestig waren de vergelijkende studies van Idenburg (1961, 1962, 1964 en 1971) naar de schoolstelsels van diverse landen in de wereld toonaangevend. Hij betrok bij deze studies heel nadrukkelijk het ongelijkheidsvraagstuk. Idenburgs belangstelling ging vooral uit naar het voortgezet onderwijs en hij vergeleek in dat verband de high school in de V.S. met vormen van geïntegreerd voortgezet onderwijs in andere landen, waaronder Nederland met zijn mammoetwet. Bij latere onderzoekers als Bourdieu (1970, 1979) en Ringer (1979) en in ons land bij Frijhoff (1981, 1983), gaat de belangstelling primair uit naar de universiteit en naar het secundair onderwijs dat op de universiteit voorbereidt. Onderzocht worden de houding van de sociale klassen in een bepaalde historische periode tegenover de universiteit en de ontwikkelingen in deze verhoudingen door onder meer de opkomst van nieuwe sociale klassen en door technologische ontwikkelingen. Een belangrijke uitkomst van dit onderzoek is dat in ons land, evenals in Engeland, geen consensus tot stand gekomen is over de betekenis van onderwijs in het algemeen en van de universiteit in het bijzonder tussen de adel en het oude patriciaat enerzijds en de opkomende industriële bourgeoisie anderzijds. In ons land en in Engeland, hoezeer zij ook verschillen in tijdstip van aanvang en in tempo van industrialisering, heeft de industriële bourgeoisie zich afgekeerd van de universiteit en is de bedrijfsopleiding, de 'praktijk' de belangrijkste eigen leerschool geworden. Dit in tegenstelling tot een land als Frankrijk waar de diverse elites (de culturele, de economisch-administratieve en de technische) elk

hun eigen vorm van universiteit hebben gerealiseerd. Onderzoek van Beekenkamp en Dronkers (1984) is te zien als een bevestiging van deze bevindingen voor de actuele situatie in ons land⁶.

Deze per land verschillende situatie heeft ook gevolgen voor de houding van andere sociale klassen tegenover onderwijs in het algemeen en het vwo (in de V.S. binnen de high school de 'academic track') in het bijzonder. Er zijn in ons land twee zeer van elkaar verschillende manieren om - het schaarse goed van - een leidinggevende positie te bereiken, afhankelijk van het maatschappelijk veld. Eén loopt via de universiteit, de ander loopt via het bedrijf. Ringer (1979) bekijkt bij zijn onderzoek naar de historische betekenis van de universiteit en het VWO in diverse landen, tevens het democratisch gehalte, de toegankelijkheid van de universiteiten. Nederland behoort met West-Duitsland tot de landen die van oudsher een relatief laag percentage handarbeiderskinderen en vrouwen op de universiteiten kennen (Girls and Women, 1986). Een probleem dat verklaring behoeft.

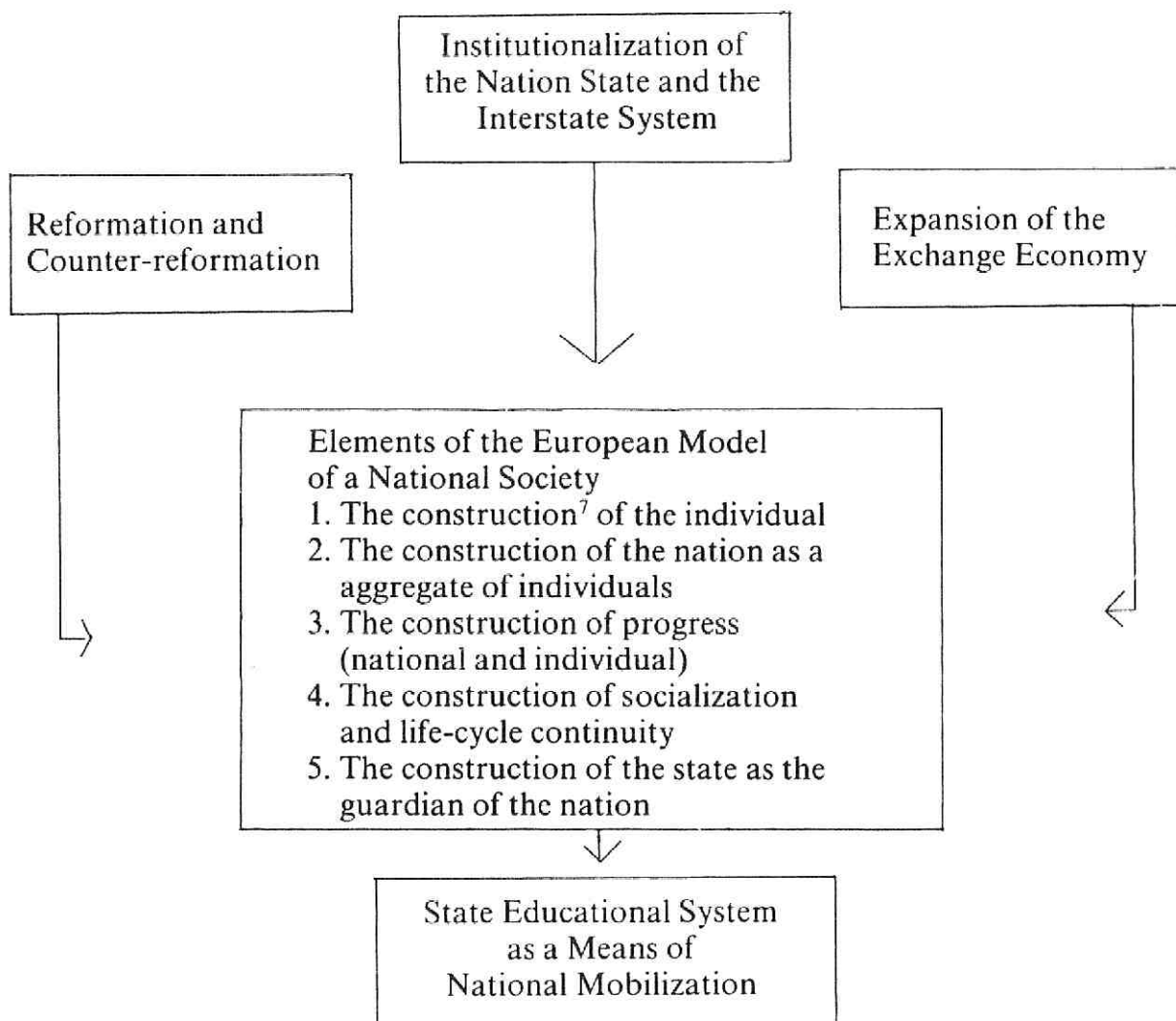
De discussie en het onderzoek naar ongelijkheid en onderwijs krijgen meer betekenis wanneer men deze gegevens legt naast het feit van het ontstaan van het massa-onderwijs en het daarmee gepaard gaande eenheidsstreven. De vraag onder welke voorwaarden massa-onderwijs in de Westerse landen is ontstaan en zich heeft uitgebreid over de gehele wereld, heeft in toenemende mate in de belangstelling gestaan. Maar zolang het volksonderwijs alleen wordt gezien als een vorm van disciplinering van de arbeidersbevolking (cf. Kruithof, 1980) en de lagere volksklassen in het algemeen, voegt het aan de discussie weinig nieuws toe. Boli, Ramirez en Meyer (1985) komen in dit verband tot de stelling dat niet de opkomende industrialisering (zoals Dronkers nog stelt, 1986, p. 42), noch de komst van stromen immigranten, noch de veronderstelde wil tot ideologische onderwerping van de arbeidersklasse door een heersende klasse, de opkomst van het massa-onderwijs kan verklaren. Het is volgens hen een samenspel tussen de volgende factoren: de opkomst van de nationale staat en het daarbij horende individu, reformatie en contra-reformatie, concurrentie tussen landen en veranderingen in het economisch stelsel. Zij zetten dat in schema (zie p. 246).

De constructie van het 'individu' en van de natie als een verzameling individuen, het individuele en nationale vooruitgangsgeloof, de denkbeelden over de opvoedbaarheid van de mens en over de staat als hoedster van de natie; al deze denkbeelden ontstaan tegen de achtergrond van de Reformatie en de Contra-reformatie enerzijds en de ontwikkeling van de kapitalistische (vrije markt)economie anderzijds. De Reformatie betekent dat de overheersende positie van één kerk - en daarmee van het samenbindend beginsel -in Europa vervalt. De geloofsverdeeldheid schept de behoefte aan een nationale staat.

Het tijdperk van de Verlichting in Europa geeft steun aan een (op rationalistische leest geschoeid) geloof in de vooruitgang van de menselijke samenleving onder de hoede van de nationale staat. Met behulp van het volksonderwijs zou het voor de staat mogelijk zijn het welzijn van de kinderen te dienen, de armoedecyclus te doorbreken waarin veel gezinnen in de loop van de achttiende eeuw terecht waren gekomen en actieve productieve staatsburgers te vormen. Onderwijs wordt van cruciaal belang, ook in de concurrentie met andere staten.

Onderzoek in ons taalgebied van Van Setten (1982) en Van Damme (1987) en ook wel van Lenders (1988) spoort met deze these. Deze theorie lijkt uiteindelijk veelbelovender dan die van Archer (1979,1982; Rupert en Wolthuis, 1985), die ten onrechte de industrialisatie als grondvoorwaarde aanwijst. Massa-onderwijs wordt historisch gezien niet als tegengesteld aan standenonderwijs opgevat. Het gaat in het onderwijs voor allen

6. De onderzoekers beweren zelf ten onrechte dat hun onderzoek kritiek inhoudt op de theorie van Bourdieu.



Uit: F.O. Ramirez and J. Boli, 1987, p. 10)

om wat - aangepast aan rang en stand - de burger moet weten om zich als burger te kunnen gedragen. Het onderwijs heeft op deze manier ook een sterk opvoedend karakter gekregen. Daar de eisen aan de voor de burger noodzakelijk geachte kennis steeds opgeschroefd worden, gaat het algemene kennisniveau van de bevolking omhoog, maar blijven de relatieve verschillen tussen de diverse sociale milieus bestaan. Historici als Boekholt (1978), De Booy (1980), Knippenberg (1985) en Boekholt en De Booy (1987) hebben een schat aan materiaal bijeengebracht over de inhoud van het massa-onderwijs en de deelname eraan in de negentiende eeuw in Nederland. Hun analysekader beperkt zich jammer genoeg tot de schoolstrijd. Het zou zeer de moeite waard zijn hun materiaal en ook materiaal uit de moderne geschiedenis te heranalyseren vanuit het referentiekader van Ramirez en Boli op het punt van de opvoeding tot staatsburgerschap in een naast in zuilen ook in standen verdeelde en als zodanig te handhaven staatsvorm.

Verschillen tussen scholen

Het onderzoek van Coleman (1966) heeft lange tijd de toon gezet. Diens conclusie dat de

7. Ramirez en Boli spreken van 'The myth of...' in plaats van 'the construction of...'. Wij prefereren 'construction' omdat dit een algemener begrip is dat ook de ideologische functie omvat.

school er niet of weinig toe doet en dat het sociaal milieu doorslaggevend is, is pas sinds de jaren tachtig genuanceerd. In navolging van onderzoek in Engeland en de USA is hier onderzoek gedaan naar de effectiviteit van scholen, onder andere door Meijnen (1984) en door Van der Wolf (1984). Het gaat hier om de vraag of scholen, bij gelijk gehouden leerlingsamenstelling, verschillen in schoolvorderingen bij leerlingen realiseren. Een vraag die ook bij de evaluatie van het onderwijsvoorrangsbeleid aan de orde komt. Bij dit onderzoek worden vooral aspecten van de organisatie van de school en van het onderwijs betrokken, afkomstig uit organisatiekundige theorieën. Jammer genoeg is er sinds Marx (1975) nog relatief weinig organisatie-theoretisch werk gedaan met betrekking tot het onderwijs. Kritiek op de onderzoeksmethode is o.a. geleverd door Kalmijn en Swanborn (1987). Zij wijzen er terecht op dat in het schooleffectiviteitsonderzoek veel te weinig aandacht wordt besteed aan de vraag naar de stabiliteit van de schoolvariabelen.

Een tweede vrij recente benadering is gebaseerd op het werk van Bourdieu. Uitgangspunt is dat ook voor scholen het marktbegin-sel geldt. Op elk niveau van het onderwijs zoekt een school zich een positie op de markt van scholen en tracht deze positie te handhaven. In ons land is deze markt voor wat betreft het voortgezet onderwijs evident (van VWO tot LBO) en binnen deze totaalmarkt bestaan ook weer deelmarkten, zoals binnen het VWO een Montessorilyceum tegenover een traditioneel lyceum (Kalmijn en Batenburg, 1986), en openbaar tegenover bijzonder onderwijs (Coleman e.a., 1982; vgl. ook Van Laarhoven e.a., 1986 en 1987, en Dronkers, 1989b). Gegeven hun marktpositie recruterende scholen leerlingen van bepaalde sociale groepen; bepaalde sociale groepen zoeken een school uit die past bij hun toekomstperspectief. Dit perspectief zal volgens Bourdieu per klasse verschillen en binnen elke klasse tussen frakties van die klasse met een meer economische dan wel een meer culturele oriëntatie (vgl. ook Bernstein, 1977; Collins, 1971). Volgens onderzoek van Rupp e.a. (1989) bestaat de marktpositie van basisscholen uit het opleidingsdomein van de school, dat zijn de scholen voor voortgezet onderwijs waarvoor de basisschool specifiek wil opleiden (de een voor VWO en HAVO; de ander voor MAVO en LBO enz.). Het opleidingsdomein is het belangrijkste aspect van wat genoemd wordt de opleidingsstatus van een basisschool. Het opleidingsdomein van een school is een zaak van voortdurende onderhandelingen tussen schoolleiding en ouders.

Verschillen binnen scholen

Scholen kunnen behalve van elkaar verschillen, ook intern verschil maken in het onderwijs dat zij kinderen geven, afhankelijk van de herkomst en de sekse van die leerlingen. Gezien de lange traditie die de USA met de High School hebben, bestaat daar ook de langste onderzoekstraditie op het gebied van milieugebonden interne differentiatie binnen de High School (Oakes, 1985; vgl. diverse artikelen in het tijdschrift *Sociology of Education*, Vol. 1986 en 1987).

De laatste jaren is dit type onderzoek weer sterk in opkomst. In ons land heeft vooral Jungbluth (1985) de aandacht gevestigd op deze kant van de onderwijsongelijkheid. Zijn stelling is dat in elke school verschil wordt gemaakt in geboden onderwijs, afhankelijk van het leerlingbeeld dat onderwijsgevers hanteren, een beeld dat weer samenhangt met sekse en sociaal milieu van de leerlingen. Het belangrijkste verschil in het leerlingbeeld is de verwachting die men heeft van wat de leerling kan presteren, resp. welke vorm van voortgezet onderwijs de leerling zal kunnen bereiken. In het basisonderwijs, in elke basisschool tekenen zich in de loop van de jaren diverse stromen af voor het voortgezet onderwijs en deze stroomverdeling is klasse- en seksegebonden. Op deze manier wordt verschil in sociaal milieu vertaald in onderwijsongelijkheid (Zijlmans en Van der Kley, 1985).

Verschillen tussen onderwijsgeevenden

Na de studie van Renes (1969) over de recrutering van onderwijsgeevenden voor het lager onderwijs is er, in tegenstelling tot wat andere landen laten zien, geen onderzoek verricht naar de sociale achtergronden van onderwijsgeevenden van welk niveau van onderwijs dan ook. De achterliggende gedachte bij dergelijk onderzoek is dat een open recrutering van docenten, in het bijzonder aan de universiteit, een teken van democratisering is; verder dat verschil in sociale achtergrond van leerling en docent de emancipatie van de leerling zal belemmeren. De laatste tijd is wel onderzoek gedaan naar het kennisniveau van aanstaande onderwijzers (Koster 1987) in het kader van een algemeen onderzoek naar de kwaliteit van het onderwijs. Dergelijk onderzoek komt in Nederland moeizaam van de grond omdat de meningen over wat goed onderwijs is hier sterk verschillen en dan met name onder de diverse elites. Een teken hiervan is de grote belangstelling in deze bevolkingsgroepen voor de Vrije scholen, Montessorischolen, de Algemeen Byzondere Schoolverenigingen en dergelijke. Een dergelijke differentiatie kent men in andere landen niet. Ook in andere landen lopen de meningen uiteen maar het beeld is daar minder gefragmenteerd. In Engeland gaat het debat grofweg gesteld tussen twee stromingen: de liberaal-rationalistische ('core-curriculum') en de 'popular culture' traditie ('really useful knowledge') (zie Veugeliers en Wendrich, 1986). In de USA gaat de discussie tussen klassieke vorming en modernisering (Bloom, 1987).

Van diverse kanten (Meijnen, 1988) is gewezen op de waarschijnlijkheid dat de verdeling van de meest bekwame docenten over de scholen ongelijk gebeurt, maar onderzocht is dit nog niet.

Opmerkelijk is de belangstelling voor de geschiedenis van het beroep van onderwijzer, in het bijzonder van de 'volksonderwijzers' en hun organisaties (DuBois-Reymond, 1981; Wolthuis, 1982; Karsten, 1986; Veld, 1987). In het voortgezet onderwijs wijzen we op onderzoek naar de invloed van de bonden op het onderwijsbeleid (Leune, 1976). Hopelijk groeit deze interesse uit tot een traditie. Het onderwijzersberoep is in ons land altijd een belangrijk middel geweest tot sociale stijging; het was de enige weg die naast het VWO toegang gaf tot de universitaire studie (in de pedagogiek). De historische studies geven een steeds nauwkeuriger beeld van de dilemma's waarvoor de 'volksonderwijzer' zich gesteld ziet. Een van die dilemma's is de vraag of werken aan de 'verheffing van het volk' niet altijd gepaard gaat met disciplinering (Lenders, 1988).

Verschillen tussen leerlingen

In veel van de gangbare theorieën en van het lopende onderzoek is het 'individu' impliciet het passieve resultaat van omgevingsinvloeden. Dat geldt niet voor de rationele keuzetheorie, waarin het individu als actor gezien wordt die voor- en nadelen van bepaalde keuzen rationeel afweegt en waarin het 'sociale' niet een entiteit is onafhankelijk van het individuele, maar het (soms onverwachte) gevolg van individuele handelingen. Deze theorie kan echter niet de bestaande normen verklaren noch de bestaande veranderingen daarin, en geeft ook geen inzicht in onderwijsstelsels en de veranderingen daarin.

Er is nog een theoretische stroming waarin aandacht wordt gevraagd voor het handelen van individuen. In die stroming wordt de ongelijkheid in en door het onderwijs opgevat als de resultante van strijd (onderhandelingen) in de klas over de definitie van de situatie. Tegenover de dominante definitie van de omstandigheden waaronder geleerd kan worden, wat belangrijke kennis is en wat een goede leerling, worden soms door leerlingen en soms door docenten andere definities gesteld (vgl. Van der Kley, 1983; Zijlmans en Van der Kley, 1985). De mogelijkheden om kollektief in verzet te komen tegen de dominante definitie worden op steeds ingenieuzere manieren verkleind, manieren die ontleend zijn aan de organisatie van de arbeid in fabriek en kantoor. Het is met name in de Amerikaanse

(Apple, 1979, 1982; Wexler, 1976, 1987; Giroux, 1983) en de Engelse onderwijssociologie (Willis, 1977) dat voor deze processen aandacht gevraagd wordt. In Nederland noemen we het onderzoek naar voortijdige schoolverlaters (Bijlsma, 1981) en naar de diverse definities van 'school' 'leren' en 'werken' zoals die onder leerlingen van LBO en MAVO worden aangetroffen (Geurts en Tesser, 1976).

4. KENMERKEN VAN GEZINNEN

In onderzoek naar de samenhang tussen sociaal milieu en schoolloopbaan wordt als maat voor 'sociaal milieu' bijna altijd het beroep van de ouder(s) - meestal van de vader - en/of het opleidingsniveau en/of het inkomen genomen. Met andere woorden: de sociaal-economische status wordt gemeten. Stratificaties verschillen sterk per land en veranderen in de loop van de geschiedenis. Zo wordt er op het ogenblik in ons land gediscussieerd over een mogelijke (nieuwe) tweedeling van onze samenleving: tussen betaald werkenden en uitkeringsgerechtigden. De consequenties van een dergelijke tweedeling voor het onderwijs zijn nog onvoldoende helder (Bosman en Louwes, 1988). Behalve misschien in het geval van samenhang tussen de hoogte van het inkomen en gevolgd, niet-kosteloos onderwijs, verklaart een gevonden samenhang nog niets, want waarom heeft de opleiding of het beroep van de ouders invloed? In ons land heeft deze vraag zich altijd in de belangstelling van de onderzoekers mogen verheugen. Theoretisch en empirisch is door een reeks van onderzoekers (o.a. Vervoort, 1968; Rupp, 1969; Claeys, 1972; Meijnen, 1977; De Graaf, 1987), gezocht naar verschillen in opvoedingspatronen en -praktijken tussen de diverse sociale milieus, die mede verschillen in schoolloopbanen kunnen helpen verklaren. Onderzoek is gedaan naar de volgende aspecten (zie ook het overzicht bij Meijnen, 1984):

- a. de betekenis van de sociaal-economische omstandigheden, resp. hulpbronnen van de ouders;

- b. de betekenis van de sociaal-culturele omstandigheden van het gezin;

Vervoort spreekt in dit verband van gezinsculturele factoren, Rupp van het cultuur-pedagogisch patroon van de opvoeding, Tesser noemt het intellectueel thuisklimaat en De Graaf spreekt in navolging van Bourdieu van culturele hulpbronnen;

- c. verschillen in normen- en waardenpatronen; meer of minder prestatiegericht; een conformistische versus een positionele rolstructuur; meer of minder conformistisch op sociaal gebied; hoger of lager zelfbeeld. De algemene stelling van de onderzoekers komt er op neer dat een succesvolle schoolloopbaan een 'moderner' gezinsleven vraagt en dat moderne gezinnen vooral in de midden- en hogere klassen worden aangetroffen (vgl. Van Setten, 1987).

Geleidelijk aan is er een duidelijk consistent beeld ontstaan over de vraag welke opvoedingspatronen de samenhang tussen sociaal milieu en schoolloopbaan kunnen helpen verklaren. Het zijn hoofdzakelijk de culturele opvoedingspatronen door welke het milieu invloed uitoefent. De financiële hulpbronnen zijn in betekenis afgenomen. Bourdieu heeft laten zien dat per sociaal milieu onderscheid gemaakt moet worden tussen meer cultureel georiënteerde en meer economisch georiënteerde gezinnen. Deze verschillende orientaties leiden tot verschillende schoolkeuzen en tot andere opleidingsroutes. De opleiding van de ouders lijkt ook belangrijker te worden dan het beroep. Ook dit wijst in de richting van de onder b. genoemde factoren.

Het betreft dan de volgende activiteiten in en rond het gezin:

- a. de culturele participatie van de ouders: deelname aan concerten, museumbezoek, leesgewoonten enzovoort, activiteiten die belangrijker zijn voor de schoolloopbaan van de kinderen naar de mate waarin zij een avant-gardistischer karakter dragen: het

- luisteren naar een uitvoering van een strijkkwartet van Bartok is belangrijker dan het luisteren naar een pianoconcert van Beethoven enz.
- b. het inwijden van de kinderen in de sfeer van het culturele erfgoed: hen voorlezen en later bekend maken met de literatuur; hen meenemen naar concerten en musea.
- c. het direkt ondersteunen van de schoolloopbaan door het bezoeken van ouderavonden, het tonen van belangstelling voor het schoolwerk, zorgen voor de begeleiding van huiswerk zonodig enz.

Naast bovengenoemd onderzoeksprogramma dat de verklaring zoekt in hulpbronnen, is er het programma van de 'rational choice'. Boudon (1973) heeft dit uitgewerkt voor het probleemgebied onderwijsongelijkheid. Hij is voor de belangrijkste beslismomenten in de schoolloopbaan van kinderen nagegaan hoe elke ouder een afweging maakt van baten en kosten bij elk van de mogelijke keuzen en ook hoe deze afweging beïnvloed wordt door het sociaal milieu van de ouders. De potentiële baten van een keuze voor een VWO-opleiding kunnen variëren met de vraag naar academisch opgeleiden op de arbeidsmarkt. Daarbij kan ook prestige sterk meetellen, speciaal bij de economisch rijken zonder veel cultureel kapitaal. Kosten kunnen zijn de economische kosten of direkte kosten van de studie (hoewel daar een beurs tegenover kan staan), maar ook de indirecte kosten, de inkomsten die ouders derven doordat hun kind langer op school blijft. Maar er zijn ook geheel andere kosten. Boudon noemt hier de kosten van het arbeiderskind dat een school voor VWO zou gaan bezoeken en daarmee van zijn vriendjes en mogelijk ook van zijn ouders zou kunnen vervreemden (vgl. de kosten-baten analyse van sociale mobiliteit door Goldthorpe, 1980; Dronkers e.a., 1982). De kosten liggen hier op psychologisch vlak. Op grond van dergelijke afwegingen kunnen twee VWO-leerlingen met verschillende milieu-achtergrond verschillende onderwijsroutes gaan volgen, ook al verschillen zij niet in schoolprestaties (vgl. Peschar, 1975).

De kracht van Boudons theorie is dat zij preciese beweegredenen onderzoekt en keuzen kan verklaren. Het Nederlandse onderzoek à la Boudon heeft interessante gegevens opgeleverd, ook op individueel niveau.

5. MAATSCHAPPELIJKE ONGELIJKHEID, ONDERWIJS EN TAAL

In de beginperiode van het ongelijkheidsonderzoek in het onderwijs lag een van de accenten op taalontwikkeling en taalgebruik van (jonge) kinderen. De beheersing van de standaardtaal of de taal van de dominante groepen in de samenleving werd geacht een *conditio sine qua non* te zijn voor het leveren van goede schoolprestaties en daarmee voor het bereiken van een hoog opleidingsniveau. De sociale omgeving, het sociale milieu waarin men als kind opgroeit bepaalt voor een groot deel over welk taalgebruiksreservoir men uiteindelijk beschikt en daarmee de aansluiting naar de fase van de secundaire socialisatie in het onderwijs.

Onder invloed van de sociolinguïstiek (Appel e.a., 1976; Lenders e.a., 1978) die zich vooral in de jaren zestig sterk heeft ontwikkeld en zich uitgebreid heeft beziggehouden met de relatie tussen taal(gebruik) en sociale klasse, is veel aandacht besteed aan dit aspect van de relatie tussen onderwijs en maatschappelijke ongelijkheid (cf ook Huls, 1982). De theorie van Basil Bernstein (1971) is in dit verband van belang. Hij stelde dat de sociale omgeving leidt tot het al of niet beheersen van de standaardtaal en het ontwikkelen van vermogens (codes) om een bepaald taalgebruik te hanteren. Kinderen uit de hogere klassen zouden dergelijke vermogens in sterker mate 'van huis' meekrijgen dan kinderen uit de arbeidersklasse. De discussie die - voor en na Bernstein - de gemoederen bezig hield, ging

eigenlijk over de vraag of men zou moeten uitgaan van één standaardtaal en afwijkingen daarvan (deficietopvatting) of dat er in feite meerdere talen op gelijke voet naast elkaar bestaan (differentieopvatting). Bernstein zelf is er - blijkens tegenspraken in zijn eigen teksten - niet helemaal uitgekomen (vgl. Jackson, 1974; Atkinson, 1987), maar deze tegenstrijdigheden zijn inhaerent aan het multidimensionele concept onderwijsongelijkheid.

De definitie van het probleem waar het nu eigenlijk om ging, verschoof in de loop der jaren. Tot in de jaren zeventig speelde het probleem eigenlijk alleen maar voor 'inheemse' groepen met van het standaard Nederlands afwijkende talen of dialecten: de 'anderstaligen' bevonden zich binnen de grenzen van de ene nationale Nederlandse cultuur (Fries, Kerkraads enz., vgl. Wijnstra, 1976). Toen echter ook in ons land de immigratiegolven voor een multiculturele samenleving gingen zorgen, veranderde de situatie. Onze tot dan toe betrekkelijk eenvoudige oplossing van het probleem van de meertaligheid (extra lessen Fries bijvoorbeeld) werkt nu niet meer: andere oplossingen moeten worden gevonden: multicultureel of intercultureel onderwijs; onderwijs in eigen taal en cultuur. De zaak is aanzienlijk gecompliceerder geworden. Taalverwerving van leerlingen van etnische minderheden (o.a. het probleem van de tweetaligheid) is nu een belangrijk topic geworden. Scholen krijgen steeds meer te maken met leerlingen van verschillende etnische herkomst, vooral in de grote steden en worden zo multi-etnisch van samenstelling.

Over de aanpak van het probleem van de etnische groepen en hun taalontwikkeling dreigen evenals ten tijde van de sociolinguïstiek uit de jaren zestig en zeventig, twee kampen te ontstaan die op essentiële punten van elkaar verschillen. Het voert te ver die verschillen hier uitgebreid te vermelden. We volstaan met op te merken dat het lijkt alsof de scheidslijn loopt tussen diegenen die met enige nadruk toch acculturatie - zeg aanpassing - voorstaan aan de Nederlandse cultuur, opgevat als mono-cultuur ook in het belang van de bepaald niet tijdelijke immigrantenstatus (De Jong 1987 bijv.) en diegenen die de problematiek zien in termen van etnische verhoudingen in de Nederlandse samenleving (Leeman, 1985).

Sociolinguïsten houden zich nu uitgebreid bezig met het probleem van de tweetaligheid. De reeks proefschriften van de laatste jaren, waarin dat probleem wordt behandeld, (o.a. Appel, 1984, J. Teunissen, 1987) is aanzienlijk en de resultaten wijzen grotendeels in één richting: tweetaligheid zou de beste aanpak zijn voor de taalverwerving, ook van het Nederlands, van immigrantenkinderen. In deze kringen en in de publieke discussie vinden we dan ook voor- en tegenstanders van tweetaligheid (de methode van de 'onderdompeling'), maar deze standpunten blijven betrekkelijk 'technisch' zolang niet de kwestie van de monocultuur of het multiculturele karakter van de samenleving wordt gesteld inclusief de problemen rond de etnische verhoudingen die daarmee gepaard gaan. Het probleem vraagt derhalve om een zekere mate van 'sociologisering'. Een veelbelovende nieuwe stroming (Koot en Rath, 1987) gaat uit van culturele verschillen tussen etnische groepen met betrekking tot het belang dat de groep hecht aan onderwijs als middel tot emancipatie. Sommige etnische groepen hebben een traditie zich in een nieuwe samenleving via onderwijs op te werken, andere kiezen een weg in de politiek, weer anderen bouwen een zaak op. We wijzen hier op een recente ontwikkeling: de oprichting van Islamitische scholen.

6. ONDERWIJS EN LEVENSKANSEN (LIFE-CHANCES)

Sinds Ph.J. Idenburg in 1958 de term 'sleutelmacht der school' reserveerde voor wat later de allocatiefunctie van het onderwijs is gaan heten, heeft deze toeleveringstaak van het

onderwijs vrijwel onafgebroken in de belangstelling gestaan⁸. Zeker in het (onderwijs) sociologisch onderzoek bestaat op dit terrein een zekere traditie, die voortkomt uit het onderkende belang van de school en van scholing voor het bereiken van maatschappelijke posities, in het bijzonder de hogere. Onderwijs is van belang bij het toedelen van 'Lebenschancen' (Weber, 1956): leden van verschillende sociale klassen hebben ongelijke toegang tot schaarse middelen en zullen als gevolg daarvan een verschillende maatschappelijke positie innemen. Onze maatschappij is daarbij zo ingericht dat het onderwijs een 'bijna exclusieve rol heeft verkregen bij de verdeling van materiële en immateriële goederen en posities (...)', aldus Dronkers (1986, p.112) in een echo van Schelsky (1957, p.18). Het onderwijs vormt daarmee een belangrijke schakel in de continuering van maatschappelijke verschillen over generaties heen, zonder overigens die verschillen zelf te genereren. Het is kortom geen systeemkenmerk of onderwijskundig gevolg van de pedagogische structuur van het onderwijsbestel, maar een gevolg van de inrichting van de samenleving als geheel en de plaats van het onderwijsstelsel daarbinnen.

Dergelijke uitspraken over de rol van de school in de maatschappij zijn de laatste decennia zo zeer gemeengoed geworden in de onderwijssociologie dat de broodnodige relativisering meestal achterwege blijft. Toch wordt de rol van het onderwijs op die manier nogal overschat: de school heeft op dit punt bepaald geen monopoliepositie, want andere socialiserende instanties leveren hun aandeel (gezin, peergroup, media enz.). Bovendien heeft het onderwijs niet voor elke maatschappelijke groep dezelfde plaats en betekenis. De geschiedenis laat zien bijvoorbeeld dat economische elites een zeer selectief gebruik maken van de beschikbare onderwijsvoorzieningen. Dat betekent dat *de rol van het onderwijs bij de vorming van elites in de maatschappij* een belangrijk thema in dit verband kan zijn. In Nederland is dit een onderontwikkeld thema vergeleken met de situatie in het buitenland. Bourdieu (1964, 1984) is opnieuw het meest sprekende voorbeeld, maar ook in Engeland bestaat voor dit thema ruime aandacht vooral waar het gaat om de rol van de traditionele elite-scholen in dat land (vgl. *The Education of Elites*, 1977). De sleutelmacht van de school en de monopoliepositie van dit instituut dienen derhalve gerelativeerd: dat verschilt naar tijd en plaats voor uiteenlopende maatschappelijke groepen.

In Nederland bestaat een duidelijke traditie op het terrein van onderzoek en theorievorming rond de relatie tussen *onderwijs en arbeid(smarkt)*, een van de belangrijkste aspecten van het vraagstuk 'onderwijs en levenskansen'⁹. Op dit terrein is vooral werk verricht door Vervoort (1981 oa.) en door Van Hoof en Dronkers (1980; vgl. ook Van Hoof, 1987). Het werk van Van Hoof en Dronkers heeft voor een deel ook zijn sporen achter gelaten in het hiervoor besproken loopbaanonderzoek. De bedoelde studie werd verricht in het kader van het Nationaal Programma Arbeidsmarkt Onderzoek, waarbij verschillende universitaire instellingen en departementen waren betrokken. In dit verband moet ook het onderzoek van Diederix (1983 o.a.) worden genoemd, een rapportage van de 3e fase van het longitudinale project school- en beroepsloopbanen 'Van jaar tot Jaar' (zie ook Diederix, 1989). Maar ook buiten de kringen van loopbaanonderzoekers is er ontwikkeling te constateren op dit terrein. Het theoretisch werk van Vervoort (zie hiervoor) geïnspireerd door de Engelse socioloog Antony Giddens en ondersteund door Wesselingh (1985a), bevat een pleidooi de reproductietheorie theoretisch te verbreden in een meer Weberiaanse richting en uit te breiden naar de relatie onderwijsarbeid. Dat

8. Dat lijkt nu vanzelfsprekend maar dat was het toen allerminst. 'Het schoolwezen als economisch verschijnsel' (de titel van Idenburgs oratie aan de Universiteit van Amsterdam in 1967) is lange tijd verwaarloosd in de onderwijswetenschappen. Het thema is nog steeds marginaal ten opzichte van de traditionele onderwijskundige thema's.

9. Dronkers wijst in zijn overzichtsartikel ook nog op de 'huwelijksmarkt' en de relatie tussen bereikt opleidingsniveau en de kansen op deze markt (1986, 113 ev.).

betekent een verdieping van de 'sleutelmacht' van de school en een versterking van de rol van de arbeidsmarkt, omdat daar enkele vitale toewijzingsprocessen plaats vinden, die de functies van de school in die zin verhelderen dat naar de achtergronden van de 'sleutelmacht' wordt gekeken. Soortgelijke opvattingen zijn neergelegd in de rapportage van de analyse van de Volkstelling 1971, voorzover het de onderwijsgegevens betreft (Vliegen en De Jong, 1981). In deze lijn werken ook de sociologen Glebbeek en Mensen (1986). Hun eveneens Neo-Weberiaanse oriëntatie lijkt vooral geïnspireerd door Parkin (1971). Werk in deze lijn kan het herstel bevorderen van de relatie onderwijs, arbeid en stratificatie c.q. ongelijkheid. De Stichting voor Onderzoek van het Onderwijs (SVO) tenslotte, heeft prioriteit gegeven aan de uitbouw van een programmalijn Onderwijs en Arbeidssituatie. In een publikatie van Hövels en Peschar (1985) wordt de draad van het in 1980 afgesloten Nationaal Programma (zie hiervoor) gedeeltelijk weer opgepakt en een programma voor onderzoek geschetst.

Naast 'onderwijs en arbeid' als een thema binnen het aandachtsveld onderwijs en levenskansen, is er het thema *onderwijs en deelname aan de cultuur*. Dit thema is sterk verbonden met het werk van de Franse (cultuur)socioloog Pierre Bourdieu. Eerder in dit overzichtsartikel is al aangegeven welk soort onderzoek in Nederland in deze lijn is ontwikkeld. In ons land bestond op dit terrein trouwens een zekere traditie door de publikaties van het Sociaal en Cultureel Planbureau met beschrijvende gegevens over deelname aan culturele activiteiten in Nederland, vaak met commentaar van sociologen. Vermeldenswaard is tenslotte in dit verband het onderzoek van Gadourek naar de relatie tussen onderwijs en levensstijl (1986).

Een vierde thema dat resorteert onder 'levenskansen' is het thema *onderwijs en gezondheid*. Ook hier is de aansluiting met de problematiek van de maatschappelijke ongelijkheid van belang. Maatschappelijke ongelijkheid kent verschillende dimensies (bezit, macht, kennis enz.) die vaak met elkaar samenhangen en zo leiden tot duurzame verschillen tussen sociale groepen in de maatschappijstructuur. Die verschillen zijn bovendien van relatief permanente aard doordat generatiegewijze steeds dezelfde sociale klassen bevoordeeld respectievelijk benadeeld worden. Dat is in het kort de theorie van de maatschappelijke reproductie. Tot voor kort bleef 'gezondheid' meestal buiten de discussie in de onderwijswetenschappen. Ten onrechte. Gezondheid en het bezit van een dergelijk schaars goed is evenzeer - en misschien nog wel meer - bepalend voor de levenskansen die mensen (soms letterlijk) geboden worden. We weten nog onvoldoende over deze verschillen in realiseringsmogelijkheden op het terrein van ziekte en gezondheid. We weten ook nog onvoldoende hoe belangrijk mensen deze zaken achten, hoeveel waarde aan gezondheid wordt toegekend in vergelijking met andere maatschappelijke goederen en diensten (vgl. Suurmeyer, 1986)¹⁰. Suurmeyer (o.c.) geeft hiervan een eerste uitwerking, maar het zal duidelijk zijn dat het hier een nog praktisch onontgonnen terrein betreft, zeker de verbinding tussen onderwijs en gezondheid.

Bij de onderzoekslijn van 'onderwijs en levenskansen' past de kanttekening dat theoretisch-verklarend in het Nederlandse onderzoek nog maar weinig werk is verricht; het meeste onderzoek is beschrijvend van karakter en opzet. Over het terrein van onderwijs en levenskansen zijn, dat zal na het voorgaande duidelijk zijn, ook nauwelijks absolute uitspraken te doen. Het belang van het onderwijs op dit punt is evident, maar is tijd- en plaatsgebonden en verschilt voor diverse maatschappelijke groepen en samenlevingen. Voor onze tijd en voor onze samenleving geldt dat op de sleutelmacht van de school weliswaar het nodige valt af te dingen maar dat anderzijds de vitale positie van onderwijs als instantie die in deze maatschappij posities toewijst grotendeels intact blijft.

10. Er is bijvoorbeeld geen onderzoek gedaan naar de mogelijke relatie tussen de kwaliteit van medici (en klinieken), de sociale herkomst van hun cliënteel en de levensduur.

LITERATUUR

- Abercrombie, N., Hill, S. & Turner, B.S. (1980). *The dominant ideology thesis*. London.
- Appel, R. (1984). *Immigrant Children Learning Dutch. Sociolinguistic and Psycholinguistic Aspects of Second-language Acquisition*. Amsterdam (diss.).
- Appel, R., Hubers, G. & Meijer, G. (1976). *Sociolinguistiek*. Utrecht.
- Appel, R. & Wesselingh, A.A. (1987). Sociale verschillen in het onderwijs. In: P. Span et al. (red.), *Onderwijzen en Leren* (pp. 77-100). Groningen.
- Apple, M.W. (1979). *Ideology and curriculum*. London.
- Apple, M.W. (1982). *Education and power*. London.
- Archer, M.S. (1979). *The social origins of educational systems*. London.
- Archer, M.S. (Ed.) (1982). *The sociology of educational expansion*. London.
- Atkinson, P. (1987). *Language, Structure and Reproduction. An Introduction to the Sociology of Basil Bernstein*. London.
- Beekenkamp, G.G. & Dronkers, J. (1984). De plaats van het onderwijs in de recrutering van president-directeuren. In J. Dronkers & F.N. Stokman (red.), *Nederlandse élites in beeld. Recrutering, samenhang en verandering* (pp. 85-100). Deventer.
- Bernstein, B. (1971). *Class, Codes and Control, Vol. 1. Theoretical Studies towards a Sociology of Language*. London.
- Bernstein, B. (1977). *Class, Codes and Control, Vol. 3. Towards a Theory of Educational Transmissions (revised edition)*. London.
- Blok, H. & Saris, W.E. (1980). Relevante variabelen bij het doorverwijzen na de lagere school, een structureel onderzoek. *Tijdschrift voor Onderwijsresearch*, 5, 2, 63-79.
- Bloom, A. (1987). *The closing of the American mind*. New York.
- Boekholt, P.Th.F.M. (1978). *Het lager onderwijs in Gelderland 1795-1858*. Zutphen.
- Boekholt, P.Th.F.M. & De Booy, E.P. (1987). *Geschiedenis van de school in Nederland*. Assen.
- Bois-Reymond, M. du (1981). *Onderwijzersleven*. Nijmegen.
- Boli, J., Ramirez, F.O. & Meyer, J.W. (1985). Explaining the origins and expansion of mass education. *Comparative Education Review*, 29, 145-169.
- Boon van Ostade, A.H. (1972). Het milieu en de overgang van het basis- naar het voortgezet onderwijs in Nederland. *Sociologische Gids*, 19, 219-228.
- Booy, E.P. de (1980). *Kweekhoven der wijsheid. Basis en vervolg onderwijs in de steden van de provincie Utrecht van 1580 tot het begin der 19e eeuw*. Utrecht.
- Bosman, R. & Louwes, W. (1988). Eenouder- en tweecoudergezinnen en schoolloopbanen. *Mens en Maatschappij*, 63, 1, 5-23.
- Boudon, R. (1973). *L'inégalité des chances. La mobilité sociale dans les sociétés industrielles*. Paris.
- Bourdieu, P. (1979). *La distinction. Critique sociale du jugement*. Paris.
- Bourdieu, P. (1984). Espace sociale et genèse de 'classe'. *Actes de la recherche en sciences sociales*, 52/53.
- Bourdieu, P. & Passeron, J.Cl. (1964). *Les Héritiers. Les Etudiants et la Culture*. Paris.
- Bourdieu, P. & Passeron, J.Cl. (1970). *La reproduction. Eléments pour une théorie du système d'enseignement*. Paris.
- Bowles, S. and Gintis, H. (1976). *Schooling in Capitalist America. Educational Reform and the Contradictions of Economic Life*. New York.
- Brands, J., Egas, G., Karsten, J. & Wendrich, E. (1977). *Andere wijs over onderwijs*. Nijmegen.
- British Journal of Sociology of Education* (1986), 7, 2. Special Number 'Whatever happened to inequality?'
- Bijlsma, K.M. et al. (1981). Leerlingculturen: verzet tegen een etiket. In *Jeugdcultuur en onderwijs*. Amsterdam.
- Calcar, C. van, Soutendijk, S. & Tellegen, B. (1968). *School, milieu, prestatie*. Enschede.
- Chalmers, A. (1981). *Wat heet wetenschap?* Meppel.
- Clacys, P.C. (1972). *Gezin en schoolsucces*. Antwerpen.
- Coleman, J.S. & Campbell, E.S. (1966). *Equality of educational opportunity*. Washington.
- Coleman, J.S., Hoffer, T. & Kilgore, S. (1982). *High school achievement; public catholic and private schools compared*. New York.

- Collins, R. (1971). Functional and conflict theories of educational stratification. *American Sociological Review*, 36, 1002-1119.
- Connolly, W.E. (1983). *The terms of political discourse*. New York.
- Damme, D. van (1987). Contradicties van alfabetisering. *Comenius*, 25, 3-26.
- Diederén, J. (1983). *De keuze van een beroep. Van Jaar tot Jaar derde fase*. Nijmegen.
- Diederén, J. (1989). *Meer of minder gelijk. Over sexe-verschillen in schoolloopbaan en beroepskeuze en over de gedeeltelijke konditionering van die sexe-verschillen door sociaal milieu en schoolprestatie*. Nijmegen.
- Dronkers, J. (1986). Onderwijs en sociale ongelijkheid. In J.A. van Kemenade et al. (red.) *Onderwijs: bestel en beleid*. Deel 2A, pp.42-151. Groningen.
- Dronkers, J. (1989a). Veranderingen in samenleving en economie en de groei van de onderwijsdeelnemers sinds 1960. *Comenius*, 9, 1, 47-72.
- Dronkers, J. (1989b). Schoolkenmerken en individuele factoren. In P. Vogel et al. (red.) *Onderwijs: keuzen en kansen*. Muiderberg (ter perse).
- Dronkers, J. & de Jong, U. (1978). Jencks en Fägelind op zijn Hollands. Een aanzet tot de studie van de relaties tussen milieu, intelligentie, onderwijs, beroep en inkomen. *Sociologische Gids*, 25, 1, 4-30.
- Dronkers, J., de Graaf, P. de Haan, D. Peek, J. en Swanborn, P. (1982). Veranderingen in ongelijke onderwijskansen. *Comenius*, 5, 97-113.
- The Education of Elites* (1977). Block V, Culture and Class. Open University Course E 202: Schooling and Society. Milton Keynes.
- Faasse, J.H., Bakker, B., Dronkers, J. & H. Schijf (1986). Voor en na de Mammoetwet. *Mens en Maatschappij*, 61, 75-94.
- Frijhoff, W.Th.M. (1981). *La société néerlandaise et ses gradués, 1575-1814*. Amsterdam.
- Frijhoff, W.Th.M. (1983). Van onderwijs naar opvoedend onderwijs. Ontwikkelingslijnen van opvoeding en onderwijs in Noord-Nederland in de achttiende eeuw. In *Werkgroep Achttiende Eeuw, Onderwijs en opvoeding in de achttiende eeuw* (pp. 3-40). Amsterdam.
- Gadourek, I. (1986). Verschillen in levensstijl bij de voornaamste geledingen in Nederland. In Th. Mensen & G.W. Meijnen (red.) *Sociale Stratificatie en Onderwijs* (pp. 63-74). Lisse.
- Ganzeboom, H. & de Graaf, P. (1983). Beroepsmobiliteit tussen generaties in Nederland in 1954 en 1977. *Mens en Maatschappij*, 58, 28-52.
- Ganzeboom, H. et al. (1987). Intergenerationele klassenmobiliteit in Nederland tussen 1970 en 1985. *Mens en Maatschappij*, 62, 1, 17-43.
- Geurts, J. & Tesser, P. (1976). *Werkende jongeren en hun onderwijs*. Nijmegen.
- Giroux, H. (1983). *Theory and resistance in education*. London.
- Girls and women in education. A cross-national study of sex inequalities in upbringing and in schools and colleges* (1986). Paris (OECD).
- Glebbeeck, A. & Mensen, Th. (1986). Het onderwijs tussen meritokratie en reproductie. In Th. Mensen & G.W. Meijnen (red.) *Sociale Stratificatie en Onderwijs*. (pp. 111-132). Lisse.
- Goldthorpe, J.A. (1980). *Social mobility and class structure in modern Britain*. Oxford.
- Graaf, P.M. de (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen*. Nijmegen.
- Habermas, J. (1983). *Moralbewusstsein und kommunikatives Handeln*. Frankfurt.
- Halsey, A.H., Heath, A. & Ridge, J.M. (1980). *Origins and Destinations. Family, Class and Education in modern Britain*. Oxford.
- Harbers, J. (1986). *Sociale wetenschappen en hun speelruimte*. Groningen.
- Heek, F. van, et al. (1968). *Het Verborgen Talent. Milieu, Schoolkeuze en Schoolgeschiktheid*. Meppel.
- Heek, F. van, Vervoort, C.E. & Weeren, P. van (1973). Nogmaals: Verborgten talenten? *Sociologische Gids*, 4, 309-316.
- Hoof, J.J. van (1987). *De arbeidsmarkt als arena*. Amsterdam.
- Hoof, J.J. van & Dronkers, J. (1980). *Onderwijs en Arbeidsmarkt*. Deventer.
- Hövels B. & Peschar, J.L. (1985). *Tussen wal en schip. Een onderzoeksprogramma op het terrein van onderwijs en arbeids-situaties*. Nijmegen.
- Huls, H.A. (1982). *Taalgebruik in het gezin en sociale ongelijkheid. Een interaktioneel-sociolinguïstisch onderzoek*. Nijmegen.

- Idenburg, Ph.J. (1961). Onderwijs en welvaart. *Economisch-Statistische Berichten*, mei, 472-478.
- Idenburg, Ph.J. (1962). *Het ideaal van de optimale ontwikkeling der talenten en de pedagogische structuur van het schoolwezen*. Groningen.
- Idenburg, Ph.J. (1964). *Schets van het Nederlandse schoolwezen*. Groningen.
- Idenburg, Ph.J. (1967). *Het schoolwezen als economisch verschijnsel*. Amsterdam.
- Idenburg, Ph.J. (1971). *Theorie van het onderwijsbeleid*. Groningen.
- Jackson, L.A. (1974). The Myth of Elaborated and Restricted Code. *Higher Education Review* (Spring 1974), pp.65-81.
- Jencks, Chr. et al. (1972). *Inequality. A reassessment of the effect of family and schooling in America*. New York.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs*. Rotterdam.
- Jungbluth, P. (1982). *Docenten over onderwijs aan meisjes: 'positieve discriminatie' met een dubbele bodem*. Nijmegen.
- Jungbluth, P. (1985). *Verborgene differentiatie*. Nijmegen.
- Kalmijn, M. & Batenburg, R. (1986). Reproductie van cultureel en economisch kapitaal op een traditioneel en een Montessorilyceum. *Tijdschrift voor Onderwijsresearch*, 11, pp. 149-163.
- Kalmijn, M. & Swanborn, P.G. (1986). Van der Wolf's onderzoek naar schooluitval. Verslag van een secundaire analyse. *Pedagogische Studien*, 63, 435-444.
- Karsten, S. (1986). *Op het breukvlak van opvoeding en politiek. Een studie naar socialistische volksonderwijzers rond de eeuwwisseling*. Amsterdam.
- Kemenade, J.A. van & Kropman, J.A. (1972). Verborgene talenten? Kritische kanttekeningen bij een onjuiste interpretatie. *Sociologische Gids*, 19, 219-228.
- Kley, P. van der (1983). *Zeg na jij! Over regels, patronen, beurten, selectie en reproductie in het lager onderwijs*. Purmerend.
- Knippenberg, H. (1986). *Deelname aan het lager onderwijs in Nederland in de negentiende eeuw*. Amsterdam.
- Koot, W. & Rath, J. (1987). *Ethnicity and emancipation. International Migration*, 25/4, 426-440.
- Koster, K.B. (1987). Reken- en wiskunde-onderwijs in de basisschool. In J.J. van Kuyk (red.). *Basisvorming in de basisschool*. (pp.-44-52). Tilburg.
- Kreft, G.G. (1987). *Models and methods for the measurement of school effects*. Amsterdam.
- Kruithof, B. (1980). De deugdzame natie. *Symposion*, 2, 22-37.
- Laarhoven, P. van, Bakker, B., Dronkers, J. & Schijf, H. (1986). Some aspects of school careers in public and non-public primary schools. *Tijdschrift voor Onderwijsresearch*, 11, 83-96.
- Laarhoven, P. van, Bakker, B., Dronkers, J. & Schijf, H. (1987). Richting van de school en schoolloopbanen in het voortgezet onderwijs. *Tijdschrift voor Onderwijsresearch*, 12, 23-40.
- Lagerweij, N.A.J. (1986). De dramatische strijd om de Nederlandse middenschool. In W. Wielemans (red.) *Vernieuwing in het secundair onderwijs. Een Europese situering*. (pp. 63-80). Leuven.
- Leeman, Y. (1985). Onderwijs en etnische verhoudingen. In J.L. Peschar & A.A. Wesselingh (red.). *Onderwijssociologie, een inleiding* (pp. 417-435). Groningen.
- Lenders, J. (1988). *De burger en de volksschool. Culturele en mentale achtergronden van een onderwijshervorming in Nederland 1780-1850*. Nijmegen.
- Lenders, J., Mazeland, H. & Nieuwstadt, M. van (1978). *Taal, arbeidersklasse en ongelijkheid. Een kritiek op de theoretische en praktische veronderstellingen van de sociolinguïstiek*. Amsterdam.
- Leune, J.M.G. (1976). *Onderwijsbeleid onder druk. Een historisch-sociologisch onderzoek naar het opereren van lerarenverenigingen in het Nederlandse onderwijsbestel*. Groningen.
- Luijckx, R.A., Doorne-Huiskes, A. van & Ultee, W.C. (1986). Het verband tussen onderwijs en beroepshoogte voor vrouwen en mannen in Nederland in 1979. *Mens en Maatschappij*, 61, 21-33.
- Marx, E.C.H. (1975). *De organisatie van scholengemeenschappen in onderwijskundige optiek*. Groningen.
- Matthijssen, M.A.J.M. (1971). *Klasseonderwijs. Sociologie van het onderwijs*. Deventer.
- Matthijssen, M.A.J.M. (1975). Sociale ongelijkheid van onderwijs kansen; vijf interpretaties van sociale werkelijkheid. *Sociologische Gids*, 6, 427-445.

- Meester, A.C. & De Leeuw, J. (1983). *Intelligentie, sociaal milieu en de schoolloopbaan*. Leiden.
- Meijers, F.J.M. (1983). *Van Ambachtsschool tot LTS*. Nijmegen.
- Meijnen, G.W. (1977). *Maatschappelijke achtergronden van intellectuele ontwikkeling*. Groningen.
- Meijnen, G.W. (1984). *Van zes tot twaalf*. Harlingen.
- Meijnen, G.W. (1988). Het rendement van cultureel kapitaal (Oratie Univ. van Amsterdam). *Comenius*, 31, 8, 323-342.
- Mullard, C. (1986). Pluralisme, etnicisme en ideologie: implicaties voor een transformatieve pedagogiek. *Comenius*, 6, 4, 407-431.
- Oakes, J. (1985). *Keeping Track. How schools structure inequality*. New Haven.
- Olson, M. (1971). *The logic of collective action*. Cambridge Mass.
- The Oxford Review of Education* (1987), 13. Special Issue, 'Plowden: history and prospect'.
- Parkin, F. (1971). *Class inequality and political order*. London.
- Peschar, J.L. (1975). *Milieu, school, beroep. Een achterafexperiment over de periode 1958-1973*. Groningen.
- Peschar, J.L. (1987). *Zo vader-zo zoon; zo moeder-zo dochter? Vergelijkende analyse naar de processen van statusverwerving en onderwijsmobiliteit in Nederland, Hongarije en Polen*. Lisse.
- Ramirez, F.O. & Boli, J. (1987). The political construction of mass schooling. *Sociology of Education*, 60, 1, 2-17.
- Renes, P.B. (1969). *Recruterings en selectie van leerkrachten voor het lager onderwijs in Nederland*. Groningen.
- Ringer, F. (1979). *Education in modern society*. London.
- Rupert, H.G.A. & Wolthuis, J. (1985). Ontstaan en ontwikkeling van het Nederlandse onderwijssysteem. In J.L. Peschar & A.A. Wesselingh (red.) *Onderwijssociologie, een inleiding* (pp. 279-306). Groningen.
- Rupp, J.C.C. (1969). *Opvoeding tot schoolweerbaarheid*. Groningen.
- Rupp, J.C.C. (1983). Ins and outs van de Middenschoolevaluatie-discussie. *Onderwijs en Opvoeding*, 35, 3, 150-154.
- Rupp, J.C.C. (1988). Het pedagogisch optimisme van de jaren zeventig. Terugblik op het tijdperk van de compensatie- en activeringsprogramma's, de voorlopers van het onderwijsvoorrangsbeleid. *Comenius*, 29, 90-106.
- Rupp, J.C.C., Lange, R. de & Gernler, T. (1989). De markt van basisscholen. Het belang van de opleidingsstatus van basisscholen voor de schoolkeuze. In P. Vogel e.a. (red.), *Onderwijs: Keuzen en Kansen*. Muiderberg (ter perse).
- Schelsky, H. (1957). *Schule und Erziehung in der industriellen Gesellschaft*. Würzburg.
- Setten, H. van (1982). Opvoedend onderwijs. De vernieuwing van het Nederlandse volksonderwijs in het begin van de vorige eeuw. *Comenius*, 2, 1, 5-36.
- Setten, H. van (1987). *In de schoot van het gezin. Opvoeding in de Nederlandse gezinnen in de twintigste eeuw*. Nijmegen.
- Sewell, W.H. & Hauser, R.M. (1980). The Wisconsin longitudinal study of social and psychological factors in aspirations and achievement. In A.C. Kerckhoff (Ed.) *Research in Sociology of Education and Socialization*, Vol. I. Greenwich (Conn.).
- Sixma, H. & Ultee, W.C. (1983). Een beroepsprestigeschaal voor Nederland in de jaren tachtig. *Mens en Maatschappij*, 58, pp. 360-382.
- Suurmeijer, Th. (1986). Sociale stratificatie en gezondheid. In Th. Mensen & G.W. Meijnen (red.). *Sociale Stratificatie en Onderwijs* (pp. 49-62). Lisse.
- Swaan, A. de (1988). *In Care of the State*. Oxford.
- Tesser, P. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen.
- Teunissen, J.M.F. (1986). *Eén school, twee talen: een onderzoek naar de effecten van een tweetalig intercultureel onderwijsprogramma voor Marokkaanse en Turkse leerlingen op basisscholen te Enschede*. Utrecht.
- Tijn, Th. van (1977). Achtergronden van de ontwikkeling van het lager onderwijs en van de schoolstrijd in Nederland, 1862-1905. Een poging tot sociaal-historische analyse. In L. Box, J. Dronkers, M. Molenaar & J. de Mulder (red.), *Vrijheid van onderwijs* (pp. 25-36). Nijmegen.
- Veld, Th. (1987). *Volksonderwijs en leerplicht. Een historisch sociologisch onderzoek naar het ontstaan van de Nederlandse leerplicht 1860-1900*. Delft.

- Veneberg, A. (1982). De overleving van 'vrouwelijke' onderwijsinhouden. *Comenius*, 2, 1, 37-57.
- Vervoort, C.E. (1968). *Gezin en schoolkeuze bij handarbeiders*. Leiden.
- Vervoort, C.E. (1981). Interventiestaat, onderwijs en ongelijkheid. In P. v.d. Kley et al. (red.). *Interventiestaat en Ongelijkheid* (pp. 117-134). Nijmegen.
- Veugeliers, W. en Wendrich, E. (1986). Neokonservatisme, democratisering en pedagogiek. Recente Engelstalige publicaties over onderwijspolitiek. *Comenius*, 23, 360-375.
- Vliegen J.M. & Jong, U. de (1981). *Onderwijs in Nederland*. Monografieën Volkstelling 1971, nr. 12. Den Haag.
- Vos, J.F., Koning, P. de & Blom, S. (1985). *Onderwijs op de tweesprong*. Den Haag.
- Vroon, P. (1980). *Intelligentie. Over het meten van een mythe en de politieke, sociale en onderwijskundige gevolgen*. Baarn.
- Weber, M. (1956). *Wirtschaft und Gesellschaft*. Zweiter Halbband. Tübingen.
- Wendrich, E. (1981). De pedagogiese conceptie van Antonio Gramsci. *Comenius*, 1, 101-117.
- Wesselingh, A.A. (1985a). *Onderwijs en reproductie van maatschappelijke ongelijkheid 1975-1985*. Nijmegen.
- Wesselingh, A.A. (1985b). Onderwijs tussen reproductie en emancipatie. In J.L. Peschar & A.A. Wesselingh (red.). *Onderwijssociologie, een inleiding* (pp. 343-371). Groningen.
- Wexler, Ph. (1976). *The sociology of education; beyond inequality*. Indianapolis.
- Wexler, Ph. (1987). *Social analysis of education. After the new sociology*. London.
- Willis, P. E. (1977). *Learning to labour. How working class kids get working class jobs*. Westmead.
- Wolf, J.C. van der (1984). *Schooluitval; een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het regulier onderwijs*. Lisse.
- Wolthuis, J. (1982). De onderwijsvernieuwing in Groningen van 1800 tot 1820 en de rol daarin van de onderwijzers en hun gezelschappen. *Info*, 13, 272-315.
- Wijnstra, J.M. (1976). *Het onderwijs aan van huis uit Friestalige kinderen: verslag van een evaluatie-onderzoek in een meertalige regio*. Den Haag.
- Zijlmans, W.C. & Kley, P. van de (1985). Een sociologie van de schoolklas. In J.L. Peschar en A.A. Wesselingh (red.). *Onderwijssociologie, een inleiding* (pp. 207-231). Groningen.

Manuscript ontvangen 28-11-1988

Definitieve versie ontvangen 3-5-1989.

Notities en Commentaren

A NONLINEAR CORRECTION ON LINEAR EQUATING

Dato N.M. de Gruijter

University of Leiden, Educational Research Center

ABSTRACT

Linear equating is bound to fail at the extreme score levels. Several proposals for correction have been suggested in the literature. Here a new correction is proposed, based on logits.

INTRODUCTION

Let X and Y be two test forms which must be equated. In linear equating to the score scale of Y a linear conversion equation

$$y^* = ax + b \quad (1)$$

is obtained. Linear equating is bound to fail at the extremes: negative scores y^* or scores y^* exceeding the maximum score on test Y , y_m , are possible.

The problems at the extremes frequently are solved ad hoc, for example with the 'dogleg' (Braun & Holland, 1982). Recently Divgi (1988) proposed another solution in which a cubic is used which coincides with the linear function for $x = \bar{x}$, and which passes through the points $(-.5, -.5)$ and $(x_m + .5, y_m + .5)$.

Actually, two cubic functions - one for $y^* = f(x)$ and another for $x^* = g(y)$ - are obtained and the two conversion lines are averaged in order to preserve the symmetry property of equating.

Divgi also proposed to use the cubic as an alternative to equipercentile equating in cases where linear equating is inadequate. A disadvantage of equipercentile equating is its inaccuracy, unless some smoothing is done (Kolen, 1984). However, in our opinion, Divgi's approximation method is too simple for really nonlinear cases. We will restrict ourselves to the proposal to approximate linear equating nonlinearly in order to satisfy the constraints at the extremes.

The purpose of this note is to suggest an alternative to the cubic.

THE LOGIT METHOD

In the method we propose, it is assumed that the relation between X and Y is more or less linear in logits

$$x' = \ln[x/(x_m - x)] \quad (2a)$$

and

$$y' = \ln[y/(y_m - y)]. \quad (2b)$$

Address: Boerhaavelaan 2, 2334 EN Leiden

We obtain the relation

$$y'^* = a'x' + b' \quad (3)$$

which can be written in terms of the original scales as

$$y^* = y_m / \{1 + d [x^{-1} (x_m - x)]^e\} \quad (4)$$

($x > 0$). Equation 4 has two advantages over the cubic. First, the relation is simpler. Secondly, Equation 4 defines a family of conversion functions. One of the consequences is that the symmetry property of equating is retained.

The coefficients d and e in Equation 4 have to be chosen in such a way that the function coincides with the linear conversion line for a value of, say, x_c . This value might be equal to the average on x , $\bar{x}$, when the equating design has two random examples of examinees, one of which obtains test X and the other test Y . In other cases x_c might be the estimated mean for the synthetic population (Braun & Holland, 1982). In other words, the nonlinear conversion line should go through the point (x_c, y_c) with $y_c = ax_c + b$, and its slope at this point should be equal to a . The resulting values of d and e are

$$d = y_c^{-1} (y_m - y_c) [x_c (x_m - x_c)^{-1}]^e \quad (5)$$

and

$$e = ax_m^{-1} x_c (x_m - x_c) y_m y_c^{-1} (y_m - y_c)^{-1} \quad (6)$$

REFERENCES

- Braun, H.I. & Holland, P.W. (1982). Observed-score test equating: a mathematical analysis of some ETS equating procedures. In P.W. Holland & D.B. Rubin (Eds.): *Test equating*. New York: Academic Press.
- Divgi, D.R. (1988). *Constrained cubic: combining the desirable features of linear and equipercentile equating*. Paper presented at the annual meeting of the AERA, New Orleans, April 5-9.
- Kolen, M.J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9, 25-44.

Ontvangen 25-11-1988

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. 13, nr. 4

Oktober 1989

Inhoudsopgave

Convocatie ledenvergadering VOR	262
Congressen, Symposia, Cursussen, Trainingen	263
Ontvangen publikaties	264

Convocatie ledenvergadering VOR

De ledenvergadering van de VOR zal worden gehouden op 30 november om 14.00 uur in de Godebaldzaal van het Stiltecentrum, Hoog Cathelijne Utrecht. Op de agenda staan enkele voor de leden belangrijke punten. Leest u vooral de toelichting bij de agendapunten 5 en 6. Het bestuur rekent op een grote opkomst.

Agenda ledenvergadering 30 november

1. Opening/mededelingen.
2. Verslag vorige ledenvergadering (zie VOR-bulletin no. 3 juni '89)
3. Begroting 1989 (op te vragen bij de penningmeester, dr. E. Roede. Tel. 020-5550325).
4. Bestuurssamenstelling voor 1990.
Aftredende bestuursleden: Prof. dr. J. Claessen.
Dr. E. Roede.
Door het bestuur worden als nieuw lid voorgedragen.
Dr. P. Wijnen.
Drs. K. de Glopper.
Tegenkandidaten kunnen tijdens de vergadering worden ingebracht.
5. VOR-gedragscode (zie toelichting; de volledige code ligt ter kopiëring bij de instituten c.q. vakgroepen of kan worden opgevraagd bij de secretaris, dr. M. P. C. v.d. Werf, tel. 050-636657).
6. Relatie VOR-OTG's-SVO (zie toelichting).
7. ORD-bundels.
8. International Journal for Education Research (IJER).
9. W.v.t.t.k.
10. Sluiting.

Toelichting bij agendapunt 5: VOR-gedragscode

In het najaar van 1987 heeft het VOR-bestuur een commissie ingesteld ten behoeve van een beroepscode voor onderwijsonderzoekers. De aanleiding daartoe werd gevormd door problemen op het RION en mogelijk ook op andere instituten ontstaan bij in uitvoering zijnde onderzoeken en de daarbij gehanteerde onderzoeksmethoden.

De werkzaamheden van deze commissie, bestaande uit twee VOR-bestuursleden een vertegenwoordiger van de onderzoeksinstituten, en een waarnemer van de SVO hebben in oktober 1988 geleid tot een concept VOR-code. Na verwerking van het commentaar van het VOR-bestuur heeft de commissie de conceptcode voor becommentariëring rondgestuurd naar de instituten, vakgroepen en anderen die mogelijk met de code geconfronteerd zouden kunnen worden. De reacties op de conceptcode waren zeer uiteenlopend. De commissie heeft deze reacties verwerkt hetgeen heeft geleid tot de versie die nu aan de leden wordt voorgelegd.

Het bestuur is van plan de code, na eventuele verwerking van de reacties van de leden, bindend te verklaren voor de VOR-leden. In eerste instantie zal het daarbij gaan om een gedragscode en niet om een beroepscode met centrale registraties en externe erkenning. Wel zal er een kamer van Toezicht moeten komen.

Aangezien de gedragscode dus belangrijke consequenties kan hebben voor de VOR-leden verzoekt het bestuur u de code serieus te bekijken en uw reactie kenbaar te maken.

Toelichting bij agendapunt 6

Zoals u waarschijnlijk bekend is heeft de SVO besloten de gelden die voorheen via de

Onderzoeks Themagroepen (OTG's) besteed werden op een andere wijze te kanaliseren. In kringen van onderwijsonderzoekers heeft dit besluit de nodige reacties opgeroepen. Het VOR-bestuur heeft uitvoerig met de SVO overlegd om de gemeenschap van onderwijsonderzoekers actief te betrekken bij de nieuwe opzet. Aan de SVO is het volgende voorgesteld:

- de voorzitters van respectievelijk het VOR-bestuur, de OTG's en OMO vormen een werkverband waarlangs de participatie van de gemeenschap van onderwijsonderzoekers gestructureerd wordt;
- dit werkverband fungeert voor de SVO als aanspreekpunt;
- de SVO draagt zorg voor de continuering van de (bescheiden) subsidie aan OTG's en OMO, zodat de bestaande infrastructuur (nieuwsbrief, porti, repro, vergaderkosten e.d.) in stand kan worden gehouden;
- nadat de thema's voor fundamenteel strategisch onderzoek door het SVO-bestuur zijn vastgesteld, wordt het werkverband door de SVO uitgenodigd om suggesties te doen voor personen/instellingen die in aanmerking zouden komen om de thema's uit te werken tot een samenhangend pakket van onderzoeksschetsen. Het zou daarbij aanbeveling verdienen wanneer meer dan één thema tegelijkertijd zou worden aanbesteed;
- de uiteindelijke keuze is zaak voor SVO;
- het samenhangend pakket van onderzoeksschetsen wordt van advies voorzien door het wetenschappelijk forum via een door het werkverband te initiëren procedure;
- SVO beslist in laatste instantie en gaat over tot aanbesteding van de concrete onderzoeksprojecten;
- in de uitvoeringsfase van de concrete projecten kan via het werkverband een forum van geïnteresseerde onderzoekers georganiseerd worden dat de functie van wetenschappelijk forum kan vervullen.

Het VOR-bestuur heeft de verwachting dat zich via deze opzet, waarbij een aantal positieve punten uit de OTG-structuur behouden blijven, in de komende periode zal uitkristalliseren hoe de participatie van onderwijsonderzoekers verder vorm kan krijgen. Op dit moment onderzoekt het bestuur in hoeverre de VOR en OTG's een divisiestructuur vergelijkbaar aan die van de AERA kan gaan vormen. Op de ledenvergadering zal hierover nadere informatie kunnen worden verstrekt.

Congressen, Symposia, Cursussen, Trainingen

STUDIEDAG PRAKTIJKGERICHT ONDERZOEK

Het Amsterdams Pedologisch Centrum (A.P.C.) organiseert op 17 november 1989 een studiedag over praktijkgericht onderzoek.

Wie onderzoek in de praktijk wil doen, wordt ermee geconfronteerd niet alleen rekening te moeten houden met voorwaarden van methodologische aard. Vele andere factoren, gerelateerd aan het specifieke praktijkveld waar het onderzoek wordt verricht of van waaruit de onderzoeks-vraagstelling is geformuleerd, moeten in het onderzoeksplan worden betrokken. Daarbij stelt het veelal omvangrijke materiaal dat wordt verzameld (observatieverslagen, mondelinge interviews) zijn eigen eisen ten aanzien van verwerking en analyse.

Op de studiedag zullen onderzoekers hun ervaringen met praktijkgericht onderzoek toelichten en de strategieën die zij hebben ontwikkeld demonstreren.

De kosten van de studiedag, inclusief lunch, bedragen f 100,-. Voor studenten (HBO/WO) is de prijs f 60,-. Tevens ontvangen de deelnemers enkele weken voor de studiedag het boek 'Praktijkgericht onderzoek. Debat, design, data.', (uitgever Acco, Amersfoort). Voor informatie of aanmelding kunt u zich richten tot het secretariaat van het Orthopedagogisch Instituut, IJsbaanpad 9, 1076 CV Amsterdam, tel. 020-664 43 31.

Ontvangen publikaties

- Bakker, B.F.M., Dronkers, J. & Meijnen, G.W. (1989). Educational opportunities in the welfare state. Longitudinal studies in educational and occupational attainment in the Netherlands. Nijmegen: ITS (OOMO-reeks).
- Gheysens, L., Janssens, D., Vanhecke, L. & Vanthuyne, L. (1989). Analytische meetkunde. Amersfoort/Leuven: ACCO. f 30,75.
- Graaff, E. de (1989). Simulation of initial medical problem-solving. Studies on the development of a new measure for the assessment of medical problem-solving ability. Haarlem: Thesis.
- Klarus, R. & Dool, P. van den (1989). Ontwerpen van leerprocessen. Ervaringsleren en cultuurhistorische leerpsychologie binnen vorming en onderwijs. Baarn: Anthos.
- Klerk, L.F.W. de, Schouten, A. & Sanden, J.M.M. van der (1989). Het leren van praktische vaardigheden. Aanzetten tot een didactiek van technisch onderwijs. Amersfoort/Leuven: ACCO. f 25,-.
- Kuhlemeier, H. & Bergh, H. van den (1989). De proefpeiling Nederlands: een onderzoek naar de haalbaarheid van peilingsonderzoek in het Voortgezet Onderwijs + bijlage. Arnhem: CITO (Bulletinreeks nr. 74).
- Meyer Viol, W.P.M. (1989). De oorsprong van gedrag. Wolters-Noordhoff Academieprijs. Groningen: Wolters-Noordhoff.
- Stokking, K.M. & Eberg, J.W. (1989). Opzet, uitvoering en resultaten van informatica-opleidingen in het kader van het leerlingwezen. Eindrapport van het evaluatieonderzoek van het landelijk experiment Instroomproject Schoolverlaters Informatiseringsberoepen. Utrecht: ISOR.
- Tan, L.H.C. (1989). Tekorten in de opleiding van huisartsen: ziektebeelden en medisch-technische vaardigheden. Proefschrift Universiteit van Amsterdam.
- Voogt, J.C. (1989). Scholen doorgelicht. Een studie over schooldiagnose. Proefschrift RU Utrecht. De Lier: Academisch Boeken Centrum.

Schrijfvaardigheid en het schrijf-denkproces

Iris Breetvelt

Stichting Centrum voor Onderwijsonderzoek

ABSTRACT

This review of literature is concerned with 25 empirical researches in which writing ability is related to the composing process. The differences in the writing process which were found with competent/skilled writers as against basic/unskilled writers relate to the planning and revising. Competent writers not only derive the content-ideas from the topic but also from the goal and the audience which the text aims to reach. Before they start to write, an outline is being made. While formulating plans competent writers not only plan on a local level, the next sentence or clause, as basic writers do, but also on a broad level, aimed at the total meaning of the text. Over and above that, their plans are not just concerned with the content but also with the way of dealing with the writing task: procedural plans. To good writers revising is not so much tracing errors as developing the text by means of different versions: their revisions are not only concerned with words and parts of a sentence but also with larger parts of a text.

Ofcourse results from various researches are not always exactly comparable because of differences in operationalisation of the variables writing ability, writing task and writing process. The findings are always concerned with the frequency with which subprocesses appear within the writing process, but not with sequence patterns with which subprocesses appear. For the rest the variable writing process is almost always related to the level of writing skills of the writer in general and not directly to characteristics of quality of the written text.

1. INLEIDING

Probleemstelling

In dit artikel wordt verslag gedaan van een literatuuronderzoek naar het verband tussen schrijfvaardigheid en kenmerken van het denkproces tijdens het schrijven van een tekst. Deze literatuurstudie is verricht in het kader van een empirisch onderzoek naar de relatie tussen tekstkwaliteitskenmerken en aspecten van het bijbehorende schrijfproces.

Achtergrond

In de loop van de jaren zeventig is binnen het onderzoek naar schrijven de aandacht verschoven van het schrijfprodukt, de tekst, naar het eraan voorafgaande proces: men vroeg zich af welke cognitieve processen zich afspelen tijdens het schrijven van een tekst. Het uitvoeren van een schrijftaak werd opgevat als een cognitief probleemoplossingsproces. Een onderzoeksmethode die veel wordt toegepast bij onderzoek naar de wijze waarop mensen problemen oplossen, is het registreren en vervolgens analyseren van het door de proefpersoon hardop geverbaliseerde denkproces.

Kennis over de cognitieve processen die zich afspelen tijdens het uitvoeren van een schrijftaak in het algemeen en in het bijzonder over die schrijfstrategieën en -processen waarin goede en slechte schrijvers van elkaar verschillen is om twee redenen van groot belang. Ten eerste levert dat informatie op voor een model van het schrijfproces dat individuele verschillen in schrijfgedrag verklaart. Ten tweede kunnen pas dan goede didaktieken die gericht zijn op schrijfaanpak worden ontwikkeld, wanneer duidelijk is

welke deelprocessen binnen het schrijfproces positief gerelateerd zijn aan tekstkwaliteit.

Uit het literatuuronderzoek waarover hieronder verslag gedaan wordt blijkt dat er nog erg weinig onderzoek is gedaan naar het directe verband tussen tekstkwaliteit en schrijfproces. Voorzover dergelijk onderzoek wel is gedaan zijn er geen relaties onderzocht tussen het schrijfproces en specifieke tekstkwaliteitskenmerken, zoals bijvoorbeeld argumentatie, alineaopbouw of taalgebruiksaspecten. Aangezien de tekstkwaliteit binnen een en dezelfde schrijver sterk kan variëren is het noodzakelijk didactische instructies te baseren op de directe relatie tussen het schrijfproces en het uit dat proces voortvloeiende schrijfprodukt.

Ten aanzien van de onderzoeksliteratuur die is bekeken zijn de volgende selectiecriteria gehanteerd:

- onderzoek waarin het schrijfproces wordt geanalyseerd en gerelateerd aan verschillen in schrijfvaardigheidsniveau;
- empirisch onderzoek waarin het schrijven van een zakelijke tekst als schrijftaak heeft gediend; de overweging hierbij is dat zakelijke teksten in de schoolse en naschoolse praktijk het meest beoefende genre zijn (Janssen, 1987);
- onderzoek vanaf 1970.

Op basis van deze selectiecriteria zijn 25 studies uitgekozen die aan alle drie de criteria voldoen (zie tabel 1). Over de bevindingen van deze studies wordt in het onderstaande verslag gedaan.

Tabel 1: Onderzoekskenmerken van 25 studies waarin het verband tussen schrijfvaardigheid en schrijfproces is nagegaan.

auteurs	proefpersonen	n	experimentele interventie(s)	onderzochte relatie(s)	procesfase
1. Stallard '74	15 goede schrijvers (senior high school students) crit.: score op essayschrijftest 15 middelmatige schrijvers (senior high school) crit.: willekeurige selectie	30		schrijfvaardigheid en schrijfproces	hele proces
2. Perl '79	5 slechte schrijvers (unskilled college writers) crit.: writing samples	5		schrijfvaardigheid en schrijfproces	hele proces
3. Pianko '79	7 goede schrijvers (traditional college writers) 10 slechte schrijvers (remedial college writers)	17		schrijfvaardigheid en schrijfproces	hele proces
4. Warters '79	3 slechte schrijvers (college basic writers) crit.: advies voor schrijfcursus gebaseerd op writing sample 1 goede schrijver (superior writer) crit.: faculteitsbeoordeling	4		schrijfvaardigheid en schrijfproces	hele proces

5.	Atlas '79 (Exp.1)	10 goede schrijvers (student expert writers) crit.: ervaring 20 slechte schrijvers (student novice writers) crit.: verbal SAT-score	30	schrijfvaardigheid en tekstkwaliteit / schrijfsproces	hele proces
6.	Flower, Hayes '80	goede schrijvers (expert writers) crit.: teachers in writing / rhetoric slechte schrijvers (student novice writers) crit.: schrijfsproblemen		schrijfvaardigheid en schrijfsproces	retorische probleemdefinitie
7.	Rose '80 1)	5 blokkerende schrijvers (undergraduate students) crit.: blokkering gebleken in schrijfcursus 5 niet-blokkerende schrijvers (undergraduate students)	10	schrijfblokades en schrijfsproces	hele proces
8.	Sweeder '81	6 slechte schrijvers (remedial student writers) crit.: deficient in basic writing skills, gebaseerd op testscore	6	schrijfvaardigheid en schrijfsproces	hele proces
9.	Bank '82	134 schrijvers van laaggemiddeld nivo (high school seniors) crit.: Bronx en compensatiecursus schrijven	134	tekstkwaliteit en schrijfsproces	hele proces
10.	Schumacher '82	20 ervaren schrijvers (upperclassmen) crit.: advanced composition course 22 onervaren schrijvers (freshmen)	42	schrijfvaardigheid en tekstkwaliteit / schrijfsproces	pauzege-drag
11.	Wright '83	30 slechte schrijvers (unskilled college freshmen) crit.: score op schrijfvaardigheidstest 39 middelmatige schrijvers (average college freshmen) crit.: willekeurige selectie	69	schrijfvaardigheid en tekstkwaliteit / schrijfsproces	prewrite revisie
12.	Selfe '84	4 schrijfangstigen 4 niet schrijfangstigen (freshmen composition course students) crit.: score op schrijfangsttest	8	schrijfangst en schrijfsproces	prewrite

13.	Taylor '84	5 goede schrijvers (high school students) crit.: 10% beste leerlingen 4 slechte schrijvers (remedial English classes) 9 professionele schrijvers crit.: schrijfervaring	18		schrijfvaardigheid en schrijfproces	hele proces
14.	Kennedy '85	3 leesvaardigen (fluent readers; students) crit.: score op Language skill test 3 niet-leesvaardigen (non-fluent readers) crit.: score op Language skill test	6		leesvaardigheid en tekstkwaliteit / schrijfproces	hele proces
15.	Root '85 1)	6 professionele schrijvers crit.: journalist	6		schrijfvaardigheid en schrijfproces	ideeën genereren
16.	Beach '76	11 extensive revisers 15 non-extensive revisers (English teachers) crit.: revisiefrequentie	26		revisiefrequentie en schrijfproces	zelfevaluaties
17.	Sommers '80	20 ervaren schrijvers (experienced adult writers) crit.: schrijfervaring 20 onervaren schrijvers (student freshmen) crit.: verbale SAT-score	40		schrijfvaardigheid en schrijfproces	revisie
18.	Bridwell '80	100 schrijvers (twelfth grade students)	100		tekstkwaliteit en schrijfproces	revisie
19.	Hays '81	7 goede schrijvers (upper-level students) crit.: schrijfervaring 6 slechte schrijvers (basic student writers) crit.: testscore, writing sample	13		schrijfvaardigheid en schrijfproces tekstkwaliteit en schrijfproces	revisie
20.	Garrett '81	70 goede studenten en slechte schrijvers (students)	70	exp.cond.: reviseren zonder 1e versie erbij contr.cond.: reviseren met 1e versie erbij	schrijfvaardigheid en tekstkwaliteit / schrijfproces	revisie
21.	Faigley '81	6 onervaren schrijvers (students) crit.: schrijfcursus 6 gevorderde schrijvers (students) crit.: expository writing class	18		schrijfvaardigheid en schrijfproces	revisie

		6 ervaren schrijvers (professionals) crit.: schrijfervaring			
22.	Monahan '82	4 goede schrijvers (competent writers, students) 4 slechte schrijvers (basic writers, students) crit.: score schrijfvaardigheidstest, aantal revisies	8		schrijfvaardigheid en schrijfproces revisie
23.	Hull '84	11 goede editors (student expert editors) 13 slechte schrijvers (student novice editors) crit.: writing sample	24	exp.cond.: met foutlocatie contr.cond.: zonder fout locatie	schrijfvaardigheid en schrijfproces editeren
24.	Gee '84	1372 schrijvers (twelfth grade students)	1372		tekstkwiteit en schrijfproces revisie
25.	Hayes '85	7 onervaren schrijvers (students) crit.: cijfers / docentbeoordeling 7 ervaren schrijvers (professionals / writing instructors) crit.: ervaring	14		schrijfvaardigheid en schrijfproces revisie
26.	Roen, Willey '88	studenten (freshmen)	60	exp. cond. 1 publiekgerichtheidsvragen tijdens versies schrijven exp. cond. 2 publiekgerichtheidsvragen tijdens reviseren controle conditie	tekstkwiteit en schrijfproces hele proces
27.	Kellogg, '87	psychologiestudenten	18	exp. cond. - outline - rough first draft contr. cond. - geen outline - polished first draft	schrijfstrategie en schrijfproces/ tekstkwiteit hele proces

- 1) Alle onderzoeken waarin twee of meer onderzoeksgroepen worden vergeleken, zijn opgezet als een 'static group comparison'. In alle onderzoeken behalve die van Rose en Root worden metingen verricht naar aanleiding van de wijze waarop de proefpersonen een schrijf taak uitvoeren.

2. ONDERZOEKSVARIABLEN EN OPZET

De variabelen in elk van deze onderzoeken zijn: 'schrijfvaardigheid', 'schrijfproces' en 'schrijftaak'.

'Schrijfvaardigheid' kan op twee manieren geoperationaliseerd zijn, namelijk als een kenmerk van schrijvers of als kenmerk van het schrijfprodukt, de tekst. Wanneer 'schrijfvaardigheid' als schrijverskenmerk is geoperationaliseerd zijn meestal twee niveau's onderscheiden, bijvoorbeeld goede tegenover slechte schrijvers of ervaren tegenover onervaren schrijvers. In dat geval is 'schrijfvaardigheid' als onafhankelijke variabele in het onderzoek opgenomen. Wanneer 'schrijfvaardigheid' daarentegen als kenmerk van het schrijfprodukt is geoperationaliseerd heeft het de status van afhankelijke variabele in de onderzoeksopzet. 'Schrijfproces' is vrijwel altijd afhankelijke variabele. Uitzondering hierop vormen de onderzoeken van Roen en Willey (1988), Atlas (1979) en Kellogg (1987). De eerste auteurs manipuleerden middels een interventie het procesaspect 'publieksbewustzijn'. In het onderzoek van Atlas worden enkele fasen van het schrijfproces middels specifieke deeltaken geoperationaliseerd: het betreft een genereertaak, een organiseertaak en een formuleertaak. In het onderzoek van Kellogg worden twee aspecten van het schrijfproces middels twee schrijfstrategieën gemanipuleerd.

De variabele 'schrijftaak' is steeds onafhankelijke variabele; doorgaans zijn er één of meer identieke schrijftaken voor alle (groepen) schrijvers. Alleen in de onderzoeken van Garrett (1981) en Hull (1984) is sprake van een experimentele manipulatie middels de schrijftaak: het gaat hier om het schrijven met bepaalde gegevens tegenover het schrijven zonder die gegevens. De schrijftaken lopen op een aantal relevante opdrachtvariabelen sterk uiteen: naast verschillen in het tekstdoel en het beoogde lezerspubliek zijn er ook grote verschillen wat betreft het onderwerp en de mate van vrijheid t.a.v. het onderwerp. De opdracht kan variëren van 'schrijf een essay over een onderwerp waarin je een ander wilt overtuigen' tot 'beantwoordt deze brief over (...) en richt je reactie aan (...).' In vier onderzoeken is niet alleen het publiek aangegeven maar is zelfs een hele communicatieve situatie gedefinieerd, dat wil zeggen de schrijver in zijn rol of functie, het publiek en het medium waarin de tekst zou moeten verschijnen zijn nader gespecificeerd (Atlas, 1979; Hays, 1981; Hayes, 1985; Kellogg, 1987).

Hoe zijn nu de variabelen 'schrijfvaardigheid als schrijverskenmerk', 'schrijfvaardigheid als produktkenmerk' en 'schrijfproces' gemeten?

Wanneer 'schrijfvaardigheid' als schrijverskenmerk is geoperationaliseerd is het selectie criterium gebaseerd op een norm voor schrijfvaardigheid, schrijfervaring of schrijfcompetentie. In de betreffende 25 onderzoeken zijn dat (zie tabel 1):

- verbale of schrijfcompetentietestscores (Atlas 1979; Sommers 1980; Hays 1981; Sweeder 1981; Monahan 1982; Wright 1983; Selfe 1984; Taylor 1984; Kennedy 1985; Hayes 1985);
- kwaliteits- of kenmerkbeoordeling van een geschreven tekst (Stallard 1974; Beach 1976; Perl 1979; Warters 1979; Hays 1981; Monahan 1982; Hull 1984);
- deel uitmaken van een bepaalde beroepsgroep met schrijfervaring (Atlas 1979; Flower en Hayes 1980; Sommers 1980; Faigley 1981; Taylor 1984; Flower en Hayes 1985);
- deel uitmaken van een cursusgroep die door een bepaald schrijfvaardigheidsniveau gekenmerkt wordt (Flower en Hayes 1980; Faigley 1981; Hays 1981; Bank 1982; Schumacher 1982; Hull 1984).

Overigens valt op dat bijna alle onderzoeksgroepen, uitgezonderd de 'professionele schrijvers', uit studenten (en in een enkel geval middelbare scholieren) bestaan. De leeftijden van de proefpersonen, die vrijwel nergens expliciet vermeld worden, variëren dus tussen ongeveer 16 en 25 jaar.

Wanneer 'schrijfvaardigheid' als een produktkenmerk is geoperationaliseerd zijn de

door de proefpersonen geschreven teksten op kwaliteit beoordeeld. Dat betreft steeds een holistisch oordeel op een aantal aspecten (bijvoorbeeld: inhoud, organisatie, coherentie, taalgebruik) waarbij de beoordelingen op de afzonderlijke aspecten uiteindelijk tot een totaalscore per tekst worden gecombineerd (Atlas 1979; Bridwell 1980; Hays 1981; Garrett 1981; Schumacher 1982; Wright 1983; Gee 1984; Kennedy 1985; Kellogg 1987; Roen & Willey 1988). In het onderzoek van Kellogg (1987) zijn naast produktkwaliteit tevens efficiency en produktiviteit als indices gekozen.

Registratie van het 'schrijfproces', het denkproces tijdens het schrijven van een tekst, vindt plaats via gedragsobservaties tijdens het schrijven en/of aan de hand van zelfrapportage door de proefpersoon (Ericson & Simon 1984). Dit laatste kan zijn in de vorm van een introspectief of retrospectief verslag over het eigen denkproces, tijdens het schrijven of achteraf pas, al dan niet gestuurd door interviewvragen. Een tweede, meer directe vorm van zelfrapportage, die geen reflectie op het denkproces vraagt maar slechts weergave van het denkproces, is het hardopdenken tijdens het schrijven: een in de cognitieve psychologie veel gehanteerde methode. Deze laatste methode verschaft het meeste inzicht in de manier waarop zich tijdens het schrijven allerlei subprocessen in het denken afspelen; bijvoorbeeld: welke procedurele plannen hanteert de schrijver?; welke inhoudsideeën worden gevormd en na evaluatie verworpen?

In de meeste onderzoeken worden verschillende methoden naast elkaar (in aanvulling op elkaar) gebruikt. Bijvoorbeeld de combinatie van hardopdenkend schrijven en introspectief of retrospectief verklaren of becommentariëren van bepaalde procesaspecten of -momenten. De methode 'hardopdenkend schrijven' wordt in tien van de 25 studies toegepast.

Soms worden stadia van het schrijfprodukt, d.w.z. tekstversies, tevens als bron van informatie over het schrijfproces gebruikt; dat geldt vooral voor onderzoek naar het reviseerproces.

De gegevens die resulteren uit de registratie van het schrijfproces, zoals hardopdenkprotokollen, gedragsobservaties en retrospectieve zelfrapportages, worden vervolgens aan de hand van categoriseringsschema's geanalyseerd. De categorieën uit dergelijke schema's zijn meestal gebaseerd op een schrijfprocesstheorie en zo ontworpen dat elke proceseenheid door één categorie en ook niet meer dan een categorie gedekt wordt. In de meeste studies (Perl 1979; Sweeder 1981; Bank 1982; Schumacher 1982; Wright 1983; Selfe 1984; Kennedy 1985) zijn de categorieën gegroepeerd onder bepaalde hoofdgroepen. Zo vallen bij Perl de subcategorieën 'reading directions' en 'reading statement' onder de hoofdcategorie 'reading'; andere hoofdcategorieën zijn hier 'talking' en 'writing'. Een andere indeling is die van Schumacher (1982), in cognitieve activiteiten (bijvoorbeeld 'global planning'; 'review topic'; 'review on experience') en grammaticale activiteiten (bijvoorbeeld 'spelling'; 'sentence structure'). Selfe (1984) maakt een ten dele op Flower en Hayes (1983) gebaseerd onderscheid in 'planning-', 'translating-', 'reviewing behavior' plus een restgroep 'miscellaneous behavior'. Weer een ander type categorieënordening is die van Bank (1982) naar fase in het schrijfproces: 'prewrite', 'write' en 'postwrite'.

De categorieënschema's of delen van categorieënschema's die speciaal betrekking hebben op reviseren zijn vaak twee- of meerdimensionaal opgebouwd. In de tweedimensionale systemen worden revisies getypeerd als een combinatie van een bepaald type tekstwijzigingsoperatie en het tekstniveau waarop deze wordt toegepast (Bridwell 1980, Sommers 1980, Faigley 1981, Hays 1981, Monahan 1982, Wright 1983). Deze laatste vier auteurs verwijzen daarbij allen naar Sommers die als eerste tekstniveaus en tekstoperaties met elkaar kruiste. Voorbeeld van een aldus te categoriseren revisie is het vervangen van een woord door een ander woord (= 'substitutie op woordniveau').

Hays (1981) en Faigley (1981) zijn, als het revisies betreft ook nog geïnteresseerd in de

mate waarin de tekstbetekenis wezenlijk veranderd wordt. Hays spreekt van 'semantic concern' (in contrast tot wat hij aanduidt met de term 'formal concern') en Faigley spreekt van 'textbase changes' (in tegenstelling tot 'surface changes'). Monahan voegt als derde dimensie toe wat hij aanduidt als 'level of revision'. Daarmee wordt bedoeld het moment van revisie: binnen een eerste of tweede tekstversie of op moment van overgang van de eerste naar de tweede tekstversie. De vierde dimensie die Monahan onderscheidt, betreft het doel van de revisie (bijvoorbeeld 'stylistic', 'informative').

Dat auteurs onder dezelfde categorieën soms niet hetzelfde verstaan blijkt uit de niet altijd gelijklopende omschrijvingen die eraan gegeven worden. Zo definieert Pianko (1979) bijvoorbeeld 'planning' als 'the setting of parameters, general or specific, for the composition to be written: mental, written, or both'. Kennedy (1985) omschrijft 'planning' als: 'strategies like questioning or disagreeing with source authors: organizing source information, deliberating about word choice, grammar and punctuation; telling oneself how to write or proceed and expressing concern with content, ideas or organization'. Enkele onderzoekers maken een onderscheid tussen 'general planning' en 'local planning' (Perl 1979, Selfe 1984) of tussen 'global planning' en 'immediate planning' (Schumacher 1982). Schumacher geeft geen omschrijving, Perl een korte: onder 'local planning' verstaat zij 'talking what idea comes next' en onder 'general planning': 'organizing one's thoughts for writing, discussing how one will proceed'. Selfe (1984) geeft hieraan de meest uitgewerkte definitie. 'General planning', ofwel het identificeren van mogelijke topics, wordt door haar omschreven als: 'het identificeren van een idee om het te noteren, het ordenen ervan m.b.t. tot andere, zowel hoger- als lagergeordende ideeën.' Deze zelfde definitie, die aan Flower ontleend is, laat zij ook gelden voor 'local planning': een en ander betreft dan echter lokaal niveau. Daarmee is dan natuurlijk nog niet duidelijk volgens welk criterium 'general' en 'local' te onderscheiden zijn.

In de schrijfprocesanalyse-schema's zoals die in de geanalyseerde studies gebruikt zijn, komen verschillende typen analysecategorieën voor:

1. cognitieve activiteitscategorieën, bijvoorbeeld 'global/local planning' (Perl 1979), 'rereading' (Pianko 1979), 'interpreting' (Perl 1979);
2. categorieën die gedefinieerd zijn naar doel of effect van een activiteit, bijvoorbeeld 'effect on reader' (Schumacher 1982), 'audience awareness/stylistic concerns' (Stallard 1974);
3. categorieën die vooral het moment van een activiteit aangeven, bijvoorbeeld 'prewrite' (Pianko 1979, Bank 1982);
4. categorieën die het niveau, de omvang of locatie van het tekstgedeelte waarop een bepaalde activiteit wordt uitgevoerd, aangeven, bijvoorbeeld 'phrase', 'multisentence' (Bridwell 1980).

De procesanalyse categorieën uit de diverse onderzoeken verschillen wat betreft mate van gedetailleerdheid, theoretische achtergrond en betekenis.

Uitgaande van de twee operationalisaties van 'schrijfvaardigheid', nl. als persoonskenmerk en als produktkenmerk, kunnen de onderzoeken in twee groepen worden verdeeld:

1. Onderzoek waarbij kenmerken van het schrijfproces in verband worden gebracht met het schrijfvaardigheidsniveau, als persoonskenmerk van de schrijver. In dergelijk onderzoek wordt het schrijfproces van een wat niveau betreft extreme groep bekeken of wordt het schrijfproces van verschillende groepen met elkaar vergeleken.
2. Onderzoek waarbij kenmerken van het schrijfproces in verband worden gebracht met kenmerken en kwaliteit van het produkt, de tekst.

Tot de tweede groep behoren slechts 7 studies (Bridwell 1980; Hays 1981; Garrett 1981; Bank 1982; Gee 1984; Kellogg 1987; Roen & Willey 1988). De overige 18 studies behoren tot de eerste groep. In deze studies worden de schrijfprodukten soms weliswaar op

kwaliteit beoordeeld, maar deze beoordeling wordt dan niet aan het schrijfproces gerelateerd, maar aan de indeling van de proefpersonen naar niveau van schrijfvaardigheid, de produktbeoordelingen dienen in dat geval dus slechts ter validatie van de proefpersonenindeling. (Atlas 1979; Schumacher 1982; Wright 1983; Kennedy 1985).

3. RESULTATEN

In verband met de overzichtelijkheid is bij de bespreking van de onderzoeksbevindingen getracht de studies zoveel mogelijk te clusteren in groepjes die onderling sterk vergelijkbaar zijn wat betreft enkele relevante aspecten in doel en opzet van het onderzoek. De verwachting is immers dat binnen zulke clusters de onderzoeksresultaten elkaar voor een belangrijk deel zullen bevestigen.

Allereerst kan dus een onderscheid gemaakt worden tussen de onderzoeken waarin schrijfvaardigheid als produktkenmerk is geoperationaliseerd en onderzoeken waarin schrijfvaardigheid als schrijverskenmerk is geoperationaliseerd. Vervolgens kan hiermee een ander onderscheid gekruisd worden, namelijk dat tussen studies waarin het hele schrijfproces is onderzocht en studies waarin alleen het reviseerproces onderwerp van onderzoek is.

Voor bespreking van de resultaten uit de studies vormt de schrijfproces theorie van Flower en Hayes (1983) een geschikt uitgangspunt. De pretentie die de auteurs met dit cognitief psychologische model van het schrijfproces hebben is dat allerlei individuele variaties in schrijfstrategieën ermee beschreven kunnen worden.

Volgens dit model bestaat het gehele schrijfproces uit de processen 'plannen', 'formuleren' en 'reviseren', waarbij geldt dat deze processen elkaar voortdurend, in diverse volgorde kunnen afwisselen: het formuleren van een zin kan dus weer gevolgd worden door plannen maken voor een volgend tekstdeel. Elk van deze drie processen wordt verder ontleed in subprocessen. Zo bestaat plannen uit het genereren van ideeën voor de tekstinhoud, het selecteren en ordenen van ideeën en het stellen van doelen: dat kunnen doelen zijn die betrekking hebben op de manier waarop de tekst eruit moet zien en wat er in moet staan, maar het kunnen ook doelen zijn die het verloop van het schrijfproces zelf betreffen, zoals 'eerst een tekstschema maken'. Ook het reviseren is door Flower en Hayes nader uitgewerkt in enkele subprocessen: het herlezen, detecteren en analyseren van eventuele tekstproblemen en het herschrijven van problematische tekstpassages.

Een interne monitor bepaalt in welke volgorde en op welke momenten tijdens het hele schrijfproces de diverse subprocessen elkaar afwisselen.

Hierin ligt een klein zwaktepunt van de theorie: er wordt in zo'n geval een zogenaamde homunculus gepostuleerd, een subsysteem dat het functioneren van het gehele systeem stuurt en controleert. Datgene wat verklaard moet worden, namelijk het functioneren van het gehele systeem, wordt daarmee zelf weer als verklarend principe aangeroepen. Wolters en Phaf (1988) vinden zo'n homunculus als subsysteem, monitor, daarom niet aanvaardbaar.

Schrijfvaardigheidsniveau en schrijfproces

De resultaten uit de onderzoeken waarin schrijfvaardigheid gerelateerd wordt aan het schrijfproces zijn te groeperen naar de door Flower en Hayes (1983) onderscheiden processen 'planning' en 'revisie'.

Planning:

In de onderzoeken van Stallard (1974), Atlas (1979), Pianko (1979), Schumacher (1979),

Warters (1979) en Wright (1983) worden de schrijfprocessen van studenten van verschillend schrijfvaardigheidsniveau met elkaar vergeleken.

Stallard (1974), Schumacher (1979) en Wright (1983) vergelijken een extreme groep met een willekeurige of gemiddelde groep uit dezelfde populatie; Atlas (1979), Pianko (1979), Warters (1979) vergelijken twee extreme groepen uit dezelfde populatie. Men kan dus aannemen dat het contrast tussen de beide onderzoeksgroepen in deze laatste gevallen groter is dan in de eerste drie gevallen. Nog sterker is het contrast tussen de onderzoeksgroepen in de studies van Flower en Hayes (1980) en Taylor (1984). Flower en Hayes (1980) vergelijken docenten in tekstschrijven met student-schrijvers die problemen hebben op dit gebied. Taylor (1984) vergelijkt professionele schrijvers die een ruime schrijfervaring hebben met goed schrijvende tegenover slecht schrijvende studenten. In de studies van Perl (1979) en Sweeder (1981) wordt slechts één onderzoeksgroep bekeken, namelijk slechte schrijvers. Selfe (1984) en Kennedy (1985) onderzochten schrijvers die op een ander kenmerk waren geselecteerd, namelijk schrijfangst, respectievelijk vaardigheid in 'begrijpend lezen'.

Wat betreft de prewrite fase, dat wil zeggen het tijdsverloop tussen het inzien van de opdracht en het moment waarop het eerste woord op papier wordt gezet, constateren Sweeder (1981) en Perl (1979) dat deze bij de minder goede schrijvers van korte duur is: de proefpersonen herformuleren de opdracht en associëren naar aanleiding van sleutelwoorden in de opdracht. Stallard (1974) en Pianko (1979) constateren een significant verschil in prewritetijd tussen goede en slechte schrijvers. Warters (1979) constateert eveneens een dergelijk verschil. Volgens Wright (1983) bestaat er echter op dit punt geen significant verschil. Warters (1979) meent dat achter dit verschil in tijdsbesteding ook een kwalitatief verschil schuilt, namelijk een verschillende mate van actieve mentale betrokkenheid van de schrijver bij zijn onderwerp: 'basic writers purposely exclude prewriting influences from their writing, developing their pieces without contributions from their personal web of meaning'. De geringere investering van 'basic writers' in prewriting heeft volgens Warters (1979) te maken met het ontbreken van een proces van betekenisgeving en interpretatie.

Goede schrijvers ('experts') reageren op een groot aantal aspecten van wat Flower en Hayes (1980) het retorische probleem noemen: naast zich een representatie te vormen van de schrijfopdracht en het publiek, vormen ze zich ook doelen met betrekking tot zichzelf (bv. 'I'll look like an idiot if I say ...'), met betrekking tot de werkelijkheid of de tekstbetekenis (bv. 'so if I compare those two attitudes ...') en met betrekking tot de tekststructuur (bv. 'first we'll want an introduction').

Goede schrijvers ontleen nieuwe inhoudsideeën voor meer dan de helft aan het retorische probleem en voor een veel geringer deel aan het onderwerp (bv. 'I'll want an introduction that pulls you in'); bij slechte schrijvers is die verhouding precies omgekeerd.

Selfe constateert dat schrijfangstige schrijvers minder gebruik maken van de retorische informatie uit de schrijfopdracht. Zowel lezerspubliek als cues voor de opbouw van de tekst die in de opdracht worden geïmpliceerd, worden door hen veronachtzaamd. Niet-schrijfangstige schrijvers komen tot een meer complexe en gedetailleerde uitwerking van het publiek waarop hun tekst gericht is (bv. de policy-making board van de high school wordt voorgesteld als 'a bunch of really excited, booster type people (...) and parents that really care'). Ook zijn niet-schrijfangstigen bewuster gericht op het aanbrengen van een logische ordening in de tekstinhoud: hiërarchisch, chronologisch of volgens een ander principe.

Stallard (1974) concludeert uit zijn onderzoek dat goede schrijvers meer gericht zijn op het inperken van doel en onderwerp van hun tekst.

Goede schrijvers ontwikkelen het beeld dat ze van de lezer, de retorische situatie en de eigen doelen hebben ook tijdens het schrijven steeds verder. (Flower en Hayes 1980)

Stallard (1974) constateert dat van de goede schrijvers een groter percentage vooraf een tekstplan ('mental plan') maakt dan van de willekeurig gekozen schrijvers. Wright (1983) constateert echter dat beide groepen in zijn onderzoek geen outline maken. Uit het onderzoek van Atlas (1979) blijkt dat goede schrijvers vaker eigen, nieuwe ideeën toevoegen aan een gegeven outline dan slechte schrijvers dat doen. Alvorens de eerste zin te produceren hebben niet-schrijfangstige schrijvers volgens Selfe (1984) een algemeen schrijfplan, schrijfangstige schrijvers daarentegen starten met de eerste zin zonder een dergelijk plan. Voordat ze met het schrijven van de eerste tekstversie beginnen besteden ze meer tijd aan 'local planning'; niet schrijfangstige schrijvers besteden op dat moment juist meer tijd aan 'global planning', zij maken ook meer notities vooraf die betrekking hebben op ideeënontwikkeling, planning en organisatie van de te schrijven tekst.

Het eigenlijke schrijven verloopt volgens Warters (1979) bij slechte schrijvers als een lineair proces: er wordt vanaf de eerste zin doorgeschreven tot aan de conclusie, waarbij de eerste zin in belangrijke mate de rest van de tekst bepaalt.

De schrijfplannen van slechte schrijvers zijn niet flexibel; zicht op de hiërarchie in de plannen ontbreekt; er wordt weinig tijd besteed aan planning; (Sweeder 1981; Perl 1979) Het schrijven bij deze minder goede schrijvers wordt voortdurend afgewisseld door het herlezen van de geschreven tekst ('retrospective structuring'); dit herlezen dient het verdere plannen, formuleren en editeren. Bij dit herlezen wordt overigens vaak gelezen wat er zou moeten staan in plaats van wat er werkelijk in de tekst staat.

Schumacher (1979) constateert een significant verschil in hoeveelheid cognitieve activiteiten tussen schrijvers met en schrijvers zonder een schrijfcursus voor gevorderden. Vooral de activiteit 'immediate planning' komt minder voor bij de schrijvers die geen cursus voor gevorderden hebben gehad. Atlas (1979) signaleert bij goede schrijvers meer expliciete planning, die tevens een meer tentatief karakter heeft (bv. 'why not a mixed plan like this?')

Schrijvers die goed zijn in begrijpend lezen ('fluent readers') plannen, wanneer ze een gedocumenteerd schrijven opdracht uitvoeren anders dan 'non-fluent readers'. (Kennedy 1985)

'Fluent readers' herlezen de documentatie en maken notities in de 'postreading fase'; 'non-fluent readers' doen dat vooral tijdens de eigenlijke schrijffase, alleen om er rechtstreeks uit te kunnen citeren. Ook de planning vindt bij deze laatste groep in het algemeen pas tijdens de eigenlijke schrijffase plaats in plaats van voorafgaand aan het schrijven.

'Fluent readers' gebruiken bepaalde leesstrategieën significant frequenter: het betreft de zogenaamd hogerniveau verwerking (het relateren van documentatie aan eigen kennis en ervaring en het evalueren ervan) en het studerend lezen (markeren en onderstrepen van belangrijke fragmenten uit de gelezen teksten).

'Fluent readers' doen meer aan planning. Voorzover 'nonfluent readers' plannen doen ze dat vooral op het niveau van de zin in plaats van op niveau van grotere tekstdelen.

Zowel Schumacher (1979) als Wright (1983) constateren een significant verschil in pauzegedrag tussen beide onderzoeksgroepen. Goede schrijvers houden kortere pauzes (Schumacher 1979) en pauzeren minder, maar stoppen daarentegen vaker dan slechte schrijvers (Wright 1983). Pianko (1979) geeft een kwalitatieve analyse van het verschil in pauzegedrag: 'traditional students were pausing to plan what to write next, rescanning to see if their plans fit and then pausing again to reformulate. Remedial students during their pauses were glancing around the room as a diversion, sometimes hoping the correct spelling or word or something else to say next would suddenly appear to them'.

Wat publiekgerichtheid betreft zijn de resultaten niet helemaal eenduidig. Atlas (1979) constateert bij goede schrijvers significant meer, vooral publiekgericht commentaar

tijdens de uitvoering van een ideeën genereertaak dan bij slechte schrijvers (bv. 'invite his contributions on this point').

In de eerste alinea van hun tekst zouden goede schrijvers meer personaliseren via het gebruik van persoonlijke voornaamwoorden. Stallard (1974) constateert echter geen enkel verschil wat publieksbewustzijn betreft tussen beide groepen schrijvers in zijn onderzoek.

Revisie:

In zes onderzoeken wordt een relatie gelegd tussen schrijfvaardigheidsniveau en manier van reviseren: goede en slechte schrijvers verschillen mogelijk van elkaar in de manier waarop ze wijzigingen en verbeteringen in hun tekst aanbrengen.

In drie van deze onderzoeken wordt een groep studenten afgezet tegen een groep professionele- of ervaren volwassen schrijvers (Sommers 1980, Faigley 1981, Hayes 1985). In de overige drie onderzoeken worden twee groepen 'student writers' van verschillend niveau van schrijfvaardigheid met elkaar gecontrasteerd (Beach 1976, Monahan 1982, Hull 1984).

In het merendeel van deze studies wordt het revisieproces onderzocht aan de hand van verschillende versies van een door de proefpersonen geschreven tekst. Alleen bij de studies van Hull en Hayes, die gericht zijn op specifieke deelprocessen binnen het revisieproces, is de schrijftaak een meer ingeperkte en specifieke deelopdracht. In het onderzoek van Hull (1984) moeten in gegeven teksten drie typen fouten gecorrigeerd worden. De opdracht uit het onderzoek van Hayes (1985) betreft het reviseren van een gegeven tekst en het analyseren van tekstproblemen aan de hand van los aangeboden tekstfragmenten uit diezelfde tekst.

Wat de relatieve frequentie van het reviseren betreft zijn de resultaten niet eenduidig. Faigley (1981) komt tot de conclusie dat het grootste aantal revisies wordt aangebracht door de 'onervaren' en de 'gevoerde schrijvers'; 'experts' en professionele schrijvers variëren sterk in de mate waarin ze reviseren. Uit het onderzoek van Monahan (1982) komt naar voren dat competente schrijvers meer reviseren dan 'basic writers'. Ook Hull (1984) vindt een significant verschil tussen 'experts' en 'novices' in de mate waarin ze tekstfouten corrigeren: 'expert writers' corrigeren meer fouten. Ook uit de resultaten van Stallard (1974) blijkt dat goede schrijvers significant vaker reviseren dan slechte schrijvers.

Wat betreft het tekstniveau waarop revisie plaatsvindt verschillen goede en slechte schrijvers nogal. Terwijl slechte schrijvers bijna uitsluitend wijzigingen aanbrengen op oppervlakte niveau (spelling, interpunctie) en in kleine delen van de tekst (woorden, zinnen), wijzigen goede schrijvers ook op het niveau van de totale tekstbetekenis: wijzigingen betreffen dan niet alleen woorden of zinnen maar ook alinea's of grotere tekstdelen (Beach 1976, Sommers 1980, Faigley 1981, Monahan 1982, Hayes 1985, Wright 1983). Faigley (1981) constateert dat onervaren schrijvers slechts gemiddeld 12% 'tekstbase changes', dat wil zeggen wijzigingen met effect op de betekenis van de tekst, aanbrengen, terwijl dit voor de ervaren of volwassen schrijvers 35% bedraagt. Hulls bevinding dat goede tegenover beginnende schrijvers niet verschillen in de mate waarin ze tekstbetekenisfouten ('comprehending errors') corrigeren, lijkt niet in overeenstemming met de bevindingen van Faigley.

Met betrekking tot doel en strategie van reviseren zijn er ook een aantal differentiële effecten van de factor schrijfvaardigheid te bespeuren.

Overwegingen van doel- en publiekgerichtheid bij revisie gelden meer voor goede schrijvers dan voor slechte schrijvers (Sommers 1980, Faigley 1981, Hayes 1985).

De 'extensive revisers' uit het onderzoek van Beach (1976) zien het reviseren als een proces waarbij grote inhoudelijke wijzigingen aangebracht kunnen worden; de 'non-revisers' beschouwen reviseren slechts als een proces van vormwijzigingen, waarbij de tekst

op een aantal mogelijke problemen wordt gecheckt (bv. 'needs some smoother transitions', 'make it more concrete'). De strategieën van beide typen reviseerders verschillen eveneens: de 'extensive revisers' zijn beter in staat om zich los te maken van hun tekst en deze vanuit een holistisch standpunt te bezien; het ontwikkelen van ideeën vindt bij deze groep schrijvers als het ware over de verschillende tekstversies heen plaats. 'Non-extensive revisers' kunnen zich daarentegen minder goed losmaken van de tekst, bezien deze vanuit een egocentrisch standpunt en werken aan elke tekstversie als aan een aparte eenheid.

Uit het onderzoek van Sommers (1980) komt naar voren dat studenten reviseren opvatten als een proces van opnieuw verwoorden; ze zien woorden als de eenheid van geschreven tekst. Zo letten zij bijvoorbeeld op lexicale herhalingen, maar zien conceptuele herhalingen over het hoofd. Hun strategieën betreffen woord- en zinsniveau, terwijl procedures die gericht zijn op redeneerlijnen, doel en publiek ontbreken. De regels die ze hanteren zijn bijvoorbeeld: 'never end a sentence with a proposition'. De ervaren schrijvers daarentegen zien als het doel van revisie: 'finding the form or shape of the argument'. Een tweede doel betreft het lezerspubliek. In hun streven naar het creëren van betekenis ('discovering meaning') passen ervaren schrijvers, in tegenstelling tot studentenschrijvers alle vier revisieoperaties (verwijdering, vervanging, toevoeging, herordening) op elk van de vier tekstniveaus (woord, zinsdeel, zin, tekstthema) toe. Bij ervaren schrijvers is reviseren een recursief proces: tijdens de eerste fase worden onderwerp en ideeën ingeperkt ('form of the argument'), in latere revisiefasen wordt meer aandacht besteed aan lexicale en stilistische aspecten.

Faigley (1981) meent dat 'expert writers' een grotere diversiteit aan revisiestrategieën aan de dag leggen dan onervaren schrijvers doen. Wanneer experts revisies aanbrengen in teksten die door onervaren schrijvers geschreven zijn blijken dat vooral macrostructurele veranderingen te zijn (dat wil zeggen veranderingen die de betekenis van de tekst zodanig wijzigen dat dit ook z'n effect zou hebben op een samenvatting van de tekst). Na de intentie van de tekst te hebben geanalyseerd, wordt de oorspronkelijke tekst gecondenseerd en wordt er veelal informatie nader uitgewerkt of aan toegevoegd.

Ook Monahan (1982) constateert dat competente schrijvers een groter scala aan revisies toepassen en dat bovendien op meer diverse momenten in het schrijfproces doen dan 'basic writers'. Wanneer het publiek waarop de tekst zich moet richten uit medeleerlingen bestaat, reviseren goede schrijvers meer; wanneer echter de docent het publiek vormt, wordt door de 'basic writers' juist meer gereviseerd. De revisiestrategie van basic writers kan getypeerd worden als het nalopen van een checklist voor bepaalde tekstproblemen. Competente schrijvers daarentegen ontwikkelen de tekst middels revisies inhoudelijk, dat wil zeggen bij hen staat revisie vooral in het teken van tekstontwikkeling in plaats van in het kader van foutencorrectie. Ook reviseren zij meermalen in ketens van onderling gerelateerde revisies: de ene revisie is aanleiding tot de volgende revisie.

Uit het onderzoek van Hull (1984) naar foutencorrecties komt naar voren dat 'consulting errors' (fouten in verband met taal- en tekstconventies, bv. spellingsfouten) en 'intuition errors' (fouten die weliswaar geen regelovertreding betreffen maar toch niet goed in het gehoor liggen) door goede schrijvers meer gecorrigeerd worden dan door slechte schrijvers. Wanneer is aangegeven op welke plaats in de tekst de fout zich bevindt, verschillen goede en slechte schrijvers in hun oplossingsstrategie: goede schrijvers proberen de betreffende passage te analyseren en hun kennis van regels toe te passen. Slechte schrijvers trachten op onsystematische wijze naar een goede oplossing te raden, daarbij meestal beginnend met spelling en interpunctie; de regelkennis van slechte schrijvers is minder en bovendien zijn ze minder goed in staat tot een juiste toepassing van deze kennis. Hayes (1985) deed het meest uitgebreide onderzoek naar het differentieel presteren van goede en slechte schrijvers op een aantal deelprocessen binnen het

revisieproces. Naast statistische toetsing van de kwantitatieve verschillen tussen 'expert-' en 'novice writers' op een aantal revisiesubprocessen werkt Hayes deze verschillen in meer kwalitatieve, op case studies gebaseerde analyses nader uit.

Alvorens aan het feitelijke reviseren te beginnen zijn goede schrijvers significant meer bezig met de tekst op allerlei manieren te inspecteren, bijvoorbeeld het inventariseren van tekstproblemen; het stellen van globale, op het tekstdoel en op de communicatieve context gerichte revisiedoelen. Ook detecteren goede schrijvers significant meer tekstproblemen, met name meer globale tekstproblemen, die met de inhoudelijke betekenis van de tekst te maken hebben.

De resultaten van Stallard (1974), Pianko (1979) en Wright (1983) wijzen dan ook uit dat goede schrijvers significant vaker stoppen om hun tekst over te lezen dan slechte schrijvers. Pianko benadrukt dat een essentieel onderscheid tussen goede en slechte schrijvers gelegen is in de mate van reflectie op de zelf geschreven tekst. Wright constateert bij 'average writers' dan ook een frequenter voorkomen van de categorie 'contemplation of final product' dan bij 'unskilled writers'.

Ten aanzien van niet zo scherp te definiëren problemen, die een categorisering en nadere specificering in de vorm van een diagnose vragen, geldt eveneens een significant verschil tussen 'experts' en novices'. Dit verschil betreft niet alleen het aantal diagnoses, maar ook de aard van de diagnoses. De diagnose van de 'expert writer' is een hiërarchische structuur van tekstproblemen; een globaal tekstprobleem wordt gerepresenteerd als een netwerk van onderling gerelateerde problemen die zich vaak op de diverse tekstniveaus (woord, zin, alinea, tekst) bevinden: zo wordt bijvoorbeeld de lokale incoherentie tussen twee zinnen voorgesteld als ondergeordend aan het probleem van de globale tekstincoherentie, die bijvoorbeeld weer gevolg is van het feit dat de schrijver er niet in is geslaagd zich steeds aan z'n tekstdoel te houden.

Hayes (1985) onderscheidt aan diagnoses twee aspecten: de tekstomvang, ofwel het tekstniveau waarop de diagnose betrekking heeft, en de aard van het geanalyseerde probleem. Dit laatste betreft dan tekstdoel en/of publiek ('intentionality'), weinig uitgesproken richtlijnen ('maxims') of echte regels ('rules'). Beginnende schrijvers zijn bij het diagnostiseren van tekstproblemen geneigd tot twee typen tekortkomingen. Ofwel ze zien de betrokken tekstomvang als te beperkt. Een voorbeeld hiervan is een diagnose waarbij problemen m.b.t. het doel van de tekst op zinsniveau worden geprojecteerd in plaats van op het niveau van grotere tekstdelen (tekstcoherentieproblemen zullen bij een dergelijke visie gemakkelijk aan de aandacht ontsnappen). Ofwel de aard van het probleem wordt als te beperkt opgevat, bijvoorbeeld een 'maximbased' probleem wordt als regelovertreding opgevat; een conditionele richtlijn wordt dan bijvoorbeeld verengd tot een altijd geldende regel. 'Expert writers' slagen er in het algemeen in betere specificaties te geven van de tekstkenmerken die bij 'maximbased' problemen betrokken zijn en tevens hanteren ze heuristieken die meer opening tot concrete tekstwijzigingsoperaties bieden.

Ook de tekstwijzigingsstrategie verschilt tussen beide typen schrijvers. Op basis van de gestelde diagnose zullen 'expert writers' de ene keer de tekst herschrijven ('rewrite') en de andere keer wijzigen met behoud van een belangrijk deel van het betreffende tekstfragment ('revise'). Reviseren doen beginnende schrijvers nauwelijks, wel herschrijven ze tekst, doch meestal zonder de betekenis ingrijpend te veranderen, ze parafraseren slechts.

Tekstkwiteit en schrijfproces

In zeven onderzoeken is het verband tussen de kwaliteit van de geschreven tekst en kenmerken van het bijbehorende schrijfproces geanalyseerd. (Bridwell 1980; Garrett 1981; Hays 1981; Bank 1982; Gee 1984; Kellogg 1987; Roen & Willey 1988). In de onderzoeken van Bridwell (1980), Bank (1982) Gee (1984), Kellogg (1987) en Roen & Willey (1988) is

binnen de groep proefpersonen geen onderscheid naar niveau van schrijfvaardigheid gemaakt. In deze onderzoeken, waarvan er enkele bij relatief grote groepen proefpersonen zijn uitgevoerd, zijn verschillen in schrijfproces die samenhangen met verschillen in tekstkwaliteit getoetst.

In de studie van Bridwell (1980) wordt nagegaan welke revisievariabelen de beste voorspellers van tekstkwaliteit zijn. Bij de tien als laagst beoordeelde teksten is veel gereviseerd tijdens het produceren van een tekstversie maar juist erg weinig op het moment van overgang van de eerste naar de tweede versie. De meeste revisies betreffen oppervlakteniveau en woordniveau. Bij de tien hoogst beoordeelde teksten is over het geheel genomen niet vaker gereviseerd dan bij de tien laagst beoordeelde teksten, maar wel zijn relatief meer revisies op de overgang van eerste naar tweede versie aangebracht; de hoge scoorders waren meer geneigd tot wijzigen of verwijderen van tekstdelen uit de eerste versie. Verder is onder hen een grote variatie in het aantal revisies. Lage scoorders reviseren doorgaans weinig en als ze het doen alleen op oppervlakte- en woordniveau.

In het onderzoek van Gee (1984) wordt voor eenzelfde, doch veel grotere groep proefpersonen dan in de studie van Bridwell (1980), namelijk 'twelfth graders', een relatie gelegd tussen enkele schrijfproceskenmerken en de tekstscores.

De resultaten wijzen uit dat het maken van een tekstschema en een eerste tekstversie significant positief samenhangen met tekstkwaliteit. Het experimentele onderzoek van Kellogg (1987) ondersteunt deze bevinding. Kellogg vond hogere tekstkwaliteitscores in de conditie waarin schrijvers de instructie hadden gekregen om voorafgaand aan het schrijven eerst een tekstschema te maken. Dit effect deed zich het sterkst gelden op de kwaliteitsdimensies 'ideeënvorming' en 'effectiviteit'. Tevens bleek dat in de tekstschema conditie met een grotere schrijfsnelheid langere teksten geschreven werden. Uit een ter validering van deze laboratorium resultaten uitgevoerd survey onderzoek kwam naar voren dat het maken van tekstschema's samenging met een grotere produktiviteit.

In het onderzoek van Gee is geen significante relatie gevonden tussen aard en aantal revisies enerzijds en tekstkwaliteit anderzijds. In laag gewaardeerde teksten zijn veel stilistische en retorische wijzigingen aangebracht ter uitbreiding van de tekst, ook zijn er meer grote betekenisveranderingen aangebracht. In hoog gewaardeerde teksten zijn veel lexicale en stilistische wijzigingen aangebracht.

Bij kwalitatief goede teksten hebben revisies vaak een kwaliteitsverhogend effect; dat geldt niet altijd voor revisies in kwalitatief slechte teksten. Verder is in dit onderzoek een experimentele manipulatie aangebracht: de experimentele groep reviseert z'n tekst zonder z'n eerste versie erbij te hebben, de controlegroep doet dat met de eerste versie voorhanden. De controlegroep blijkt weinig te reviseren: vrijwel alleen op lexicaal- en zinsniveau, zelden worden grote clusters van ideeën opnieuw geordend of toegevoegd. In de experimentele groep ontstaat een differentiatie tussen wel en niet schrijfvaardige schrijvers. Bij de niet-schrijfvaardige schrijvers is gemiddeld 71% verschil in inhoud tussen beide tekstversies; bij de schrijfvaardige schrijvers is dat 28%. Wanneer tijdens het reviseren de eerste versie voorhanden is, gaat bij niet-schrijfvaardige schrijvers de kwaliteit van de tekst er na revisie op achteruit (1,25 punt op een negenpuntsschaal). Wanneer niet-schrijfvaardige schrijvers bij het reviseren echter genooddaakt zijn vanuit hun geheugen tewerk te gaan, scoort de tweede versie gemiddeld 1 punt hoger dan de eerste. Blijkbaar heeft de aanwezigheid van een vorige tekstversie bij niet-schrijfvaardige schrijvers slechts een inperkend effect, weerhoudt deze versie hen van uitgebreid en produktief reviseren.

Hays (1981) en Roen en Willey (1988) onderzochten de relatie tussen een specifiek schrijfprocesaspect, namelijk publiekgerichtheid, en tekstkwaliteit. In het onderzoek van Hays wordt tevens het verband tussen schrijfvaardigheid en schrijfproces en tussen

publiekgerichtheid en overige procesvariabelen nagegaan.

Roen & Willey (1988) manipuleerden publieksgerichtheid als procesaspect door middel van vragen die de aandacht moesten vestigen op het lezerspubliek. In de ene conditie werden deze vragen voorafgaand aan het schrijven van de tekst aangeboden; in de andere conditie werden de vragen pas aangeboden op het moment dat de schrijver aan het reviseren begon. Alle tekstversies gingen onder invloed van de publiekgerichtheids vragen in kwaliteit vooruit. Het grootste effect werd echter bereikt wanneer het besef van publieksgerichtheid tijdens de reviseerfase werd gestimuleerd.

Ook Hays (1981) concludeert uit zijn materiaal dat een sterke mate van publiekgerichtheid samengaat met een hogere tekstkwaliteit. De teksten van schrijvers die weinig publiekgericht schrijven ontbreekt het aan 'a genuine focus': 'the writer conveyed a lot of information but to no particular end (...) wrote a poorly organized paper devoid of any real point until near the end.'

Zoals ook al uit een aantal eerder besproken studies naar voren kwam, betreffen de revisies van ervaren schrijvers een hoger semantisch niveau. Onervaren schrijvers houden zich vooral bezig met woord- en spellingswijzigingen. Een sterke mate van publiekgerichtheid lijkt samen te gaan met minder oppervlakkige en meer substantiële revisies.

4. DISCUSSIE

De onderzoeksresultaten samenvattend kan gesteld worden dat verschillen tussen goede en slechte schrijvers wat betreft aanpak van de schrijftaak zowel het plannings- als het reviseerproces betreffen (Hillocks, 1986). Goede schrijvers destilleren uit de schrijfpdracht behalve het onderwerp ook andere aspecten van het retorische probleem, nl. het doel van de tekst en het publiek waarvoor deze bestemd is.

Alvorens met het feitelijke schrijven te beginnen staat de goede schrijver een inhoudelijk schema van de tekst voor ogen. Van begin af aan wordt de schrijfaanpak van de goede schrijver geleid door onderling gerelateerde, meestal hiërarchisch geordende plannen. Deze schrijfplannen hebben niet alleen betrekking op de te produceren tekstinhoud en presentatie, maar ook op het schrijfproces zelf, op de volgorde waarin verschillende deelprocessen in werking moeten treden. Er is dan ook een doelgericht en flexibel wisselen van het ene proces naar het andere mogelijk (bijvoorbeeld van herlezen naar genereren van nieuwe inhoud).

Goede schrijvers analyseren hun eigen tekst beter en reviseren op zowel oppervlakkig niveau (bijvoorbeeld grammaticaal of stilistisch) als ook op meer diepgaand niveau van de betekenis van de tekst. Reviseren vindt bij hen tijdens het gehele schrijfproces plaats: al reviserend wordt de tekst ontwikkeld.

Wat opvalt is dat er enige analogie lijkt te bestaan tussen een aantal aspecten van het planningsproces en aspecten van het reviseerproces:

- Bij de planning ontbreekt het slechte schrijvers aan een hiërarchisch netwerk van onderling gerelateerde doelen en procesplannen. Bij het reviseren ontbreekt het hen aan een hiërarchische structuur van onderling gerelateerde tekstproblemen.
- Bij slechte schrijvers heeft zowel planning als revisie een lokaal karakter: het vindt vooral plaats op het niveau van woorden en zinnen en niet op het niveau van tekstbetekenissen die zinsoverstijgend zijn en die met tekstdoel en lezerspubliek te maken hebben.
- Bij slechte schrijvers werkt de eenmaal geproduceerde tekst beperkend: dat geldt voor de eerste zin ten aanzien van de verdere planning. Evenzo geldt het voor de reeds geschreven versie ten aanzien van het produceren van een nieuwe, verbeterde

tekstversie. Hays (1981) suggereert dat slechte schrijvers 'fielddependent' zijn, dat wil zeggen dat ze niet in staat zijn zich los te maken van de eigen referentiekaders.

- Slechte schrijvers hebben een voorkeur voor het toepassen van algoritmische regels in plaats van heuristische regels of richtlijnen; dat geldt inzake de planning (Rose 1980) en eveneens voor het stellen van diagnoses in het kader van het revisieproces (Hayes 1985).

Hoewel de conclusies uit de onderzochte studies elkaar voor een belangrijk deel bevestigen is een exacte vergelijking van de resultaten uit verschillende onderzoeken niet goed mogelijk, ten gevolge van de variatie in operationalisatie van de variabelen 'schrijfvaardigheid', 'schrijftaak' en 'schrijfproces'.

Wat betreft het niveau van schrijfvaardigheid is er de variatie in de bepaling hiervan als persoonskenmerk, dan wel als produktkenmerk. In de resultaten zijn, afhankelijk van dit verschil in operationalisatie, geen inconsistenties te zien. Wel biedt de operationalisatie van schrijfvaardigheid als produktkwaliteit de mogelijkheid om schrijfprocessen met meer specifieke kwaliteitskenmerken van de tekst in verband te brengen.

Het feit dat er weinig variatie in onderzoeksgroepen is doet de vraag rijzen in hoeverre de resultaten ook naar andere groepen gegeneraliseerd kunnen worden. De resultaten uit de onderzoeken waarin ook professionele volwassen schrijvers van meer diverse leeftijden proefpersoon waren (Flower & Hayes 1980; Sommers 1980; Faigley 1981; Hayes 1985) wijken echter niet duidelijk af van de resultaten uit de overige onderzoeken. De resultaten en conclusies uit de uitgebreide onderzoeken van Scardamalia en Bereiter (1987) wijzen daarentegen wel op de invloed van de factor leeftijd, d.w.z. cognitieve ontwikkelingsfase, op schrijfvaardigheid. Volgens deze auteurs kunnen de schrijfprocessen van jeugdige tegenover volwassen schrijvers het best door middel van twee verschillende modellen gerepresenteerd worden. De jeugdige schrijver (t.m. ongeveer 15 jaar) genereert inhouden, uitgaande van kennis over het onderwerp en over de beoogde tekstsoort (knowledge telling strategy). Bij hen zijn variaties in tekstkwaliteit het gevolg van variatie in onderwerpskennis, kennis van het betreffende tekstgenre en formuleervaardigheid. Verder hangt de variatie af van de mate waarin de jonge schrijver in staat is om bij het genereren niet slechts aan te sluiten bij het onderwerp of het laatste inhoudsitem, maar vanuit de combinatie van het onderwerp en meerdere inhoudsitems tegelijk verdere inhouden te genereren. De volwassen schrijver, vooral de expert, gaat uit van zowel een inhoudelijke als een retorische probleemruimte (Flower & Hayes 1981). Typerend voor het schrijven van de volwassen schrijver volgens de knowledge transforming strategy is, dat er naast inhoudsplanung ook veel conceptuele planning plaatsvindt. Inhoudsideeën hebben diverse niveau's van abstractie en worden vanuit gestelde doelen geëvalueerd en gestructureerd (b.v. combinatie, opsplitsing, herordening, nadere uitwerking). Wanneer men het schrijfproces van jeugdige schrijvers onderzoekt zal dus de variatiebreedte in planningsprocessen kleiner zijn dan bij een onderzoeksgroep volwassen schrijvers, ten gevolge van het nog grotendeels ontbreken van conceptuele planning.

Dat de aard van de schrijfopdracht van invloed kan zijn op het schrijfproces tonen de resultaten van het onderzoek van Jansen en Schilperoort (1987) aan. In dit onderzoek werd het schrijven van een persoonlijke brief met het schrijven van een zakelijke brief vergeleken.

Ook de procesregistratiemethode, de procesanalyse methode en de combinatie van beide verschillen per onderzoek. Voor de meermalen gehanteerde procesregistratiemethode van het hardop-denken geldt, juist inzake onderzoek naar verschillen tussen goede en slechte schrijvers, een validiteitskwestie: bestaat er geen gevaar van contaminatie van schrijfvaardigheid en vaardigheid in het hardop-verbaliseren van het eigen denkproces?

Hayes (1985) haalt enkele kritiekpunten aan van Cooper en Holzman (1985) die betrekking hebben op het gebruik van hardopdenkprotocollen bij schrijfprocesonder-

zoek. Een van deze kritiekpunten betreft het mogelijke verschijnsel dat goede of ervaren schrijvers zich ook van slechte of onervaren schrijvers zouden onderscheiden in hun vermogen tot verbaliseren van schrijfsprocesactiviteiten. Dat zou betekenen dat studies zoals die welke in dit literatuurverslag besproken zijn niet alleen verschillen in het feitelijke schrijfsproces van goede tegenover slechte schrijvers weerspiegelen, maar tevens verschillen in het vermogen om het denkproces te verwoorden.

Enerzijds zijn (volgens Hayes 1985) bij goede schrijvers bepaalde subprocessen zodanig routine geworden dat ze niet meer verwoord worden, anderzijds zullen goede schrijvers dankzij een groter verbaal vermogen juist beter in staat zijn bepaalde denkstappen te verwoorden. Hayes (1985) verdedigt zich met de bewering dat hoger niveau-processen, zoals het selecteren van relevante inhoud of reviseren, afgezien van iemands schrijfvaardigheid, nooit geautomatiseerd zullen kunnen verlopen. Verder vat Hayes (1985) dit bezwaar op als een onvermijdelijk gegeven dat inherent is aan dergelijk differentieel schrijfsprocesonderzoek. Naar mijn mening is de veronderstelling dat de verbalisaties van denkprocessen niet helemaal identiek zijn aan de denkprocessen zelf met deze verdediging van Hayes (1985) niet helemaal ontkracht. Bovendien wijst Hayes' commentaar erop dat het verband tussen de verbalisaties en het feitelijke denkproces niet eenduidig is: het verband verschilt per type schrijver en per deelproces. Het bezwaar van Cooper en Holzman (1985) lijkt me overigens ook te gelden voor retrospectief schrijfsprocesonderzoek waarin goede en slechte schrijvers worden vergeleken.

Met betrekking tot de categoriseringssystemen die gebruikt zijn voor analyse van het schrijfsproces zijn de besproken studies niet alle even duidelijk.

Bij een aantal auteurs blijkt uit hun verslaggeving niet welke categorieën ze hebben gebruikt voor analyse van het schrijfsproces (Beach 1976, Garrett 1981, Taylor 1984). In deze gevallen ontbreekt ook de statistische data-analyse, zodat de bespreking van de onderzoeksresultaten niet geheel controleerbaar is. In veel onderzoeksrapportages zijn wél de analysecategorieën vermeld, doch zonder nadere uitleg, en dat maakt de conclusies soms slecht interpreteerbaar en onderling vergelijkbaar. Een voorbeeld hiervan is het onderzoek van Bank (1982) waaruit naar voren komt dat de procesactiviteit 'discovering' positief gerelateerd is aan tekstkwaliteit. Wat onder deze procesactiviteit verstaan wordt, wordt echter nergens in het verslag vermeld. Soms verklaren auteurs wel de constructiewijze van het door hen toegepaste systeem, maar niet de betekenis van elk der resulterende categorieën apart. Dit geldt voor een aantal revisie-analysesystemen (Sommers 1980, Faigley 1981, Hays 1981, Monahan 1982). Omdat in dit geval revisies meestal getypeerd worden naar het tekstniveau en naar de daarop toegepaste operatie ontstaat enige onduidelijkheid wanneer de mogelijke combinaties van tekstniveaus en operaties niet apart worden benoemd en uitgelegd: wat is bijvoorbeeld de substitutie van een 'clause' met een stilistisch doel, aangebracht op de overgang van de eerste naar de tweede tekstversie?

Onduidelijkheid over de exacte betekenis van veel procesanalysecategorieën maakt het moeilijk te bepalen in hoeverre ongeveer gelijklopende categorieën hetzelfde betekenen en in hoeverre categorieën met verschillende benamingen verschillende procesactiviteiten aanduiden.

De auteurs die de meest duidelijke uitleg geven aan de door hen gehanteerde categorieën zijn Perl (1979), Pianko (1979), Bridwell (1980), Selfe (1984), Hayes (1985), Kennedy (1985).

In een beperkt aantal gevallen wordt ook de betrouwbaarheid van het categoriseringssysteem vermeld.

In de meeste onderzoeken zijn frequentieverschillen per schrijfsprocesactiviteit (analyse-categorie) tussen de verschillende onderzoeksgroepen en/of niveaus van tekstkwaliteit nagegaan. In minder dan de helft van de geanalyseerde studies zijn statistische

significantietoetsen gebruikt. Doorgaans zijn dit toetsen op frequenties (Chi-kwadraattoets) of op gemiddelde waarden (Variantie analyse, t-toets, Mann-Whitney). In enkele gevallen is regressie-analyse of discriminant analyse uitgevoerd.

In de overige studies, waarin geen statistische significantietoetsing plaatsvindt (Beech 1976, Perl 1979, Warters 1979, Flower en Hayes 1980, Sommers 1980, Faigley 1981, Garrett 1981, Hays 1981, Sweeder 1981, Selfe 1984, Taylor 1984), zijn uitspraken slechts gebaseerd op op het oog beoordeelde verschillen in frequenties.

Om de mate van wetenschappelijke verantwoording van de diverse studies te beoordelen zou als criterium kunnen gelden de operationalisatie van 'schrijfvaardigheid' en 'schrijfproces' en de keuze van statistische analysetechnieken. 12 Van de 25 studies voldoen dan aan het criterium dat:

- a. er m.b.t. het niveau van schrijfvaardigheid van de proefpersonen of de tekstkwaliteit de gehanteerde normen of criteria geëxpliciteerd zijn (soms met vermelding van interbeoordelaarsbetrouwbaarheid bij tekstkwaliteitsbeoordelingen) en
- b. de toegepaste procesanalysecategorieën geëxpliciteerd zijn, al dan niet voorzien van achterliggende theorie en betekenis (soms met vermelding van de intercodeursbetrouwbaarheid) en
- c. de verbanden tussen schrijfvaardigheid / tekstkwaliteit en schrijfproces aan statistische toetsing zijn onderworpen.

De studies die hieraan voldoen zijn: Stallard (1974), Pianko (1979), Atlas (1979), Bridwell (1980), Schumacher (1982), Monahan (1982), Wright (1983), Hull (1984), Gee (1984), Kennedy (1985), Hayes (1985), Kellogg (1987), Roen en Willey (1988).

In geen van de onderzoeken worden opeenvolgingen van processen geanalyseerd; er wordt in slechts enkele gevallen gekeken naar de duur van bepaalde procesactiviteiten en naar de procesfase waarin een bepaalde procesactiviteit voorkomt. Toch zijn er enkele bevindingen die er op wijzen dat dit een interessante invalshoek is; het gaat hier om bevindingen die aangeven dat het moment in het hele schrijfproces waarop een bepaalde activiteit optreedt gerelateerd kan zijn aan schrijfvaardigheid. Zo vond Kennedy (1985) dat goede schrijvers documentatie herlezen voordat ze gaan schrijven terwijl slechte schrijvers dat pas doen tijdens het schrijven. Een ander voorbeeld is de bevinding van Bridwell (1980) dat reviseren op het moment van overgang van de eerste naar de tweede tekstversie positief gerelateerd is aan de mate van schrijfvaardigheid en dat naarmate revisies later in het schrijfproces worden aangebracht ze op meer tekstniveaus betrekking hebben. Roen en Willey (1988) vonden dat het vestigen van de aandacht op aspecten van publieksgerichtheid tijdens de reviseerfase effectiever was dan wanneer dit tijdens het gehele schrijfproces gebeurde.

De monitorfunctie die door Flower en Hayes wordt verondersteld (zie p. 19) zou zich moeten manifesteren in patronen van processen, waarachter bepaalde configuraties van doelen liggen. Via analyse van proces-opeenvolgingen, patronen van processen, zouden eventuele verschillen in configuraties van schrijfdoelen tussen goede en slechte schrijvers aan het licht kunnen komen.

Er zijn relatief weinig onderzoeken waarin procesactiviteiten of -kenmerken gerelateerd worden aan de kwaliteit van uit het betreffende schrijfproces resulterende teksten. Zoals reeds eerder is opgemerkt is schrijfvaardigheid meestal als een schrijverskenmerk en niet als een produktkenmerk geoperationaliseerd. In de enkele studies waarin dit wél het geval is, is de tekstkwaliteit steeds getypeerd door een totaalscore, die overigens resultaat is van beoordelingen op aparte kwaliteitsaspecten. Helaas zijn deze aspecten (zoals organisatie, inhoud, publiekgerichtheid) tamelijk globaal en zijn de beoordelingen per aspect nergens vermeld.

In geen enkel van de geanalyseerde studies wordt dus een relatie gelegd tussen schrijfprocesactiviteiten en specifieke kenmerken van tekstkwaliteit, terwijl dat een zuiverder en meer uitgewerkte operationalisatie van tekstkwaliteit zou vormen dan de schrijfvaardigheid van de betreffende schrijver in het algemeen.

De resultaten uit onderzoek naar directe relaties tussen bepaalde schrijfprocesaspecten en specifieke tekstkenmerken zouden aanknopingspunt kunnen vormen voor meer gedetailleerde en effectieve instructies voor de wijze van aanpak van schrijftaken.

LITERATUUR

- Atlas, M.A. (1979). *Adressing an audience: a study of expertnovice differences in writing*. (ERIC Document Reproduction Service No. ED 192-338)
- Bridwell, L.S. (1980, October). Revising strategies in twelfth grade students transactional writing. *Research in the Teaching of English*, 197-222.
- Bank, S. (1982). *Basic writers perceive the process of composing*. (ERIC Document Reproduction Service No. ED 214 188)
- Beach, R. (1976). Selfevaluation strategies of extensive revisers and nonrevisers. *College Composition and Communication*, 27, 160-164.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. New Jersey: Lawrence Erlbaum Associates.
- Cooper, M. & Holzman, M. (1985). Counterstatement reply. *College Composition and Communication*, 36, 97-100.
- Ericson, K.A., & Simon, H.A. (1984). *Protocol analysis. Verbal reports as data*. Cambridge: MIT press.
- Faigley, L. (1981). *Using text structure models for analyzing revision*. (ERIC Document Reproduction Service No. ED 200-978)
- Flower, L.S., & Hayes, J.R. (1980a). Identifying the organization of writing processes. In L.W. Greg, Steinberg, *Cognitive processes in writing*. New Jersey.
- Flower, L.S., & Hayes J.R. (1983). *A cognitive model of the writing process in adults*. (final report) (ERIC Document Reproduction Service No. ED 240-608)
- Garrett, P.W. (1981). *Revision, analysis of the revision strategies of college writers*. (ERIC Document Reproduction Service No. ED 199 760)
- Gee, Th.W. (1984). *Drafting and revising processes in Grade twelve students examination writing*. (ERIC Document Reproduction Service No. ED 269770)
- Hayes, J.R. (1985). *Cognitive processes in revision*. (ERIC Document Reproduction Service No. ED 267 396)
- Hays, J.N. (1981). *The effect of audience considerations upon the revisions of a group of basic writers and more competent junior and senior writers*. (ERIC Document Reproduction Service No. ED 204-802)
- Hillocks, G.Jr. (1981). *Research on written composition*. NCRE ERIC.
- Hull, G.A. (1984). *The editing process in writing: a performance study of experts and novices*. (ERIC Document Reproduction Service No. ED 245-254).
- Janssen, D., & Schilperoord, J. (1988). Schrijfprocessen en schrijftaken. *Tijdschrift voor Taalbeheersing*, 10, 105-120.
- Janssen, T. (1987). *Kenmerken van schoolse en naschoolse schrijftaken*. Amsterdam: SCO.
- Kellogg, R.T. (1987). Writing preformance: effects of cognitive strategies. *Written Communication*, 4, 269-298.
- Kennedy, M.L. (1985). The Composing process of college students writing from sources. *Written Communication*, 2, 434-456.
- Monahan, B.D. (1982). *Revision strategies of basic and competent writers as they write for different audiences*. (ERIC Document Reproduction Service No. ED 229 756)
- Perl, S. (1979). The composing processes of unskilled college writers. *Research in the Teaching of English*, 13, 317-336.

- Pianko, S. (1979). A description of the composing processes of college freshmenwriters. *Research in the Teaching of English*, 13, 5-22.
- Roen, D.H. & Willey, R.J. (1988). The effects of audience awareness on drafting and revising. *Research in the Teaching of English*, 22 (1), 75-88.
- Root, R.L. (1985). *Assiduous stringsavers: the ideagenerating strategies of professional expository writers*. (ERIC Document Reproduction Service No. ED 258-205)
- Rose, M. (1980). Rigid rules, inflexible plans and the stiffling of language: a cognitivist analysis of writers block. *College Composition and Communication*, 31, 389-400.
- Schumacher, G.M. et al. *Cognitive processes during pauses in writing*. (ERIC Document Reproduction Service No. ED 236 624)
- Selfe, C.L. (1984). The predrafting processes of four high- and four lowapprehensive writers. *Research in the Teaching of English*, 18, 45-64.
- Sommers, N. (1980) Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31, 378-388.
- Stallard, C.K. (1974). An analysis of the writing behavior of good student writers. *Research in the Teaching of English*, 8, 206-218.
- Sweeder, J.J. (1981). A descriptive study of six adult remedial writers: their composing processes and heuristic strategies. *Dissertation Abstracts International*, 42/05-A, 2004.
- Taylor, K.K. (1984). Eighteen writers writing: the processes of inexperienced and experienced writers. *Teaching English in the Two Year College*, 10, 219-225.
- Taylor, K.K. (1984). The different summary skills of inexperienced and professional writers. *Journal of Reading*, 27, 691-699.
- Warters, S. (1979). The writing processes of college basic writers. (ERIC Document Reproduction Service No. ED 175 008)
- Wolters, G., R.H. Phaf (1988). Connectionisme: een nieuwe uitdaging voor de psychologie. *De Psycholoog*, 23, 361-369.
- Wright, J.R. (1983). The writing processes of college frehsmen. *Dissertation Abstracts International*, 44/04-A, 967.

Manuscript ontvangen 10-2-1989

Definitieve versie ontvangen 9-6-1989

BEKROOND ORD PAPER 1989

Dit artikel is een bewerking van het door de Vor bekroonde ord-paper

Effecten van een metacognitief trainingsprogramma

Harm J.A. Biemans¹

Sectie Onderwijs- en Opleidingspsychologie, Katholieke Universiteit Brabant²

ABSTRACT

In this study the effects of a training in the use of metacognitive skills on the basis of metacognitive knowledge were examined. The training was embedded in extra exercises in a mathematics course of a comprehensive school. The study makes use of a pretest-posttest-control-group design. Half of 26 low-achieving pupils (average age 13) were trained during nine sessions. The training had a positive effect on the metacognitive knowledge as measured by a situational questionnaire. From the analysis of thinking-aloud protocols it could be derived that, as a result of the intervention, more metacognitive (especially testing) activities took place when the trained pupils did a problem-solving task. There was no significant effect of the training on the performances on a corresponding test, but there was a trend in that direction. During several training sessions the experimental group performed better and more homogeneous than the control group. A longer training with a greater emphasis on monitoring and directing and with a greater integration in the regular curriculum is recommended to increase the possibility of internalisation and transfer taking place.

INLEIDING

Een belangrijk gedeelte van de mentale capaciteit wordt bij leerprocessen in beslag genomen door transformatie-activiteiten die aangewend worden om de informatie uit de leertaak te internaliseren. Derry en Murphy (1986, p. 2) omschreven deze cognitieve activiteiten als 'specific skills one uses in service of the strategy'. Met specifiek wordt in deze bedoeld dat er sprake is van een bepaalde mate van taakafhankelijkheid. In de loop der jaren zijn er vele trainingsprogramma's in transformatie-activiteiten ontwikkeld voor leerlingen met leerproblemen. Men is er echter volgens velen niet in geslaagd om door middel van dergelijke interventies door te dringen tot de kern van de problematiek van vele leerlingen. Campione (1987, p. 125) spreekt in dit verband van 'attacking the symptom, but not the cause of the problem'. De oorzaak van vele leerproblemen lijkt volgens hem en anderen eerder gelegen op het niveau van de metacognitie.

Reeve en Brown (1985, p. 344) omschreven metacognitie als 'the individuals' abilities to understand what is required, to understand their own capacities, to plan strategies that will allow them to reach the goal and to monitor and coordinate these activities'. In deze

¹ Mijn dank gaat uit naar Frank de Jong en Johan van der Sanden

² Adres: Katholieke Universiteit Brabant, Sectie Onderwijs- en Opleidingspsychologie (P 62), Postbus 90153, 5000 LE Tilburg

omschrijving wordt metacognitie enerzijds gezien als kennis en opvattingen van leerlingen omtrent de variabelen die in interactie hun leerprocessen beïnvloeden (metacognitieve kennis) en anderzijds als vaardigheden die leerlingen in staat stellen om controle uit te oefenen over hun leerprocessen ('executive control').

Zowel waar het metacognitieve kennis als waar het metacognitieve vaardigheden betreft, blijken er grote verschillen tussen leerlingen te bestaan die gerelateerd zijn aan leerprestaties. Zo vond Schneider (1984) een gemiddelde correlatie van .41 tussen metacognitieve zelfbeoordelingen en prestaties voor verschillende categorieën van taken en leeftijdsgroepen. Pressley (1986) constateerde dat minder succesvolle leerlingen zichzelf een passieve rol toebedelen in onderwijsleersituaties in tegenstelling tot leerlingen die goede leerresultaten boeken. Verder blijken minder succesvolle leerlingen relatief minder vaak de effecten van de door hen gehanteerde transformatie-activiteiten te bewaken. Ze hebben bijvoorbeeld minder oog voor de (tussen)resultaten die ze boeken (Brown, Campione en Barclay, 1979). Ook stemmen minder succesvolle leerlingen de aard van hun toetshandelingen minder af op het te bereiken doel (Vermunt, 1984).

Daar metacognitie blijkbaar samenhangt met leerresultaten, werpt de vraag zich op hoe men minder succesvolle leerlingen kan leren om op basis van een adequaat metacognitief kennisbestand hun cognitieve activiteiten te bewaken en indien nodig bij te sturen ('leren leren').

Volgens Brown, Palincsar en Armbruster (1984) moeten trainingsprogramma's die gericht zijn op het 'leren leren' niet alleen oefeningen in transformatie-activiteiten bevatten. Ze zouden tevens expliciet gericht moeten zijn op de metacognitieve kennis van leerlingen alsmede op hun daadwerkelijke controle-activiteiten. Dergelijke trainingsprogramma's waarin aandacht wordt geschonken aan deze drie facetten bestaan al geruime tijd en een aantal ervan is zeer succesvol gebleken (Brown, Palincsar en Armbruster, 1984; Scardamalia, Bereiter en Steinbach, 1984).

Het leeuwedeel van dit metacognitieve trainingsonderzoek richtte zich echter op leerlingen met ernstige leerproblemen. Men kan zich dan ook afvragen in hoeverre de resultaten van dergelijke trainingsstudies te generaliseren zijn naar leerlingen die niet echt met zware leerproblemen kampen. Is het, met andere woorden, ook mogelijk om de kwaliteit van de metacognitie van dergelijke leerlingen te verbeteren?

Om dit na te gaan werd een trainingsprogramma ontwikkeld waarin overdracht van metacognitieve kennis en oefening in metacognitieve vaardigheden aan de hand van extra wiskunde-opgaven centraal stonden. Concreet golden de volgende vraagstellingen met betrekking tot het effect van de training:

- I Heeft de training een effect op de metacognitieve kennis van de leerlingen over activiteiten die gebruikt kunnen worden in verschillende leersituaties bij het zelfstandig maken van huiswerk?
- II Heeft de training een effect op het aantal metacognitieve activiteiten dat door de leerlingen aangewend wordt bij het zelfstandig maken van een probleem-oplostaak?
- III Heeft de training een effect op de leerprestaties op een toets over de probleem-oplostaak?
- IV Heeft de training een effect op het aantal opgaven dat tijdens de vaardigheidstrainings-sessies goed opgelost wordt?

METHODE

Proefpersonen

Aan het onderzoek namen 26 leerlingen uit twee verschillende eerste-brugklassen van een HAVO/VWO-scholengemeenschap deel. Op basis van hun paasrapportcijfers voor het

vak wiskunde (criterium: < 7) en de klas waarin zij zaten, werden zij gematched in paren. De leerlingen die een paar vormden, werden vervolgens at random aan de experimentele of aan de controlegroep toegewezen om te controleren voor een systematische invloed van andere, mogelijk relevante variabelen. Het gemiddelde op het paasrapport voor de trainingsgroep was gelijk aan 5.67 ($sd = 1.03$) terwijl dit voor de controlegroep 5.68 ($sd = 0.98$) bedroeg. In de beide groepen zaten evenveel leerlingen uit beide klassen.

Design

Het gehanteerde design was een zogenaamde voortest-natest-controlegroep-onderzoeksopzet. Kenmerken hiervan zijn zowel de aanwezigheid van een trainingsgroep en een controlegroep als van een voortest en een natest.

De metacognitieve kennis van de leerlingen werd zowel bij de voor- als bij de nameting op dezelfde wijze geregistreerd met behulp van de Vragenlijst Metacognitieve Kennis. In deze vragenlijst werden tien situaties geschetst waarin de leerlingen terecht konden komen bij het zelfstandig maken van wiskundehuiswerk. Deze situaties correspondeerden met de verschillende leersituaties uit de trainingsheuristiek. De leerlingen werd gevraagd of ze in die situaties iets zouden doen (b.v. 'Zou je iets doen om er achter te komen hoe je een som moet oplossen wanneer je niet weet hoe je dit moet doen?') en zo ja, welke leeractiviteiten ze in die situaties zouden aanwenden. De antwoorden van de leerlingen werden gescoord op de aanwezigheid van activiteiten die gerelateerd waren aan de heuristiek.

De effecten van de training op de metacognitieve activiteiten van de leerlingen werden opgespoord middels een analyse van hardop-denkprotocollen. Zowel bij de voor- als bij de nameting dienden alle leerlingen individueel en hardop denkend een probleem-oplostaak op het gebied van de kansberekening te maken. Er werd voor deze transfertaak gekozen omdat de taak in diagnostische studies die voorafgingen aan dit onderzoek, eveneens gebruikt was (De Jong, 1987 en 1988). Van de probleem-oplostaak bestonden twee versies die gecounterbalanced waren over personen, groepen en voor- en natest. De ene versie had betrekking op 'gelijke kansen' en de andere op 'ongelijke kansen'. In beide versies was theorie over het onderwerp gekoppeld aan oefenopgaven die de leerlingen dienden op te lossen. Ter illustratie volgt hier een oefenopgave uit versie I: 'Een schriftelijk proefwerkje bestaat uit 2 vierkeuzevragen. Een leerling die niets geleerd heeft, vult de vragen willekeurig, op goed geluk in. Hoe groot is nu de kans dat hij beide vragen goed heeft?'. Bij het maken van de probleem-oplostaak dienden de leerlingen zoals gezegd hardop te denken. Hun verbalisaties werden op audio-tape opgenomen, letterlijk uitgetypt en geanalyseerd op het voorkomen van metacognitieve activiteiten met behulp van een daarvoor ontwikkeld analyseschema (De Jong, 1987; De Jong en Biemans, 1988). Het analyseschema kende 6 categorieën die elk verschillende concrete activiteiten omvatten. De categorieën uit dit analyseschema waren transformeren (b.v. lezen), oriënteren (b.v. aangeven van de studiewijze: 'thuis lees ik de theorie altijd eerst door'), proces bewaken (b.v. constateren van negatieve tussenresultaten: 'ik snap die som niet'), sturen (b.v. kiezen van een activiteit: 'ik ga nu de som maken'), toetsen (b.v. controleren van de oplossingen van de oefenopgaven: 'de uitkomst is 1/16, ja') en 'off-task' opmerkingen (b.v. procedurele vragen stellen). Uit eerder onderzoek was gebleken dat het analyseschema een hoge interbeoordelaarsbetrouwbaarheid kende: voor de kansberekeningstaak golden Pearson-correlaties van .97 tot .99 tussen 4 codeurs wat betreft de frequenties van de verschillende gecodeerde activiteiten (De Jong, 1988).

Bij wijze van prestatiemeting werd zowel bij de voor- als bij de natest een toets afgenomen die betrekking had op de stof uit de bewuste versie van de kansberekeningstaak. Beide toetsversies bestonden uit 10 kansberekeningssommen die vergelijkbaar waren met de oefenopgaven uit de taak. De leerlingen dienden de toetssommen op te lossen na het maken van de probleem-oplostaak.

Voor de laatste 5 trainingsbijeenkomsten werd per sessie het aantal wiskunde-opgaven, dat door de leerlingen goed opgelost werd, bepaald. Op deze manier werden de prestaties van de leerlingen gedurende de training gemeten. Het gemiddelde aantal opgaven per sessie was 12. De opgaven hadden direct betrekking op de leerstof die in de voorafgaande wiskundelessen aan de orde geweest was, met uitzondering van de laatste bijeenkomst, waar leerstof als basis werd genomen die nog niet behandeld was in de klas. Ter illustratie volgt hier een opgave uit de voorlaatste sessie: 'Bereken beide kruisprodukten van de volgende evenredigheden: $2/5 = 8/20$ '.

Procedure en training

Alle leerlingen werden voorafgaande aan de voortest gedurende 25 minuten individueel getraind in het verbaliseren van hun gedachten bij het maken van een probleem-oplostaak. Dit geschiedde, net zoals alle andere bijeenkomsten, op de school van de leerlingen na afloop van de reguliere schooldag. Deze eerste bijeenkomst had verder als doel om een vertrouwelijke, ongedwongen sfeer te scheppen.

Zoals eerder vermeld, dienden alle leerlingen zowel bij de voor- als bij de nameting individueel en hardop denkend een probleem-oplostaak op het gebied van de kansberekening te maken. De twee groepen ontvingen bij beide metingen dezelfde instructie en kregen dezelfde hoeveelheid tijd toebedeeld (maximaal 30 minuten).

Gedurende de 9 bijeenkomsten tussen de voor- en de nameting vond de daadwerkelijke training van de leerlingen uit de experimentele groep plaats. Het trainingsprogramma, dat de naam 'Huiswerk maken kun je zelf' droeg, had ten doel metacognitieve kennis over te dragen en metacognitieve vaardigheden te trainen. Oefening in transformatie-activiteiten vond in de training niet plaats aangezien hier in de reguliere wiskundelessen al aandacht aan besteed werd.

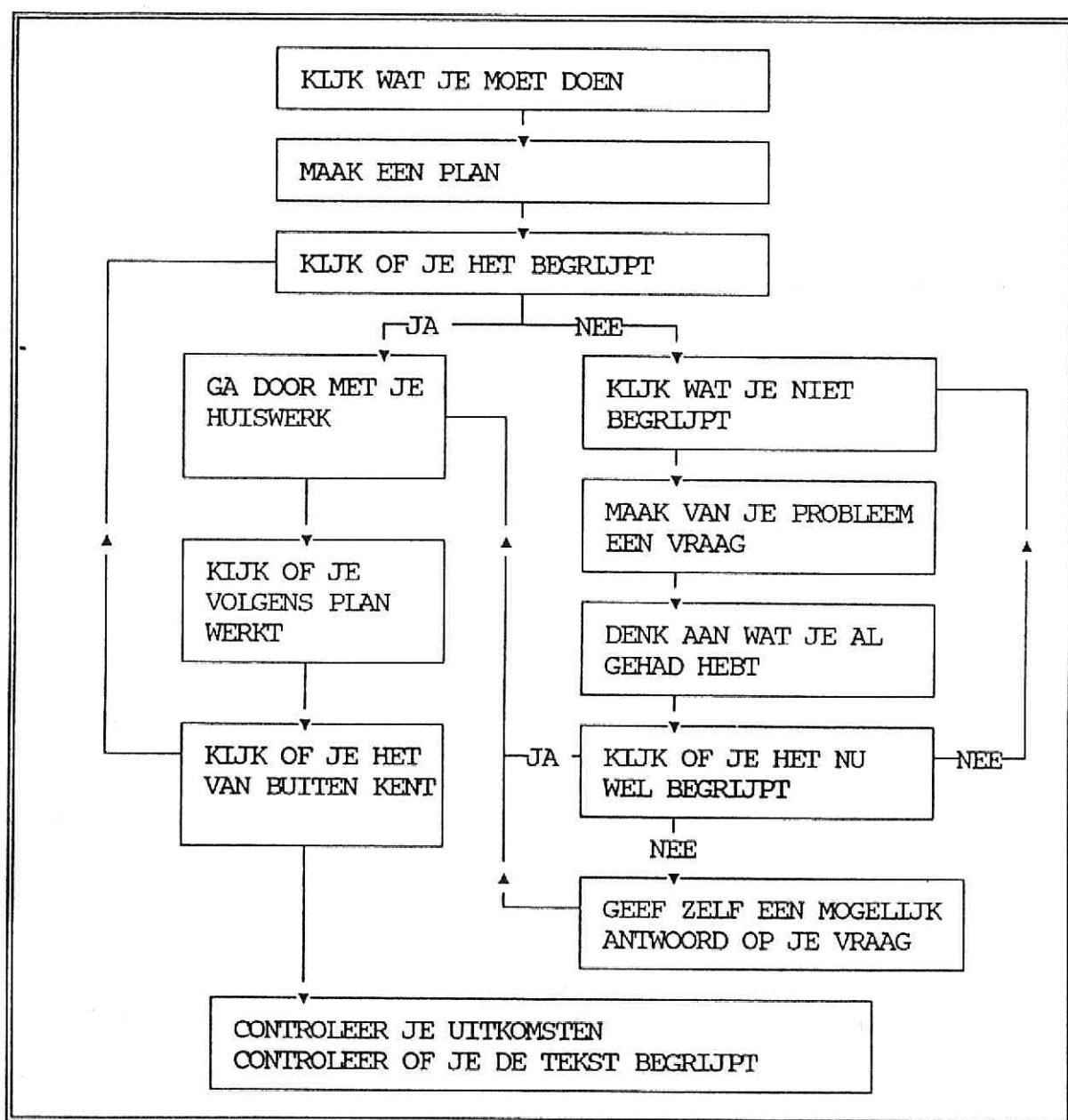
In de eerste 3 bijeenkomsten van de training, die gemiddeld 45 minuten duurden, lag het accent op het overdragen van metacognitieve kennis aan de leerlingen. Het centrale doel van deze groepsgewijze bijeenkomsten was dat de leerlingen zich bewust werden van de verschillende metacognitieve activiteiten die ideaal gesproken, afhankelijk van de leersituatie, door een succesvolle leerling gehanteerd worden. Om dit te bewerkstelligen moest elke leerling zijn eigen leeractiviteiten, zoals gerapporteerd op de Vragenlijst Metacognitieve Kennis, vergelijken met de leeractiviteiten die een prototypische succesvolle leerling in dezelfde leersituatie hanteerde. Wanneer door een leerling geen of ongeschikte activiteiten genoemd waren als antwoord op de geschetste leersituaties, werden de nadelen hiervan door de proefleider aan de betreffende leerling uitgelegd. De leeractiviteiten van een tweede prototypische leerling die model stond voor een minder succesvolle leerling en die derhalve een soort 'spiegelfunctie' voor de leerlingen moest vervullen, werden hierbij groepsgewijs besproken. Voor geschikte alternatieven werd verwezen naar de leeractiviteiten van het succesvolle prototype. Het gebruik van deze leeractiviteiten werd gerechtvaardigd door het aanvoeren van een aantal redenen die ook in de trainingsheuristiek opgenomen waren. Aldus werd getracht om de leerlingen te doordringen van de relatie tussen het hanteren van geschikte leeractiviteiten en het boeken van goede leerresultaten. Het belang van het hanteren van een actieve aanpak en het dragen van een eigen verantwoordelijkheid voor het leerproces en het leerresultaat werden hierbij benadrukt. Dit alles geschiedde in de vorm van een geleide discussie waarbij de leerlingen frequent aangespoord werden om hun eigen werkwijze te beargumenteren en te reageren op hetgeen door de proefleider gezegd werd.

De werkwijze van het succesvolle prototype was rechtstreeks afgeleid van de heuristiek die de kern van de training vormde. Deze heuristiek was enerzijds gebaseerd op een

verzameling van metacognitieve activiteiten die in eerder uitgevoerd onderzoek veel variantie in de leerprestaties op de toets over de kansberekeningstaak verklaarde (De Jong, 1987) en anderzijds op een aantal bestaande heuristiek op het gebied van probleem-oplossen (Polya, 1973; Charles en Lester, 1984; Derry en Hawkes, 1985). De in de heuristiek opgenomen leeractiviteiten waren: oriënteren, plannen, begrip toetsen, werkwijze controleren, memoriseren, probleemgerichte actie ondernemen (probleem identificeren, probleem als vraag formuleren, voorkennis actualiseren en hypothesen vormen) en evalueren.

De heuristiek werd op vier verschillende niveau's in gebundelde vorm aan de leerlingen aangeboden:

- als stroomdiagram waarbij de genoemde leeractiviteiten en de relaties ertussen gevisualiseerd waren (zie figuur 1);



Figuur 1. Het stroomdiagram met de getrainde metacognitieve activiteiten en hun onderlinge relaties (niveau 1 van de heuristiek)

- als opsomming waarbij de leeractiviteiten tekstueel weergegeven waren als opeenvolgende stappen (b.v. 'Maak een plan');
- als vragenlijst waarbij de leeractiviteiten omgezet waren in vragen die de leerlingen zichzelf konden stellen (b.v. 'Wat ga ik allemaal doen om mijn huiswerk te maken?');
- als uitgebreide vragenlijst die dezelfde vragen bevatte maar nu met informatie omtrent de bedoeling van de vragen ('WAT'-blok), de wijze waarop ze beantwoord konden worden ('HOE'-blok) en het belang van de corresponderende activiteiten ('WAAROM'-blok). Bij wijze van voorbeeld volgen hieronder het 'WAT'-blok, het 'HOE'-blok en het 'WAAROM'-blok behorende bij de vraag 'Wat ga ik allemaal doen om mijn huiswerk te maken?' (zie figuur 2).

- WAT:** Stel een plan op voor het leren of het maken van je huiswerk. Bepaal dus wat je eerst gaat doen, wat daarna, wat daarna enzovoorts. Probeer ook te zeggen op welke manier je dat allemaal wilt gaan doen.
- HOE:** Denk er eens over na hoe je normaal gesproken een dergelijke taak zou bestuderen of hoe je dit in het verleden al eens gedaan hebt. Kijk eens hoe de taak eruit ziet en waar hij over gaat.
- WAAROM:** Zo'n plan geeft je al een houvast voordat je begint. Je kunt dan tijdens je werk kijken of je nog steeds bezig bent met datgene wat je vooraf wilde gaan doen. Dit wil niet zeggen dat je je altijd aan je plan moet houden. Je kunt natuurlijk je plan bijstellen als je dat nodig vindt.

Figuur 2. Het 'WAT'-blok, het 'HOE'-blok en het 'WAAROM'-blok behorende bij de vraag 'Wat ga ik allemaal doen om mijn huiswerk te maken?'

Gedurende de volgende 6 trainingsbijeenkomsten die elk 25 minuten duurden, oefenden de leerlingen individueel onder begeleiding van de proefleider in het gebruik van de vaardigheden uit de heuristiek. Ze deden dit aan de hand van de wiskunde-opgaven die ze bij wijze van huiswerk moesten maken. De leerlingen uit de controlegroep moesten gedurende de periode van de training zelfstandig dezelfde opgaven maken. Aangezien de training van de leerlingen uit de experimentele groep gestoeld was op verbalisaties en interventies, werd er door de leerlingen uit de controlegroep bij het maken van de opgaven niet hardop gedacht en werd er ook niet geïntervenieerd.

Beide groepen ontvingen steeds hetzelfde type feedback over hun leerprestaties bij de voorgaande bijeenkomsten. Er werd hen medegedeeld welke opgaven ze fout gemaakt hadden en welke paragrafen uit hun wiskundeboek daarop betrekking hadden. De leerlingen moesten bij elke verkeerd gemaakte opdracht aangeven wat de reden van de gemaakte fout was. De leerlingen uit de trainingsgroep dienden dit, indien mogelijk, in termen van leeractiviteiten uit de trainingsheuristiek te doen.

Tijdens de eerste van deze 6 bijeenkomsten deed de proefleider voor hoe de leeractiviteiten uit de heuristiek bij het maken van wiskunde-opgaven aangewend konden worden. De leerling diende hierbij steeds aan te wijzen waar de proefleider zich in het stroomdiagram (zie figuur 1) bevond. Daarna volgde een rolwisseling waarna de leerling zelf hardop denkend wiskunde-opgaven maakte aan de hand van het stroomdiagram en de proefleider het leerproces volgde en indien nodig op de hieronder beschreven wijze intervenieerde. In tegenstelling tot de overige vaardigheidstrainingssessies was er bij deze bijeenkomst geen sprake van een prestatiemeting.

Bij de volgende 5 vaardigheidstrainingssessies dienden de leerlingen wederom zelf wiskunde-opgaven te maken aan de hand van de heuristiek, zoals weergegeven in het stroomdiagram. Dit gebeurde ook weer hardop zodat de proefleider het verloop van het

leerproces van de leerling kon volgen. Hij intervenieerde indien uit de verbalisaties van de leerling bleek dat deze zijn leerproces niet op adequate wijze in het oog hield en waar nodig bijstuurde. Als criterium golden hierbij de stappen uit het stroomdiagram. De wijze van interveniëren was vastgelegd in een interventieschema. De proefleider probeerde de leerling te leiden naar een geschikte activiteit door het stellen van vragen zoals 'Heb je nu aandacht besteed aan dit vakje?' en 'Welke vraag hoort er bij dit vakje?'. De vragen en de bijbehorende informatieblokken uit de uitgebreide vragenlijst vormden hierbij een hulpmiddel waarvan indien nodig gebruik werd gemaakt. Op deze manier werd getracht de internalisatie van de metacognitieve vaardigheden uit de heuristiek te bevorderen. De proefleider gaf nimmer vakspecifieke aanwijzingen.

RESULTATEN

Om de invloed van de training op de metacognitieve kennis van de leerlingen te onderzoeken, werd een ANOVA uitgevoerd met als factoren 'Conditie' en 'Tijd'. 'Conditie' gold hierbij als 'between'-factor (niveau's: training en controle) en 'Tijd' als 'within'-factor (niveau's: voortest en natest). De afhankelijke variabele was 'Score op de Vragenlijst Metacognitieve Kennis'. Deze variabele werd dus zowel op de voortest als op de natest gemeten (repeated measurement). Er was sprake van een hoofdeffect van 'Conditie' ($F(1,24)=11.02; P \leq 0.01$). De trainingsgroep scoorde gemiddeld hoger dan de controlegroep (resp. marginalen: 10.8 en 7.8). Verder werd er een hoofdeffect van 'Tijd' geconstateerd ($F(1,24)=8.97; P \leq 0.01$). Op de natest werd gemiddeld hoger gescoord dan op de voortest (resp. marginalen: 10.3 en 8.3). Tussen beide factoren werd een significant interactieëffect gevonden ($F(1,24)=15.35; P \leq 0.001$). Dit interactieëffect bleek ten gunste van de trainingsgroep te zijn (zie tabel 1).

Tabel 1. Gemiddelden (X) en standaarddeviaties (SD) van de variabele 'Score op de Vragenlijst Metacognitieve Kennis' uitgesplitst naar 'Conditie' en 'Tijd'

Conditie	Trainingsgroep		Controlegroep	
	X	SD	X	SD
Tijd				
Voortest	8.5	2.4	8.1	2.8
Natest	13.1	2.4	7.5	3.7

Ter bepaling van de effecten van de training op de door de leerlingen gehanteerde metacognitieve activiteiten werden de data uit de hardop-denkenprotocollen geanalyseerd. Allereerst werd een MANOVA uitgevoerd met 'Conditie' als 'between'-factor (niveau's: training en controle), 'Tijd' als 'within'-factor (niveau's: voortest en natest) en de totale frequenties van de activiteiten per categorie, te weten 'Oriënteren', 'Proces bewaken', 'Sturen' en 'Toetsen', als variaten. De activiteiten die behoorden tot de categorieën 'Transformeren' en 'Off-task opmerkingen' werden hierbij niet meegenomen aangezien zij vanwege hun aard niet beschouwd werden als metacognitieve activiteiten. Er werd geen significant hoofdeffect gevonden van de factor 'Conditie' ($F(4,21)=1.01; P=0.43$) maar wel van de factor 'Tijd' ($F(4,21)=9.21; P \leq 0.0002$). Op de voortest werden gemiddeld meer metacognitieve activiteiten gehanteerd dan op de natest (resp. marginalen: 116.2 en 107.0). Tussen de factoren 'Conditie' en 'Tijd' werd een significant interactieëffect vastgesteld ten gunste van de trainingsgroep ($F(4,21)=2.89; P \leq 0.05$) (zie tabel 2).

Om deze effecten verder te onderzoeken, werden de univariate ANOVA's bekeken. Er bleek zich alleen een significant hoofdeffect van de factor 'Tijd' voor de categorie 'Sturen' voor te doen ($F(1,24)=4.18; P \leq 0.05$). Op de voortest kwamen dergelijke verbalisaties meer

voor dan op de natest (resp. marginalen: 41.3 en 34.6). Het multivariate interactieëffect werd teruggevonden voor de categorie 'Toetsen' ($F(1,24)=4.83; P \leq 0.04$). Dit univariate interactieëffect voor 'Toetsen' bleek wederom ten faveure van de Trainingsgroep te zijn (zie tabel 2).

Om het gevonden univariate hoofdeffect op het niveau van de activiteiten na te gaan, werd een MANOVA uitgevoerd met 'Conditie' als 'between'-factor, 'Tijd' als 'within'-factor en de frequenties van de activiteiten behorende tot de categorie 'Sturen' als variaten. Er werd geen significant hoofdeffect gevonden van de factor 'Conditie' ($F(12,13)=0.75; P=0.69$) en ook niet van de factor 'Tijd' ($F(12,13)=2.28; P=0.08$). Er was op het niveau van de activiteiten dus geen significant multivariaat hoofdeffect van 'Tijd'. Het interactieëffect was evenmin significant ($F(12,13)=0.93; P=0.55$).

Teneinde het gevonden univariate interactieëffect op het niveau van de activiteiten te traceren, werd een MANOVA uitgevoerd met 'Conditie' als 'between'-factor, 'Tijd' als 'within'-factor en de frequenties van de activiteiten behorende tot de categorie 'Toetsen' als variaten. Er werd geen significant hoofdeffect gevonden van de factor 'Conditie' ($F(11,14)=0.81; P=0.63$) maar wel van de factor 'Tijd' ($F(11,14)=4.90; P \leq 0.004$). Op de natest kwamen gemiddeld meer toetsende activiteiten voor dan op de voortest (resp. marginalen: 35.7 en 29.8). Tussen de factoren 'Conditie' en 'Tijd' werd een significant interactieëffect vastgesteld ten gunste van de Trainingsgroep ($F(11,14)=2.84; P \leq 0.04$).

Om het gevonden hoofdeffect en het gevonden interactieëffect verder te onderzoeken, werden wederom de univariate ANOVA's bekeken. Er was alleen sprake van een significant hoofdeffect van de factor 'Tijd' voor 'Controleren van de antwoorden op de oefenopgaven' ($F(1,24)=7.81; P \leq 0.01$). Op de natest kwamen dergelijke verbalisaties meer voor dan op de voortest (resp. marginalen: 7.1 en 5.1). Het interactieëffect werd teruggevonden voor de activiteiten 'Parafraseren in overeenstemming met de tekst' ($F(1,24)=7.55; P \leq 0.02$) en 'Controleren van de antwoorden op de oefenopgaven' ($F(1,24)=5.59; P \leq 0.03$). Ook hier betrof het effecten ten gunste van de Trainingsgroep (zie tabel 2).

Tabel 2. Gemiddelden (X) en standaarddeviaties (SD) van de activiteitvariabelen waarvoor een significant interactieëffect ($\alpha \leq 0.05$) werd gevonden, uitgesplitst naar 'Conditie' en 'Tijd'

Variabele	Conditie Tijd	Trainingsgroep		Controlegroep	
		X	SD	X	SD
Metacognitieve activiteiten	Voortest	123.5	68.2	108.8	50.2
	Natest	125.5	50.2	88.5	39.0
Toetsen	Voortest	29.7	20.9	29.9	19.0
	Natest	43.5	16.3	27.8	10.0
Parafraseren (+)	Voortest	1.6	2.0	4.2	6.7
	Natest	5.5	4.7	2.3	3.0
Controleren van antwoorden	Voortest	5.0	4.7	5.2	5.4
	Natest	8.7	4.4	5.5	3.8

Teneinde de invloed van de training op de prestaties op de toets over de probleemoplostaak na te gaan, werd een ANOVA uitgevoerd op de toetsscores met 'Conditie' als 'between'-factor (niveau's: training en controle) en 'Tijd' als 'within'-factor (niveau's: voortest en natest). De afhankelijke variabele was 'Score op de toets over de probleemoplostaak'. Er waren geen significante hoofdeffecten ('Conditie': $F(1,24)=0.26; P=0.61$ en 'Tijd': $F(1,24)=1.84; P=0.19$) en er was evenmin een significant interactieëffect tussen 'Conditie' en 'Tijd' ($F(1,24)=1.84; P=0.19$) (zie tabel 3). Wat betreft de scores op de toets

over de probleem-oplostaak konden derhalve geen effecten van de training op de leerprestaties vastgesteld worden. Het bleef in deze bij een niet-significante tendens.

Tabel 3. Gemiddelden (X) en standaarddeviaties (SD) van de variabele 'Score op de toets over de probleem-oplostaak' uitgesplitst naar 'Conditie' en 'Tijd'

	Conditie	Trainingsgroep		Controlegroep	
		X	SD	X	SD
Tijd					
Voortest		4.1	1.6	4.5	1.5
Natest		5.4	1.7	4.5	1.8

Er waren echter nog andere prestatiemetingen namelijk de metingen bij de vaardigheids-trainingsbijeenkomsten. Bij de 5 laatste bijeenkomsten werd steeds gekeken hoeveel opgaven de leerlingen goed beantwoord hadden. Aangezien het aantal te maken opgaven per bijeenkomst verschilde, werden de scores omgezet naar een 12-puntsschaal om een vergelijking tussen de bijeenkomsten mogelijk te maken. Vervolgens werd een ANOVA uitgevoerd over alle bijeenkomsten heen met als 'between'-factor 'Conditie' (niveau's: training en controle) en als 'within'-factor 'Tijd' (niveau's: voortest en natest). Bij deze analyse werd een significant hoofdeffect gevonden van 'Conditie' ($F(1,24)=8.29; P \leq 0.01$).

De trainingsgroep presteerde gemiddeld beter dan de controlegroep (resp. marginalen: 8.9 en 7.8). Verder was het hoofdeffect van 'Tijd' significant ($F(4,96)=12.09; P \leq 0.0001$). Het interactieëffect tussen 'Tijd' en 'Conditie' was niet significant ($F(4,96)=0.96; P=0.43$).

Per sessie werd vervolgens een ANOVA uitgevoerd met 'Conditie' als 'between'-factor. Bij deze analyses werden bij de zesde, de zevende en de achtste trainingsbijeenkomst

Tabel 4. Gegevens betreffende de scores bij de diverse vaardigheidstrainingbijeenkomsten als functie van de factor 'Conditie' (X = gemiddelde; SD = standaarddeviatie)

Sessie	Conditie	Trainingsgroep		Controlegroep		F	df	P
		X	SD	X	SD			
5		8.6	1.2	8.5	1.6	0.01	1,24	0.92
6		9.2	1.2	7.7	2.1	5.02	1,19	0.04*
7		9.8	0.9	8.7	1.5	4.63	1,20	0.04*
8		9.8	1.1	8.4	1.9	5.48	1,24	0.03*
9		7.2	1.6	6.1	2.5	1.94	1,24	0.18

*significant op ≤ 0.05 - niveau

Tabel 5. Gegevens betreffende de 'Levene's Test For Variances', uitgevoerd op de scores bij de diverse vaardigheidstrainingbijeenkomsten met als 'between'-factor 'Conditie'

Conditie	Trainingsgroep		Controlegroep		F	df	P
Sessie	Variantie		Variantie				
5	1.5	2.6	1.26	1,24	0.27		
6	1.4	4.3	5.36	1,24	0.03*		
7	0.9	2.4	7.60	1,24	0.01*		
8	1.3	3.8	3.53	1,24	0.07**		
9	2.7	6.2	0.73	1,24	0.40		

* significant op ≤ 0.05 - niveau

**significant op ≤ 0.10 - niveau

significante verschillen gevonden ten gunste van de trainingsgroep (zie tabel 4). Uit een 'Levene's Test For Variances' bleek dat er bij de zesde en de zevende bijeenkomst significante verschillen tussen de beide groepen optraden in de variantie in de scores (zie tabel 5). De trainingsgroep presteerde bij deze sessies homogener dan de controlegroep.

CONCLUSIES EN DISCUSSIE

Het is duidelijk dat de training een positief effect heeft op de metacognitieve kennis van de leerlingen: ze weten meer geschikte metacognitieve activiteiten aan te geven als antwoord op de hypothetische leersituaties uit de Vragenlijst Metacognitieve Kennis. Het is kennelijk voor deze leerlingen mogelijk om bepaalde geschikte leeractiviteiten te koppelen aan leersituaties.

Daarnaast is er sprake van gunstige effecten van de training op de metacognitieve activiteiten die door de leerlingen aangewend worden. De training oefent echter niet op alle in de heuristiek aangereikte metacognitieve activiteiten een even groot effect uit. De invloed van de training beperkt zich met name tot de toetsende activiteiten. Op deze activiteiten ligt ook duidelijk het accent in de trainingsheuristiek. Begripstoetsing neemt bijvoorbeeld in het stroomdiagram (zie figuur 1) een essentiële plaats in. Met het oog op de andere aspecten van de metacognitie verdient het aanbeveling om in toekomstige trainingsprogramma's meer aandacht te schenken aan procesbewakende en sturende activiteiten (met name activiteiten die gericht zijn op probleemidentificatie en probleem-aanpak).

De toename in metacognitieve kennis en vaardigheden manifesteert zich niet zonder meer in een verhoging van de leerprestaties, zeker niet op een transfertaak als de probleem-oplostaak op het gebied van de kansberekening. Er blijkt een discrepantie te bestaan tussen het verwerven van kennis en vaardigheden en het daadwerkelijk succesvol toepassen van het geleerde. Toch lijkt het aannemelijk te veronderstellen dat dit verschil gereduceerd kan worden. Er zijn immers wel aanwijzingen dat van de training een positieve invloed kan uitgaan op de leerprestaties. Vanaf de zesde bijeenkomst maken de leerlingen uit de trainingsgroep immers gemiddeld meer opgaven goed dan de leerlingen uit de controlegroep. Tevens neemt de variantie in de scores van de trainingsgroep in vergelijking met de controlegroep af. De training heeft niet alleen een prestatieverhogend effect maar tevens een reducerende invloed op de individuele verschillen tussen de leerlingen.

Een uitzondering hierop vormt de laatste bijeenkomst waar geen significante verschillen in gemiddelde en variantie geconstateerd konden worden. Een mogelijke verklaring hiervoor is dat er bij de laatste bijeenkomst geoefend wordt met opgaven die betrekking hebben op niet behandelde leerstof. De leerlingen zijn er waarschijnlijk nog niet aan toe om het geleerde succesvol toe te passen wanneer ze geconfronteerd worden met nieuwe leerstof. Wellicht is het mogelijk om met behulp van een langdurigere training de gevonden effecten op de leerprestaties uit te breiden naar nieuwe (transfer)taken.

Tenslotte zou de training meer geïntegreerd moeten worden in de normale wiskundelessen zodat de metacognitieve training meer hand in hand kan gaan met oefening in transformatie-activiteiten. Deze verschillende facetten van het leerproces zijn immers zeer sterk met elkaar verbonden. Een ander voordeel van een grotere integratie in het normale curriculum is dat de metacognitieve training door de leerlingen in dat geval minder als extracurriculaire activiteit gezien wordt, hetgeen het belang dat eraan toegekend wordt, zal doen toenemen. Dit zal waarschijnlijk de effecten van het trainingsprogramma nog vergroten.

LITERATUUR

- Brown, A. L., Palincsar, A. S. & Armbruster, B. B. (1984). Instructing comprehension-fostering activities in interactive learning situations. In H. Mandl, N. Stein & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 255-286). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brown, A. L., Campione, J. C. & Barclay, C. R. (1979). Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. *Child Development*, 50, 501-512.
- Campione, J. C. (1987). Metacognitive Components of Instructional Research with Problem Learners. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, Motivation and Understanding* (pp. 117-140). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Charles, R. I. & Lester, F. K., Jr. (1984). An evaluation of a process-oriented instructional program in mathematical problem solving in grades 5 and 7. *Journal for Research in Mathematics Education*, 15 (1), 15-34.
- Derry, S. J. & Hawkes, L. (1985). *Training arithmetic problem solving (TAPS): Proposal for a prototype model utilizing intelligent CAI*. Ongepubliceerd manuscript. Florida State University.
- Derry, S. J. & Murphy, D. A. (1986). Designing systems that train learning ability: from theory to practice. *Review of Educational Research*, 56, 1-39.
- Jong, F. P. C. M. de (1987). *Interimrapport I SVO-project 'Zelfdiagnostiserend Leren'*. Tilburg: Katholieke Universiteit Brabant.
- Jong, F. P. C. M. de (1988). *Interimrapport II SVO-project 'Zelfdiagnostiserend Leren'*. Tilburg: Katholieke Universiteit Brabant.
- Jong, F. P. C. M. de & Biemans, H.J.A. (1988). Valt metacognitie te meten? Hardop denken als onderzoeksinstrument. *Tijdschrift voor Onderwijswetenschappen*, 19 (2), 49-64.
- Polya, G. (1973). *How to solve it (2nd ed.)*. Princeton, NJ: Princeton University Press.
- Pressley, M. (1986). The relevance of the good strategy user model to the teaching of mathematics. *Educational Psychologist*, 21 (1&2), 139-161.
- Reeve, R. A. & Brown, A. L. (1985). Metacognition reconsidered: implications for intervention research. *Journal of Abnormal Child Psychology*, 13 (3), 343-356.
- Scardamalia, M., Bereiter, C. & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science*, 8, 173-190.
- Schneider, W. (1984). Developmental trends in the metamemory behavior relationship: an integrative review. In D. L. Forrest-Pressley, G. E. MacKinnon & T. G. Waller (Eds.), *Metacognition, Cognition and Human Performance (Vol. 1, pp. 57-109)*. New York: Academic Press.
- Vermunt, J. (1984). *Handelingsbegeleidende processen bij het zelfstandig leren: Een analyse met behulp van hardop-denken protocollen*. Intern rapport. Tilburg: Katholieke Universiteit Brabant.

Manuscript ontvangen 23-6-1989

Definitieve versie ontvangen 1-11-1989

Vergelijking van scoringsprocedures bij schaalconstructie geïllustreerd met de vragenlijst van Slade en Jenner

P.M. Kroonenberg, B.C.S. Pangemanann¹ en D. Bakker²
Vakgroep Algemene Pedagogiek, Rijksuniversiteit Leiden

ABSTRACT

Several different scoring rules for a Dutch translation of a questionnaire by Slade and Jenner (1978) to measure attitudes to females' social role are compared. Correlations between different scoring alternatives, external validity and reliability of the scorings alternatives are investigated. Cross-validation with another sample is carried out, as well as a brief comparison with other scales. The overall conclusion is that the scoring rules lead to essentially the same total scores and external validity, but that reliability seems more variable, and sensitive to inhomogeneities in the samples.

INLEIDING

Het oorspronkelijke doel van het hier gerapporteerde onderzoek was het vergelijken van alternatieve scoringsprocedures voor de Nederlandse versie van een Engelse vragenlijst die bedoeld is om de houding van mensen ten opzichte van de sociale rol van vrouwen te bepalen (Slade & Jenner, 1978). Hierbij werd gebruik gemaakt van standaard principale componentenanalyse en van niet-lineaire componentenanalyse programma's (HOMALS en PRINCALS) ontwikkeld door de vakgroep Datatheorie in Leiden. De resultaten van dit onderzoek waren zodanig dat we ons afvroegen (met name geïnspireerd door een anonieme beoordelaar van een eerdere versie) in hoeverre deze resultaten gekruisvalideerd konden worden en in hoeverre soortgelijke vragenlijsten, zoals in onderwijskundig en sociaal-wetenschappelijk onderzoek in het algemeen gebruikelijk zijn, vergelijkbare resultaten zouden opleveren. De resultaten lijken erop te wijzen dat het nut van alternatieve scoringsprocedures, zoals die van HOMALS, bij het soort vragenlijsten als hier beschreven wordt relatief gering is. Veel effectiever lijkt HOMALS te zijn voor het opsporen van inhomogeniteiten, uitbijters, en schendingen van de ordinaliteit. In eerste instantie zullen we ons concentreren op de vragenlijst van Slade en Jenner, om later kort de algemeenheid van de resultaten af te tasten.

SLADE EN JENNER VRAGENLIJST

De vragenlijst bestaat uit 25 items en is ontwikkeld en getest voor een Britse populatie. In een eerder onderzoek (Hubbard, Van IJendoorn & Tavecchio, 1982) is de vragenlijst vertaald in

Adres: Vakgroep Algemene Pedagogiek, Rijksuniversiteit Leiden, Postbus 9555, 2300 RB Leiden

¹ Thans werkzaam bij: Instituut voor Psychologisch Marktonderzoek, Rotterdam

² Thans werkzaam bij: Werkgroep COJ, Vakgroep Orthopedagogiek, RU Leiden

het Nederlands en is, na validering voor een Nederlandse populatie, het aantal items teruggebracht tot twintig.

Slade en Jenner hebben in hun antwoordcategoriën de 'neutrale' categorie weggelaten om zodoende de respondent te dwingen de antwoorden 'zeer mee oneens', 'min of meer mee oneens', 'min of meer mee eens' of 'zeer mee eens' te gebruiken. Deze antwoordcategoriën zijn door Slade en Jenner gewogen met respectievelijk 0, 1, 3 en 4 of omgekeerd, afhankelijk van de richting waarin het item is geformuleerd. De score van een persoon op de vragenlijst als geheel is dan de som van de itemscores met een mogelijk waarde van 0 tot 80, wanneer men uitgaat van de gereduceerde lijst met 20 items. Een score in de buurt van nul is een indicatie van een attitude die 'passende' mannelijke en vrouwelijke functies sterk onderscheidt en een score in de buurt van de 80 indiceert een houding zonder sexetypering.

Hubbard et al. zijn echter na analyse van hun data met het computerprogramma HOMALS (Gifi, 1981a) tot de conclusie gekomen dat de categoriegewichten niet 0, 1, 3, en 4 moeten zijn, maar 0, 1, 2 en 4, omdat het gemiddelde interval tussen de HOMALS schaalgewichten voor de categoriën 3 en 4 tweemaal zo groot is als dat voor 0 en 1 en voor 1 en 3. Voor deze laatste twee waren de gemiddelde intervallen nagenoeg gelijk.

Centraal blijft nu de vraag of deze twee scoringsprocedures bij het gebruik van de schaal andere resultaten opleveren wat betreft de ordening van de totaalscores van de proefpersonen en de relatie van de totaalscores met andere variabelen. Indien dit het geval is, kan men zich afvragen of er nog andere zinvolle scoringsprocedures te bedenken zijn die weer van deze twee verschillen en welke van zulke scoringsprocedures het meest betrouwbaar is. Verder kan men zich afvragen of herziening van categoriewaarden op basis van optimale schaaltechnieken wel zin heeft. Immers, de data-afhankelijke aanpassingen zouden wel eens minder robuust kunnen zijn dan schaalwaarden bepaald op inhoudelijke gronden. Gedeeltelijk is hier al aan tegemoet gekomen door Hubbard et al. door niet de HOMALS kwantificaties zelf te gebruiken maar geheeltallige waarden afgeleid van deze kwantificaties.

In ons onderzoek is uitgegaan van de vertaalde vragenlijst gereduceerd tot 20 items zoals Hubbard et al. die gebruikt hebben (zie aldaar voor de preciese benaming van de items), zodat een vergelijking met dit laatste onderzoek mogelijk wordt. Daarnaast is een vragenlijst met 30 items gebruikt, die door ons zelf is opgesteld, om enig inzicht in de achtergrond van de respondenten te krijgen. De vragen hierin hebben betrekking op leeftijd, sexe, houding ten aanzien van ouders, van politiek, van vrouwenemancipatie en dergelijke.

METHODE

Homals en Princals

Voor de analyse van de Slade & Jenner vragenlijst is veelvuldig gebruik gemaakt van het computerprogramma HOMALS (Gifi, 1981a). Dit programma is een implementatie van Guttman's principale componenten analyse voor nominale schalen. Categoriegewichten worden zo gekozen dat een kleinste kwadraten verliesfunctie geminimaliseerd wordt. Nishisato (1980, p. 100) merkt op dat Lord (1958) heeft laten zien dat dit overeenkomt met het maximaliseren van zowel de gekwadrateerde gemiddelde item-totaal correlatie als ook Cronbach's α . Naast de analyse met HOMALS is een analyse met het programma PRINCALS (Gifi, 1981b, p. 184 e.v.) gedaan. Ook hierbij wordt Cronbach's α gemaximaliseerd (De Leeuw, 1986, pers. comm.)¹, maar nu (in ons geval) onder de restrictie dat de categoriën ordinaal zijn. Met PRINCALS kan dus een optimale schaling van gerangordende scores bereikt worden. Beide programma's zijn thans ook opgenomen in het statistische pakket SPSS, versie 4.0.

Beide analyses proberen tot optimale schaling te komen middels een alternerend kleinste

kwadraten (ALS) algoritme en wijzen optimale scores aan proefpersonen toe. Elementaire uiteenzettingen over de gebruikte niet-lineaire analyse technieken zijn te vinden in Meerling (1988, p. 145-149, 252) en vele technische en praktische details zijn te vinden in Gifi (1981b, in druk). Niet-lineaire analysetechnieken worden onder andere gebruikt om inzicht te krijgen in de invloed van andere (d.w.z. 'optimale') scoringsmethoden voor de categorieën. Tevens kan nagegaan worden in hoeverre de intervalaannahes van Slade en Jenner en van Hubbard et al. terecht waren. Daarnaast wilden we inzicht krijgen in de ordinaliteit van de categorieën binnen de items. Met andere woorden of voor elk item de antwoordcategorieën monotoon waren ten opzichte van de onderliggende één-dimensionale schaal.

Respondenten

De twee vragenlijsten zijn tegelijk afgenomen aan 211 tweede- en derdejaars studenten pedagogiek aan de Rijksuniversiteit te Leiden. Van deze 211 vielen er 18 af wegens het niet volledig invullen van de Slade & Jenner vragenlijst en na een HOMALS analyse op de vragenlijst naar de kenmerken van de proefpersonen zijn een viertal proefpersonen afgefallen wegens hun extreme scores. Hierdoor bleef er een groep over bestaande uit 151 vrouwen en 36 mannen (Van twee personen was het geslacht niet bekend). Bij de antwoorden van de 189 respondenten op de Slade & Jenner vragenlijst waren bij acht items twee of minder personen die in een uiterste categorie scoorden. Deze enkele extreme scoorders domineerden de HOMALS analyses zodanig dat de oplossing niet meer representatief geacht werd voor de groep als geheel, zodat de homogeniteit van de vragenlijst als geheel niet meer goed beoordeeld kon worden (zie ook Gifi, 1981b, p. 115). Daarom zijn die extreme categorieën hercodeerd of 'teruggetrokken' naar de de naastbijliggende categorie, zodat er dus items ontstaan met ingedikte categorieën. Met behulp van HOMALS is daarna geconstateerd dat er geen verdere multivariate uitbijters meer waren.

De bovengeschetste procedure is duidelijk nogal subjectief en daardoor niet zonder gevaar. Dit is dan ook de reden dat we ter vergelijking ook de gewone (ongeschaalde) totaalscores hebben gebruikt zonder dat items ingedikt zijn.

Een één-dimensionale HOMALS oplossing voor de ingedikte items is berekend voor de 20 items van de Slade & Jenner vragenlijst. De scores van de proefpersonen ('objectscores') uit deze analyse zijn in de verdere analyse gebruikt; aangeduid met [HOM]. Op basis van de HOMALS categoriekwantificaties zijn een driepuntsschaal en een zevenpuntsschaal gemaakt, waarbij de schaal met kwantificaties in respectievelijk drie en zeven stukken is opgedeeld. Zo beslaat voor de driepuntsschaal categorie 1 de categorie-kwantificaties die groter dan $-.50$ zijn; categorie 2 heeft die tussen -1.50 en $-.50$; en categorie 3 heeft die kleiner dan -1.50 . Voor de zevenpuntsschaal zijn de bereiken voor de categorieën 1 t/m 7 respectievelijk $>.50$, 0 tot $.50$, $-.50$ tot 0 , -1.00 tot $-.50$, -1.50 tot -1.00 , -2.00 tot -1.50 en <-2.00 . De originele itemscores zijn op deze wijze hercodeerd naar een drie- en zevenpuntsschaal. Door optelling van deze scores is voor iedere persoon een somscore berekend, respectievelijk aangeduid met [3CAT] en [7CAT].

Daarnaast zijn natuurlijk de somscores gebruikt zoals berekend via de scoringsprocedures van Slade en Jenner [S&J], die van Hubbard et al. [HUB], die waarbij 'gewoon' de waarden 1, 2, 3, en 4 gebruikt worden met de indikking zoals boven beschreven [INDIK], en die zonder indikking [1234]. Op de ingedikte items is ook een één-dimensionale PRINCALS toegepast. Ook van deze analyse zijn de objectscores [PRIN] gebruikt als een alternatief voor somscores. (Een volledig overzicht van de per item gebruikte kwantificaties kan bij de auteurs worden opgevraagd.)

¹ Voor de goede orde dient opgemerkt te worden dat een formeel gepubliceerd bewijs van deze bewering ons niet bekend is.

Over de HOMALS oplossing kan nog vermeld worden, dat bij vijf variabelen (items 4, 6, 7, 11, en 14) de kwantificaties na optimale schaling niet monotoon zijn met de oorspronkelijke categorieën. Met andere woorden *voor deze groep* kan een hogere gemiddelde gekwadrateerde item-totaal correlatie verkregen worden door de volgorde van de oorspronkelijke categorieën te veranderen. In een inhoudelijk gericht onderzoek zou men dit verschijnsel op zijn minst aan een nader onderzoek moeten onderwerpen, ook al besluit men om er bij verdere analyses geen aandacht meer aan te schenken. De PRINCALS analyse met ordinaliteitsrestricties handhaaft de oorspronkelijke volgorde door net zoveel naburige categorieën van de problematische categorie gelijke waarden te geven tot er geen schendingen meer zijn van de monotonie.

RESULTATEN

Samenvattend zijn er dus voor elk individu zes somscores, namelijk [S&J], [HUB], [INDIK], [1234], [3CAT], [7CAT], en samen met de twee objectscores uit HOMALS [HOM] en PRINCALS [PRIN] levert dat een achttal alternatieve scores op. Deze acht zijn met elkaar gecorreleerd en om enig inzicht te krijgen in de invloed van de herkwantificaties op de externe validiteit zijn ze ook gecorreleerd met Vraag 26 [V26] uit onze eigen vragenlijst: 'Hoe staat u tegenover vrouwenemancipatie in het algemeen?', waarbij men op een zevenpuntsschaal kon scoren met een indeling van 1=zeer sympathiek, via 3=neutraal naar 7=zeer afwijzend. Tabel 1 geeft deze correlaties. Daarnaast geeft deze tabel de Cronbach α 's voor elk van de vragenlijsten berekend via de procedure Reliability uit SPSS-X (SPSS, Inc., 1985). Bij de objectscores van de HOMALS en de PRINCALS oplossing is gebruik gemaakt van de volgende formule voor de berekening van de Cronbach's α (zie Nishisato, 1980, p. 101):

$$\alpha = 1 - [(1 - \delta^2) / (n - 1) \delta^2]$$

waarbij δ^2 de eigenwaarde van de eerste dimensie is, zoals die door het programma berekend is, en n is het aantal items. (Voor het gebruik van deze formule bij PRINCALS moet enig voorbehoud uitgeoefend worden, zie ook voetnoot 1).

Tabel 1. Correlaties van scoringsalternatieven met elkaar en met de variabele 'Houding t.a.v. vrouwenemancipatie in het algemeen' (V26)

	HOM	7CAT	PRIN	3CAT	INDIK	HUB	S&J	1234	V26
[HOM]	1.00	1.00	.99	.98	.95	.93	.92	.95	-.51
[7CAT]		1.00	.99	.98	.95	.93	.92	.95	-.51
[PRIN]			1.00	.98	.96	.95	.95	.96	-.49
[3CAT]				1.00	.93	.91	.91	.92	-.47
[INDIK]					1.00	.99	.99	1.00	-.51
[HUB]						1.00	.99	.99	-.47
[S&J]							1.00	.99	-.48
[1234]								1.00	-.50
[V.26]									1.00
α	.87	.86	.86	.84	.81	.79	.79	.73	

Uit Tabel 1 valt op te maken dat de onderlinge correlaties tussen de acht somscores allemaal erg hoog zijn (tussen .91 en 1.00). Met name de correlaties tussen de HOMALS scores en de uit HOMALS ontwikkelde scores (3- en 7-puntsschaal - [3CAT] en [7CAT]) zijn erg hoog (.98 - 1.00). Zo ook de correlaties tussen deze drie scores en de PRINCALS scores. Deze vier alternatieven hebben ieder een iets hogere Cronbach's α dan de andere vier alternatieven. Het verschil tussen de hoogste waarde bij [HOM], 0.87, en die bij [1234], 0.73, is niet onaanzienlijk, en op dit verschil komen we later nog terug. De correlaties met de externe variabele [V26] zijn echter vrijwel gelijk (tussen -.47 en -.51).

De beoordelaars van een eerdere versie van dit artikel wezen er terecht op dat optimale schaalprocedures door hun optimaal gebruik van de beschikbare gegevens en dus ook van hun eigenaardigheden, gevoelig zijn voor kanskapitalisatie en dus doorgaans ook minder robuust zijn (zie ook Wainer, 1976, en de uit dit artikel voortvloeiende discussie in *Psychological Bulletin*). In het huidige geval zou dit er toe leiden dat de winst in Cronbach's α van .08 bij de HOMALS oplossing [HOM] ten opzichte van de oorspronkelijke Slade & Jenner codering [S&J] en van .14 ten opzichte van de rechttoe rechtaan 1,2,3,4 codering [1234], bij kruisvalidatie weer als sneeuw voor de zon zou verdwijnen.

Kruisvalidatie

Onafhankelijk van het vorige onderzoek, werd de Nederlandse versie van de Slade & Jenner ook gebruikt in een inventarisatie van ouderschapsmotivaties (Tavecchio & Van IJzendoorn, 1987). De Slade & Jenner vragenlijst werd uitvoeriger geanalyseerd door Bakker (1988), en de gegevens van de moeders van de door haar beschreven steekproef konden gebruikt worden voor de kruisvalidatie. Bij deze laatste steekproef is alleen gekeken naar de [HOM] en de [1234] variant, dat wil zeggen kwantificaties 1, 2, 3, en 4 zonder indikken, omdat deze de uitersten vormden in de vorige steekproef. Tegelijkertijd zijn er bij deze moeders nog enkele schalen afgenomen die later nog kort aan de orde zullen komen.

Voor de kruisvalidering zijn wederzijds de categoriekwantificaties uit de andere steekproef gebruikt, dat wil zeggen dat om de somscore van een persoon te berekenen, de score op ieder item is vervangen door de bijbehorende categoriekwantificatie uit de HOMALS analyse op de andere steekproef. Bij de steekproef moeders zijn de kruisvalidatiecorrelaties (0.88;0.86) rond .10 lager vergeleken met de correlatie (0.99) tussen de twee interne scoringsvarianten (Tabel 2). Bij de steekproef studenten is de enig beschikbare kruisvalidatiecorrelatie (0.94) vrijwel gelijk aan de 'interne' correlatie (0.95). Wat betreft de betrouwbaarheid zien we in de eerste plaats dat deze bij de steekproef moeders hoger (0.91) ligt dan bij de studenten (0.73; 0.87), en dat daar de HOMALS betrouwbaarheid geen enkele verbetering geeft op de al zeer hoge betrouwbaarheid van [1234].

Tabel 2. Kruisvalidatie van de Slade & Jenner Vragenlijst

Scoringsvariant	Steekproef Moeders			Steekproef Studenten		
	[1234]	[HOM-m]	[HOM-s]	[1234]	[HOM-m]	[HOM-s]
[1234]	-			-		
[HOM-m]	0.99	-		??	-	
[HOM-s]	0.88	0.86	-	0.95	0.94	-
Cronbach's α	0.91	0.91	0.85	0.73	??	0.87

Noten: Vet gedrukte waarden geven kruisvalidatiecorrelaties aan.

[HOM-m]=scores op basis van de categoriekwantificaties van de steekproef moeders; [HOM-s] idem voor de steekproef studenten.

[1234]=scores op basis van categoriekwantificaties 1,2,3,4 zonder indikking.

??=waarde onbekend; zie tekst.

De kruisvaliderende betrouwbaarheid bij de steekproef moeders is .85, hoger dan die van de oorspronkelijke [1234] codering en in de buurt van de [HOM] van de steekproef studenten. Zoals we later zullen betogen, vermoeden we dat dit een effect is van de (vermeende) grotere homogeniteit van de steekproef ten opzichte van de studentensteekproef. De corresponderende kruisvaliderende betrouwbaarheid en de correlatie tussen [1234] en [HOM-m] bij de studenten is helaas onbekend, daar de ruwe studentengegevens jammerlijk teloor zijn gegaan.

De over het algemeen bescheiden toename in betrouwbaarheid van de HOMALS scores ten opzichte van de oorspronkelijke score varianten (afgezien van [1234]) doet de vraag rijzen of dit soms het gevolg is van de reeds aanzienlijke tot hoge betrouwbaarheid van de Slade & Jenner vragenlijst. Om hier een - zij het vluchtig - inzicht in te krijgen hebben we voor een aantal schalen over ouderschapsmotivatie die apart aan de moeders en hun echtgenoten waren voorgelegd dezelfde procedure als hierboven toegepast. Totaalscores werden bepaald door de itemscores op te tellen [DIRECT], HOMALS objectscores werden berekend [HOM], de betrouwbaarheden voor beide scoringen, en de correlaties tussen de totaalscores en de objectscores werden bepaald. De resultaten hiervan staan in Tabel 3. Ter vergelijking zijn ook de reeds bepaalde waarden voor de Slade & Jenner toegevoegd.

Tabel 3. Correlaties tussen Scoringsalternatieven en Betrouwbaarheden uit Onderzoek naar Ouderschapsmotivatie

Ouder	Schaal	no. items	Cronbach's α			r ([DIRECT], [HOM])	no. cases
			[DIRECT]	[HOM]	Vershil		
Ma	Affectie	21	0.88	0.89	0.01	0.98	362
Ma	Normgedrag	16	0.84	0.87	0.03	0.97	373
Ma	Kosten	8	0.77	0.80	0.03	0.95	377
Ma	Reflectie	5	0.60	0.64	0.04	0.90	381
Pa	Affectie	21	0.88	0.89	0.01	0.99	333
Pa	Normgedrag	16	0.82	0.88	0.06	0.91	362
Pa	Kosten	8	0.76	0.79	0.03	0.75	378
Pa	Reflectie	5	0.61	0.66	0.05	0.93	379
Ma	Slade & Jenner	20	0.91	0.91	0.00	0.99	384
Pa	Slade & Jenner	20	0.89	0.91	0.02	??	384
*	Slade & Jenner	20	0.73	0.87	0.14	0.95	187

Noten: *=steekproef studenten; ??=waarde onbekend.

Het is allereerst duidelijk uit deze tabel dat de betrouwbaarheden vrijwel monotoon oplopen met het aantal items bij een en dezelfde steekproef. Hetgeen men zou verwachten bij redelijk geconstrueerde schalen op basis van de toename van de betrouwbaarheid bij testverlenging. Verder blijkt dat ook hier HOMALS geen overtuigende verhoging van Cronbach's α geeft, van .01 tot .06. De correlaties tussen de totaalscores en de objectscores zijn steeds bijzonder hoog, met een verrassende uitzondering van .75 bij de Kosten-vragenlijst van de vaders. Een uitgebreide bespreking van de oorzaak valt buiten het huidige bestek, maar in het kort komt het erop neer, dat op deze vragenlijst de vaders zich onderscheiden in drie duidelijke groepen, groep een met een aparte plaats op de eerste HOMALS as tegenover twee en drie, en groep drie met een aparte plaats op de tweede HOMALS as tegenover een en twee. Met andere woorden, de optimale schaling benadrukt een inhomogeniteit in de vaders, waardoor de correlatie tussen de eerste HOMALS as en de totaalscores omlaaggedrukt wordt.

DISCUSSIE EN CONCLUSIES

Door diverse vragenlijsten te analyseren, met name die van Slade & Jenner over de houding ten aanzien van de rol van de vrouw in de maatschappij, hebben we geprobeerd enig inzicht te verkrijgen over de vergelijkbaarheid van diverse scoringsprocedures, over hun betrouwbaarheid en hun relatie met een externe variabele. De algemene conclusie is dat de correlaties tussen de diverse scoringsprocedures dermate hoog zijn en dat de correlaties met de externe variabele zo weinig verschillen, dat het blijkbaar niet zo belangrijk is welke gewichten men aan de categorieën toekent. Dit sluit aan bij de bevindingen van Wilks (1938) die bewees dat onder zeer algemene condities (gemiddelde onderlinge correlaties positief en niet in de buurt van 0; gewichten positief en onafhankelijk van de intercorrelaties) de correlatie tussen twee gewogen lineaire combinaties naar 1.00 gaat als het aantal variabelen in de lineaire combinaties groot wordt. Zoals Green (1977) opmerkt: 'The theorem says what many practitioners have found from experience: differential weighting of items has little value.' (p. 264). De eerder genoemde discussie, ook opgerakeld door Wainer (1976, 1978; Keren & Laughlin, 1976; Molenaar, 1976; Pruzek & Frederick, 1976; Rozeboom (1979); zie ook Gifi, 1981b, p. 76), onderstreept dit nog eens.

Merk evenwel op dat de situatie hier iets subtieler ligt. Hier praten we niet over verschillende gewichten voor gehele items, deze blijven gelijk en we tellen ze ongewogen op, maar over verschillende gewichten die de categorieën van de items krijgen vóór de optelling. In feite zijn per item (althans voor de meeste items) de scoringsmonotone transformaties van elkaar, en als zodanig hebben de verschillende scoringsalternatieven van een item meestal een redelijke positieve correlatie met elkaar, zodat daardoor het effect van differentieel wegen gering zou kunnen zijn. De kruisvalidatie laat zien, zij het niet eenduidig, dat er sprake kan zijn van kanskapitalisatie bij het gebruik van optimale categoriekwantificaties. Immers bij de steekproef moeders correleren de eigen HOMALS-scores en de scores gebaseerd op de HOMALS-kwantificaties van de studenten 0.86 tegen een 'interne' correlatie van 0.99. Bij de steekproef studenten is de kruisvaliderende correlatie echter 0.94, geheel vergelijkbaar met de onderlinge correlaties van de diverse scoringsvarianten.

De hoogte van Cronbach's α is variabel, zowel binnen als tussen steekproeven: de coëfficiënt is sterk afhankelijk van het aantal items, meer afhankelijk van de scoringsmethode en meer afhankelijk van de steekproef. Zo is bijvoorbeeld de kruisvaliderende betrouwbaarheid bij de steekproef moeders, gebruikmakend van de studenten categoriekwantificaties, (0.85), hoger dan menige betrouwbaarheidsbepaling in de studentensteekproef zelf. De sterke indruk bestaat dat Cronbach's α niet alleen gevoelig is voor de interne consistentie van de vragenlijst, maar evenzeer voor inhomogeniteiten in de steekproef. Zo vermoeden wij dat de steekproef moeders homogener is dan die van de studenten, en dat dit tot uiting komt in de hogere betrouwbaarheid van de Slade & Jenner, ook voor de codering 1,2,3,4 zonder indikking. De relatief lage betrouwbaarheid (0.73) van de [1234] scoring bij de studenten is aantoonbaar het gevolg van enige uitbijters en wanneer de categorieën waarin zij scoren zijn samengevoegd met de naastliggende categorieën [INDIK] dan klimt de betrouwbaarheid naar (0.81). Een dergelijke observatie is natuurlijk niet nieuw. Lumsden (1978) schreef eens een artikel getiteld 'Tests are perfectly reliable', implicerend dat de personen er voor zorgen dat vragenlijsten niet betrouwbaar zijn.

Ten aanzien van het gebruik van HOMALS in dit soort situaties, lijkt het grootste profijt in de methode te zitten in het opsporen van uitbijters en inhomogeniteiten bij de steekproef in het algemeen (zie met name ook Gifi, 1981b, p. 103, 104). De Kosten-vragenlijst bij de vaders en de paar extreme scoorders bij de studenten zijn daar voorbeelden van. Hoe men deze inhomogeniteiten wil behandelen, lijkt meer een inhoudelijk dan een statistisch probleem. Met 'in dit soort situaties' bedoelen we dan schaalconstructie met Likert-achtige items,

waarbij de proefpersonen de geïmpliceerde rangorde van de antwoordcategorieën meestal als zodanig opvatten, en waar doorgaans redelijke correlaties bestaan tussen de items. Intelligentietesten die in het onderwijs gebruikt worden, vallen bijvoorbeeld ook in deze categorie. We hebben in onze vragenlijsten geen voorbeelden, waarbij de proefpersonen als groep deze ordinaliteit op zeer grove wijze schenden, en over dat soort situaties kunnen we dan ook niet veel zeggen. Ook hebben we geen informatie over situaties zoals gerapporteerd bij Van der Kloot & Brouwer (1982), die aan de hand van triaden via specifieke scoringsregels personen een score toekenden op items van de Mach-V schaal. In hun geval slaagden zij er in om met behulp van PRINCALS de betrouwbaarheid voor een 20-item schaal op te vijzelen van rond de 0.35 tot rond de 0.68. Of dit ligt aan de zeer lage betrouwbaarheid bij toch een redelijke lange schaal, is op grond van dit en hun onderzoek niet te zeggen.

Als men op basis van dit onderzoek een waarde voor de betrouwbaarheid aan de Nederlandse vertaling van de Slade & Jenner wil toekennen, die redelijk toepasbaar en (kruis)valideerbaar is, lijkt ons 0.80 een goede waarde. Al naar gelang van de homogeniteit van de steekproeven zal deze dan hoger of lager uitvallen.

Bij schaalconstructie, waar men, als het goed is, juist met een homogene verzameling items te maken heeft, kan de geciteerde literatuur als theoretische ondersteuning dienen voor de gepresenteerde empirische demonstratie dat het er niet veel toe doet welke scoringsprocedure men gebruikt om tot een somscore te komen. In ieder geval geldt voor de vragenlijst van Slade & Jenner dat men in de analyse de eenvoudigste scoringsprocedure kan gebruiken, namelijk gewoon 1, 2, 3, 4 voor opeenvolgende categorieën.

LITERATUUR

- Bakker, D. (1988). *Het moeras van oorzaak en gevolg. Een onderzoek naar het effect van de verbreding van het opvoedingsmilieu op de door ouders waargenomen veiligheid van hun kinderen*. (AP Reeks, AP 88-01-SW). Doktoraalscriptie Vakgroep Algemene Pedagogiek, Rijksuniversiteit Leiden.
- Gifi, A. (1981a). *Homals users guide*. Leiden: Rijksuniversiteit, Vakgroep Datatheorie.
- Gifi, A. (1981b). *Nonlinear multivariate analysis* (voorlopige editie). Leiden: Vakgroep Datatheorie, Rijksuniversiteit Leiden [Definitieve versie verschijnt bij Wiley].
- Green, B.F. (1977). Parameter sensitivity in multivariate methods. *Multivariate Behavioral Research*, 12, 263-287.
- Hubbard, F.O.A., Van IJzendoorn, M.H. & Tavecchio, L.W.C. (1982). Validation of a questionnaire measuring attitudes toward females' social roles for a Dutch population. *Psychological Reports*, 51, 491-498.
- Keren, G., & Newman, J.R. (1978). Additional considerations with regard to multiple regression and equal weighting. *Organizational Behavior and Human Performance*, 22, 143-164.
- Laughlin, J.E. (1978). Comment on 'Estimating coefficients in linear models: It don't make no nevermind'. *Psychological Bulletin*, 85, 247-253.
- Lord, F.M. (1958). Some relations between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika*, 23, 291-296.
- Lumsden, J. (1978). Tests are perfectly reliable. *British Journal of Mathematical and Statistical Psychology*, 31, 19-26.
- Meerling (1988). *Methoden en technieken van psychologisch onderzoek, Deel 2* (2e editie). Meppel: Boom.
- Molenaar, I.W. (1978). *Wainer on regression, some empirical results*. (Heymans Bulletins, nr. HB-78-398-EX). Groningen: Subfaculteit Psychologie, Rijksuniversiteit Groningen.
- Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Pruzek, R.M., & Frederick, B.C. (1978). Weighting predictors in linear models: Alternatives to least squares and limitations of equal weights. *Psychological Bulletin*, 85, 254-266.
- Rozeboom, W. W. (1979). Sensitivity of a linear composite of predictor items to differential weighting. *Psychometrika*, 44, 289-296.

- Slade, P. & Jenner, F.A. (1978). Questionnaire measuring attitudes to female social roles. *Psychological Reports*, 43, 351-354.
- SPSS Inc. (1983). *SPSS-X User's manual* (2nd Edition).
- Tavecchio, L.W.C., & Van IJzendoorn, M.H. (1987). Perceived security and extension of the child's rearing context: A parent-report approach. In L.W.C. Tavecchio & M.H. van IJzendoorn (Red.), *Attachment in social networks: Contributions to the Bowlby-Ainsworth attachment theory* (pp. 35-92). Amsterdam: North Holland.
- Van der Kloot, W.A. & Brouwer, P. (1982). A revised scoring procedure for the Mach V scale: a rationale and some results. *Gedrag - Tijdschrift voor psychologie*, 10, 282-293.
- Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213-217.
- Wainer, H. (1978) On the sensitivity of regression and regressors. *Psychological Bulletin*, 85, 267-273.
- Wilks, S.S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, 3, 23-40.

Manuscript ontvangen 11-5-1989

Definitieve versie ontvangen 27-9-1989

Notities en Commentaren

DE VORMING VAN DE CONTROLEGROEP BIJ EEN INTACT VELDEXPERIMENT

E.M. Scholte

*Vakgroep Orthopedagogiek, Centrum Onderzoek Jeugdhulpverlening,
Rijksuniversiteit Leiden*

1. INLEIDING

Ten behoeve van efficiënte allocatie van middelen in de jeugdgezondheidszorg en het onderwijs is het wenselijk na te gaan in hoeverre hulpverlenings- en onderwijsprogramma's effectief zijn (Posavac en Carey, 1980; Van der Ploeg, 1986; Scholte, 1988a; Scholte, 1988b). Indicaties hieromtrent kunnen worden verkregen door na te gaan of er positieve verandering is opgetreden in de doelsituatie waar een programma verandering in beoogde te brengen (Cronbach en Furby, 1970). Op grond van verbetering van de behandelde situatie alleen kan echter nog niet worden geconcludeerd dat interventie ook daadwerkelijk effect heeft gehad. Verbetering zou immers ook kunnen zijn ontstaan onder invloed van factoren die niets met de interventie te maken hebben, bijvoorbeeld rijping van respondenten (Borg en Gall, 1979). Ook hoeft het uitblijven van verbetering op zich nog niet te betekenen dat er dan ook geen effect is geweest. De positieve invloed van een behandeling kan daaruit hebben bestaan dat de doelsituatie niet verder verslechterde. Bij de bepaling van het effect van interventieprogramma's dient voor dergelijke tekorten aan interne validiteit te worden gecorrigeerd (Borg en Gall, 1979; Cook en Campbell, 1979). Dit kan door in het onderzoek een experimentele en een controlesituatie op te nemen. In het 'ideale' onderzoeksontwerp komen deze beide situaties tot stand via een op het toeval gebaseerde ('at random') verdeling van respondenten over een experimentele groep en een controlegroep (Cook en Campbell, 1979). In veldonderzoek is zo'n verdeling echter op ethische gronden vaak niet mogelijk. Het op toevalsgronden verwijzen van bijvoorbeeld hulpbehoevende jeugdigen naar een experimentele groep die wel en een controlegroep die geen hulp krijgt, staat doorgaans op gespannen voet met de beroepsethiek van zowel hulpverleners als onderzoekers. Via 'secundaire matching' is het echter mogelijk experimentele- en controlegroepen te vormen zonder de veldsituatie 'aan te tasten' (cf. Anderson, Auquier, Hauck, Oakes, Vandaele en Weisberg, 1980). In dit geval worden de experimentele groep en de controlegroep achteraf samengesteld, d.w.z. na interventie, via het 'koppelen' van jeugdigen die wel interventie hebben ondergaan aan vergelijkbare jeugdigen die geen interventie hebben gehad. In het navolgende wordt deze 'secundaire matchingsstrategie' kort uiteen gezet aan de hand van een voorbeeld uit de onderzoekspraktijk.

2. DE ONDERZOEKSSITUATIE

Het voorbeeld dat wordt gepresenteerd is de effectevaluatie van een experiment voor preventieve hulpverlening aan jeugdigen tot 16 jaar die met de Haagse politie in aanraking zijn gekomen vanwege weglopen, diefstal, vandalisme etc. (Scholte, 1988). Teneinde effectuitspraken te kunnen doen werd een vergelijking gemaakt tussen de followup situatie van een experimentele groep (jeugdigen met vroegtijdige hulp) en een controlegroep (geen vroegtijdige hulp, in verder psychosociaal opzicht vergelijkbaar met eerstgenoemde groep). Beide onderzoeksgroepen werden via secundaire matching samengesteld. Hierbij werd uitgegaan van twee groepen. De eerste groep bestond uit jeugdigen, die naar het preventieproject doorverwezen waren door de jeugdpolitie (N=86). Uit deze groep werden de experimentele jeugdigen geselecteerd. De tweede groep bestond uit een steekproef jeugdigen die niet door de jeugdpolitie naar het preventieproject doorverwezen waren (N=200)¹. Hierbij werd verondersteld dat deze groep een aantal jeugdigen zou omvatten dat te kampen had met psychosociale (gedrags)problematiek, doch die niet door de jeugdpolitie was opgemerkt en waarbij de desbetreffende jeugdigen derhalve niet waren doorverwezen naar het preventieproject. Jeugdigen derhalve die in principe geschikt zijn om te functioneren als controlejeugdigen. Voor de vorming van een experimentele groep en een controlegroep dient nu bij elke preventiejeugdige een geschikte controlejeugdige te worden gezocht. Hiertoe werden uit een steekproef niet naar het project doorverwezen jeugdigen met politiecontacten die jeugdigen geselecteerd die wat relevante psychosociale kenmerken betreft de sterkste gelijkenis vertoonden met de preventiejeugdigen.

3. HET 'MATCHINGS' PROBLEEM

Theoretisch gezien zou de controlegroep samengesteld dienen te zijn uit jeugdigen die exacte copieën zijn van de jeugdigen in de interventiegroep. Zo'n 'ideale' gelijkenis is in de onderzoekspraktijk uiteraard niet mogelijk en bovendien niet nodig. Waar het om gaat is dat de experimentele- en de controlegroep wat relevante aspecten betreft op elkaar moeten lijken (Anderson et al., 1980, p. 69-112). Om welke aspecten het hierbij gaat wordt door de vraagstelling van het onderzoek bepaald. Zo is het voor de bestudering van psychosociale hulpverlening bijvoorbeeld relevant dat de onderzoeksgroepen overeenstemmen wat betreft demografische kenmerken en psychosociale problematiek. In het voorbeeld waren in de eerste plaats verschillen te verwachten met betrekking tot hulpverlening aan jongens en meisjes, binnen de verschillende leeftijdscategorieën en met betrekking tot de hulpverlening aan de verschillende etnische groeperingen (Junger-Tas, 1983). De jeugdigen in de experimentele groep en de controlegroep dienen derhalve in ieder geval op elkaar te lijken wat betreft sekse, leeftijd en etniciteit. In de tweede plaats dienen de jeugdigen in de experimentele- en de controlegroep vergelijkbaar te zijn wat betreft de kernaspecten van de psychosociale situatie. Het betreft die aspecten van de leefsituatie van de jeugdigen waar de hulpverlening zich - conform de doelstelling van het preventieproject - op heeft gericht. In dit verband kunnen worden genoemd: de (eventuele) gezins- en schoolproblematiek, het

¹ De omvang van de steekproef waaruit controle respondenten worden geselecteerd dient groter te zijn dan die van de oorspronkelijke experimentele groep daar niet alle respondenten in deze steekproef 'potentiële controle respondenten' in voldoende mate aan de matchingscriteria zullen voldoen. Naarmate dit criterium strenger is, zullen controle respondenten moeilijker te vinden zijn. Anderson et al. (1980; 98-99) adviseren voor de steekproef potentiële controlerespondenten een omvang van 2 tot 5 keer die van de experimentele groep.

(eventuele) wetsovertredend gedrag en het aantal contacten met het politieapparaat. Deze gebieden worden in de 'matchings' procedure door de vier variabelen gerepresenteerd (vgl. voor de betrouwbaarheid en validiteit, Scholte, 1988). De 'matchings' procedure werd vervolgens uitgevoerd door bij elke preventiejeugdige een jeugdige uit de niet preventiegroep te zoeken die bij de aanvang van de hulp een maximale gelijkenis vertoont op alle vier hierboven genoemde variabelen. In technische termen vertaald betekent dit, dat bij elke preventiejeugdige een niet preventiejeugdige gevonden moet worden met een analoog scorepatroon op de 'matchings' variabelen. Voordat deze procedure uitgevoerd kan worden, dient echter eerst nog te worden vastgesteld in welke mate de (scorepatronen van) jeugdigen in de experimentele en controlegroep op elkaar moeten lijken om nog te kunnen spreken van vergelijkbaarheid. In het algemeen geldt, dat naarmate een grotere gelijkenis wordt geëist, de kans kleiner is dat er een controlejeugdige in het reservoir van potentiële controlejeugdigen voorhanden zal zijn. Rubin (1976 a,b; 1979) stelt ten aanzien van dit probleem voor net zo lang door te zoeken in het reservoir van potentiële controlejeugdigen totdat alle experimentele casussen een controletegenhanger hebben gevonden, ongeacht de mate waarin de paren van elkaar verschillen. Echter, om onvergelijkbaarheid van groepen te voorkomen, moet een gelijkeniscriterium worden gespecificeerd dat de (rand)voorwaarden geeft waaronder nog sprake is van een vergelijkbaar paar. Indien de overeenkomst tussen een jeugdige uit de experimentele groep en een jeugdige uit de potentiële controlegroep binnen dit criterium valt, wordt het betreffende paar matchbaar geacht. Indien de gelijkenis voor de potentiële controlesubjecten buiten dit criterium valt, is de desbetreffende jeugdige niet matchbaar en wordt deze uit de experimentele groep verwijderd. Door het volgen van deze strategie zal de experimentele groep een geringer aantal jeugdigen omvatten dan de oorspronkelijke preventiegroep. Dit is echter te prefereren boven het hebben van een niet vergelijkbare controlegroep. Als gelijkeniscriterium werd in het voorbeeld de volgende regel gehanteerd: 1) Een preventiejeugdige en een bijbehorende controlejeugdige dienen wat sekse en etniciteit betreft gelijk te zijn; 2) Tussen een preventiejeugdige en een bijbehorende controlejeugdige dient het leeftijdsverschil zo gering mogelijk te zijn en een eventueel verschil in leeftijd mag niet meer dan twee jaar bedragen, alsmede 3) Tussen een preventiejeugdige en een bijbehorende controlejeugdige dient een zo gering mogelijk verschil te zijn wat betreft de kwaliteit van de gezinsinteracties, de schoolmotivatie, het zelfgerapporteerde wetsovertredend gedrag en het totaal aantal eerdere contacten met het politieapparaat. Een eventueel verschil mag niet meer dan één inhoudelijke categorie op ten hoogste drie van de vier genoemde variabelen bedragen.

4. DE MATCHINGSPROCEDURE

De secundaire matching bestaat uit twee stappen, namelijk:

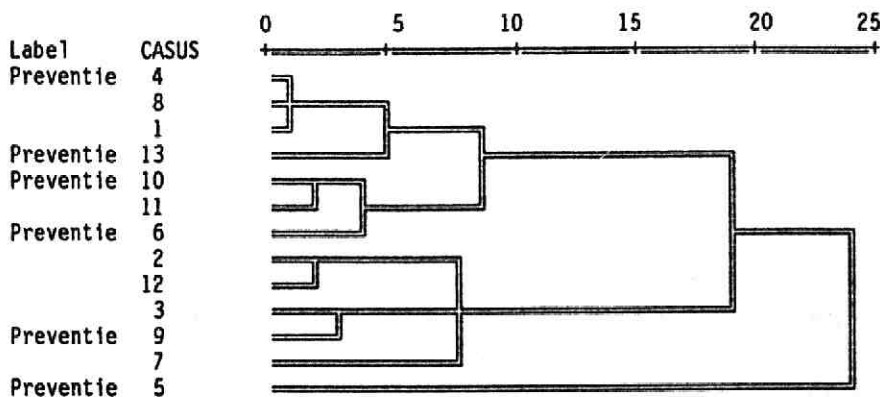
Stap 1: stratificatiematching (McKinlay, 1975):

Alle preventie- en potentiële controlejeugdigen worden tot één groep samengevoegd en via een sorteerprocedure ingedeeld naar de strata 'Sekse' en 'Etniciteit'. Hierdoor ontstaan vier groepen met experimentele en controlejeugdigen, nl. autochtone jongens en autochtone meisjes, alsmede allochtone jongens en allochtone meisjes.

Stap 2: multivariate afstandsmatching

Vervolgens worden binnen de gevormde strata alle casussen door middel van clusteranalyse (Aldenfelder, 1985) bijeen geplaatst. Clustering vindt plaats op basis van gelijksoortig-

heid in scorepatronen op de matchingsvariabelen. Deze variabelen dienen een gelijk gewicht te hebben in de clusterprocedure. Dit is het geval indien ze allen op eenzelfde wijze zijn geschaald (Anderson et al., 1980). In het onderzoek van het voorbeeld werden de matchingsvariabelen herschaald volgens een intervallschaal met een reikwijdte van 3.00. Ter bepaling van de overeenkomst in scorepatronen wordt eerst het kwadraat van de euclidische afstand tussen de scores van de casussen voor alle matchingsvariabelen apart berekend. Deze gekwadrateerde verschillen worden vervolgens voor elk paar casussen over alle matchingsvariabelen gesommeerd. Naarmate dit 'somkwadraat van de euclidische afstand over alle matchingsvariabelen' kleiner is, komen scorepatronen van casussen beter overeen en vormen de paren een betere 'match'. Als (normatieve) indicatie van de kwaliteit van de 'match' werd in het voorbeeldonderzoek de volgende regel aangehouden: goede overeenkomst ($0 \leq (\text{euclid})^2 < 1$), redelijke overeenkomst ($1 \leq (\text{euclid})^2 < 2$), matige overeenkomst ($2 \leq (\text{euclid})^2 < 3$) en slechte overeenkomst ($3 \leq (\text{euclid})^2$). Hierbij komt een somkwadraat van de euclidische afstand ter grootte van 1 overeen met een maximale schaalwaardeafstand van 0.25 op alle vier de matchingsvariabelen of een verschil van 1 schaalwaarde op een van de matchingsvariabelen in combinatie met exacte overeenkomst op de overige drie variabelen. Dit betekent dat in het voorbeeld gesproken wordt van een 'goede overeenkomst' indien 'gematchte' paren op niet meer dan één van de matchingsvariabelen minder dan 1 schaalwaarde van elkaar afwijken dan wel op alle vier de variabelen minder dan een kwart schaalwaarde van elkaar verschillen. Een 'goed gematched paar' is bijvoorbeeld: een preventiejeugdige met frequent conflicten thuis, een zeer lage schoolmotivatie, een hoge frequentie (zelf)gerapporteerd wetsovertredend gedrag en twee eerdere politiecontacten, gekoppeld aan een controlejeugdige met frequent conflicten thuis, een zeer lage schoolmotivatie, een hoge frequentie zelfgerapporteerd wetsovertredend gedrag en één eerder politiecontact. In het voorbeeld is sprake van een 'slechte overeenkomst' wanneer een verschil bestaat van 1 schaalwaarde op alle vier matchingsvariabelen of een verschil van 2 schaalwaarden op één der matchingsvariabelen in combinatie met een exacte overeenkomst in schaalwaarde op de drie overige variabelen. Figuur 1 geeft ter illustratie van de procedure van multivariate afstandsmatching de uitkomst van de clusteranalyse voor het stratum allochtone meisjes. Naarmate de afstand tussen casussen kleiner is, komen de scorepatronen van de casussen op de matchingsvariabelen meer overeen. Uit de figuur blijkt bijvoorbeeld dat casus 004 en 008 qua scorepatroon sterk op elkaar lijken. Zij vormen een goede match voor elkaar.



Figuur 1: Relatieve clusterposities

De kwaliteit van de matching kan worden beoordeeld door na te gaan in hoeverre de via secundaire matching gevormde groepen overeenkomen wat betreft hun verdeling over de relevante demografische en psychosociale kernvariabelen. Tabel 1 geeft de uitkomst van de toetsing van deze hypothese van gelijke verdeling in beide groepen voor de matchingsvariabelen van het voorbeeld. Uitspraken betreffende het effect van de interventie kunnen vervolgens worden ontleend aan pré- en posttest vergelijkingen tussen de experimentele groep en de controlegroep op de relevante variabelen.

Tabel 1: Overeenkomst tussen experimentele- en controlegroep van het voorbeeldonderzoek na secundaire matching

	N	gemid. exp.	gemid. contr.	vrijheids- graden	T- waarde	Sd- exp.	Sd- contr.	F- waarde
leeftijd	71	13.9	14.1	140	-0.79*	2.1	2.1	1.04*
gezinsproblemen	71	2.7	2.5	140	1.27*	1.1	1.1	1.04*
schoolproblemen	71	1.9	2.0	140	-0.97*	0.8	0.8	1.22*
delinquentie	71	2.2	2.4	140	-1.61*	1.0	1.0	1.18*

*pwaarde > 0.10

	N	afwezig	aanwezig
politiecontacten			
Experimentele groep	71	39 (55%)	32 (45%)
Controlegroep	71	41 (58%)	30 (42%)

$\chi^2 = 0.21$, $df = 1$, $p > 0.10$

5. CONCLUSIES

Bij intacte veldsituaties is het vanwege praktische en ethische gronden meestal niet goed mogelijk toegepaste interventies te onderzoeken aan de hand van het 'random treatment-control group' design. Hierdoor kunnen geen 'harde' uitspraken worden gedaan omtrent het effect van de onderzochte interventies. Theoretisch bestaat er echter een alternatief in de vorm van het achteraf construeren van een controlegroep. De werkwijze is hierbij als volgt. Er wordt een steekproef respondenten genomen uit een met de interventiegroep vergelijkbare populatie. Uit deze groep wordt vervolgens een controlegroep geconstrueerd door respondenten op te nemen met eenzelfde scorepatroon op relevante matchingsvariabelen als de subjecten in de interventiegroep. Uit het hier gepresenteerde voorbeeld blijkt dat deze zogenaamde 'secundaire matchingsstrategie' ook in de praktijk een bruikbare methode is om een controlegroep te vormen. Clusteranalyse blijkt een geschikte techniek te zijn om bij elkaar passende interventie-controle subjecten te vinden. Door secundaire matching ontstaat de mogelijkheid om ook bij intacte veldsituaties te kunnen komen tot intern valide uitspraken omtrent het effect van interventie of behandeling. Opgemerkt moet echter worden dat matching achteraf niet altijd goed mogelijk zal zijn. Naarmate de matchingskenmerken zeldzamer zijn en/of de matchingstolerantie strenger wordt gesteld zal een grotere steekproef 'potentiële controle respondenten' nodig zijn om nog controlerespondenten te kunnen vinden. Het zou derhalve voor kunnen komen dat controlesubjecten onvindbaar zijn.

LITERATUUR

- Aldenderfer, M.S. & Blashfield, R.K. (1985). *Cluster analysis*. London: Sage.
- Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W. & Weisberg, H.I. (1980). *Statistical methods for comparative studies*. New York: Wiley & Sons.
- Borg, W.R. & Gall, M.D. (1979). *Education Research*. London: David McKay.
- Cook, T.D. & Campbell, D.I. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Chicago: Rand and McNally.
- Cronbach, L.J. & Furby, L. (1970). How should we measure 'change' - or should we? *Psychological Bulletin*, 74(1), 68-80.
- Junger-Tas, J. (1983). *Jeugddelinquentie I; achtergronden en justitiële reactie*. Den Haag: Staatsuitgeverij.
- McKinlay, S.M. (1975). The effect of bias on estimators of relative risk for pairmatched and stratified samples. *Journal of the American Statistical Association*, 70(352), 859-864.
- Ploeg, J. D. van der (1986). Is kwaliteit in de residentiële hulpverlening een meetbaar begrip? In: W. Hellinckx (red.), *Kwaliteit en de hulpverlening m.b.t. kinderen met psychosociale problemen*. Leuven: ACCO.
- Posavac, P.E. & Carey, R.G. (1980). *Program evaluation: methods and case studies*. Englewood Cliffs: Prentice Hall.
- Rubin, D.B. (1976a). Multivariate matching methods that are equal percent bias reducing I: Some examples. *Biometrics*, 32, 109-120.
- Rubin, D.B. (1976b). Multivariate matching methods that are equal percent bias reducing II: maximums on bias reduction for fixed samples. *Biometrics*, 32, 121-132.
- Rubin, D.B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 4, 318-328.
- Scholte, E.M. (1988a). *Jeugd, politie en hulpverlening*. Leuven/Amersfoort: ACCO.
- Scholte, E.M. (1988b). *Fundamentele kwaliteitsbepaling in de psychosociale jeugdhulpverlening: identiteit en effectiviteit van hulpvormen*. Tilburg: S.A.S.T.

Ontvangen 9-1-1989

Notities en Commentaren

PROGRAMMERING VAN ONDERWIJSONDERZOEK: EEN KRITISCH COMMENTAAR OP HET BELEID VAN SVO

Wijnand Th.J.G. Hoebe

RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen

Recente ontwikkelingen

Het Instituut voor Onderzoek van het Onderwijs (SVO) heeft in 1988 en 1989 opnieuw aanleiding gegeven om kritisch na te denken over het fenomeen onderzoeksprogrammering. Bleek in 1987 de eigen programmering door SVO zodanig uit de hand te zijn gelopen, dat voornemens voor nieuw onderzoek naar 1988 moesten worden doorgeschoven, in 1988 werden onderwijsonderzoekers geconfronteerd met een aantal heroverwegingen. Onderzoeksthemagroepen (OTG's) ontvingen bericht dat het OTG-beleid van SVO zou worden heroverwogen, zodat er mogelijk vanaf 1 juli 1988 geen middelen voor OTG's beschikbaar zouden kunnen worden gesteld. Onderzoeksinstituten ontvingen bericht dat de met hen gesloten meerjarencontracten werden opgezegd. Volgens de eerste, mondelinge mededelingen zouden vanaf 1989 niet meer instituutsmeerjarencontracten worden afgesloten, maar contracten voor meerjarige onderzoeksprogramma's. Eveneens volgens de eerste berichten zou de tot dan toe voor OTG-onderzoek gereserveerde ruimte in de toekomst althans ten dele worden gereserveerd voor meerjarige strategische onderzoeksprogramma's.

In het op 24 juni 1988 gedateerde meerjarenbeleids- en onderzoeksprogramma (SVO/BR88-107) bleek nog weinig van besluitvorming naar aanleiding van de aangekondigde heroverwegingen. Meerjarige strategische onderzoeksprogramma's maken voortaan weliswaar deel uit van het beleid; in plaats van een budget hiervoor vrij te maken ten laste van de middelen die tot nu toe voor (te heroverwegen) OTG-onderzoek beschikbaar waren, werd een heroverweging aangekondigd van het niet-programmagebonden onderzoek (ANGO) om langs deze weg middelen vrij te maken.

De vraag of en hoe vanaf 1989 het OTG-beleid zou moeten worden bijgesteld, werd niet beantwoord. De daarvoor noodzakelijke evaluatie van OTG's zou 'mede moeten uitwijzen of en hoe deze werkgemeenschappen van onderzoekers betrokken kunnen worden bij de ontwikkeling, coördinatie en evaluatie van bedoelde fundamenteel strategische onderzoeksprogramma's' (pag. 16). Eind 1988 werden OTG's geïnformeerd over de uitgevoerde evaluatie van het OTG-beleid. Er werd gunstig geoordeeld over de OTG's. Het huidige stelsel van vijf OTG's zou voorlopig blijven voortbestaan. Heroverweging van de vijf themagebieden werd aangekondigd. Tevens werd de betrokkenheid van de vijf OTG's bij de ontwikkeling van meerjarige strategische onderzoeksprogramma's aanbevolen.

In 1989 werden betrokkenen vervolgens in februari verrast met de mededeling dat er in het ANGO-fonds en in de OTG-fondsen geen onderzoekssubsidies meer zouden worden toegekend. De budgettaire ruimte zou worden gereserveerd voor strategisch onderzoek. In april volgde de mededeling dat het ANGO-fonds blijft bestaan, maar dan in de vorm van bescheiden persoonsgebonden subsidies (beurzen) in plaats van in beginsel kostendekkende subsidies voor onderzoeksprojecten. In juli werd bekend gemaakt dat de keuze van

thema's en de vaststelling van procedures voor strategisch onderzoek werd uitgesteld, zodat duidelijkheid over een en ander pas in 1990 kon worden verwacht. Tevens werd duidelijk dat het in 1988 aangekondigde voortbestaan van de vijf OTG's weer op losse schroeven staat; er zal geen nieuw OTG-onderzoek door SVO worden gesubsidieerd; nagegaan wordt op welke wijze de OTG's eventueel betrokken kunnen worden bij de beoordeling van projectvoorstellen in het kader van programma's voor strategisch onderzoek (PSO's). Hoe serieus dit wordt nagegaan kan worden afgeleid uit het gegeven dat SVO in mei in overleg trad met de Vereniging voor Onderwijsresearch (VOR) over een alternatief voor het betrekken van OTG's bij de beoordeling van onderzoeksvoorstellen.

De OTG's wijken voor PSO's. Aangezien PSO's sinds 1988 definitief deel uitmaken van het onderzoeksbeleid van SVO, mag worden verwacht dat duidelijk wordt gemaakt wat met dergelijke programma's wordt bedoeld. Het meerjarenbeleidsplan van 1988 is de enige beschikbare officiële tekst. Het beleidsplan volstaat met de volgende omschrijving: 'dergelijke programma's sluiten aan bij de fundamentele vraagstukken die zich in het onderwijs voordoen en *zouden kunnen* bijdragen aan de wetenschappelijke kennisgroei en de ontwikkeling van het wetenschappelijke instrumentarium' (pag. 15) (cursivering WH). Wat hier staat is dat strategische onderzoeksprogramma's niet hoeven bij te dragen aan de wetenschappelijke kennisgroei, niet hoeven bij te dragen aan de ontwikkeling van het wetenschappelijke instrumentarium, niet hoeven aan te sluiten bij fundamentele vraagstukken van theorie-ontwikkeling in onderwijswetenschappen en niet hoeven aan te sluiten bij praktische vraagstukken die zich in het onderwijs voordoen. Wat een strategisch onderzoeksprogramma dan wel is, blijft onduidelijk. Bij 'vraagstukken die zich in het onderwijs voordoen' kan het een en ander worden gedacht; vroeger werd dit de praktische relevantie van onderwijsonderzoek genoemd. De commissie die het OTG-beleid opnieuw heeft gezien spreekt dan ook over strategische (praktische) kennis als onderscheiden van theoretische kennis. Spreken over fundamenteel strategisch onderzoek wekt dan ook meer associaties met Orwelliaanse dubbelpraat dan met serieus onderzoeksbeleid.

Ook met betrekking tot het onderzoek ten behoeve van het onderwijsbeleid en het onderwijsveld wordt het voornemen uitgesproken de 'programmeringsinspanningen mede te richten op het tot stand brengen van intern samenhangende meerjarige onderzoeksprogramma's' (pag. 14). Dit voornemen wordt geplaatst in het kader van het aanbrengen van interne samenhang in de onderzoeksactiviteiten van de SVO. Meerjarige programmacontracten zouden in de plaats komen van instituutsmeerjarencontracten en daarmee een instrumentarium vormen om de interne samenhang van onderzoeksactiviteiten binnen dit onderdeel van het budget van SVO te bevorderen. Dit voornemen heeft het einde van 1988 niet overleefd. In 1989 werd weer onderhandeld over instituutsmeerjarencontracten.

Bij de diverse soorten onderzoeksprogramma's wordt in het beleidsplan gezegd dat ze gedurende de planperiode (dat wil zeggen uiterlijk in 1992) tot stand kunnen worden gebracht. Dat zou kunnen resulteren in verlenging van de periode van de in 1987 ingetreden onzekerheid tot uiterlijk eind 1992. Voeg daarbij het ervaringsgegeven dat programmeringsconcepten het bij SVO meestal niet lang uithouden (de OTG- en ANGO-kaders zijn zeer gunstige uitzonderingen: ze hebben het sinds 1980 uitgehouden en worden pas sinds 1988 heroverwogen) en het zal duidelijk worden dat onderwijsonderzoekers moeten anticiperen op een langere periode van onzekerheid over de beleidskaders waarbinnen en de criteria aan de hand waarvan onderzoekssubsidies zullen worden toegekend.

Consistentie in het programmeringsbeleid van SVO.

Het beleid van SVO met betrekking tot onderzoeksprogrammering kent een aantal constanten. Eén van deze constanten is dat tot beleidswijziging wordt overgegaan,

alvorens het tot dan toe vigerende beleid is geëvalueerd. Het programmeringsbeleid van voorstudies in het midden van de jaren zeventig is bijvoorbeeld verlaten voordat de uitgezette voorstudies in rapporten waren uitgemond. De groepen van onderzoekers die het begeleidend onderzoek voor de grootschalige vernieuwingsprojecten in de jaren zeventig programmeerden werden snel vervangen door ministeriële programma-adviescommissies. Het voornemen om vanaf de integratie van geldstromen van SVO en departement programma-adviescommissies ten behoeve van het SVO-bestuur in te stellen, is met uitzondering van het voortgezet onderwijs eerste fase niet geïmplementeerd. In plaats daarvan is sinds 1985 het instrument van de probleemverkenning tot ontwikkeling gekomen. Recente geluiden doen vermoeden dat ook dit instrument wordt heroverwogen. Vanaf 1980 zijn de OTG's opgericht met als opdracht een op kennisgroei gericht programma te ontwerpen van onderzoeksactiviteiten op een bepaald onderzoeksthema. Na goedkeuring van deze programma's in 1981 werden onderzoeksplannen die binnen deze programmatische context pasten en voldoende kwaliteit hadden, subsidiabel gesteld. Ondanks een positieve evaluatie is deze functie van OTG's nu beëindigd.

Een andere constante is dat beoordelingscommissies van wetenschappelijke experts eerst om advies wordt gevraagd als het onderzoeksprogramma is vastgesteld, en als op basis van een onderzoeksprogramma concrete subsidie-aanvragen ter besluitvorming voorliggen. Formeel kunnen deze commissies zich slechts uitspreken over de kwaliteit van afzonderlijke onderzoeksvoorstellen. Omdat beoordelingscommissies ook dienen te adviseren over de relevantie van onderzoeksvoorstellen - een onderwerp dat eigenlijk binnen de programmatische kaders thuishoort -, krijgen zij langs deze omweg toch invloed op de inhoud van de onderzoeksprogrammering. Onderzoekers die in aanbestedingsprocedures van SVO participeren, zullen wel eens met het fenomeen zijn geconfronteerd dat hun onderzoeksplan in voorliggende vorm wordt afgekeurd omdat de beoordelingscommissie vraagtekens plaatst achter de relevantie van de in het onderzoeksprogramma van SVO opgenomen vraagstelling. Het duidelijkste voorbeeld kan worden gevonden in de evaluatie van het OTG-beleid. De evaluatiecommissie concludeert dat geen enkele OTG tot een gezamenlijke planning van onderzoeksactiviteiten rond gemeenschappelijke problemen is gekomen, en dat geen enkele OTG op basis van reconstructies en reviews zijn actuele researchfront heeft uitgetekend in de vorm van optimaal gestelde onderzoeksproblemen (Het OTG-beleid opnieuw bezien, 1988, p. 59-60). De OTG die in termen van gesubsidieerd onderzoek het meest succesvol was, lijkt het minste naar een dergelijke programmering te hebben gestreefd. Dit lijkt een effect van het functioneren van de beoordelingscommissie, die vooral criteria van wetenschappelijke kwaliteit hanteerde en niet of nauwelijks criteria ontleend aan de programmatische doelstellingen van SVO met het OTG-beleid. Andere OTG's, die in hun beginfase nog aanleunden tegen de programmatische doelstellingen van SVO, gingen zich daar uiteraard van distantiëren.

Als derde constante in het SVO-beleid mag worden genoemd het voortdurende streven om onderzoek te coördineren. Wetenschappelijke kwaliteit en interne samenhang lijken daarbij minder belangrijk dan een organisatorische greep op de ontwikkelingen. Programmeringsinitiatieven waarop door onderzoekers wordt ingespeeld en die dientengevolge een autonome ontwikkeling krijgen, worden ingedamd of zelfs beëindigd. Het OTG-beleid en de ontwikkelingen met het ANGO-fonds zijn voorbeelden. De implementatie van dit beleid was immers feitelijk in handen geraakt van wetenschappelijke beoordelingscommissies. Prioriteit geven aan coördinatie, planning en programmering maakte het mogelijk deze invloed in te dammen. Een ander voorbeeld kan worden gevonden in de in de kiem gesmoorde programma-adviescommissies. Vergeleken met dergelijke commissies is het instrument van de probleemverkenning onder verantwoordelijkheid van het SVO-bureau immers veel effectiever om greep op de ontwikkelingen te houden.

De kenmerken van het SVO-beleid met betrekking tot onderzoeksprogrammering kunnen in een onwelwillende interpretatie worden gezien als voorbeelden van de manier waarop een bureaucratisch apparaat controle kan houden op haar eigen positie. Bevorder de onzekerheid bij de betrokkenen, zorg voor belangenconflicten en verander regelmatig het beleid; dan ben je zeker van een organisatorische greep op de ontwikkelingen. Gepaard met enig cynisme over wetenschappelijke kwaliteit en interne samenhang van onderzoek kunnen deze begrippen zelfs worden gebruikt om de greep van de organisatie op het onderzoek te verstevigen. Voor deze onwelwillende interpretatie is echter niet voldoende bewijs voorhanden. SVO heeft een uitgesproken bedoeling om ontwikkelingen in onderwijsonderzoek in een (nu nog nader vast te stellen) richting te sturen. Het blijft echter de vraag of de teweeg te brengen coördinatie van onderzoek meer is dan een versterking van de invloed van de organisatie en of deze coördinatie tot een (wenselijke) sturing van de ontwikkelingen in onderwijsonderzoek leidt.

Sturing van onderwijsonderzoek

Er zijn globaal twee scholen in het denken over de sturing van ontwikkelingen in onderzoek. De school met in Nederland traditioneel de grootste invloed denkt aan de hantering van het criterium wetenschappelijke kwaliteit en het overlaten van toepassing ervan aan de georganiseerde wereld van de wetenschappelijke onderzoekers (werkgemeenschappen, beoordelingscommissies) voldoende te hebben om te bevorderen dat grensverleggend onderzoek ("onderzoek aan het actuele researchfront") daadwerkelijk plaats vindt. Het tweede geldstroomonderzoek wordt op deze manier verdeeld. Het voorwaardelijk gefinancierd onderzoek aan universiteiten wordt langs deze lijnen gestuurd. De meeste wetenschappelijke onderzoekers denken langs deze lijnen.

De SVO heeft traditioneel een andere opvatting, die wordt gekenmerkt door het hanteren van inhoudelijke criteria en door het terugdringen van de invloed van de georganiseerde wetenschap. De SVO ziet zichzelf als een organisatie die de wetenschappelijke ontwikkelingen in onderwijsonderzoek in een voor haar wenselijke richting stuurt. De vele programmatische bemoeienissen van SVO zijn evenzovele mislukte pogingen om zo'n sturing feitelijk vorm te geven. Het OTG-beleid is een pregnant voorbeeld. Gestart na een uitgebreide selectie van inhoudelijke thema's zijn vijf werkgemeenschappen op de geselecteerde themagebieden opgericht met als opdracht 'op basis van eigenhandig uitgevoerde reconstructies en/of reviews hun actuele researchfront uit te tekenen in de vorm van goed en optimaal opgestelde onderzoeksproblemen' en om 'een gezamenlijke planning van onderzoeksactiviteiten rond (deze) gemeenschappelijke problemen' tot stand te brengen (Het OTG-beleid opnieuw bezien, 1988, p. 59-60). Eerder hebben we geconstateerd dat de OTG's deze opdracht niet hebben waar gemaakt en dat ze in feite zijn teruggevallen op procedures van de andere school om ontwikkelingen te sturen. Het ingezette beleid van fundamenteel strategisch onderzoek (PSO's) is in feite een hernieuwde poging om onderwijsonderzoek in een inhoudelijke richting te sturen.

Het is de vraag of een dergelijke sturing kan lukken. Ongetwijfeld zal de daarmee belaste adviescommissie nieuwe inhoudelijke thema's voorstellen waarop naar haar mening in de nabije toekomst bij voorrang onderzoek moet worden uitgevoerd. Er hoeft niet aan te worden getwijfeld dat de bestuursraad van de SVO enkele van deze thema's zal vaststellen. Vanaf dat moment is de SVO echter genooddaakt om het ontwerp en de uitvoering van onderzoeksprogramma's (PSO's) uit te besteden aan georganiseerde onderzoekers, zoals ook bij de implementatie van het OTG-beleid noodzakelijk was: beoordelingscommissies, werkgemeenschappen, vakgroepen en instituten. Zolang de in deze verbanden georganiseerde onderzoekers er een andere sturingsfilosofie op na houden, is ook het nieuwe PSO-beleid gedoemd te mislukken.

Op korte termijn zal dit nog niet zichtbaar zijn. In dit verband kan worden gewezen op de slimheid van SVO om een nieuwe werkgemeenschapsorganisatie te willen opzetten en daarbij in de richting van de VOR met de geldbuidel te zwaaien. Ook de subsidies voor binnen PSO's uit te voeren onderzoek zullen de bereidwilligheid om op korte termijn mee te werken beïnvloeden, zeker tegen de achtergrond van krapper wordende middelen en toenemende concurrentie. Op langere termijn echter zullen, naar mag worden verwacht, de inhoudelijke criteria verwateren, zullen de thematisch afgebakende lijnen vervagen en zullen criteria van wetenschappelijke kwaliteit en procedures van peer-review weer de overhand krijgen, als gevolg van de in wetenschappelijke kringen dominante sturingsfilosofie.

Verwatering van inhoudelijke criteria en vervaging van inhoudelijke thema-afbakening mogen eveneens worden verwacht tengevolge van het gegeven dat SVO de bij haar sturingsfilosofie behorende pretentie niet waar maakt om het omvangrijkste onderdeel van haar onderzoeksprogramma, het onderzoek ten behoeve van overheid en veld, althans gedeeltelijk in samenhangende en wetenschappelijke programma's in te delen. Het voornemen wordt met een zekere regelmaat geformuleerd en weer op de lange baan geschoven. Het is niet te verwachten dat de onderwijsonderzoekers hun eigen sturingsfilosofie zullen inruilen voor de opvatting terzake van de SVO, als de SVO geen blijk geeft ernst te maken met haar eigen sturingsfilosofie. De conclusie van een toekomstig evaluatierapport: 'Het PSO-beleid opnieuw bezien' kan derhalve worden voorspeld.

Ontvangen 12-9-1989

Boekbesprekingen

D. Brugman, *Personaliseren van de leerstof. Een interventie-onderzoek naar het ontwikkelen van waardegebieden door leerlingen tijdens lessen sociale-wereldoriëntatie*. Proefschrift Rijksuniversiteit Leiden, 1988 (259 pagina's).

Brugman maakt aan het begin van zijn proefschrift duidelijk dat: "Personaliseren van het onderwijs wil zeggen dat de bevordering van de persoonlijkheidsontwikkeling van de leerlingen als een hoofddoel betrokken wordt bij het inrichten van het onderwijs". (p. 5.) De psychologische theorie die hij in dit verband met name gebruikt is de waarderingstheorie, waaraan in het eerste hoofdstuk de voornaamste aandacht gegeven wordt. Eén en ander mondt uit in de probleemstelling van het onderzoek: "Welke factoren zijn van belang voor een effectieve interventiestrategie op het gebied van de waardeneducatie in de context van vakinhouden en wat zijn de leerresultaten op het cognitieve, affectieve en sociale domein van een interventiestrategie gebaseerd op de waarderingstheorie, waarin rekening met deze factoren wordt gehouden?" (p. 19-20).

In hoofdstuk 2 worden interventieprogramma's en hun (veronderstelde) effecten op het gebied van de waardeneducatie bij vakinhouden besproken. Met name het gebruik en bespreken van dilemma's in de leerstof zou effecten bij leerlingen kunnen bewerkstelligen. Aansluitend wordt in hoofdstuk 3 een mini-theorie gepresenteerd over de ontwikkeling van waardegebieden bij vakinhouden. Deze theorie is algemeen geformuleerd.

In hoofdstuk 4 worden allereerst vier hypothesen geformuleerd (p. 80-81). De eerste drie hypothesen hebben betrekking op interactionele psychologische processen tussen onderwijsaanbod-kenmerken op klas- of leerlingniveau en effecten op leerlingniveau; de laatste betreft het effect van een klaskenmerk op leerlingkenmerken. In het onderzoeksontwerp (p. 83 e.v.) worden twee dichotome factoren (wel/geen bevordering van waarderingprocessen over de aangeboden leerstof; wel/geen feedback tijdens het onderzoek aan de leerkrachten over hun vaardigheden in klassikale interacties met de leerlingen) met elkaar gekruist. De beschrijving van de onderzoeksprocedure en het schema met interventie-activiteiten en af te nemen instrumenten (p. 87) scheppen duidelijkheid over de onderzoeksstrategie (en concretiseren de voorafgegaane theorie). Het onderzoek gebeurt in 39 groepen uit de hoogste jaargroep van het basisonderwijs. Als begintoetsen worden bij de leerlingen (slechts) het klasseklimaat en de continue motivatie voor de zaakvakken gemeten.

Vervolgens wordt hoofdstuk 5 besteed aan de uitwerking van het onderwijsleerpakket en de bepaling van de thema's ten behoeve van de uit te voeren interventie. De ontwikkeling van het personaliserend onderwijsleerpakket wordt door projectmedewerkers zelf uitgevoerd. De thema's in de lesboekjes zijn achtereenvolgens: "De strijd om de walvissen", "De indianen en blanken in Noord-Amerika, vroeger en nu", en "Honderd jaar baanbrekers". Bij elk van de lesboekjes horen boekjes met vragen en opdrachten voor de leerlingen, een algemene handleiding en specifieke handleidingen voor de leerkrachten. De feedback aan de leerkrachten wordt gebaseerd op een instrument voor systematische observatie tijdens de les.

Hoofdstuk 6 is gewijd aan de ontwikkeling van de meetinstrumenten ter scoring van de afhankelijke variabelen. Aan de hand van de facetontwerptechniek wordt als afbeeldingszin van "waardering van de leerstof" bijvoorbeeld gepresenteerd: "De waardering van de leerstof is de mate waarin leerlingen voor zichzelf instemmen met *kwaliteit* (positief, negatief) uitspraken over de *dimensie* (cognitief, affectief, gedrag) betrokkenheid voor *onderwerp* (lesboek, lesinhouden, werkwijze, presentatie)". Op deze wijze worden items

gegenereerd, door leerlingen beantwoord, en geanalyseerd. In het algemeen resulteert dit in betrouwbare schalen (p. 138-139).

De informatie over het proces van personaliseren van de leerstof *tijdens* het onderzoek (hoofdstuk 7) is veelzijdig gedocumenteerd en leidt tot vier conclusies: (1) de interventie in alle vier condities lijkt redelijk tot goed geslaagd, (2) de integratie van de bevordering van het waarderingsproces bij het lesmateriaal lijkt redelijk geslaagd, (3) de feedback heeft stimulerend gewerkt, en (4) de integratie van personaliserend onderwijs in het reguliere curriculum was tijdens het onderzoek slechts gedeeltelijk te realiseren, maar in principe is deze goed te verwerkelijken.

In hoofdstuk 8 wordt aandacht gegeven aan de *effecten* van de interventie wat betreft de variabelen "bevordering van het waarderingsproces" en "feedback". De gebruikte analyse is covariantieanalyse (met behulp van multiple regressie) op klasniveau. De resultaten ondersteunen de hypothese dat *feedback aan de leerkracht* positief bijdraagt aan de continue motivatie en kennisverwerving door de leerlingen; niet gesteund wordt de veronderstelde invloed op de waardering voor de werkwijze bij het lesmateriaal. Verder worden positieve effecten gevonden van feedback op: de mening over het thema van de leerstof, een sterker ervaren empathie van de leerkracht, en (minder) cognitieve onzekerheid bij het bepalen van een standpunt in een dilemma. Ten aanzien van de effecten van *bevordering van het waarderingsproces* werd geen van de hypothesen bevestigd; wel werd de sociale wenselijkheid verminderd. Er werden enkele interactie-effecten geconstateerd.

In het slothoofdstuk worden de toekomstperspectieven geschetst. Het eerste perspectief was ten tijde van het schrijven van het proefschrift reeds gerealiseerd: het ontwikkelings/onderzoeksproject wordt voortgezet in de vorm van een diepte-project "Waarderend leren". Ten tweede wordt gepleit voor het verder empirisch uitbouwen van een theorie over het ontwikkelen van waarden door leerlingen. Ten derde passeren suggesties voor vervolgonderzoek de revue.

Het proefschrift overziende kan gesteld worden, dat het oorspronkelijke project uitermate ambitieus was opgezet. Deze ambities zijn grotendeels gerealiseerd. Het sterkste deel van het proefschrift vind ik het onderzoeksdeel: hier komt tot uiting dat er sprake is van een consequente pragmatische aanpak van tot nu toe zeldzaam opgepakte aspecten van het onderwijs(leer)proces. Om deze reden is de dissertatie "fundamenteel" van karakter. Dit wordt ook duidelijk in de bespreking van de bestaande literatuur in de eerste hoofdstukken: terugredenerend vanuit het onderzoek blijkt er zeer weinig relevante theorie te zijn. Hier draagt Brugman interessante bouwstenen aan, die in de bestaande (onderwijs)literatuur nauwelijks voorhanden zijn en die essentieel zijn in alledaagse onderwijs(leer)processen.

Er zijn natuurlijk ook kanttekeningen te maken. Het kostte mij vanwege de schrijfstijl relatief veel moeite om de eerste hoofdstukken te lezen. Ook de poging om aan de hand van een brochure deelnemende scholen te vinden is niet sterk, maar waarschijnlijk kan er bij een dergelijk interventie-onderzoek nooit sprake zijn van een random groep deelnemende scholen. Belangrijker is de kanttekening dat de voornaamste resultaten in het onderzoek gevonden zijn op de *feedback* factor, terwijl op p. 83 blijkt dat juist ten opzichte van de feedback conditie *zelf-selectie* van leerkrachten bij de aanmelding een rol kan hebben gespeeld. In volgend onderzoek zou dit niet meer mogen gebeuren. Tenslotte: de theorie en hypothesen bevinden zich grotendeels op leerlingniveau, maar de analyse wordt op groepsniveau uitgevoerd. Om theoretische en analytische redenen is het wenselijk (alsnog) een multi-niveau analyse te doen.

Geconcludeerd kan worden dat Brugman een belangwekkend interventie-onderzoek heeft verricht naar een problematiek, waaraan in het Nederlandse onderwijs ten onrechte

te weinig aandacht wordt gegeven. Met behulp van het beschikbare onderwijsleermateriaal en de lopende activiteiten in het diepteproject kan hopelijk een adequate lijn van onderwijsontwikkeling en -onderzoek worden voortgezet. Van belang is dan wel dat de gebruikte theorie "gezuiverd" wordt van onnodige ballast, zodat deze enerzijds doelgerichter wordt en anderzijds eenvoudiger (met minder meetinstrumenten) geoperationaliseerd kan worden. Dan blijft er vanzelf ruimte over voor het ontwikkelen van volgende leerstof en het betrekken van méér (groepen) leerlingen in het onderzoek.

Ton Mooij
Katholieke Universiteit Nijmegen
Instituut voor Toegepaste Sociale Wetenschappen
Graafseweg 274
Postbus 9048
6500 KJ Nijmegen

E. G. Harskamp, *Rekenmethoden op de proef gesteld*. (Monografieën voor onderwijsonderzoek 2) Groningen: RION, Rijksuniversiteit Groningen, 1988, 142 pp.

De voorbije vijftien jaar heeft er zich een belangwekkende verschuiving voorgedaan in het Nederlandse rekendidactische landschap. Deze verschuiving wordt algemeen gekarakteriseerd als de opkomst en de doorbraak van de z.g. realistische benadering van wiskunde en wiskunde-onderwijs. Dit is vooral het werk geweest van het in 1971 opgerichte Nederlandse Instituut voor Ontwikkeling van het Wiskunde Onderwijs (IOWO), en meerbepaald van de mensen die zich binnen dit instituut met het basisonderwijs bezighielden, het z.g. Wiskobasteam.

Volgens de realistische benadering is wiskunde geen vast systeem van begrippen, regels en formules dat zo efficiënt mogelijk van de ene op de volgende generatie moet worden overgedragen, doch in de eerste plaats zinvolle, menselijke activiteit. Goed wiskunde-onderwijs wordt in deze benadering dan ook niet opgevat als het aanbieden, uitleggen en laten inoefenen van een geheel van kant-en-klare begrippen en regels, maar als het creëren van een onderwijsleersituatie waarin de leerlingen gestimuleerd worden tot en geholpen worden bij het zelf opbouwen en ontdekken van hun wiskundige kennis uitgaande van de intuïtieve begrippen en informele strategieën waarover zij reeds beschikken. Deze algemene visie op wiskunde en wiskunde-onderwijs heeft niet enkel directe implicaties voor de concrete aanpak van het onderwijsleerproces, maar ook voor de keuze van de leerinhouden, de didactische werkvormen, de wijze van differentiëren, enz.

Het Wiskobasteam heeft haar ideeën en spullen nooit vertaald in een eigen rekenmethode. Wel besloten in de loop van de jaren zeventig diverse enthousiaste auteursgroepen een realistische rekenmethode te ontwikkelen, uitgaande van het door het Wiskobasteam ontwikkelde en verspreide gedachtengoed en onderwijsleermateriaal. Anno 1989 hebben de methoden die door de pioniers van de realistische stroming als relatief goede en getrouwe afspiegelingen van hun reken/wiskunde-onderwijsvisie beschouwd worden, al meer dan 2/3 van de Nederlandse boekenmarkt veroverd en deze opmars zal zich de komende jaren allicht nog verder doorzetten (De Jong, ter perse). Ook in Vlaanderen valt er de laatste tijd een sterk toenemende belangstelling voor deze realistische stroming waar te nemen (Verschaffel, 1988).

Voorstanders van realistisch reken/wiskunde-onderwijs verwachten uiteraard dat de nieuwe rekenmethoden tot betere rekenprestaties en tot meer plezier in rekenen zullen

leiden dan de traditionele methoden, tenminste wanneer er in de klas mee gewerkt wordt overeenkomstig de principes van de realistische reken/wiskundendidactiek.

Dit brengt ons meteen bij de centrale vraagstelling van het SVO-onderzoek "Vergelijking van Rekenmethoden in het Basisonderwijs" dat van 1984 to 1986 aan het RION is doorgevoerd en waarover gerapporteerd wordt in onderhavige dissertatie: in hoeverre zijn er verschillen tussen traditionele en realistische rekenmethoden in de leerresultaten van de leerlingen aan het eind van de basisschool, en in hoeverre zijn verschillen in gebruik van de rekenmethoden van invloed op de leerresultaten? (p. 3)

Bij de aanvang van de studie werd er een selectie gemaakt van de traditionele en de realistische methoden die in de analyse zouden worden betrokken. Daarvoor maakte Harskamp gebruik van de resultaten van het door De Jong (1986) uitgevoerd onderzoek naar de mate waarin de invloed van Wiskobas in de rekenmethode terug te vinden is (en waarin gebruik gemaakt is van een beoordelingslijst bestaande uit 21 Wiskobaskenmerken). Uiteindelijk werden enkel die methoden opgenomen die per cursusjaar 1984/85 *tot en met de zesde klas* (d.w.z. groep 8) volledig uitgebracht waren. Een spijtig gevolg daarvan is wel dat drie nog niet helemaal afgewerkte methoden die reeds in 1984 algemeen als veruit de meest getrouwe uitwerkingen van het Wiskobas-gedachtengoed werden beschouwd (nl. "Rekenwerk", Rekenen en Wiskunde" en "De Wereld in Getallen"), in Harskamp's onderzoek uit de boot gevallen zijn.

Harskamp's eerste onderzoeksvraag slaat op de verschillen en de overeenkomsten tussen realistische en traditionele rekenmethoden wat betreft de nadruk die op de *diverse leerstofonderdelen* wordt gelegd. Om deze vergelijking te kunnen maken, heeft hij een nieuw analyseschema ontwikkeld, dat naast de traditionele leerstofcategorieën uit de CITO-eindtoets basisonderwijs (bijv. getallen, hoofdrekenen, enz.) ook enkele bijkomende leerstofcategorieën bevat (bijv. meten met natuurlijke maten, schatten, combinatoriek, meetkunde, enz.). Een analyse door één beoordelaar laat zien dat er inhoudelijk gezien tussen de methoden over het geheel genomen geen grote verschillen zijn wat betreft de aandacht voor bepaalde leerstofcategorieën. Er was enkel een lichte trend van de realistische methoden om meer aandacht te besteden aan de vier laatstgenoemde onderdelen. Een bijkomende analyse van enkele makkelijk te detecteren *verschillen in instructiewijze* tussen de traditionele en de realistische methoden (zoals leerstofdifferentiatie, variatie in leerstof), leverde evenmin grote verschillen op; over het geheel genomen bleken de moderne methoden iets meer variatie in leerstof (m.n. meer redactie-opgaven) te bieden dan de traditionele methoden.

Voor het tweede gedeelte van het onderzoek werd een instrument geconstrueerd bestaande uit items uit de CITO-eindtoets aangevuld met een aantal toetsen die speciaal opgesteld waren in samenspraak met de auteurs van de in het onderzoek betrokken realistische methoden, die aanvullende doelstellingen getoetst wilden zien. Daarnaast werd er van de leerlingen ook nog een intelligentietest, een rekenattitudevragenlijst en een vragenlijst met achtergrondgegevens afgenomen. Tot slot werden er ook van de leerkrachten via vragenlijsten gegevens ingewonnen over het gebruik van de methode in de klas evenals enige achtergrondgegevens betreffende henzelf.

Een aselechte gestratificeerde steekproeftrekking van scholen die aan de CITO-eindtoets voor het basisonderwijs deelnemen, leverde 118 klassen (niveau zesde leerjaar of groep 8) op. Bekijken we het aantal klassen en leerlingen per methode (p. 46), dan valt op dat de realistische methode die door de experts uit De Jong's (1986) onderzoek veruit de laagste "getrouwheidscore" kreeg (nl. "Operator Rekenen", dat een getrouwheidscore van

slechts 48% kreeg, tegenover meer dan 90% voor de drie eerder genoemde, meer recente realistische methoden), bijna de helft van de "realistisch gevormde" leerlingen uit onderhavige studie geleverd heeft.

Een eerste serie resultaten heeft betrekking op de relatie tussen het gebruik van de acht methoden enerzijds en een aantal implementatiekenmerken van de methode en achtergrondkenmerken van leerkrachten en leerlingen anderzijds. Daaruit onthouden we vooreerst dat de methoden zoals bedoeld van invloed zijn op de implementatie van twee *vakdidactische implementatiekenmerken* (m.n. leerstofdifferentiatie en variatie in leerstof). Wat de leerlingkenmerken betreft, valt vooral het verschil in *intelligentiescores* op tussen klassen waarin met een traditionele en realistische methode gewerkt is. De vanzelfsprekendheid waarmee de auteur deze merkwaardige bevinding aan vooraf bestaande, (reken)onderwijs-onafhankelijke factoren toeschrijft (en dus de mogelijkheid afwijst dat de gevonden intelligentieverschillen iets te maken zouden hebben met het rekenonderwijs dat deze leerlingen gedurende zes jaar hebben ontvangen), is enigszins verrassend. Bezien we tot slot de leerkrachtkenmerken, dan onthouden we vooral het verschil in *het aantal jaren ervaring met de methode* tussen de leerkrachten die werken met een traditionele en met een realistische methode.

De belangrijkste vraag is echter of er verschillen gevonden werden tussen de methoden in rekenprestaties en rekenattitudes van de leerlingen. Op basis van diverse statistische analyses komt de auteur tot de conclusie dat de verwachting dat leerlingen die met realistische methoden hebben leren rekenen betere rekenprestaties en meer positieve rekenattitudes behalen dan leerlingen van traditionele methoden, niet kan worden weerhouden. Zelfs op die onderdelen die speciaal ontworpen zijn op vraag van de ontwikkelaars van de realistische methoden (combinatoriek, meten, schatten en verhoudingen), blijkt de factor "methode" geen significante invloed uit te oefenen op de prestatie- en attitudescores van klassen. De enige uitzondering hierop vormt het onderdeel "combinatoriek".

In het laatste hoofdstuk van het proefschrift wordt er op basis van de beschikbare resultaten enerzijds en een aantal algemene veronderstellingen omtrent effectief onderwijs anderzijds, een exploratief model voorgesteld van de relatie tussen onderwijsfactoren en rekenprestaties, dat vervolgens getoetst wordt door middel van de analysetechniek LISREL.

Samenvattend kunnen we stellen dat de resultaten van Harskamp's onderzoek suggereren dat er inhoudelijk gezien maar weinig verschil bestaat tussen traditionele en realistische methoden, en dat laatstgenoemde methoden als zodanig geen betere resultaten opleveren dan traditionele methoden. Op het einde van zijn dissertatie noemt de auteur trouwens nog een aantal andere recente studies die deze laatste conclusie zouden ondersteunen. De enkele onderzoeken waarin *wel* betekenisvolle verschillen werden gevonden, zitten volgens hem methodologisch zo zwak in elkaar, dat op grond daarvan geen verantwoorde uitspraken mogelijk zijn (p. 89). In dit verband vraag ik me echter af of er aan de studies waarin *geen* verschil tussen traditionele en realistische methoden gevonden werd, dan geen methodologische zwakheden kleven.

Tot zover een overzicht van de opzet en de resultaten van dit groots opgezet evaluatie-onderzoek. In wat voorafging heb ik reeds enkele kritische opmerkingen gemaakt. Tot slot van deze bespreking plaats ik nog enkele bedenkingen van meer algemene aard.

Een eerste problematisch aspect van onderhavige studie is dat geen van de drie methoden die thans algemeen erkend worden als de meest getrouwe uitwerkingen van het realistisch gedachtengoed, in onderhavige studie betrokken zijn. De auteur geeft als argument hiervoor dat deze methoden bij de aanvang van het onderzoek nog niet helemaal

beschikbaar waren, laat staan gebruikt werden in de zesde klas. Dit lijkt een redelijk argument. Doch had het niet de moeite geloond om deze nieuwe methoden toch al op te nemen in de inhoudsanalyse? In elk geval heeft deze beslissing ertoe geleid dat de actualiteitswaarde van Harskamp's dissertatie reeds bij het ter perse gaan eerder gering was. In dit verband kan trouwens verwezen worden naar het recent peilingsonderzoek van het onderwijsniveau (PPON), waarin duidelijke aanwijzingen te vinden zijn dat de realistische methoden wel degelijk betere resultaten opleveren dan traditionele methoden, en wel op die leerstofonderdelen waar de leergangen het meest ingrijpend veranderd c.q. verbeterd zijn (De Jong, ter perse; Treffers, 1988).

Ten tweede kleven er aan onderhavige studie toch ook een aantal problemen van onderzoekstechnische aard, die door de voorstanders van de realistische stroming kunnen ingeroepen worden ter verklaring van de afwezigheid van betekenisvolle verschillen tussen de methoden. Ik wijs in dit verband op de manier waarop de variabele "methoden" in het onderzoek is geoperationaliseerd, op de wijze waarop de verschillende methoden over de steekproef zijn verdeeld, op de manieren waarop de verschillen in het leerstofaanbod en in de implementatie werden gemeten, en op het gebruik van de CITO-eindtoets om de opbrengsten van rekenmethoden te meten (voor meer details zie Treffers, 1988; Grave-meijer, persoonlijke mededeling).

Tot slot mag niet uit het oog verloren worden dat er vanuit realistische hoek nooit beweerd is dat een nieuwe methode *op zich* een betekenisvol positief effect zou hebben op de rekenprestaties en het rekenplezier van leerlingen. Integendeel, reeds in 1986 waarschuwde o.m. De Jong ervoor dat een toenemende verspreiding van op realistische leest geschoeide rekenmethoden geenszins zou garanderen dat het door de methodenschrijvers beoogde onderwijs ook feitelijk in de klas zou worden gerealiseerd. "Er is de cruciale rol van de leerkracht die dit alles nog moet waarmaken", zo waarschuwde hij. "Zonder voldoende achtergrondkennis en ondersteuning bestaat het gevaar dat de vernieuwingsideeën worden ingekapseld in de bestaande routines en opvattingen, en nooit tot volle ontplooiing komen" (De Jong, 1986). Daarom wordt er de laatste jaren vanuit verschillende hoeken zoveel aandacht besteed aan de opleiding, de nascholing en de begeleiding van leerkrachten die met deze nieuwe methoden aan de slag willen gaan. Ook op dit punt lijken de aanbevelingen op het einde van Harskamp's dissertatie te laat te komen.

REFERENTIES

- De Jong, R. (1986). *Wiskobas in methoden*. (Doctoraatsproefschrift.) Utrecht: Vakgroep Wiskunde-onderwijs en Onderwijs Computercentrum.
- De Jong, R. (ter perse). Actuele informatie over reken/wiskunde-onderwijs. *Basisschoolmanagement*, ter perse.
- Treffers, A. (1988). Over de merkbare invloed van onderwijsmethoden op leerprestaties. In J. Wijnstra (Red.), *Balans van het rekenonderwijs op de basisschool. Opbrengst van de periodieke peiling van het onderwijsniveau (PPON)* (pp. 120-132). Arnhem: CITO.
- Verschaffel, L. (1988). Realistisch reken/wiskundeonderwijs in Nederland. Een kennismaking. *Pedagogische Periodiek*, 94, 322-338.

L. Verschaffel
K.U. Leuven
Vesaliusstraat 2
3000 Leuven

Pieter de Koning, *Programmadifferentiatie in het voortgezet onderwijs*. Een studie naar de achtergronden, ontwikkelingen en problemen van programmadifferentiatie in de eerste fase van het voortgezet onderwijs. Lisse: Swets en Zeitlinger, 1987 (Proefschrift Universiteit van Amsterdam).

Pieter de Koning geldt in de Nederlandse onderwijskunde als bij uitstek deskundig op het terrein van de differentiatie in het onderwijs. Verschillende bijdragen, empirische zowel als theoretische, getuigen daarvan. De voorliggende publikatie heeft onder meer als doel de in verschillende deelstudies verkregen inzichten samen te brengen in één studie: beschrijving en analyse van ontwikkelingen op het terrein van de differentiatie in het onderwijs. Bovendien is daarvoor een meer praktische en actuele aanleiding: de betrokkenheid van De Koning bij de vormgeving en inrichting van de eerste fase van het voortgezet onderwijs, de basisvorming, die vraagt om een heroriëntatie van het onderzoek naar differentiatie. Het boek heeft dus ook nog een meer programmatisch doel, namelijk een betere conceptualisering en programmering van de differentiatieproblematiek, teneinde de volgens de auteur segmentarische en statische wijze van behandelen van deze thematiek te kunnen verlaten. De Koning beperkt zich in deze studie overigens tot de differentiatie in de eerste fase van het voortgezet onderwijs.

De auteur gaat daarbij als volgt te werk. Na een inleidend hoofdstuk wordt in hoofdstuk 2 het begrip differentiatie nader omschreven: "het aanbrengen van verschillen tussen delen van een onderwijs(sub)systeem ten aanzien van een of meer aspecten." (p. 8/9). Het wordt gezien als een planmatige, expliciet bedoelde activiteit, resulterend in een reeks van maatregelen die leiden tot verschillen in onderwijsinhouden (programma) en instructie onder andere. De relatie met het fenomeen selectie wordt hier behandeld evenals verschillende soorten van differentiatie. Programmadifferentiatie is een vraagstuk dat zowel organisatorische kanten, selectie-aspecten als curriculumproblemen bevat. Historisch gezien, aldus de auteur, is een ontwikkeling zichtbaar van externe, voornamelijk interscholair, naar interne differentiatie, dus binnen klassen.

In hoofdstuk 3 wordt die ontwikkeling in het kort geschetst. De Koning laat zien dat de programmadifferentiatie vorm krijgt door middel van interscholair, interklassikale en interne (= intra-klassikale) differentiatie. Elk van deze wordt grondig behandeld. Hij konkludeert dat bij alle drie sprake is van op zijn minst drie problemen die voortdurend om aandacht vragen: het probleem van het onderscheid, het probleem van de selectie en het probleem van de sociale en culturele segregatie. Bovendien stelt hij vast dat combinaties van differentiatievormen de structuur van de eerste fase van het voortgezet onderwijs bepalen. Die combinaties zijn echter zelden object van onderzoek; het gebruikelijke onderzoek werkt met modellen die statisch zijn en ten opzichte van elkaar geïsoleerd en gepolariseerd.

Volgt een lang maar boeiend hoofdstuk 4, waarin wordt nagegaan welke plaats programmadifferentiatie inneemt in het onderwijskundig onderzoek. Vier invalshoeken zijn daarbij van belang: programmadifferentiatie als prediktieprobleem ("welke leerlingen zijn al dan niet geschikt voor verschillende programma's?"), als aansluitingsprobleem ("hoe didactisch en methodisch de aansluitingen het best te realiseren?"), als reproductieprobleem ("wat zijn de gevolgen van differentiatie en selectie voor de maatschappelijke structuren en verhoudingen?") en tenslotte als effectiviteitsprobleem ("welke vorm van differentiatie is het meest effectief?"). Het kost weinig moeite in te zien dat deze vier vraagstellingen min of meer samenvallen met disciplinaire invalshoeken in de onderwijswetenschappen, resp. de psychologie (prediktie), de opvoedkunde (aansluiting) en de sociologie (reproductie). Het probleem van de effectiviteit onttrekt zich aan een dergelijke afbakening en is betrekkelijk theorieeloo. De Koning laat zien wat de betekenis van elk van

deze vier invalshoeken is geweest voor de praktijk van het onderwijs(beleid) en wat hun betekenis is voor de praktijk van het onderzoek. Met name bij dit laatste punt voert hij een pleidooi voor een doorbreken van de waterdichte schotten rond elk van deze benaderingen, die van elkaars bevindingen blijkbaar niet veel moeten hebben. Zo schenken bijvoorbeeld reproductieonderzoekers weinig aandacht aan programmadiifferentiatie als aansluitings- of als prediktieprobleem en andersom. De Koning bepleit theorievorming rond programmadiifferentiatie die de verschillende probleemformuleringen gelijktijdig omvat. Dat gebeurt in hoofdstuk 6.

Maar eerst is er nog hoofdstuk 5 waarin een relatie wordt gelegd tussen differentiatie en 'schoolconcept', zijnde het geheel aan opvattingen over onderwijs en opvoeding die de achtergrond vormen waartegen programmadiifferentiatie gestalte krijgt. Sociologen zouden spreken over de kulturele dimensie van het onderwijs: opvattingen over doelen en functies van het onderwijs voor de samenleving als geheel en voor de ontwikkeling van kinderen in het bijzonder. Volgens de auteur spelen dergelijke concepten in de onderwijskunde en in de schoolpraktijk een belangrijke rol. Hij onderscheidt de volgende drie schoolconcepten: de meritokratische opvatting (onderwijs als geselecteerde vorming), de naturalistische (onderwijs als vrije vorming) en de egalistische opvatting (onderwijs als basisvorming). Voor elk van deze drie werkt hij de konsekwenties voor programmadiifferentiatie nader uit, naar curriculum, selectieproces, afsluitingen, organisatie en ondersteuning o.a.

Hoofdstuk 6 bevat uiteindelijk een voorstel voor een beschrijvingsmodel van differentiatiestructuren. De hiervóór behandelde elementen krijgen hierin een plaats. Ook is er veel aandacht voor achtergrond- en kontekstvariabelen (onderwijsbestel, regio, vormingsconcepten, schoolorganisatie, leerlingenmerken, schoolgeschiedenis enz.) die in het gebruikelijke schoolkenmerkenonderzoek (waartoe De Koning het onderzoek naar differentiatiestructuren rekent) vaak ontbreken. Dat geldt trouwens ook voor het 'teacher-effectiveness' onderzoek. Het model is sterk beïnvloed door Lundgren (1972) die variabelen uit macro-, meso- en microniveau op soortgelijke wijze in een model bijeenbrengt ter verklaring van het onderwijsproces.

In een nabeschuiving (hoofdstuk 7) wordt een perspectief geboden op verder onderzoek en ontwikkeling met behulp van de opbrengsten van deze studie.

Ik heb De Konings dissertatie met genoeg gelezen, al is het een vrij massieve en omvangrijke tekst om door te nemen. De stijl is sober en zakelijk. Hoofdstuk 5 over de verschillende schoolconcepten vind ik het minst geslaagd, vermoedelijk omdat dit weinig nieuws biedt vergeleken met wat op dit terrein al bekend is, zeker in de onderwijssociologie. In het voorgestelde model vind ik deze elementen ook nauwelijks terug.

Alles bij elkaar mag het resultaat er zijn: een goede en degelijke overzichtsstudie van iemand die op dit terrein zijn sporen heeft verdiend. De Koning is onderwijskundig breed georiënteerd, kent de literatuur en maakt veelvuldig gebruik van inzichten van andere disciplines uit de sociale wetenschappen zoals die van psychologen en sociologen. Hij gaat kortom interdisciplinair te werk en is duidelijk afkerig van een smalle, verkokerde aanpak binnen de onderwijskunde. En terecht. Hij is ook een knap syntheticus die ideeën van verschillende origine samen weet te voegen in een soms uitdagende stellingname voor een nieuwe aanpak op dit terrein. Maar ook is hij niet bang voor grondige analyses en gaat hij pogingen tot theorievorming niet uit de weg, twee zaken waar het de onderwijskunde in haar ijveren voor empirie nogal eens aan ontbreekt. Niet voor niets bepleit hij ruimte voor "discursieve kennisontwikkeling" naast empirische confrontaties met de sociale werkelijkheid. De discussie over differentiatie (die naar mijn mening weer wat 'terugkomt', zie bijvoorbeeld de speciale aflevering van het Pedagogisch Tijdschrift sept/okt. 1988) zal er

mee gebaat zijn. Ook de discussie over de toekomstige basisvorming in de eerste fase van het voorgezet onderwijs zal gefundeerder worden gevoerd na kennisneming van zijn ideeën.

Een boek dat meer aandacht verdient dan het tot nu toe heeft gekregen.

Anton Wesselingh
K.U. Nijmegen
Vakgroep Interdisciplinaire Onderwijskunde
Erasmusplein 1-20
6525 GG Nijmegen.

A. Schouten, *Denken en doen in het technisch onderwijs: Leerlingstrategieën bij het uitvoeren van praktijkopdrachten*. De Lier: Academisch Boeken Centrum (Proefschrift K.U. Brabant), (1988).

Onderwijspsychologisch onderzoek naar het aanleren van vaardigheden is schaars in Nederland. Onderwijspsychologisch onderzoek naar het leren binnen het technisch onderwijs is eveneens schaars in Nederland. Ten onrechte, aangezien er genoeg redenen zijn zulk onderzoek in dat deel van het veld uit te voeren. Niet slechts om verbetering van leerresultaten na te streven, maar tevens omdat bij het aanleren van vaardigheden in het technisch onderwijs het denken en het doen elkaar zeer dicht naderen. Een uit onderwijspsychologisch oogpunt zeer interessant probleem: mensen voeren vele taken uit waarbij vele handelingen in het geding zijn en velen kunnen daarbij ook aangeven wat ze doen. Een interessant punt hierbij is in hoeverre het denken over de handeling plaatsvindt vóór of ná de uitvoering van de handeling. Voor vele onderzoekers in de psychologie een vraag die het onderzoeken meer dan waard is. In het verleden zijn bespiegelingen gewijd aan en theorieën opgesteld over de epifenomenen van ons gedrag: het vormen van cognitieve representaties van uitgevoerde handelingen, die 'after the fact' een rechtvaardiging voor het waarom van het gedrag leveren. Tegenwoordig spreken we van mentale modellen, beelden of representaties van handelingen, die als basis dienen voor niet alleen de uit te voeren handeling, maar ook voor de waarneming van het resultaat van die handeling of van andere gelijksoortige handelingen door dezelfde persoon of door een ander uitgevoerd. De bekende onderzoeker naar motorische controle Michael Turvey geeft daarvan een simpel, maar treffend voorbeeld. De letter *A* kan op vele manieren worden geschreven of worden gedrukt in velerlei oriëntaties, wat blijft is dezelfde gewaarwording van een *A*. Deze constantie in de waarneming treffen we ook aan in het produceren van de letter *A*. De letter kan worden geschreven met veel verschillende spieren of spiergroepen, niet noodzakelijkerwijs van alleen de hand. Voor al deze geschreven en waargenomen *A*'s bestaan niet exacte copieën in ons geheugen. Dat zou schier onmogelijk zijn. De oplossing is gelegen in het veronderstellen van een dieptestructuur, een schema, of interne representatie.

Tony Schouten heeft zich in haar overzichtelijk en goed geschreven proefschrift (126 pagina's en bijlagen; promotores: prof. dr. L. F. W. de Klerk en prof. dr. M. Groen) niet druk gemaakt over deze zaken. Zij gaat ervan uit, op basis van een rationele analyse van praktijkopdrachten in het lager technisch onderwijs, dat aan het uitvoeren van een technische handeling altijd het vormen van een mentale representatie van deze handeling dient vooraf te gaan. In het algemeen kent volgens Schouten een efficiënte taakuitvoering de volgende processen: beeldvorming, analyse, planning, uitvoering en evaluatie. Vandaar

dat zij aan het einde van hoofdstuk 2 stelt dat '...de uitvoering van een praktijkopdracht niet alleen zal worden bepaald door de kwaliteit van de psychomotorische deelhandelingen, maar in grote mate mede wordt bepaald door de kwaliteit van de voorafgaande en begeleidende cognitieve processen'.

Het proefschrift bevat een verslag van zes deelonderzoeken die zijn uitgevoerd naar de relaties tussen de R(esultaat)-, O(nderwijssituatie)- en L(eerling)-factoren en de wijze waarop hiermee bij het arrangeren van optimale onderwijsleersituaties moet worden rekening gehouden. Alvorens de deelonderzoeken te bespreken wijdt Schouten een hoofdstuk aan de analyse van praktijkopdrachten in het lager beroepsonderwijs. De rationele taakanalyses zijn in een eerdere fase van het onderzoek uitgevoerd en daarover is elders reeds samen met Johan van der Sanden gepubliceerd. Schouten resumeert slechts het bekende quintet van processen: beeldvorming, analyse, planning, uitvoering en evaluatie.

De resultaten van de zes deelonderzoeken zijn gebruikt om een veelvoud aan vragen te beantwoorden.

De eerste vragen hebben betrekking op motorische vaardigheden en motorische vermogens. In hoeverre zijn motorische vermogens van belang als het gaat om het verklaren van verschillen tussen leerlingen in de uitvoering van praktijkopdrachten? Het antwoord van Schouten is: nauwelijks. Voor de hand ligt te vermoeden dat verschillen verklaard kunnen worden op grond van de beheersingsgraad van relevante motorische basisvaardigheden. Ook hier moet een negatief antwoord worden gegeven, zodat de conclusie volgens Schouten moet zijn dat '... een verklaring voor de grote verschillen in prestaties gezocht moet worden in andere dan de (psycho)motorische aspecten van praktijkopdrachten'.

De volgende vragen dienen zich dan aan. Welke (niet-motorische) leerlingkenmerken zijn van belang? En welke relaties zijn er met de kwaliteit van de vervaardigde producten? Schouten gaat uit van de koppeling van leerlingkenmerken aan de hierboven reeds genoemde processen. Onderzocht zijn derhalve leerlingkenmerken als ruimtelijk inzicht, veldafhankelijkheid en prestatiemotivatie. In totaal zijn 14 tests afgenomen en de resultaten op deze tests zijn vergeleken met de resultaten bij de praktijkopdrachten. Opvallend is de grote mate van overeenkomst tussen de bij de praktijkopdracht behaalde resultaten en de leerlingkenmerkende technisch- en ruimtelijk inzicht, veldonafhankelijkheid en subjectieve competentie. Aanzienlijk minder sterk is de relatie met prestatiemotivatie, faalangst en impulsiviteit.

Schouten onderbreekt in haar proefschrift de lijn van leerlingkenmerken naar instructie met het besteden van aandacht aan het analyseren van leerlingstrategieën. Zij begint aan dit onderwerp met enige reserve. Het analyseren van kwalitatieve gegevens is niet zonder risico. Schouten kiest voor een brede benadering waarin plaats is voor observatie, vragenlijst, hardop denk protocollen, en retrospectie. Wijselijk zijn geen fysiologische metingen verricht. Afgezien van de grote technische problemen leveren deze metingen betrekkelijk weinig inhoudelijke procesinformatie tegenover het vele werk dat er mee gemoeid is. Terecht merkt Schouten op dat grote voorzichtigheid moet worden betracht met het interpreteren van de kwalitatieve data. Alhoewel zij in haar proefschrift niet ingaat op de vraag in hoeverre een proefpersoon in staat moet worden geacht mentale processen die de uitvoering van handelingen begeleiden, te verbaliseren, geschiedt de verzameling van de gegevens met grote zorgvuldigheid. Daarbij wordt ter ondersteuning van de indeling in mentale processen die een rol spelen bij de taakuitvoering, een nieuwe dichotomie geïntroduceerd: handelings- en toestandsgerichte leerlingen. Leerlingen worden toegewezen aan de ene groep of aan de andere groep op basis van uitspraken die zij doen tijdens het bekijken van video-beelden van hun eigen taakuitvoering. Schouten

maakt echter geen melding van de precieze operationalisatie en scoring van de handelingsgerichte en de toestandsgerichte aanpak van leerlingen. Uit de resultaten kan worden opgemaakt dat een zeer gering percentage leerlingen (ongeveer 12%) tijdens *beeldvorming*, *analyse* en *planning* een handelingsgerichte aanpak hanteert. Overigens moet worden vermeld dat het aantal niet te classificeren uitspraken van leerlingen in de fasen *uitvoering* en *evaluatie* erg hoog ligt. Een verklaring daarvoor wordt niet gegeven. Tevens blijft de lezer in het ongewisse of er belangrijke individuele strategische verschillen te constateren zijn tijdens beeldvorming, planning etc. Een uit onderwijspsychologisch oogpunt niet oninteressante vraag.

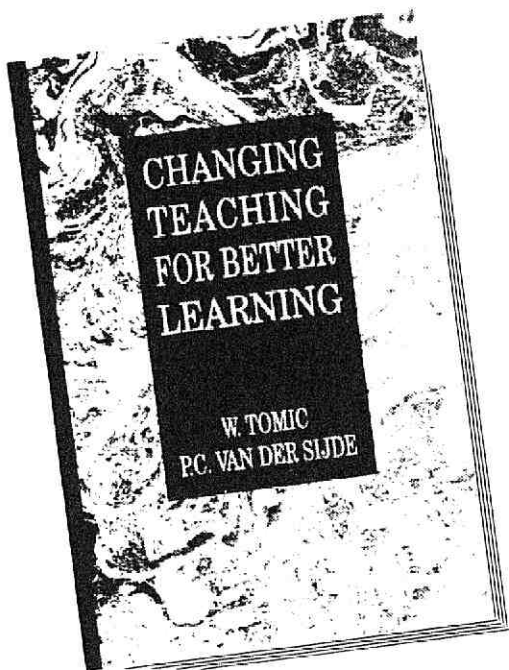
Schouten is vervolgens nagegaan in hoeverre er een relatie bestaat tussen technische intelligentie en handelingsgerichtheid. Deze relatie blijkt zich wel degelijk te openbaren op grond van de correlatie-metingen, waarmee overigens niet gezegd kan worden dat alle technisch-intelligente leerlingen zich bedienen van een handelingsgerichte aanpak.

Hoe moet de instructie worden vormgegeven voor de leerlingen bij het uitvoeren van praktijkopdrachten? Geheel in de traditie van het ATI-onderzoek komt Schouten met de vergelijking tussen technische intelligentie, voorstructurering (aanwijzingen) en resultaten. Verrassende gegevens levert deze vergelijking niet op. Dat relatief intelligente leerlingen de beste prestaties leveren indien zij geen of slechts globale aanwijzingen krijgen, was reeds bekend. Onderzoek naar 'learner control' toont aan dat met name intelligente leerlingen geneigd zijn vast te houden aan eigen strategieën ongeacht in welke richting zij tijdens de instructie zijn gestuurd. Dat minder intelligente leerlingen het beste presteren indien zij veel, gedetailleerde aanwijzingen krijgen omtrent de *uitvoering*, is interessanter. Vooral als dit gegeven wordt gecombineerd met het hierboven vermelde gegeven dat in de categorie *uitvoering* het hoogste aantal niet te classificeren uitspraken voorkomt. Schouten gaat daar overigens niet op in.

Tenslotte heeft Schouten de rol van de docent onderzocht. Lessen van docenten mechanische technieken en electrotechniek zijn geobserveerd en interviews afgenomen. De lessenregistratie en de interviews betroffen de rol van de docent bij de taakuitvoering door de leerlingen. De conclusie van Schouten is, weergegeven in mijn woorden, dat de docent in het algemeen een toestandsgerichte aanpak praktizeert: hij besteedt weinig aandacht aan beeldvorming en nodigt in zijn instructie de leerling niet uit tot het zoeken naar eigen oplossingen.

De conclusie is dat dit proefschrift een fraaie afsluiting vormt van een reeks onderzoeken die het cognitief-psychologisch denken over het aanleren van psychomotorische vaardigheden in daden heeft omgezet. Niet alleen voor de onderzoeker maar ook voor de ontwerper van instructiesystemen van belang.

J. M. Pieters
Universiteit Twente
Toegepaste Onderwijskunde
Postbus 217
7500 AE Enschede



What is it that teachers do that leads to student learning? How can we increase students' achievement and improve their attitudes? In order to improve the quality of education, the authors focus on teaching/learning processes taking place in the traditional classroom situations. The study reported in this book is the Dutch contribution to the IEA Classroom Environment study: Teaching for Learning, whose general objective is to identify alterable teaching behaviors that correlate with desirable cognitive and affective student learning outcomes, and to develop a training

Changing Teaching for Better Learning

W. Tomic, P.C. van der Sijde
Open University
Heerlen, Holland

program that will recommend ways of organizing teaching/learning processes. For their research the authors used the so called "observational-correlational-experimental loop". They have conducted both a correlational and an experimental study in regular classrooms with regular teachers teaching regular subject matter.

The results of this research show that even a short training course can successfully change the teachers' teaching script, and subsequently their teaching behavior, which in turn influences student achievement.

Verkrijgbaar in de boekhandel.

**Changing Teaching for
Better Learning**
W. Tomic, P.C. van der Sijde
ISBN 90 265 1019 5
f 29.95, 128 blz.

S&Z

SWETS & ZEITLINGER B.V.

Heereweg 347, 2161 CA Lisse, Tel. 02521-35111

Schoolkenmerken en leerlingfunctioneren; een replicatie-onderzoek*

Anja W.M. Knuver

RION: Instituut voor Onderwijsresearch

ABSTRACT

The research discussed here is a replication of a study on the relationship between school factors and pupil functioning. The data were gathered in two types of Dutch secondary schools, LBO and MAVO schools. The aim of this replication was to compare the results of the main study with those of a similar sample. We found that many significant relations between school factors and pupil functioning found in one sample failed to replicate in the second. These findings stress the danger of capitalizing on chance in correlational research.

INLEIDING

In 1983/1984 werd door het RION, Instituut voor Onderwijsresearch, een onderzoek opgezet, met als probleemstelling welke schoolkenmerken er samenhangen met verschillen in schooloutput. Voor de uitvoering van dit onderzoek werden twee onafhankelijke steekproeven getrokken, beiden zowel van MAVO- als LBO-scholen. In beide steekproeven werd vergelijkbare informatie verzameld. De analyse van de gegevens uit de eerste steekproef (het hoofdonderzoek), heeft indicaties opgeleverd over welke schoolkenmerken relevant zijn met betrekking tot het functioneren van leerlingen (Stoel, 1986). Deze indicaties zijn verkregen met behulp van correlatieve technieken, waarbij een groot aantal schoolkenmerken in verband werd gebracht met een aantal aspecten van het leerlingfunctioneren. Karakteristiek voor dit soort analyses, waarbij een groot aantal variabelen met elkaar in verband wordt gebracht zonder dat daaraan een hecht theoretisch kader ten grondslag ligt, is dat een aantal van de gevonden correlaties op toeval kan berusten. Een mogelijkheid om dit probleem te ondervangen is het uitvoeren van replicatie-onderzoek. Het doel van de replicatie van de analyses van Stoel is na te gaan of de daar gevonden verbanden terug te vinden zijn in een tweede, onafhankelijke steekproef.

THEORETISCHE ACHTERGROND

Onderzoek naar de samenhang tussen schoolkenmerken en leerlingfunctioneren heeft in de loop der jaren een ontwikkeling doorgemaakt. In de zestiger jaren leidde onderzoek van Coleman e.a. (1966) tot de conclusie dat scholen niet kunnen compenseren voor

* Met dank aan Wouter G.R. Stoel, die de data voor deze replicatie welwillend ter beschikking heeft gesteld. (De data zijn verzameld in het kader van een project dat is gesubsidieerd door het Instituut voor Onderzoek van het Onderwijs (SVO) te Den Haag, project nr. SVO 0610). Voorts dank aan Margo G.H. Jansen en Hennie P. Brandsma voor de begeleiding bij het schrijven van mijn doctoraalscriptie, waarvan dit artikel een samenvatting is.

leerlinginstroomkenmerken. In de zeventiger jaren kwamen onderzoeksresultaten beschikbaar, waaruit zou blijken dat scholen wel degelijk verschillen vertonen met betrekking tot het functioneren van leerlingen, ook nadat gecontroleerd was voor leerlinginstroomkenmerken als intelligentie en sociaal milieu (Blom e.a., 1986). In de jaren tachtig zijn ook in Nederland onderzoeken uitgevoerd naar de relatie tussen schoolorganisatiefactoren en het functioneren van leerlingen (Stoel, 1980; Van Marwijk Kooy-Von Baumhauer, 1984; Meijnen, 1984; Van der Wolf, 1984; Brandsma & Stoel, 1987; Van der Grift, 1987; Vermeulen, 1987; Brandsma & Knuver, 1988).

De meeste onderzoeken naar de relatie tussen schoolkenmerken en leerlingfunctioneren hebben een aantal potentieel relevante schoolkenmerken opgeleverd. Bij vergelijking van de onderzoeksresultaten blijkt echter al snel dat niet telkens dezelfde schoolkenmerken belangrijk lijken voor het verklaren van verschillen in leerlingfunctioneren. Sommige studies leveren zelfs tegenstrijdige resultaten op. De verklaring hiervoor kan gezocht worden in de complexiteit van de onderzochte materie en de problemen bij de opzet en uitvoering van de onderzoeken. In het algemeen zijn de onderzoeken niet zodanig opgezet, dat er een vantevoren opgesteld theoretisch model wordt getoetst. Een groot aantal correlaties wordt tegelijkertijd op significantie getoetst, waardoor enkele correlaties toevalligerwijs significant zullen zijn. Hier doet zich dan een interpretatieprobleem voor. Het is niet duidelijk aan welke verbanden betekenis gehecht mag worden. In het algemeen wordt dit probleem van kanskapitalisatie door onderzoekers wel onderkend; het leidt er echter vaak niet toe dat men hieruit consequenties trekt voor het kiezen van het significantieniveau en het interpreteren van de resultaten. Een oplossing voor de problemen met betrekking tot kanskapitalisatie zou gevonden kunnen worden in het regelmatig uitvoeren van replicatie-onderzoeken. Als in verschillende onderzoeken eenzelfde patroon van correlaties gevonden wordt, houdt dat een bevestiging in van de eerstverkregen gegevens.

PROBLEEMSTELLING

In het onderzoek van Stoel zijn een aantal kenmerken gevonden die op LBO- en MAVO-scholen een (positief of negatief) verband vertoonden met variabelen die indicaties vormen voor gemiddelde schoolprestaties en welbevinden van leerlingen. De gevonden correlaties zijn over het algemeen echter laag en het is waarschijnlijk dat een aantal van de gevonden verbanden op toeval berust. Het doel van dit replicatie-onderzoek is uitsluitend te geven over de waarde van de door Stoel gevonden verbanden, om daarmee een bijdrage te leveren aan het opsporen van relevante schoolkenmerken.

De centrale probleemstelling van het hoofdonderzoek was (Stoel, 1986):

'Welke verschillen in schoolkenmerken hangen samen met verschillen in schooloutput?'

De probleemstelling van het replicatie-onderzoek luidt als volgt:

'In welke mate worden de in het hoofdonderzoek gevonden verbanden tussen schoolkenmerken en outputvariabelen teruggevonden in het replicatie-onderzoek?'

Als het replicatie-onderzoek andere verbanden te zien geeft dan het hoofdonderzoek kunnen we veronderstellen, dat de door Stoel gevonden verbanden deels op toeval berusten. Als in het replicatie-onderzoek duidelijke overeenkomsten gevonden worden met de resultaten van Stoel, is dat een bevestiging van die bevindingen. De gevonden verbanden kunnen dan als betekenisvol beschouwd worden.

METHODE

Replicatie-onderzoek

Onder het repliceren van onderzoek wordt verstaan het herhalen van dit onderzoek met het doel vast te stellen in hoeverre de resultaten van het replicatie-onderzoek overeenkomen met de resultaten van eerder uitgevoerde onderzoeken (Van de Ende en Verhoef, 1973). Op het nut van replicatie-onderzoek is o.a. gewezen door de Groot (1975): 'In de praktijk is geen enkel confirmatieargument sterker dan dit: dat een bepaald voorspeld verband telkens opnieuw, zonder uitzondering, werd gevonden'. In de sociale wetenschappen wordt weliswaar het belang van replicatie-onderzoek onderkend, maar het wordt veel minder vaak uitgevoerd dan in de exacte wetenschappen. De uitvoering van replicatie-onderzoek kan op verschillende manieren gebeuren. Een onderzoek kan zo getrouw mogelijk gerepliceerd worden, waarbij het onderzoek wordt overgedaan met een andere steekproef, maar verder precies zoals het eerder uitgevoerde onderzoek. Men kan ook kiezen voor een steekproef uit een andere populatie, een andere onderzoeksmethode of andere analysetechnieken om de reikwijdte van de eerder gevonden onderzoeksresultaten na te gaan (Van de Ende & Verhoef, 1973).

De replicatie van het onderzoek van Stoel is een zo getrouw mogelijke afspiegeling van het hoofdonderzoek. In het hoofdonderzoek zijn correlaties berekend tussen alle predictor- en criteriumvariabelen. De correlaties die in deze analyse significant bleken te zijn op 1%-niveau werden genoteerd en als betekenisvol beschouwd. Op basis hiervan werden de conclusies geformuleerd.

Er wordt echter een zeer groot aantal onderling afhankelijke correlaties tegelijk berekend. Per correlatie ($H_0: \rho=0$) is een significantie-niveau van 1% gekozen. Naar aanleiding van Van de Ende en Verhoef (1973) kunnen we berekenen dat wanneer zo 10 nulhypothese tegelijk getoetst worden, de kans dat tenminste één nulhypothese ten onrechte wordt verworpen gelijk is aan 10% in plaats van 1%. De kans op het maken van één fout van de eerste soort is dus veel groter dan het gekozen significantieniveau. In zijn algemeenheid kan de verhouding als volgt worden geschetst. Als n hypothesen worden getoetst, waarin r van de n significant blijken te zijn, met name wanneer r klein is ten opzichte van n , is er sprake van een misleidende situatie. Als alle n nulhypothese juist zijn en op α niveau getoetst wordt is het meest waarschijnlijke resultaat dat er $\alpha \times n$ 'significante' toetsingsresultaten zijn. Pas als r groter is dan $\alpha \times n$ kan men waarde hechten aan de gevonden toetsingsresultaten (van de Ende en Verhoef, 1973).

Nu worden in het onderzoek van Stoel wel meer dan $\alpha \times n$ significante resultaten gevonden; problematisch blijft echter welke van deze resultaten betekenisvol zijn en welke toevallig significant. Hierbij moet bovendien nog opgemerkt worden dat bovenstaande beschouwingen gelden voor het toetsen van een serie onafhankelijke hypothesen. Zowel de criterium- als de predictorvariabelen zijn in het onderhavige onderzoek onderling afhankelijk, zodat ook de hypothesen waarmee het bestaan van relaties tussen deze variabelen getoetst wordt, afhankelijk van elkaar zijn. Het gevaar voor het vinden van toevallig significante resultaten wordt hierdoor nog groter. Een vuistregel voor het toetsen van meerdere afhankelijke nulhypothese is het kiezen van een significantieniveau van $\alpha/2m$, waarbij m het aantal hypothesen is (Morrison, 1976). Als deze conservatieve regel gehanteerd zou worden, zouden we in het onderzoek van Stoel nauwelijks significante correlaties overhouden.

Analyse

Als eerste stap wordt de toets van Bartlett uitgevoerd. Met deze toets wordt nagegaan of een correlatiematrix afwijkt van de eenheidsmatrix. Als deze veronderstelling bevestigd

wordt betekent dat, dat de gevonden correlaties toevallig zijn. Als de correlatiematrix niet gelijk is aan de eenheidsmatrix wil dat zeggen, dat minstens één van de gevonden correlaties niet toevallig is. Dit hoeft nog geen reden tot optimisme te zijn, omdat de hoge correlaties niet gelokaliseerd kunnen worden. De van 0 afwijkende correlaties kunnen predictor-criteriumcorrelaties zijn, maar ook correlaties tussen predictoren of criteriumvariabelen onderling.

Om na te gaan in hoeverre de resultaten die gevonden zijn in hoofd- en replicatie-onderzoek overeenstemmen worden de beide correlatiematrices met elkaar vergeleken. Hiervoor worden slechts die variabelen in de analyses opgenomen, die in het hoofdonderzoek verbanden lieten zien (bij $\alpha=.01$). De gebruikelijke procedure om de overeenstemming tussen twee correlatiematrices te berekenen is het uitvoeren van de Box-M-toets (Morrison, 1976) waarbij de hypothese getoetst wordt dat twee covariantiematrices (correlatiematrices zijn gestandaardiseerde covariantiematrices) gelijk zijn. Als deze toets een significante waarde oplevert is dat aanleiding om te veronderstellen dat de matrices en daarmee de gevonden verbanden in de twee steekproeven verschillen. Als de nulhypothese geaccepteerd wordt (dat wil zeggen de matrices verschillen niet) is dit een ondersteuning voor de empirische waarde van de in het hoofdonderzoek gevonden verbanden.

Wanneer de Box-M-toets uitwijst dat de correlatiematrices in de beide steekproeven verschillen, resteert de mogelijkheid een inventarisatie te maken van de correlaties die wel in beide onderzoeken worden gevonden. Hierbij is geen sprake meer van toetsing, maar van vergelijken 'op het oog'.

De steekproeven

Voor de uitvoering van het hoofdonderzoek (Stoel, 1986) werd een vragenlijst verstuurd naar een groot aantal LBO- en MAVO-scholen. In totaal werkten 255 van deze scholen mee aan dit onderzoek. De analyse van deze gegevens resulteerde in een RION-onderzoeksrapport (Stoel, 1986). Een vergelijkbare vragenlijst is ingevuld door 145 scholen voor voortgezet onderwijs, aselekt gekozen uit de populatie waaruit de scholen uit de eerste steekproef verwijderd zijn. De analyse van de gegevens betreffende LBO- (N=67) en MAVO-scholen (N=43) uit dit databestand ('replicatie-onderzoek') dient als replicatie van de analyses op de eerste steekproef van LBO- en MAVO-scholen.

Variabelen

Zowel in het hoofdonderzoek als in het replicatie-onderzoek zijn alle variabelen op schoolniveau gemeten. Als criteriumvariabelen zijn de volgende variabelen gemeten:

- het percentage leerlingen dat de school voortijdig heeft verlaten in het voorgaande schooljaar (DROPOU);
- het percentage geslaagden in het voorafgaande schooljaar (GESLAA);
- het percentage herexamens in het voorafgaande schooljaar (HEREXA);
- het percentage leerlingen dat is blijven zitten in de klassen 1 en 2 (ZITB1, ZITB2).

De meting van de genoemde indicaties voor leerprestaties en welbevinden is bij het hoofdonderzoek en het replicatie-onderzoek op gelijkwaardige wijze gebeurd; de betreffende vragen zijn in beide vragenlijsten gelijk. Een t-toets op de gemiddelden van de outputvariabelen tussen hoofd- en replicatie-onderzoek (uitgevoerd per schooltype) levert geen significante verschillen op. Er is dus aanleiding te veronderstellen dat het hier steekproeven betreft, die uit dezelfde populatie getrokken zijn.

De predictorvariabelen, de schoolkenmerken, zijn onderscheiden in drie groepen, geordend van moeilijk naar eenvoudig te beïnvloeden. De eerste groep bevat contextvariabelen als de grootte van de school, het bouwjaar van de school en de leerkracht-leerlingratio. De tweede groep bevat stafopbouwvariabelen zoals het aantal parttime

leerkrachten en het aantal jonge leerkrachten. De derde groep wordt gevormd door de vormgevings- of schoolorganisatievariabelen.

Onder deze variabelen van de derde groep vallen de meeste predictorvariabelen. Het betreft hier onder andere variabelen als: in hoeverre wordt gebruik gemaakt van individualiserende didactische werkvormen, hoeveel tijd wordt door docenten besteed aan vergaderingen, wordt lesmateriaal kant en klaar gekocht of door docenten gemaakt, komt er regelmatig een schoolkrant uit, is er een regeling voor de aanpak van verzuim? Deze predictorvariabelen zijn in het algemeen op dezelfde wijze gemeten in hoofd- en replicatie-onderzoek. Ook de schaalconstructie heeft zover mogelijk op dezelfde wijze plaatsgevonden. Bij sommige variabelen doen zich echter kleine verschillen in vraagstelling voor. Naar onze mening zijn deze verschillen nergens zodanig groot dat verondersteld kan worden dat niet hetzelfde construct gemeten wordt.

Voor een volledige beschrijving van de variabelen wordt verwezen naar Stoel (1986).

RESULTATEN

De correlatiematrices van hoofd- en replicatie-onderzoek

Of de correlatiematrices verschillen van de eenheidsmatrix is getoetst met de Bartlett-toets. De toetsingsgrootte is chi-kwadraat verdeeld en ziet er in formulevorm als volgt uit (Morrison, 1976):

$$\chi^2 = -(N-1-(2p+5)/6) \log |R| \text{ met } df = \frac{1}{2} p(p-1)$$

waarbij: N = steekproefgrootte

p = het aantal variabelen opgenomen in de matrix

|R| = de determinant van de steekproef correlatiematrix

Tabel 1. Resultaten van de Bartlett-toets op de correlatiematrices uit hoofd- en replicatie-onderzoek.

		N	χ^2
HOOFDOZ	LBO	123	1357.59*
	MAVO	136	1236.77*
REPLIOZ	LBO	65	770.16*
	MAVO	43	683.16*

* significante waarde bij $\alpha = .05$ en $df = 406$

Uit Tabel 1 valt af te lezen dat voor alle vier correlatiematrices geldt dat ze verschillen van de eenheidsmatrix.

Er geldt dus voor alle matrices dat zij minstens één correlatie bevatten die niet toevallig is. We kunnen hier echter niet uit afleiden of de betekenisvolle correlaties criterium-predictorcorrelaties zijn of dat de criterium- en predictorvariabelen onderling hoog correleren.

Toetsing vergelijking hoofd- en replicatie-onderzoek

Voor de vergelijkende analyse zijn per schooltype de volledige correlatiematrices van hoofd- en replicatie-onderzoek vergeleken. Hiervoor zijn de ruwe scores voor alle variabelen per steekproef (MAVO en LBO afzonderlijk) getransformeerd naar Z-scores. Missende waarden hebben de gemiddelde Z-score 0 gekregen.

Per schooltype zijn de correlatiematrices gebaseerd op dezelfde variabelen. De hypothese, dat de correlatiematrices in hoofd- en replicatie-onderzoek aan elkaar gelijk zijn, is per schooltype getoetst met de Box-M-toets (Morrison, 1976). De toetsingsgrootheid is bij benadering chi-kwadraat verdeeld. De resultaten van de Box-M-toets staan vermeld in Tabel 2.

Tabel 2. Resultaten van de vergelijking van de correlatiematrices uit hoofd- en replicatie-onderzoek op MAVO- en LBO-scholen met de Box-M-toets (Morrison, 1976).

	chi-kwadraat	df	p
LBO	812.37	435	.00
MAVO	892.87	435	.00

Voor beide schooltypen levert deze toets een significant resultaat op, wat betekent dat voor zowel LBO- als MAVO-scholen geldt dat de correlatiematrix in het hoofdonderzoek meer verschilt van die in het replicatie-onderzoek dan op grond van toeval verwacht zou mogen worden. Om na te gaan of het verschil tussen de correlatiematrices wellicht veroorzaakt wordt door verschillen in varianties van de predictorvariabelen, is de Box-toets ook uitgevoerd op de covariantiematrices uit hoofd- en replicatie-onderzoek. Ook nu leverde de toets een significant resultaat op.

Beschrijving vergelijking hoofd- en replicatie-onderzoek

Uit de overall-vergelijking van de correlaties in beide onderzoeken blijkt, dat in beide onderzoeken een aantal correlaties toevalligerwijs zijn gevonden. Er bestaat hierdoor verwarring over de situatie in de empirie. Een visuele inspectie van de overeenkomsten en verschillen tussen de correlaties in beide onderzoeken kan aanwijzingen geven over de werkelijke situatie. In totaal zijn voor de analyses voor zowel LBO- als MAVO-scholen 29

Tabel 3. Overzicht overeenstemmende positieve en negatieve correlaties in hoofd- en replicatie-steekproef (α hoofdonderzoek = .01, α replicatie-onderzoek MAVO = .10, α replicatie-onderzoek LBO = .05).

	%GESLAA	%HEREXA	%ZITB1	%ZITB2	%DROPOU
LBO					
aantal leerlingen		+			
urbanisatiegraad				—	
% lk > 20 lesuren		+		+	
% lk < 15 lesuren				—	—
indiv.did. hulp					+
vergadertijd directie				+	
freq. overleg leraren-raad			+		
indiv.did. werkvormen			+		
huiswerkdruk					—
MAVO					
indiv.did. hulp			—	—	
lesmateriaal kant en klaar					—
externe contacten				—	

variabelen (schoolkenmerken en outputvariabelen samen) meegenomen. De schoolkenmerkvariabelen geven in het hoofdonderzoek allemaal een verband te zien met één of meer outputvariabelen. In het replicatie-onderzoek worden een aantal overeenstemmende positieve of negatieve correlaties gevonden, en veel correlaties die anders of zelfs tegenovergesteld zijn aan die in het hoofdonderzoek. Een overzicht van de overeenstemmende positieve en negatieve correlaties van hoofd- en replicatie-onderzoek is vermeld in Tabel 3.

Op MAVO-scholen worden slechts enkele verbanden teruggevonden in het replicatie-onderzoek. De aanwezigheid van individuele niet-vakspecifieke didactische hulp en de frequentie externe contacten hangen negatief samen met het percentage zittenblijvers. Gebruik maken van kant en klaar lesmateriaal hangt negatief samen met het percentage voortijdig schoolverlaters.

Op het LBO wordt een groter aantal verbanden teruggevonden. Positief met leerprestaties hangen in beide onderzoeken samen de urbanisatiegraad van de plaats waarin de school gevestigd is en het aantal parttime leerkrachten. Negatief met leerprestaties hangen samen: de vergaderfrequentie van leerkrachten en directie, individualiserende didactische werkvormen, het aantal leerkrachten dat meer dan 20 uur per week les geeft en het leerlingaantal van de school. Het geven van huiswerk en het aantal parttime leerkrachten hangen negatief, het aanwezig zijn van didactische hulp positief samen met het percentage drop-out.

DISCUSSIE EN CONCLUSIES

Uit de beschrijving van de resultaten van de correlationele analyses blijkt dat zowel voor LBO- als voor MAVO-scholen in beide steekproeven verbanden tussen de schoolkenmerken en leerlingoutputvariabelen gevonden worden. De gevonden verbanden zijn echter in het hoofdonderzoek voor een groot deel andere dan in het replicatie-onderzoek.

Opvallend is dat op MAVO-scholen minder correlaties teruggevonden worden dan op LBO-scholen. De kleine omvang van de steekproef ($N=34$) lijkt hiervoor niet de oorzaak, daar relatief veel significante correlaties worden gevonden (op een aangepast significantieniveau van .10); de verbanden zijn in het replicatie-onderzoek alleen anders dan in het hoofdonderzoek.

Een consistent beeld van verbanden tussen schoolfactoren en leerlingfunctioneren heeft deze studie dan ook niet opgeleverd. Er is een aantal significante correlaties dat in het ene onderzoek wel, en in het andere onderzoek niet wordt gevonden. Wat betreft deze bevindingen bestaat er twijfel over de situatie in de empirie. Variabelen die in beide onderzoeken geen verband met outputvariabelen vertoonden hangen in werkelijkheid waarschijnlijk ook niet samen met de outputvariabelen. Wanneer eenzelfde verband twee keer teruggevonden is in de verschillende steekproeven, is dat een extra aanwijzing voor wat in de populatie de werkelijke verbanden kunnen zijn. Er bestaat echter veel onzekerheid over de interpretatie van deze correlaties. Het is bijvoorbeeld moeilijk de causaliteit van de verbanden te duiden; is het relatief hoge percentage zittenblijvers bijvoorbeeld oorzaak of gevolg van het regelmatig vergaderen, of ligt er een andere verklarende variabele aan dit verband ten grondslag?

Dat de beide onderzoeken relatief weinig consistente verbanden opgeleverd hebben, kan een aantal oorzaken hebben. Het is bijvoorbeeld mogelijk dat er maar erg weinig verbanden bestaan tussen schoolkenmerken en leerlingfunctioneren. Het lijkt ons echter onjuist om met deze verklaring genoegen te nemen aangezien in recente, geavanceerde onderzoeken steeds verbanden worden gevonden tussen schoolkenmerken en het functioneren van leerlingen (Mortimore e.a., 1988; Brandsma & Knuver, 1988).

Er is een aantal problematische punten aan te wijzen in zowel het hoofdonderzoek als in het replicatie-onderzoek. Ten eerste is het moeilijk de begrippen op een juiste wijze te operationaliseren. Het veronderstellen van verbanden tussen schoolkenmerken en het functioneren van leerlingen lijkt plausibel. Het blijkt echter niet eenvoudig deze veronderstelling te toetsen aan de empirie (vgl. theoretische achtergrond). Ook de operationalisering in de hier beschreven onderzoeken is geen geslaagde meting van de concepten 'schoolkenmerken' en 'leerlingfunctioneren'. De vragen zijn vaak multi interpretabel, en bij het meten op schoolniveau wordt een grote afstand ingebouwd tot het dagelijks schoolgebeuren. De gegevens worden op deze manier aangevuld met een hoeveelheid ruis, waardoor zij in betekenis afnemen.

Een beperking van het replicatie-onderzoek ligt in de onderzoeksmethoden. Om de resultaten uit beide onderzoeken te vergelijken is gebruik gemaakt van 'overall' toetsen en beschrijvingen. Hierbij worden de complete correlatiematrices uit beide onderzoeken vergeleken, waardoor slechts de conclusie getrokken kan worden: de onderzoeken leveren niet dezelfde correlaties op. Een vergelijking van de correlaties tussen schoolkenmerken en outputvariabelen in beide onderzoeken is alleen mogelijk via een beschrijving; een andere procedure hiervoor is ons niet bekend.

Samenvattend kunnen we concluderen dat het hier besproken onderzoek twijfels doet rijzen over de waarde van de gevonden correlaties. Mogelijke oorzaken van dit resultaat kunnen liggen in operationalisatie van concepten en het gebrek aan geschikte methoden voor het vergelijken van onderzoeken.

Een verdienste van deze replicatie is dat eens te meer het belang van replicatie-onderzoek wordt benadrukt. De gevolgen van kanskapitalisatie worden bovendien geïllustreerd. Het is aan te bevelen bij correlatieve studies daadwerkelijk rekening te houden met kanskapitalisatie. Dit kan door een zeer laag significantieniveau te kiezen bij de toetsing, maar ook door een replicatie standaard in te bouwen in de onderzoeksopzet.

LITERATUUR

- Blom, M.J.T., Brandsma, H.P. & Stoel, W.G.R. (1986). *Effectieve scholen in het basisonderwijs*. Een voorstudie. Basisschoolkenmerken als determinanten van het functioneren van leerlingen. Groningen: RION.
- Brandsma, H.P. & Stoel, W.G.R. (1987). Schoolleiders in het voortgezet onderwijs. In J. Scheerens en W.G.R. Stoel (red.). *Effectiviteit van onderwijsorganisaties* (pp. 99-114). Lisse: Swets & Zeitlinger.
- Brandsma, H.P. & Knuver, J.W.M. (1988). Organisatorische verschillen tussen basisscholen en hun effect op leerlingprestaties. *Tijdschrift voor Onderwijsresearch*, 13, 4, 201-212.
- 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.
- Ende, H. van de & Verhoef, M. (1973). *Inductieve statistiek voor de gedragswetenschappen*. Amsterdam: Elsevier.
- Grift, W. van de (1987). Zelfpercepties van onderwijskundig leiderschap en gemiddelde leerlingprestaties. In J. Scheerens en W.G.R. Stoel (red.). *Effectiviteit van onderwijsorganisaties*. Lisse: Swets & Zeitlinger.
- Groot, A.D. de (1975). *Methodologie*. Grondslagen van onderzoek en denken in de gedragswetenschappen. 's-Gravenhage: Mouton.
- Marwijk Kooy-Von Baumhauer, L. (1984). *Scholen verschillen*. Een verkennend vergelijkend onderzoek naar het intern functioneren van vijftientig scholengemeenschappen VWO-HAVO-MAVO. Groningen: Wolters-Noordhoff.
- Meijnen, G.W. (1984). *Van zes tot twaalf*. Een longitudinaal onderzoek naar de milieu- en schooleffecten van loopbanen in het lager onderwijs. Den Haag: SVO (SVO-reeks no. 79).

- Morrison, D.F. (1976). *Multivariate statistical methods*. (second edition). New York: Mc. Graw Hill.
- Mortimore, P. e.a. (1988). *School Matters. THE JUNIOR YEARS*. Somerset: Open Books.
- Stoel, W.G.R. (1986). *Schoolkenmerken en het gedrag van leerlingen en docenten in het voortgezet onderwijs*. Groningen: RION, interne publicatie.
- Vermeulen, C.J. (1987). De effectiviteit van onderwijs bij 17 Rotterdamse stimuleringscholen. *Pedagogische Studiën*, 64, 49-58.
- Wolf, J.C. van der (1984). *Schooluitval*. Een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het regulier onderwijs. Lisse: Swetz & Zeitlinger.

Manuscript ontvangen 20-7-1988

Definitieve versie ontvangen 6-10-1989

Een onderzoek¹ naar de effecten van auditieve feedback op leesfouten

G.W.G. Spaai², P. Reitsma³ en H.H. Ellermann²

²*Instituut voor Perceptie Onderzoek (IPO), Eindhoven*

³*Paedologisch Instituut (PI), Duivendrecht*

ABSTRACT

An experiment was conducted to investigate the effects of two forms of verbal feedback on oral reading errors made by beginning readers. In one condition the whole correct word sound was provided when a child was not able to read the word correctly within a certain time limit (the whole-word condition). In a second condition the sound of the word was produced phoneme-by-phoneme when an error was made or if the child did not read the word at all (the segmented-feedback condition). Learning effects as measured by differences in reading times and reading errors made in the pretest and posttest, were compared with a control condition in which children received no feedback at all (the control condition). The results of training over three sessions showed that children in the whole word condition learned more than children in the control and segmented-feedback conditions. There was no significant difference between the control condition and the segmented-feedback condition. Possible explanations for the results are given and the results are discussed in terms of their implications for oral reading practice.

INLEIDING

In nagenoeg alle hedendaagse methoden voor het aanvankelijk leesonderwijs wordt de beginnende lezer geleerd letters van links naar rechts te verklanken en deze letterklanken samen te voegen tot de klank van het woord welke kan leiden tot de herkenning van het woord (z.g. fonologische strategie, Bus, 1984). De leesmethoden verschillen echter onder andere in de manier waarop de koppeling tussen lettertekens en letterklanken geleerd moet worden en hoe de lezer geïnstrueerd moet worden indien er fouten gemaakt worden (Kooreman, 1977; Caesar, 1980; zie ook Chall, 1983). In dit artikel wordt dieper ingegaan op het corrigeren van leesfouten.

Op het belang van zelfcorrecties door lezers is veelvuldig gewezen in de literatuur (D'Angelo, 1982; Kusters, 1987). Niet alle leesfouten worden echter door de lezer zelf gecorrigeerd hetgeen correctie van buitenaf noodzakelijk maakt (Perkins, 1988). Verschillende factoren spelen een rol bij het corrigeren van leesfouten van buitenaf, waaronder het *type* leesfout dat gecorrigeerd moet worden, het *tijdstip* waarop de correctie plaatsvindt en

¹ Het hier gerapporteerde onderzoek werd door de eerste auteur als paper gepresenteerd tijdens de Onderwijs Research Dagen 1986 te Groningen. Het onderzoek maakt deel uit van het onderzoek-sproject (LEESBORD-project) dat met financiële ondersteuning van het Ministerie van Onderwijs en Wetenschappen werd uitgevoerd op het Instituut voor Perceptie Onderzoek (IPO) te Eindhoven in samenwerking met het Paedologisch Instituut (PI) te Duivendrecht en de Pedagogisch Technische Hogeschool Nederland te Eindhoven. Met dank aan prof. dr. D. G. Bouwhuis voor diens commentaar op een eerdere versie van dit artikel.

de wijze waarop de lezer gecorrigeerd moet worden (Hoffman, O'Neal, Kastler, Clements, Segel & Nash, 1984). De literatuur overziend concluderen Hoffman et al. (1984) dat het corrigeren van leesfouten met enige terughoudendheid en tolerantie dient te geschieden zodat de lezer zelf ook de mogelijkheid geboden wordt leesfouten te corrigeren en hij tot een actieve leeshouding wordt aangezet. Over de manier waarop de lezer gecorrigeerd moet worden indien deze niet tot zelfcorrectie in staat is, bestaat echter veel onduidelijkheid.

In het algemeen wordt onderscheid gemaakt tussen twee verschillende procedures voor het corrigeren van leesfouten namelijk de *woord-analyse methode* en de *woord-totaal methode* (zie o.a. Jenkins & Larson, 1979). Indien een leesfout gemaakt wordt, wordt bij de woord-analyse methode de aandacht van de lezer gericht op de structuur van het woord door verklanking van de afzonderlijke woordsegmenten. De leerling moet deze vervolgens aaneen rijgen tot de klank van het woord. Wanneer bij de tweede correctieprocedure, de *woord-totaal methode*, een leesfout gemaakt wordt, wordt de gehele correcte woordklank verstrekt.

In een onderzoek van Meyer (1982) werden twee correctie-procedures onderzocht. Wanneer de lezer bij de woord-analyse methode een leesfout maakte werd hij ondersteund bij het verklanken van de woordsamenstellende lettertekens en het samenvoegen van deze letterklanken tot de gehele woordklank. Bij de woord-totaal methode werd de lezer bij het maken van een leesfout de gehele correcte woordklank aangeboden. In dit onderzoek participeerden kinderen met leerstoornissen in de leeftijd van 7 tot 14 jaar. Meyer (1982) vond geen verschil in leereffect tussen de beide correctie-procedures: het simpelweg verstrekken van de woordklank indien een leesfout werd gemaakt bleek net zo effectief te zijn als het volledig doorlopen van de decodeerstrategie.

In een onderzoek van Bouwhuis en Truin (1986) werd een conditie waarin de gehele woordklank werd aangeboden indien een leesfout werd gemaakt, vergeleken met een conditie waarin de klank van het woord lettergreep voor lettergreep werd aangeboden. Opnieuw werden geen verschillen gevonden tussen diverse correctie-procedures. Opmerkelijk was dat er in condities waarin de leerling gecorrigeerd werd niet meer geleerd werd dan in een conditie waarin dit niet gebeurde. Mogelijk is dit laatste toe te schrijven aan het feit dat de proefpersonen die aan dit onderzoek hebben deelgenomen ernstig leesgestoorde kinderen waren die veelal geen of slechts weinig vooruitgang boekten tijdens het experiment.

In tegenstelling tot de bovengenoemde studies waarin de taak bestond uit het hardop lezen van losse woorden, werden in het onderzoek van Olson, Foltz en Wise (1986) teksten (stil) gelezen. De lezer die problemen ondervond bij het op eigen kracht identificeren van een woord had de mogelijkheid de gesproken vorm hiervan op te vragen. De teksten werden aangeboden op het beeldscherm van een microcomputer en de gesproken vorm van ieder woord kon geheel kunstmatig door de computer gegenereerd worden. Het onderzoek vond plaats met het oog op de ontwikkeling van spraakondersteunde, gecomputeriseerde leesoefeningen voor beginnende en zwakke lezers. Afhankelijk van de oefenconditie werd de woordklank in zijn geheel, lettergreep voor lettergreep of lettertekens voor lettertekens uitgesproken. De onderzoeksresultaten lieten geen verschillen in voorde- ringen tussen de drie verschillende correctieprocedures zien. Soortgelijke resultaten werden gevonden in ander onderzoek (Haupt, Cohen, Magee, Axelrod & Price, 1977 en meer recentelijk Perkins, 1988).

In de hierboven beschreven onderzoeken werden geen verschillen gevonden tussen de diverse correctie-procedures. In deze onderzoeken participeerden altijd min of meer ernstig leesgestoorde kinderen. Het is mogelijk dat deze kinderen niet in ruimere mate profiteren van de woord-analyse strategie dan van een woord-totaal strategie omdat ze

moeite hebben met het auditief analyseren en synthetiseren (Kochnowar, Richardson & DiBenedetto, 1983).

Onbekend is echter in hoeverre normale lezers, in tegenstelling tot kinderen met leesproblemen, wél meer leren van een woord-analyse strategie dan van een woord-totaal strategie. In een onderzoek van Spaai, Reitsma en Ellermann (1986) werden de leereffecten van twee verschillende correctieprocedures bij het leren lezen van losse woorden onderzocht. Bij de eerste correctieprocedure werd de gehele woordklank aangeboden indien de lezer niet in staat was het woord correct te identificeren terwijl bij de tweede correctieprocedure het woord werd voorgesped. De leereffecten van beide correctieprocedures werden vergeleken met een controle conditie waarin de leerlingen in het geheel geen hulp ontvingen. Bovendien werd nagegaan in hoeverre er een relatie bestaat tussen de aard van de leesfout en de wijze waarop de leesfout gecorrigeerd werd. De resultaten wezen uit dat in condities waarin de gehele woordklank werd aangeboden dan wel het woord werd voorgesped indien een leesfout gemaakt werd, meer geleerd werd dan wanneer in het geheel geen hulp werd verstrekt. Er werden echter geen betrouwbare verschillen gevonden tussen de beide correctieprocedures noch werd een relatie gevonden tussen de aard van de correctieprocedure en de aard van de leesfout. In het onderzoek van Spaai et al. (1986) participeerden lezers met ongeveer 8 maanden leeservaring. Mogelijk dat lezers met minder leeservaring meer gebaat zijn bij een woord-analyse strategie aangezien deze strategie mogelijk beter aansluit bij de gehanteerde woordidentificatiestrategie. Verder onderzoek naar de effecten van de verschillende correctieprocedures bij normale lezers is derhalve noodzakelijk. De informatie die zodoende verkregen wordt kan van groot belang zijn in verband met de manier waarop de leerkracht in het regulier (lees)onderwijs leesfouten dient te corrigeren en informatie over leesfouten in gecomputeriseerde lees oefeningen moet worden verstrekt zodat het leereffect zo groot mogelijk is (Reitsma, Ellermann & Spaai, 1987; Spaai, Reitsma & Ellermann, 1987).

Met normale, beginnende lezers met ongeveer vier maanden leeservaring is een experiment uitgevoerd waarin de leereffecten van twee verschillende correctie-strategieën werden onderzocht. Bij de eerste correctie-procedure werden, wanneer de leerling een leesfout maakte of het woord niet binnen de gestelde tijdlimiet las, de lettertekens successievelijk verklankt. Bij de tweede correctie-procedure werd, wanneer de lezer niet in staat was het woord correct te identificeren, de gehele correcte woordklank gepresenteerd. In bovengenoemde studies waarin ofwel leergestoorde kinderen ofwel relatief ervaren, normale lezers participeerden, werden geen verschillen in leereffect tussen de woord-totaal strategie en de woord-analyse strategie gevonden. De verwachting was nochtans dat normale beginnende lezers meer zouden leren zodra het woord werd voorgesped dan wanneer de gehele woordklank werd aangeboden indien de beginnende lezers niet in staat zouden zijn het woord correct te identificeren. Deze hypothese is gebaseerd op het idee dat de leerling het basale leesproces (n.l. het identificeren van woorddelen, het verklanken van deze woorddelen en samenvoegen van deze deelklanken tot de woordklank) niet doorloopt indien de totale woordklank gepresenteerd wordt. Als de gehele woordklank gepresenteerd wordt besteedt de lezer vermoedelijk geen aandacht aan de afzonderlijke grafeem-foneem samenhangen en is het zelf construeren van de woordklank niet noodzakelijk. Dientengevolge is het mogelijk dat de lezer minder kennis omtrent het systematisch verband tussen de sequentie van lettertekens en de fonologische kenmerken van dat woord verwerft (Reitsma, 1985). Deze kennis is van belang om te kunnen komen tot een directe herkenning van het woord (Ehri, 1980).

ONDERZOEKSOPZET

Onderzoeksgroep

Aan het experiment hebben 66 leerlingen deelgenomen (32 jongens en 34 meisjes) afkomstig van drie reguliere basisscholen te Eindhoven. Alle leerlingen zaten in groep drie van het basisonderwijs en hadden vier maanden leesonderwijs gehad op het moment dat het experiment van start ging (december). Er hebben geen zittenblijvers aan het experiment deelgenomen. Alle scholen hanteerden dezelfde leesmethode (Caesar, 1980). De gemiddelde leeftijd van de leerlingen bedroeg 6.5 jaar. De gemiddelde score van de totale groep leerlingen op de Caesar-1-minuut test (Mommers, 1983) bedroeg 18.3 (sd= 8.7). Deze leerlingen konden dus als normale lezers beschouwd worden (de normscore bedraagt ongeveer 19, Van Dongen & Mommers, 1980).

Stimulusmateriaal

Het stimulusmateriaal bestond uit drieëndertig doelwoorden die volgens Kohnstamm, Schaerlaekens, Vries, Akkershuis en Frooninckx (1981) tot de passieve woordschat van de leerlingen behoren. De woorden waren nog niet in het reguliere leesonderwijs aan de orde geweest d.w.z. de woorden waren nog niet in geschreven vorm aan de orde geweest en dus nieuwe, onbekende woorden. De woorden waren bovendien zodanig gekozen dat ze voor de gemiddelde leerling die vier maanden leesonderwijs had genoten, relatief moeilijk leesbaar waren. Het merendeel van de woorden bevat vijf of zes fonemen (23 woorden in totaal) terwijl een gering aantal woorden slechts zeven (vijf woorden) of acht (vijf woorden) fonemen telt. Achtereenvolgens de volgende doelwoorden werden gebruikt: blind, brand, kwast, slurf, staart, vriend, speld, straks, herfst, strand, sprong, tikken, zagen, moeilijk, lelijk, wormpje, boodschap, gezond, altijd, prettig, zeepsop, bedtijd, hoofdpijn, veertje, armband, kapstok, badmuts, broekspijp, gepraat, poppenhuis, spelletje, onthouden en vergissen. Drie woorden die wat moeilijkheidsgraad betreft met de doelwoorden vergelijkbaar waren werden ter introductie van het experiment aangeboden (b.v. plant) om zodoende de leerlingen te laten wennen aan de experimentele condities. Terwille van de motivatie van de leerlingen werden regelmatig (vijf woorden in totaal) gemakkelijk leesbare woorden aangeboden (b.v. boom, roos en vis). In totaal werden dus 41 woorden aangeboden.

Procedure

Op basis van de scores op de Caesar-1-minuut test werd een leesvaardigheidsrangorde van de 66 leerlingen opgesteld. Vervolgens werden de leerlingen uit een groep van drie opeenvolgende rangnummers op basis van toeval aan een van de drie condities toegewezen. Doordat sommige leerlingen (5 in totaal) niet in de gelegenheid waren de hele experimentele procedure te doorlopen (door ziekte e.d.) was het aantal leerlingen in de drie verschillende condities niet gelijk, namelijk 22 (woord-conditie), 19 (letterklank-conditie) en 20 (controle-conditie). De gemiddelde scores op de Caesar-1-minuut test in de woord-conditie, letterklank-conditie en de controle-conditie bedroegen respectievelijk: 17.7 (sd=8), 19.1 (sd=10.6) en 18.2 (sd=8). Er bestonden geen significante verschillen tussen deze scores ($F < 1$).

Het experiment werd uitgevoerd in een aparte, rustige ruimte in de school terwijl tijdens de uitvoering van het experiment voortdurend twee proefleiders aanwezig waren. De taakstelling was in alle condities gelijk: een woord dat op het beeldscherm van een computer werd aangeboden, moest hardop worden gelezen. Tevens werd de leerling afhankelijk van de conditie waaraan deze was toegewezen medegedeeld of en in welke vorm hulp zou worden verleend bij het lezen van de woorden.

In de eerste conditie (controle-conditie) verkreeg de leerling in het geheel geen feedback over het gelezene. Nadat het woord (correct of incorrect) gelezen was, werd door de proefleider een toets van het toetsenbord ingedrukt waarna het woord 1.5 seconde later van het scherm verdween. Indien de leerling in het geheel niet in staat was het woord te verklanken, verdween het woord na het verstrijken van de maximale leestijd (10 sec) automatisch van het scherm. Twee seconden nadat het woord van het scherm was verdwenen werd een nieuw woord op het beeldscherm van de computer gepresenteerd.

De tweede conditie (woord-conditie) was gelijk aan de controle conditie en verschilde alleen door het feit dat de gehele woordklank werd aangeboden zodra bleek dat de leerling niet in staat was het woord correct te verklanken.

In de derde conditie (letterklank-conditie) werd niet de gehele correcte woordklank aangeboden indien een fout gemaakt werd maar werd het woord voorgespeeld. Het tempo waarin de letterklanken werden aangeboden bedroeg ongeveer een letterklank per 750 msec. De vocalen en fricatieve consonanten (b.v. /f/, /v/, /s/ en /z/) werden geïsoleerd aangeboden, terwijl de stopconsonanten (b.v. /d/, /p/, /g/, /k/ en /t/) altijd een segment van de klank 'sjwa' bevatten.

De leerlingen hebben individueel aan drie trainingssessies op drie opeenvolgende dagen deelgenomen. Respectievelijk een dag voor de trainingen en een dag na de trainingen werden de woorden in een afzonderlijke test ter lezing aangeboden (z.g. pretest en posttest) waarbij eenzelfde procedure gehanteerd werd als in de controle-conditie en de leerlingen geïnstrueerd werden de op het scherm gepresenteerde woorden zo snel en correct mogelijk hardop te lezen. Tijdens de trainingen en de pretest c.q. posttest werden de woorden in een wisselende, random volgorde aangeboden. Op leesfouten werd door de proefleider niet gereageerd. Alle niet zelf-gecorrigeerde leesfouten werden beschouwd als leesfouten (vervangingen, toevoegingen, weglatingen etc.). Er werd geen onderscheid gemaakt naar de soort leesfout aangezien in eerder onderzoek van Spaai et al. (1986) geen relatie ontdekt werd tussen de soort leesfout en de aard van de correctieprocedure. Haperingen of gedeeltelijke herhalingen werden niet als incorrect beschouwd zolang het gelezen woord maar correct was. Klemtoonfouten werden niet als leesfouten beschouwd (Leu, 1982). Tijdregistratie startte op het moment dat het woord op het beeldscherm van de computer verscheen. Onmiddellijk nadat het woord hardop gelezen was, werd de spatie-balk van het toetsenbord ingedrukt waarmee de leestijd werd geregistreerd. Vervolgens werd de correctheid van het gelezene beoordeeld en geregistreerd via een andere toets van het toetsenbord. Indien een woord niet binnen de maximale leestijd gelezen werd, werd dit automatisch geregistreerd als een omissie. Omissies werden in de analyses beschouwd als leesfouten. Aan de hand van het trainingsverloop en de verschillen in scores bij de pretest en posttest werden de leereffecten van de verschillende condities bepaald.

Apparatuur

De woorden werden één voor één op het beeldscherm van een Apple MacIntosh microcomputer aangeboden. De woorden werden gepresenteerd in een lettergrootte en lettertype die bij de standaard leesmethoden gebruikelijk zijn, terwijl de kijkafstand ongeveer 30 cm bedroeg. De voor de letterklank- en woord-conditie benodigde spraak was door een ervaren mannelijke spreker op een geluidsband ingesproken. Deze spraak was vervolgens gedigitaliseerd met een bemonsteringsfrequentie van 11 kHz en een nauwkeurigheid van 8 bits en werd tenslotte gereproduceerd via de ingebouwde digitaal-analoog omzetter van de computer. De spraak werd door middel van de luidspreker van de computer aangeboden. Een 20 Mb hard-disk werd als geheugenmedium voor de opslag van de benodigde spraak gebruikt.

RESULTATEN

Pre- en posttest

Leesfouten. Nagenoeg 79% van de woorden was foutief gelezen op de pretest, terwijl dat bij 61% van de woorden het geval was op de posttest. Hieruit blijkt dat de woorden voor de leerlingen die aan het onderzoek hebben deelgenomen behoorlijk moeilijk waren (zie Tabel 1). Nagenoeg alle leerlingen maakten echter minder fouten bij de posttest dan bij de pretest hetgeen de veronderstelling dat het geselecteerde stimulusmateriaal niet te moeilijk was voor de lezers, ondersteunt. Sommige lezers (zes in totaal) maakten zelfs slechts een of twee fouten op de pretest. In alle condities maakten de leerlingen minder fouten bij de posttest dan bij de pretest ($F(1,58)=60.89$; $p < 0.01$).

Er is een covariantie-analyse, met als covariaat de score op de Caesar-1-minuut test, op de afname van het aantal fouten uitgevoerd. Deze heeft een significant effect van de covariaat aan het licht gebracht ($F(1,57)=6.21$; $p < 0.05$) in die zin dat naarmate de score op de Caesar-1-minuut test hoger was er meer geleerd werd. De condities verschilden onderling niet wat betreft het verband tussen de afname van het aantal fouten en de score op de Caesar-1-minuut test.

Voorts bleek het effect van conditie op de afname van het aantal leesfouten significant ($F(2,57)=5.35$; $p < 0.01$). Uit een a posteriori uitgevoerde Scheffé contrast analyse bleek dat er een significant verschil was tussen de woord-conditie en de letterklank-conditie ($p < 0.05$). Er werden bij de posttest in de woord-conditie minder fouten gemaakt dan in de letterklank-conditie. Ook bestond er een significant verschil tussen de woord-conditie en de controle-conditie ($p < 0.05$). Het verschil tussen de letterklank-conditie en de controle-conditie was niet significant ($p > 0.05$).

Om na te gaan in hoeverre er een verband bestaat tussen de aard van de correctieprocedure en de woordlengte (d.w.z. het aantal woordsamenstellende fonemen) werd een variantie-analyse uitgevoerd met twee factoren te weten correctieprocedure en woordlengte. Er werd geen significante interactie gevonden tussen de correctieprocedure en de woordlengte ($F < 1$).

Leestijden. De gemiddelde leestijd van de woorden welke bij de pretest en de posttest correct werden gelezen bedroeg respectievelijk 5.5 sec ($sd=2.2$) en 4.2 sec ($sd=2$), zie Tabel 2. In alle condities vereiste het correct lezen van de woorden bij de posttest minder tijd dan tijdens de pretest ($F(1,31)=36.14$; $p < 0.01$). Er is een covariantie-analyse uitgevoerd op de vermindering in leestijd met als covariaat de score op de Caesar-1-minuut test. Het effect van de covariaat was significant in die zin dat naarmate de score op de Caesar-1-minuut test hoger was de vooruitgang groter werd. Verder bleek dat er geen differentieel effect van oefenconditie was op de afname in leestijd ($F < 1$).

Trainingsresultaten

Bij nagenoeg 64% van de woorden werden de leerlingen drie keer gecorrigeerd terwijl slechts bij 13.2% van de woorden in het geheel geen correctie noodzakelijk was.

Tabel 3 laat - uitgesplitst naar experimentele conditie- het gemiddeld aantal leesfouten over de drie trainingen zien. Er is een variantie-analyse op het aantal fouten tijdens de trainingen uitgevoerd met trainingssessie als binnenproefpersoon-factor en de conditie als tussenproefpersoon-factor. Uit deze analyse is gebleken dat de trainingssessie een significant effect heeft gehad ($F(2,116)=35.82$, $p < 0.01$) en dat er een significante interactie tussen conditie en trainingssessie bestaat ($F(4,116)=14.72$; $p < 0.05$). In de woord-conditie werden tijdens de tweede training minder fouten gemaakt dan tijdens de eerste training ($p < 0.05$) terwijl tijdens de derde training minder fouten werden gemaakt

Tabel 1. Gemiddelde* (met de spreiding tussen haakjes) van leesfouten per experimentele conditie in de pretest en posttest.

Conditie	Pretest	Posttest	Pretest-Posttest
letterklank-conditie	27 (9.3)	22.4 (11.8)	4.5 (3.4)
Woord-conditie	26.6 (8.9)	18.5 (9.7)	8.2 (5.3)
Controle-conditie	24.7 (9.8)	21.3 (11.1)	3.4 (3.1)
Totale Gemiddelde	26.1 (8.9)	20.6 (10.8)	5.5 (5.2)

* Het maximum aantal leesfouten bedraagt 33.

Tabel 2. Gemiddelde* (met de spreiding tussen haakjes) leestijden in sec per experimentele conditie in de pretest en posttest.

Conditie	Pretest	Posttest	Pretest-Posttest
Letter-conditie	5.5 (2.2)	4.1 (1.9)	1.4 (2.1)
Woord-conditie	5.5 (2.2)	4.2 (2)	1.4 (2.3)
Controle-conditie	5.5 (2.1)	4.3 (2)	1.2 (2.1)
Totale Gemiddelde	5.5 (2.2)	4.2 (2)	1.3 (2.1)

* De maximale leestijd bedraagt 10 sec.

Tabel 3. Het gemiddeld aantal leesfouten (met de spreiding tussen haakjes) tijdens de drie trainingen uitgesplitst naar experimentele conditie.

Conditie	Training 1	Training 2	Training 3
letterconditie	24.5 (11.2)	23.5 (11.3)	23.1 (11.9)
Woord-conditie	25.5 (8.4)	22.6 (9.7)	18.8 (11.1)
Controle-conditie	27.7 (9.8)	26.1 (10)	26.1 (9.4)
Totale Gemiddelde	25.9 (9.3)	24.1 (10.3)	22.7 (11.1)

* Het maximum aantal leesfouten bedraagt 33.

dan tijdens de tweede training ($p < 0.05$). In de letterklank-conditie en de controle-conditie was alleen het verschil tussen de eerste en de tweede training significant ($p < 0.05$). Uit de a posteriori uitgevoerde Scheffé contrast-analyse bleek verder dat het verschil tussen de controle-conditie en de woord-conditie alleen significant was tijdens de derde trainingssessie ($p < 0.05$). Dit betekent dat er tijdens de derde trainingssessie meer fouten werden gemaakt in de controle-conditie dan in de woord-conditie. Andere verschillen tussen de condities waren niet significant ($p > 0.05$)

DISCUSSIE

Leerlingen maakten minder fouten bij de posttest dan bij de pretest. Bovendien was er een effect van oefenconditie op de afname van het aantal fouten. Dit conditie-effect ondersteunt echter niet de eerder in dit artikel geformuleerde hypothese dat leerlingen in de letterklank-conditie meer zouden leren dan de leerlingen in de woord-conditie. Integendeel, lezers in de woord-conditie maakten minder fouten bij de posttest dan de lezers in de controle-conditie en de letterklank-conditie terwijl het verschil tussen de beide laatstgenoemde condities niet significant was. Een mogelijke verklaring voor dit relatief gunstige leereffect in de woord-conditie is dat de presentatie van de woordklank als het

ware kan leiden tot een 'impliciete heranalyse van het woord door middel van de juiste strategie' (Meyer, 1982).

In vergelijking tot de woord-conditie wordt in de letterklank-conditie een geringer leereffect gevonden. Twee mogelijk verklaringen voor dit geringer leereffect zullen hier besproken worden. Daarna zal de generaliseerbaarheid van deze onderzoeksbevindingen kort worden besproken.

Een eerste mogelijke verklaring voor het relatief geringe leereffect van de woord-analyse strategie is dat de leerlingen niet vertrouwd zijn met de woord-analyse strategie aangezien in het dagelijks leesonderwijs veelal de gehele, correcte woordklank wordt verstrekt onafhankelijk van het tijdstip waarop de leesfout wordt gemaakt, de aard van de leesfout en de vaardigheid van de lezer (Spiegel & Rogers, 1980). Voorts sluit de woord-analyse strategie niet volledig aan bij de gangbare lees oefeningen waarbij zowel auditieve analyse als auditieve synthese plaatsvindt (zogenaamde auditieve structurerings oefeningen, zie Caesar, 1980). Het is mogelijk dat leerlingen derhalve meer moeite ondervinden bij het werken met de woord-analyse methode en derhalve minder gebaat zijn met het gebruik van de woord-analyse procedure en meer gebaat zijn met het gebruik van een combinatie van de woord-analyse methode en de woord-totaal methode waarbij eerst het woord wordt voorgespeld en vervolgens de gehele woordklank wordt aangeboden.

Een andere meer voor de hand liggende verklaring gezien de grote verschillen tussen de letterklank-conditie en de woord-conditie is dat lezers in de letterklank-conditie niet in staat zijn de afzonderlijke letterklanken te combineren tot de totale woordklank aangezien het proces van 'sound blending' wordt belemmerd zodra de afzonderlijke letterklanken met lange tussenpauzes worden gepresenteerd waarmee hoge eisen worden gesteld aan het korte-termijn-geheugen in verband met de opslag van de afzonderlijke lettertekens. Voorts wordt het korte-termijn-geheugen mogelijk extra belast zodra langere woorden worden voorgespeld (Wise, 1987).

De onderzoeksresultaten laten voorts zien dat in alle condities woorden sneller en nauwkeuriger gelezen werden op de posttest dan op de pretest. Dit is in overeenstemming met eerder onderzoek dat aantoonde dat het herhaaldelijk lezen van dezelfde tekst de snelheid en nauwkeurigheid waarmee die tekst gelezen wordt, verhoogt (Carver & Hoffman, 1981; Moyer, 1982; Reitsma, 1988). Uit de resultaten blijkt echter geen significant effect van conditie op de afname in leestijd. Waarschijnlijk wordt dit veroorzaakt door de lage frequentie (6.54% van de woorden) waarmee hulp verstrekt werd bij woorden die zowel op de pretest als op de posttest correct gelezen waren. De woorden die correct gelezen waren op zowel de pretest als de posttest werden waarschijnlijk reeds beheerst door de kinderen.

Voor de ontwikkeling van computerondersteunde lees oefeningen voor beginnende lezers (Olson, Foltz & Wise, 1986; Reitsma, Ellermann & Spaai, 1987), waarbij gebruik wordt gemaakt van kunstmatig gegenereerde spraak, hebben de hier beschreven onderzoeksbevindingen als belangrijkste implicatie dat het successievelijk verklanken van woordsamenstellende letters vooralsnog geen geschikte strategie lijkt voor het corrigeren van leesfouten.

Alhoewel generalisatie van de onderzoeksbevindingen naar de praktijk van het leesonderwijs met enige voorzichtigheid dient te geschieden omdat de leerlingen slechts gedurende een korte periode geïnstrueerd werden en leereffecten onmiddellijk nadien bepaald werden lijkt het erop dat lezers met weinig leeservaring niet of nauwelijks gebaat zijn bij het gebruik van een correctie-procedure waarbij het woord wordt voorgespeld indien een leesfout gemaakt wordt. Aangezien de woord-analyse procedure de lezer echter

mogelijk aanzet tot een actievere en zelfstandige leeshouding is het onduidelijk wat de effecten van een dergelijke methode over een langere oefenperiode zullen zijn (Perkins, 1988) en of er generalisatie zal optreden van de geleerde woorden naar sterk gelijkende woorden ('transfer woorden'). Dit laatste is echter niet waarschijnlijk aangezien in de beginfase van het lezen het visuele lexicon nog zo weinig is ontwikkeld dat spellingregels en generalisatieregels weinig doeltreffend kunnen opereren (Bouwhuis & Truin, 1986; Bouwhuis & Bouma, 1979).

Voortgezet onderzoek waarin de leerlingen gedurende een langere periode geïnstrueerd worden en waarbij lange-termijn effecten gemeten worden is noodzakelijk. Voorts is het zodoende mogelijk om na te gaan in hoeverre de leereffecten van de verschillende correctieprocedures samenhangen met de leesvaardigheid van de leerling.

LITERATUUR

- Bus, A. G. (1984). *Leesproblemen en instructiemethoden*. Lisse: Swets & Zeitlinger.
- Bouwhuis, D. G. & Bouma, H. (1979). Visual recognition of three-letter words as derived from the recognition of the constituent letters. *Perception and Psychophysics*, 25, 12-25.
- Bouwhuis, D. G. & Truin, P. G. M. (1986). Woorden lezen met spraak naar keuze door dyslectische kinderen. In S. Dijkstra & P. Span (Eds), *Leerprocessen en instructie*. Bijdrage O.R.D. 1985, 1, 99-109. Lisse: Swets & Zeitlinger.
- Caesar, F. B. (1980). *Veilig Lerén Lezen*. Tilburg: Zwijzen.
- Carver, R. P. & Hoffman, J. V. (1981). The effect of practice through repeated reading on gain in reading ability using a computer based instructional system. *Reading Research Quarterly*, 16, 374-390.
- Chall, J. S. (1983). *Learning to read: The great debate*. New York: McGraw-Hill.
- D'Angelo, K. (1982). Correction behaviour: Implications for reading instruction. *The Reading Teacher*, 35, 395-398.
- Ehri, L. C. (1980). The role of orthographic images in learning printed words. In J.F. Kavanagh & R.L. Venezky (Eds), *Orthography, reading and dyslexia*. Baltimore: University Park Press.
- Haupt, E. J., Cohen, M., Magee, J. W., Axelrod, S. & Price, M. (1977). Rapid reduction of graphically similar substitution errors in oral reading with a drill procedure. *Journal of Educational Research*, 71, 114-116.
- Hoffman, J. V., O'Neal, S. F., Kastler, L. A., Clements, R. O., Segel, K. W. & Nash, M. F. (1984). Guided oral reading and miscue focused verbal feedback in second-grade classrooms. *Reading Research Quarterly*, 14, 367-384.
- Jenkins, J. R. & Larson, K. (1979). Evaluating error correction procedures for oral reading. *Journal of Special Education*, 13, 145-156.
- Kochnow, J., Richardson, E. & DiBenedetto, B. (1983). A comparison of the phonetic decoding ability of normal and learning disabled children. *Journal of Learning Disabilities*, 16, 348-351.
- Kohnstamm, G. A., Schaerlaekens, A. M., Vries, A. K. de, Akkershuis, G. W., & Frooninckx, M. (1981). *Nieuwe Streefwoordenlijst voor 6-jarigen*. Lisse: Swets & Zeitlinger.
- Kooreman, H. J. (1977). *Letterstad*. Groningen: Wolters Noordhoff.
- Kusters, E. D. M. (1987). *Self-corrections in oral reading some aspects of the reading process of good and poor readers*. Dissertatie. Enschede: Sneldruk.
- Leu, D.J. (1982). Oral reading error analysis: A critical review of research and application. *Reading Research Quarterly*, 17, 420-437.
- Meyer, L. M. (1982). The relative effects of word-analysis and word-supply correction procedures with poor readers during word-attack training. *Reading Research Quarterly*, 17, 544-555.
- Mommers, M. J. C. (1983). *Caesar-Een-Minuut test*. Handleiding en Verantwoording. Tilburg: Zwijzen.
- Moyer, S.B. (1982). Repeated Reading. *Journal of Learning Disabilities*, 45, 619-623.
- Olson, R. K., Foltz, G. & Wise, B. (1986). Reading instruction and remediation with the aid of computer speech. *Behavior Research Methods, Instruments & Computers*, 18, 93-99.

- Perkins, V. L. (1988). Feedback effects on oral reading errors of children with learning disabilities. *Journal of Learning Disabilities*, 21, Nr 4, 244-248.
- Reitsma, P. (1985). De verwerving van orthografische kennis. In E. Assink & G. Verhoeven (Eds). *Visies op spelling*. Groningen: Wolters-Noordhoff.
- Reitsma, P. (1988). Reading practice for beginners: effects of guided reading, reading-while-listening and independent reading with computer-based speech feedback. *Reading Research Quarterly*, 23, 219-235.
- Reitsma, P., Ellermann, H. H. & Spaai, G. W. G. (1987). An electronic aid for practising letter-sound relations. In J. Moonen & T. Plomp (Eds), *Eurit 86: Developments of educational software and courseware*. Oxford: Pergamon Press.
- Spaai, G. W. G., Reitsma, P. & Ellermann, H. H. (1986). Effects of two forms of sound feedback on learning to read words. *IPO-Annual Progress Report*, 22, 87-96.
- Spaai, G.W.G., Reitsma, P. & Ellermann, H.H. (1987). Effects of several feedback methods for correcting reading errors by computer-assisted instruction. *IPO Annual Progress Report*, 22, 87-96.
- Spiegel, D. L. Rogers, C. (1980). Teacher responses to miscues during oral reading by second grade students. *Journal of Educational Research*, 74, 1, 8-12.
- Van Dongen, D. & Mommers, M. J. C. (1980). *Lees- en spellingtoetsen voor leerjaar 1 tot 4 van de lagere school*. K.U. Nijmegen : Vakgroep Interdisciplinaire Onderwijskunde.
- Wise, B. W. (1987). *Word segmentation in computerized reading instruction*. Dissertatie. Universiteit van Colorado.

Manscript ontvangen 17-11-1988

Definitieve versie ontvangen 16-10-1989

Notities en Commentaren

TEST- TOETS- GEBRUIK IN PROGRAMMA-EVALUATIE: HERZIENE 'RICHTLIJNEN'

H.P.G.M. Caminada

KUB, Katholieke Universiteit Brabant

Onlangs is door het Nederlands Instituut van Psychologen (NIP) een herziene versie uitgegeven van de *Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen* (Evers, Caminada, Ter Laak, Maessen de Sombref, en Starren, 1988). De vorige uitgave stamde uit 1978. De redactie van de nieuwe Richtlijnen bestond uit een subcommissie van de COTAN (Commissie Testaangelegenheden Nederland), waar ondergetekende deel van uitmaakte.

Belangrijke veranderingen van deze herziene uitgave zijn, dat niet alleen aandacht wordt besteed aan nieuwe ontwikkelingen in de testtheorie, maar ook, dat concrete voorschriften voor gebruik van psychologische instrumenten op specifieke toepassingsgebieden worden gegeven. De Richtlijnen zijn deels een bewerking van *Standards for educational and psychological testing* (APA, 1985). Enkele hoofdstukken zijn echter geheel herschreven, waaronder een hoofdstuk over *Toetsgebruik in het onderwijs* en een hoofdstuk *Programma-Evaluatie (PE)*. Van de twee laatstgenoemde hoofdstukken, behandelt het eerstgenoemde hoofdstuk het gebruik van toetsen voor het nemen van beslissingen over leerlingen (voortgangscontrole, plaatsing/selectie en kwalificatie). Het hoofdstuk PE, dat hieronder wordt besproken, beschrijft het gebruik van diverse instrumenten bij het onderzoek naar de *effectiviteit* van programma's in het onderwijs en op andere toepassingsgebieden.

Wat is een programma en hoe wordt PE gedefinieerd?

Een programma is een *plan* voor een treatment. Een plan specificiert de middelen, dat wil zeggen de treatment-condities, alsook de doelen, evenals de wijze waarop de doelen, de beoogde of geplande effecten, via de middelen moeten worden gerealiseerd. PE is primair onderzoek naar de effectiviteit van een plan. In PE onderzoekt men de effectiviteit van een plan niet direct, maar indirect door middel van onderzoek naar de effectiviteit van het uitgevoerde plan. Het uitgevoerde plan is het programma zoals dat in feite door middel van een bepaalde treatment is geconcretiseerd.

PE wordt nu omschreven als: Nagaan of een bepaalde treatment het geplande effect sorteert en ook of dit effect gegeneraliseerd kan worden naar treatment-condities, effectmaten en onderzoekspopulaties die niet zijn onderzocht, maar die wel voor het plan van belang zijn. Omdat aan een plan een theorie ten grondslag ligt, die de relatie beschrijft tussen treatment-condities en geplande effecten, onderzoekt men in PE in feite de geldigheid van een theorie.

Soms blijkt het geplande effect niet gerealiseerd te zijn. Hoe komt dat?

Uitgangspunt van het hoofdstuk PE is, dat evaluatie gericht dient te zijn op *causale conclusies* en verder, dat niet alleen zuivere experimentele evaluatie-designs, maar ook quasi-experimentele evaluatie-designs tot causale conclusies kunnen leiden (zie ook Hofstee, 1987). Afstand wordt genomen van de opvatting van Cronbach (1982), namelijk, dat men in PE niet naar causale gevolgtrekkingen zou moeten streven vanwege de

moeilijkheid om tot extern-relevante causale uitspraken te komen.

Als kernprobleem in PE wordt gezien de interpretatie van de resultaten verkregen met behulp van de effectmaten. Bij de interpretatie van de resultaten moet men erop bedacht zijn, dat het niet vinden van een gepland treatment-effect het gevolg kan zijn van uiteenlopende factoren, zoals:

1. een ongeldige theorie over de causale relatie tussen de treatment-condities (middelen) en de geplande effecten (doelen);
2. een ondeskundig uitgevoerde treatment;
3. een treatment, die ongeschikt is voor de personen die aan de treatment deelnemen;
4. een onjuist opgezet en/of uitgevoerd onderzoek.

Het omgekeerde kan natuurlijk ook het geval zijn. Een gepland treatment-effect kan zich voordoen, ondanks het optreden van één of meer van bovenstaande factoren.

Wat zijn de kwaliteitseisen waar instrumenten in PE aan moeten voldoen en waar moet men bij een verantwoorde keuze van deze instrumenten op letten?

In het hoofdstuk PE heeft de eerste van de vijf richtlijnen betrekking op de keuze van instrumenten in PE en op de verantwoording van die keuze. Volgens deze eerste richtlijn behoren validiteits- en betrouwbaarheidsgegevens te worden verstrekt die aantonen, dat de keuze van de gebruikte instrumenten verantwoord is. Bij de keuze moet worden gelet op:

1. de specifieke gebruiksdoelen van instrumenten in PE. D.w.z., instrumenten kunnen in PE worden gebruikt als selectie- of classificatiematen, als co-variabelen of matchingsvariabelen, en als effectmaten.
2. Theoretische overwegingen.
3. De doelen van het evaluatie-onderzoek. D.w.z., de gerichtheid van het onderzoek op ontwikkeling of verbetering van programma's ofwel op effectmeting van reeds ontwikkelde programma's.
4. De onderzochte populaties.

In PE kunnen instrumenten voor verschillende doeleinden worden gebruikt:

1. Ze kunnen allereerst worden toegepast om te bepalen welke personen wel en welke niet geschikt zijn om aan een programma deel te nemen, en ook om uit te maken welke personen geschikt zijn om het programma of bepaalde onderdelen ervan uit te voeren. In beide gevallen gaat het om toepassing van instrumenten als *selectie- of classificatiematen*. Van zo'n toepassing is uiteraard ook sprake, wanneer instrumenten worden gebruikt voor de toewijzing van personen aan de treatment- en controlegroep. Voor de kwaliteitseisen van instrumenten in PE bij dit gebruiksdoel kan worden verwezen naar de richtlijnen in de hoofdstukken 8 (Toetsgebruik in het onderwijs) en 10 (Testgebruik in arbeidssituaties) van het boek 'Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen'.
2. Instrumenten kunnen in PE ook worden gebruikt als *co-variabelen en matchingsvariabelen* om na te gaan of alternatieve causale verklaringen kunnen worden uitgesloten. De betrouwbaarheid van de metingen op deze variabelen dient te worden vermeld. Behalve tests en toetsen kunnen ook sexe, leeftijd en opleidingsniveau als co-variabelen of matchingsvariabelen worden gebruikt. De keuze van deze variabelen wordt allereerst bepaald door de theorie die aan de treatment ten grondslag ligt. De keuze van deze variabelen wordt verder bepaald door de aard van de onderzochte populatie. Bijvoorbeeld: Er kunnen verschillen op toetsen in schoolvorderingen tussen jongens en meisjes optreden als gevolg van verschillen in rijping gedurende de puberteit.
3. Instrumenten kunnen in PE bovendien worden gebruikt om het geplande effect van een treatment te bepalen. De inhoud van de *effectmaten* behoort aan te sluiten bij de doelen

of geplande treatment-effecten. Er dienen dan ook gegevens te worden verstrekt die de inhoudsvaliditeit van de gebruikte effectmaten met betrekking tot de corresponderende doelen of geplande effecten ondersteunen. Wanneer twee of meer programma's worden vergeleken, dan moet men bij de keuze van effectmaten rekening houden met de doelen van de afzonderlijke programma's. Omdat het ook van belang is een schatting te maken van positieve en negatieve bij-effecten en niet-beoogde resultaten, dient men ook instrumenten te kiezen die aspecten meten die geen specifieke doelen van het te evalueren programma zijn.

Wanneer effectmaten psychologische eigenschappen meten, dan zullen ook gegevens moeten worden verstrekt die de begripsvaliditeit ondersteunen. Wanneer in PE gebruik wordt gemaakt van kwalitatieve effectmaten, zoals zelfbeschrijvingen, brede 'open' vraag- en observatiemethoden, dan dienen er gegevens te worden verstrekt over de betrouwbaarheid en validiteit van de kwalitatieve methoden.

Het onderscheid tussen de beide doelen van PE is van belang, met name voor de beleidsmatige implicaties. Onderzoek gericht op de effectmeting van reeds ontwikkelde programma's kan inhouden, dat het voortbestaan van een programma, in casu de onderzochte treatment, afhankelijk wordt gesteld van de onderzoeksresultaten. Deze implicatie is er niet in onderzoek, dat is gericht op de ontwikkeling of verbetering van programma's. Gelet op het gewicht van de beleidsbeslissingen die kunnen voortvloeien uit de resultaten van het onderzoek, is het allereerst van belang bij de effectmeting van reeds ontwikkelde programma's de geplande effecten op verschillende manieren te operationaliseren. Wanneer het effect van de treatment zich op de verschillende maten steeds op dezelfde wijze manifesteert en wanneer dat effect tevens het geplande effect is, dan is het waarschijnlijker, dat de treatment aan zijn doel beantwoordt dan wanneer er maar één maat is of wanneer de veranderingen op de verschillende maten weinig of niet met elkaar zouden overeenstemmen. De keuze van de verschillende operationalisaties van elk doel moet dan echter wel worden bepaald op basis van de theorie die aan de treatment ten grondslag ligt. Bij effectmeting van reeds ontwikkelde programma's dienen tevens gegevens te worden verstrekt die de validiteitsgeneralisatie van de effectmaten ondersteunen. Hierbij is het niet alleen van belang, dat aannemelijk wordt gemaakt, dat de validiteitsgegevens van toepassing zijn op andere condities dan de onderzochte treatment-conditie of, dat de validiteitsgegevens gelden voor andere populaties dan de onderzochte populatie, maar het is ook van belang, dat aannemelijk wordt gemaakt, dat de effectmaten in principe bruikbaar zijn voor de bepaling van het duurzame effect van de treatment. Dit laatste is onder voorbehoud, dat een duurzaam effect van belang wordt geacht.

Bij de keuze van psychologische instrumenten in PE (dus niet alleen van de effectmaten) dient men niet enkel rekening te houden met kwaliteitseisen die samenvallen met psychometrische eisen. Men dient bovendien te letten op de kosten, de uitvoerbaarheid van de afname, de beschikbare tijd en kenmerken van de onderzochte populatie.

Tenslotte: behalve het hierboven aangesneden aspect van de keuze van instrumenten, zijn er nog andere kanten aan het gebruik van instrumenten in PE te onderscheiden, zoals het definiëren van verandering of groei, alsmede het definiëren van veranderingsscores; de beperking van het gebruik van veranderingsscores; de procedure om individuele scores om te zetten in een groepsscore; en verder de verschillende wijzen waarop de effectiviteit van een programma kan worden weergegeven. Voor een bespreking van laatstgenoemde aspecten en de daarop betrekking hebbende richtlijnen, alsook van de wetenschappelijke en beleidsmatige implicaties van PE, kan worden verwezen naar het betreffende hoofdstuk PE.

LITERATUUR

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Cronbach, L.J. (1982). *Designing Evaluations of Educational and Social Problems*. San Francisco: Jossey-Bass.
- Evers, A., Caminada, H., Koning, R., Laak, J. ter, Maessen de Sombref, P. van der, Starren, J. (1988). *Richtlijnen voor Ontwikkeling en Gebruik van Psychologische Tests en Studietoetsen*. Amsterdam: Nederlands Instituut van Psychologen.
- Hofstee, W.K.B. (1987). Het causaliteitsvraagstuk in evaluatie-onderzoek. In W.K.B. Hofstee (Ed.), *Evaluatiemethodologie* (pp. 93-108). Lisse: Swets & Zeitlinger.

Ontvangen 2-6-1989

Notities en Commentaren

PEIL VAN PPON-ONDERZOEK KAN HOGER

Y.J. Pijl & E.G. Harskamp

RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen

In 1988 werd de eerste in een reeks van rapporten gepubliceerd over de uitkomsten van het project 'Periodieke Peiling van het Onderwijsniveau' (PPON). Het doel van het project is de discussie over inhoud en niveau van het basisonderwijs van een empirische basis te voorzien. Daartoe worden periodiek gegevens verzameld over de resultaten die door de leerlingen behaald worden, en over het leer- en vormingsaanbod. De analyse van deze gegevens zal gericht worden op uitspraken over het niveau van de behaalde resultaten, de inhoud van het onderwijs en de samenhang tussen inhoud en resultaat van het onderwijs. Door het herhalen van de metingen wordt inzicht in ontwikkelingen in de loop van de tijd nagestreefd.

Met de eerste peiling heeft het PPON-project zich geconcentreerd op rekenvaardigheden halverwege en aan het eind van het basisonderwijs. In PPON-publicatie nr. 1 'Balans van het Rekenonderwijs in de basisschool' wordt hiervan verslag gedaan. Met de uitgave van het voorliggende rapport hebben de auteurs geen wetenschappelijke verslaggeving nagestreefd. Het rapport is vooral bedoeld voor een inhoudelijk geïnteresseerd publiek. Voor de methodologische details verwijzen de auteurs naar een nog te verschijnen technische rapportage. Het lijkt dan ook voorbarig om een antwoord te willen geven op de vraag of de resultaten van de eerste rekenpeiling op een wetenschappelijk verantwoorde wijze geïnterpreteerd kunnen worden. Het voorliggende rapport is echter zodanig gedetailleerd dat ook op het terrein van het gebruik van wetenschappelijke methoden en technieken in het PPON-onderzoek belangwekkende uitspraken in de rapportage teruggevonden kunnen worden. Vaak geven deze uitspraken aanleiding tot informatieve vragen naar het hoe en waarom van de gekozen oplossingen. Het rapport geeft echter ook aanleiding tot twijfel aan de vraag of de in het PPON-onderzoek verzamelde gegevens geïnterpreteerd kunnen worden op een wijze die optimaal is in het licht van de doelstelling van het project.

Gezien de grote kosten van het PPON-project en de grootte van de opbrengst die door de opdrachtgevers verwacht wordt, is het noodzakelijk dat het PPON-onderzoek op wetenschappelijke wijze uitgevoerd wordt. Aangezien het PPON-onderzoek overeenkomstig de titel een periodiek karakter heeft en er nog vele peilingen op de rol staan, achten wij het zinvol in een vroegtijdig stadium enkele kritische kanttekeningen bij de eerste peiling voor te leggen aan de betreffende onderzoekers en aan het wetenschappelijk forum.

De meting van de rekenvaardigheden

Er zijn bij de peilingen halverwege en aan het eind van het basisonderwijs respectievelijk 13 en 27 toetsen afgenomen. Door deze grote aantallen was het niet praktisch om elke leerling alle toetsen volledig te laten maken. De onderzoekers hebben dit probleem opgelost door elf toetsen bij de peiling halverwege het basisonderwijs te verdelen in vijf groepen van

toetsen. De 26 toetsen van de peiling aan het einde van het basisonderwijs zijn in zes groepen verdeeld. Elke leerling maakt naast een algemene basiskennistoets de opgaven van één van deze groepen van toetsen.

Het nadeel van deze procedure ten opzichte van bijvoorbeeld een procedure waarbij aan elke leerling at random de toetsen worden toegewezen, is dat er uitsluitend gepaarde waarnemingen beschikbaar zijn voor de toetsen die onderdeel vormen van dezelfde groep van toetsen. Daardoor is er op empirische basis geen discussie mogelijk over de factorstructuur van het toetsinstrumentarium. Dit betekent dat vragen die van belang kunnen zijn voor de interpretatie van de resultaten, niet beantwoord kunnen worden. Daarbij moet bijvoorbeeld gedacht worden aan de vraag naar de eventuele redundantie van sommige toetsen binnen het grote aantal rekenvaardigheden die in het PPON-onderzoek onderscheiden zijn, en aan de vraag naar de mogelijkheid om een totaalscore te berekenen op grond van een aangetoonde algemene 'rekenvaardigheids'factor.

Een tweede probleem met het onderzoeksontwerp betreft de constructie van 2 versies van elke toets: er is een makkelijke (A) en een moeilijke (B) versie. Op grond van het oordeel van de leerkracht of een leerling meer dan wel minder rekenvaardig is, wordt bij de betreffende leerling de A of B versie afgenomen. Beide versies bestaan gedeeltelijk uit dezelfde items. Naast de gemeenschappelijke items van een middelmatige moeilijkheidsgraad bevat de A-versie relatief gemakkelijke items en de B-versie relatief moeilijke items. Omdat aan geen enkele leerling zowel de A- als de B-versie wordt voorgelegd, zijn er geen gepaarde waarnemingen voor itemparen van een gemakkelijk item dat specifiek is voor de A-versie, en een moeilijk B-item. Om nu scores op de A-versie met de B-versie te kunnen vergelijken moet een meetmodel gespecificeerd worden dat strakke eisen stelt aan het gedrag van de itemscores (bijv. Rasch).

Uit het rapport blijkt dat de gegevens niet aan het Rasch-model voldoen maar wel aan een minder restrictief model. De vraag wat dit laatste model inhoudt, en in welke mate het model op de data past, zal ongetwijfeld in de aangekondigde technische rapportage aan de orde worden gesteld en kan nog niet beantwoord worden.

Waar het hier om gaat is de vraag of met de opzet van de rekenpeilingen geen groot risico is genomen door op één en dezelfde steekproef zowel het meetmodel te toetsen als de resultaten van de peilingen te berekenen. Stel dat achteraf geen meetmodel op de data bleek te passen dat de vertaling van A- en B-scores in één schaal toelaat. In dat geval waren alle scores onbruikbaar geweest omdat A- en B-versies niet door een toevalsmechanisme maar door leerkrachten worden toegewezen.

In toekomstige peilingen lijkt het veiliger om voorafgaand aan de peiling een vooronderzoek uit te voeren naar het gedrag van de itemscores of om de peiling zodanig op te zetten dat ook minder restrictieve meetmodellen gebruikt mogen worden.

Invloed van zittenblijven op peilingsresultaten.

Een belangrijk motief voor PPON is de behoefte aan de kwaliteitsbewaking van het onderwijs geweest, onder andere in de vorm van een periodieke peiling van de kwaliteit van de resultaten van het basisonderwijs. Wanneer de baten geïnventariseerd moeten worden in de verschillende jaren voor meerdere groepen, moeten de kosten van het onderwijs eveneens op een of andere wijze verdisconteerd worden. Het steekproefontwerp van het PPON-onderzoek sluit wat betreft de definitie van de doelpopulatie niet aan bij een dergelijke doelstelling.

Bij de peiling halverwege het basisonderwijs bestaat de doelpopulatie uit leerlingen die in een groep 5 zitten. Dit betekent dat het beleid van de scholen ten aanzien van zittenblijvers van invloed is op de samenstelling van de doelpopulatie. Stel dat de ontwikkeling van de schoolprestaties onder andere afhankelijk is van de leeftijd van de

leerling, omdat met de leeftijd goeddeels het aantal jaren onderwijs dat genoten is, is gegeven. Stel verder dat maatregelen als zittenblijven geen noemenswaardige invloed hebben op de prestaties maar vooral om organisatorische redenen genomen worden. Op een school waar men in het algemeen iets eerder geneigd is de kinderen te laten doubleren, stijgt de gemiddelde leeftijd in een groep en daarmee de gemiddelde prestatie.

Veel zittenblijvers in een school kunnen de resultaten ook negatief beïnvloeden. Wanneer een school zeer terughoudend is ten aanzien van het blijven zitten van matig presterende kinderen in de groepen 3 en 4, maar niet in groep 5, zal de gemiddelde prestatie dalen in vergelijking met een school die in alle groepen evenveel kinderen laat doubleren, of een school waar niemand blijft zitten. Immers, aan alle kinderen die wat betreft leeftijd (8 of 9) in groep 5 zouden kunnen zitten, worden nu enkele matig presterende leerlingen van 9 of 10 jaar toegevoegd.

In welke richting de resultaten vertekend worden, is niet van belang. Het probleem is veel meer dat onbekend is hoe de resultaten van de peilingen vertekend zijn. Daarnaast leidt een mogelijk verband tussen beleid van scholen ten aanzien van zittenblijven enerzijds en kenmerken van scholen of het jaar waarin een peiling wordt uitgevoerd anderzijds tot onjuiste interpretaties van verschillen tussen de groepen scholen en van veranderingen van prestaties in de loop van de tijd.

De inhoudelijk minder gewenste effecten van zittenblijven op de resultaten van de peilingen kunnen vermeden worden door een doelpopulatie te definiëren in termen van kinderen die een bepaald aantal jaren basisonderwijs genoten hebben of, dit komt in de praktijk op vrijwel hetzelfde neer, in termen van de leeftijd van het kind. Hoewel de PPON-onderzoekers het nut van een dergelijke definitie van een doelpopulatie voorzien hebben en de mogelijkheid tot de 'leeftijds'-variant open hebben willen houden, is deze variant in de peiling halverwege het basisonderwijs niet toegepast. In de peiling aan het eind van het basisonderwijs is deze variant eveneens niet toegepast en kan ook achteraf niet meer toegepast worden, omdat er door onbekende oorzaken te weinig 12-jarige zittenblijvers zijn getest en ook niet bekend is hoe de 12 jarigen in de populatie verdeeld zijn over de groepen.

Selectie door uitstroom naar het speciaal onderwijs.

Het volgende kritiekpunt is verwant aan het voorgaande in die zin, dat het een selectiemechanisme betreft dat de feitelijke doelpopulatie in onbekende mate doet afwijken van de genoemde of gewenste doelpopulatie. Er is geen rekening gehouden met verwijzingen naar het speciaal onderwijs. Wanneer bij een volgende peiling het percentage verwijzingen naar het speciaal onderwijs wederom gestegen blijkt, of wellicht door bezuinigingen sterk gedaald blijkt, is er bij die volgende peiling sprake van een andere populatie dan bij de eerste peiling. Verschillen tussen peilingen kunnen in dat geval niet geïnterpreteerd worden. Ditzelfde geldt voor verschillen tussen groepen van scholen wanneer de ene groep van scholen gemiddeld minder snel verwijst naar het SO dan een andere groep.

Stel, door bezuinigingen en negatieve onderzoeksrapporten daalt de animo van ouders voor LOM en MLK. Daardoor blijven de kinderen die relatief slecht presteren, vooral in taal maar ook in rekenen, op de reguliere basisschool. Deze kleine groep zal extreme scores produceren die de gemiddelde score op de nieuwe PPON-toets naar beneden trekken. Dit probleem had voorkomen kunnen worden door de scholen in het basisonderwijs te stratificeren naar regulier en speciaal en vervolgens in beide strata de mate waarin de gestelde leerdoelen bereikt zijn eenduidig te kwantificeren.

Conclusies

De interpretatie van de resultaten wordt bemoeilijkt door een ongelukkige definitie van de doelpopulatie. Zowel het beleid van de scholen ten aanzien van zittenblijven en de verwijzing naar het speciaal onderwijs kan de uitkomst van de peilingen beïnvloeden. Aangezien het beleid van scholen op genoemde punten in de loop van de tijd kan wijzigen, is vooral het periodieke aspect van het PPON-onderzoek bij de eerste peiling nog onvoldoende uit de verf gekomen.

Ontvangen 20-7-1989

Notities en Commentaren

NORMAAL GRONINGS PEIL?

EEN REACTIE OP PIJL & HARKAMP: PEIL VAN PPON-ONDERZOEK KAN HOGER

Theo J.H.M. Eggen, Norman D. Verhelst en Johan M. Wijnstra
Instituut voor Toetsontwikkeling (Cito), Arnhem

In hun bijdrage in dit nummer betwijfelen Pijl & Harskamp (verder: P & H) of de gegevens uit de eerste rekenpeiling van het PPON-project (Wijnstra, 1988; zie ook Wijnstra, ter perse) 'geïnterpreteerd kunnen worden op een wijze die optimaal is in het licht van de doelstelling van het project'. Zoals in deze reactie zal worden betoogd, berusten de door P & H aangesneden punten voornamelijk op een eigen interpretatie van de doelstelling van het project, respectievelijk onbegrip voor de gehanteerde methodologie, wat tot overbodige opmerkingen en laakbare onzorgvuldigheden leidt. In verband hiermee is het betreuringswaardig dat P & H hun conclusies niet - opnieuw - hebben geverifieerd bij het projectteam, respectievelijk de tijd hebben genomen om de wetenschappelijke rapportage af te wachten, waarvan het belangrijkste deel inmiddels beschikbaar is (Verhelst & Eggen, 1989).¹

Doelstelling van peilingsonderzoek

In hun inleiding formuleren P & H de doelstelling van het project correct, en met name ook dat de tot nu uitgevoerde peilingen zich beperken tot het basisonderwijs. Op zichzelf hebben zij gelijk dat verschuivingen in de proporties kinderen die naar het speciaal onderwijs worden verwezen, tot complicaties bij de interpretatie van de resultaten zouden kunnen leiden bij herhaling van een peiling in het basisonderwijs. Evenwel, de opdracht (van de Minister van Onderwijs en Wetenschappen) beperkte tot medio dit jaar de peilingen nadrukkelijk tot het reguliere basisonderwijs.² Om deze reden zullen we op dit punt verder niet ingaan.

De doelstelling van het project, namelijk een inhoudelijke discussie mogelijk maken over de inhoud en het niveau van het onderwijs, heeft tot het uitgangspunt geleid 'dat de instrumenten een beschrijving van de vaardigheden van de leerlingen mogelijk moeten maken op nauwkeurig omschreven, didactisch betekenisvolle eenheden van beperkte omvang'. (Wijnstra, 1988, p. 5). Het onderzoeksontwerp is hierop afgestemd. Uiteraard leidt dit tot redundantie, in die zin dat de onderscheiden vaardigheden (sterk) gecorreleerd zullen zijn. De vragen die P & H in dit verband stellen naar de factorstructuur van de toetsen, behoren niet tot de primaire vraagstelling van het project. Beantwoording van deze vragen, in combinatie met de primaire vraagstelling in hetzelfde onderzoeksontwerp (zoals P & H voorstellen), zou tot een aanzienlijke verhoging van de complexiteit van de dataverzameling hebben geleid en in het geval van de peiling einde basisonderwijs - uitgaande van slechts 100 gepaarde waarnemingen per schaal - tot een verdubbeling van de omvang. Hoewel deze mogelijkheden wel zijn overwogen, is gekozen voor andere 'complicaties' in het onderzoeksontwerp (bijvoorbeeld een structureel onvolledig calibratiedesign) om daarmee de informatie-opbrengst ten gunste van de oorspronkelijke vraagstelling te optimaliseren.

Adres: Cito, Postbus 1034, 6801 MG Arnhem

Dat bij de keuze van een onderzoeksontwerp andere vragen, die niet primair aan de orde zijn, minder goed of niet kunnen worden beantwoord, is evident. De beantwoording van de bijkomende vragen die P & H stellen, is overigens relatief goedkoop te realiseren in een afzonderlijk onderzoekje. Door van elke (in de rekenpeiling gecalibreerde) schaal een klein aantal opgaven aan één steekproef voor te leggen, is het heel goed mogelijk de intercorrelaties tussen alle onderliggende vaardigheden te schatten. Dit is in 1988 ook gebeurd in het kader van een experimentele schoolrapportageservice met opgaven uit de peiling einde basisonderwijs. Het resultaat van dit onderzoek was een intercorrelatiematrix die essentieel door één enkele factor te reproduceren was. Dit wil echter niet zeggen dat we in de peiling met één (inhoudelijke heterogene) toets hadden kunnen volstaan, op basis waarvan een totaalscore op een algemene rekenvaardigheidsfactor zou kunnen worden berekend. Dan zou de discussie over inhoud en niveau van het rekenonderwijs wel van een erg smalle, ongedifferentieerde empirische basis worden voorzien.

Methodologie

Op zichzelf hebben P & H gelijk dat het idealiter beter is het meetmodel en het structurele model (waarin de peilingsresultaten worden ondergebracht) in afzonderlijke onderzoeken aan de orde te stellen, zoals overigens uitvoerig is betoogd in Verhelst & Eggen (1989). Onder de druk van de omstandigheden zijn beide onderzoeken in één design ondergebracht, waardoor beide niet optimaal kunnen zijn. P & H suggereren dat hiermee - met name ook door toepassing van itemresponstheorie (in een structureel onvolledig calibratiedesign) - grote risico's zijn genomen die ertoe hadden kunnen leiden dat de gehele dataverzameling waardeloos was geweest.

Nog afgezien van het feit dat de opmerking nogal als mosterd na de maaltijd komt, waren die risico's beperkt. Het risico waar P & H op zinspelen heeft te maken met het feit dat er 'strakke eisen' gesteld worden aan de itemscores. Met de strakheid van die eisen valt het echter nogal mee. De twee belangrijkste handicaps van het Rasch-model, namelijk de onmogelijkheid om 'raden' te verklaren en de gelijke discriminatie per item, werden omzeild door respectievelijk (bijna) uitsluitend open vragen op te nemen, en door een model te hanteren dat ongelijke discriminaties van de items toelaat. Het risico dat zo'n model geen acceptabele beschrijving zou geven voor een acceptabele deelverzameling van de aangeboden items per schaal (zeg tenminste 75%) was gezien de inhoudelijke homogeniteit van de schalen erg klein. De analyses hebben ons in dit opzicht gelijk gegeven: minder dan 5% van het totaal aantal items moest wegens misfit worden verwijderd. Het structureel onvolledige design, met een selectievariabele, maakte daarbij een nauwkeuriger schatting van de itemparameters mogelijk dan in een volledig design (met eenzelfde aantal item-persoon combinaties) of een onvolledig design met random toewijzing van de leerlingen aan gelijkwaardige toetsversies (vgl. Verhelst, 1989).

Het enige echte risico in de zin die door P & H is bedoeld, was de situatie waarin voor een schaal *alle* gemeenschappelijke items voor de A- en de B-versie (en dit waren er voor alle schalen tenminste acht) zouden moeten worden verwijderd. Als deze onwaarschijnlijke situatie zich zou voordoen, zou het dan alleen nog maar die ene schaal betreffen, niet zo'n vreselijke ramp gezien de grote hoeveelheid 'redundante' schalen die overbleef.

Met betrekking tot de peilingsresultaten spannen P & H zich in om aan te tonen dat we de verkeerde referentiepopulatie hebben gekozen, namelijk alle leerlingen die in jaargroep 5, respectievelijk 8 zitten, omdat door een ander zittenblijversbeleid in toekomstige peilingen de gemiddelde vaardigheid zou kunnen stijgen of dalen. Op de rationales die ze daarvoor opstellen, behoeven we hier niet in te gaan, om de eenvoudige reden dat P & H niet beseffen dat de beschrijving van de referentiepopulatie niet meer is dan een bijproduct van de schattingsmethode. Daarom in het kort een samenvatting van de schattingsprocedure.

In de calibratie worden voor de te meten vaardigheid eenheid en nulpunt *willekeurig* gekozen. Deze kunnen te allen tijde gewijzigd worden. Omdat het hier een eerste peiling betrof, zijn eenheid en nulpunt voor de referentiepopulatie zodanig gekozen dat deze een goed interpreteerbare betekenis hebben. Omdat de populatie echter niet ongedifferentieerd is gedefinieerd, blijft deze betekenis invariant ook als in de toekomst de samenstelling van de referentiepopulatie zou veranderen op de gecontroleerde zes variabelen, waaronder zittenblijven als de combinatie van leeftijd en leerjaar (de overige zijn stratum waartoe de school behoort, gewichtscategorie voor de formatiebepaling, geslacht, type rekenmethode, het wel of niet gehad hebben van de basisstof).

Veronderstel dat er twee belangrijke achtergrondvariabelen systematisch effect hebben op de vaardigheid van de leerlingen, bijvoorbeeld leeftijd (d.w.z. zittenblijver of niet) en geslacht. Dan onderscheiden we vier subpopulaties, en in elk van deze vier subpopulaties worden de gemiddelde vaardigheid en de variantie geschat (met de eenheid en het nulpunt zoals in de calibratie vastgelegd). Omdat ook de proportie van elk van deze subpopulaties bekend is, kunnen we dus ook gemiddelde en variantie in de totale populatie schatten. Vervolgens veranderen we eenheid en nulpunt van de schaal zodanig, dat in de totale (huidige) populatie het gemiddelde 250 is en de standaardafwijking 50. Dit is echter een simpele lineaire transformatie waardoor ook gemiddelde en variantie in elk van de subpopulaties vastliggen. Bijgevolg kunnen we gemiddelde en variantie bepalen in elke populatie die een willekeurige compositie is van de vier subpopulaties. Als in een toekomstige peiling de gemiddelden in die vier subpopulaties onveranderd zijn, zullen we ze ook als onveranderd vinden, ongeacht of de samenstelling van de totale populatie veranderd is of niet ten opzichte van de vorige peiling. Als omgekeerd het gemiddelde in een of meerdere van die subpopulaties verandert, constateren we dit ook. Met andere woorden, veranderingen in de vaardigheidsverdeling van een toekomstig cohort ten opzichte van het huidige cohort kunnen duidelijk en ondubbelzinnig worden toegeschreven aan één of meerdere van de geobserveerde achtergrondvariabelen en/of aan een veranderde samenstelling van de totale populatie in termen van deze achtergrondvariabelen. De kritiek van P & H snijdt dus geen hout.

In dit verband is - voor alle duidelijkheid - nog het volgende van belang. In de rapportage is, mede gelet op de onderwijspraktijk, gekozen voor een referentiepopulatie die in termen van leerjaar (5, respectievelijk 8) is gedefinieerd. De resultaten van de acht-, respectievelijk twaalfjarige leerlingen die in deze jaargroepen hadden kunnen zitten, maar in een lagere groep zijn geplaatst, worden uitgedrukt in schaaleenheden van de referentiepopulatie. Wanneer dat noodzakelijk zou zijn, is een herdefiniëring van de referentiepopulatie in termen van leeftijd ook zonder bezwaar mogelijk, althans in de peiling einde basisonderwijs.³ De twaalfjarige leerlingen in de jaargroepen 6 en 7 hebben volledig aan de peiling deelgenomen. Weliswaar hebben we het vermoeden geuit dat de proportie twaalfjarigen in de jaargroepen 6 en 7 in de steekproef kleiner is dan in de populatie, maar er zijn geen aanwijzingen dat de representativiteit van de steekproeven in deze cellen is aangetast (vgl. Wijnstra, 1988, p. 83) en de voorgaande uiteenzetting is ook zonder problemen toepasbaar bij deze definitie van de referentiepopulatie.

Het voorgaande betekent natuurlijk niet dat de achtergrondvariabelen die in de analyses worden betrokken, alle systematische variantie verklaren. Stel dat we in een volgende peiling vinden dat zittenblijvers relatief slechter presenteren dan in de huidige peiling. Dan is deze verandering mogelijk te verklaren door een ander beleid met betrekking tot verwijzing naar het speciaal onderwijs, doch er zijn oneindig veel andere verklaringen denkbaar. Over de geldigheid van deze verklaringen kunnen we pas iets te weten komen als we de verklarende variabelen observeren. In het geval van die ene hypothese die P & H aanhalen, betekent dit dat de peiling naar het speciaal onderwijs moet worden uitgebreid.

Deze uitbreiding is inmiddels in de opdracht opgenomen, echter niet met als primaire doelstelling de toetsing van deze hypothese.

Besluit

Dat het periodieke aspect van peilingsonderzoek nog onvoldoende uit de verf is gekomen, is een conclusie van P & H die kant noch wal raakt. Hoewel er nog geen herhalingen hebben plaatsgevonden, heeft in de gekozen methodologie het periodieke karakter van het project een doorslaggevende rol gespeeld. De indeling in inhoudelijk homogene leerstof-domeinen, de toepassing van itemrespons-theorie en de daarmee geassocieerde wijze waarop effecten van achtergrondvariabelen op de vaardigheden van de leerlingen worden onderzocht, bieden juist de waarborgen dat periodiek vergelijken ook inderdaad kan plaatsvinden. Dat P & H zonder grondige kennisname van de methodologie dit niet beseffen, is tot daar aan toe. Dat zij echter proberen het wetenschappelijk peil van PPON hoe dan ook verdacht te maken, is laakbaar. Het peil van PPON kan wellicht hoger, de suggesties van P & H dragen daartoe weinig bij.

LITERATUUR

- Verhelst, N.D. (1989). Informatiewinst bij vertakt toetsen. In: W.J. van der Linden & L.J.Th. van der Kamp (Red.), *Meetmethoden en data-analyse in onderwijskundig onderzoek*. Amsterdam: Swets & Zeitlinger.
- Verhelst, N.D. & Eggen, T.J.H.M. (1989). *Psychometrische en statistische aspecten van peilingsonderzoek*. Arnhem: Instituut voor Toetsontwikkeling (PPON-rapport nr. 4).
- Wijnstra, J.M. (1988). *Balans van het rekenonderwijs in de basisschool*. Arnhem: Instituut voor Toetsontwikkeling (PPON-reeks nr. 1).
- Wijnstra, J.M. (ter perse). Periodiek peilingsonderzoek: de opzet van het project PPON geïllustreerd aan de rekenpeiling einde basisonderwijs. *Pedagogische Studiën*.

Noten

¹ Deze publikatie is in de eerste week van september 1989 verspreid. P & H waren bekend met dit tijdstip.

² In juli 1989 heeft de minister het Cito verzocht de peilingen uit te breiden tot het speciaal onderwijs. Het ligt in de bedoeling de eerste ronde in het speciaal onderwijs, voorlopig beperkt tot de kernafdelingen van LOM- en MLK-scholen, te combineren met de tweede ronde in het reguliere basisonderwijs.

³ In tegenstelling tot wat P & H beweren, hebben de leerlingen die in jaargroep 5 hadden kunnen zitten, maar in jaargroep 3 of 4 zijn geplaatst, wel aan de mediopeiling deelgenomen. Bij deze kinderen is echter slechts één toets afgenomen (basisautomatismen optellen en aftrekken). Gezien hun achterstand werd het niet zinvol geoordeeld deze kinderen ook aan de andere toetsen te laten deelnemen. Om deze reden is een definitie in termen van leeftijd in deze peiling niet goed mogelijk.

Ontvangen 27-9-1989

Boekbesprekingen

A. J. M. de Jong, *Kennis en het oplossen van vakinhoudelijke problemen (een voorbeeld uit een natuurkundig domein)*. Eindhoven: Academisch proefschrift, 1986 (205 pag.).

De kern van het theoretisch uitgangspunt in dit onderzoek is gelegen in de cognitieve psychologie: de mens interpreteert en manipuleert zijn omgeving vanuit zijn kennis. De centrale vraag in deze studie is: welke kenmerken van vakinhoudelijke kennis hangen samen met succes bij het oplossen van deze problemen.

Daarbij ligt de nadruk op de empirische en niet de rationele analyse van de vakinhoudelijke kennis, die nodig is voor het oplossen van een bepaalde klasse van problemen. Daartoe maakt De Jong in een aantal experimenten vergelijkingen tussen foute en goede oplossingen en analyseert verschillen in kennis tussen zwakke en goede beginners. Het veld waarin De Jong zijn onderzoek verrichtte, betreft het oplossen van natuurkundige opgaven door beginnende studenten in het technisch wetenschappelijk onderwijs.

Tegenover de breedte van de term Probleem Oplossen geeft De Jong al direct aan duidelijke inperking namelijk:

1. Het gaat hem om het oplossen van vakinhoudelijke, en wel natuurkundige, opgaven.
2. Er wordt gekeken naar de invloed hierop van slechts één factor: de vakinhoudelijke kennis.
3. Het onderzoek betreft een vrij specifieke groep personen n.l. beginnende studenten aan technische universiteiten.

De dissertatie bevat eerst een historisch overzicht van de ontwikkelingen en theorievorming over Probleem Oplossen. Daarna volgen drie hoofdstukken met elk een verslag van een empirisch onderzoek gericht op een aspect van de kennis van beginnende studenten.

Tenslotte wordt een integratie van de resultaten van het onderzoek gegeven en worden mogelijke implicaties voor de praktijk behandeld. Ik zal eerst de verschillende hoofdstukken schetsen en tenslotte een kort commentaar op het geheel geven.

Een kernpunt in de theoretische verantwoording betreft de verschillende soorten vakinhoudelijke kennis die te onderscheiden zijn bij het oplossen van natuurkundige problemen. In het kennisrepertoire onderscheidt De Jong:

- kennis van probleemsituaties
- declaratieve kennis
- procedurele kennis
- strategische kennis.

Deze verschillende soorten kennis spelen in op de verschillende fasen van de probleemrepresentatie, die een centrale plaats inneemt in het oplossingsproces en de informatiestromen tijdens dat proces. Kennis van probleemtransformaties is vooral van belang en te onderscheiden van declaratieve kennis, omdat de eerste een rol speelt bij de selectie uit het geheugen van informatie die voor de oplossing van belang is.

Declaratieve kennis, ook wel conceptuele, propositionele of descriptieve kennis genoemd, is kennis van feiten en principes uit een bepaald vakgebied. De kwaliteiten daarvan, zoals interne integratie resp. verbondenheid, zijn vooral van belang in de tweede fase van het

oplossingsproces wanneer informatie over de oplossing aan de probleemrepresentatie moet worden toegevoegd.

Kennis van acties, manipulaties of handelingen die in een specifiek vakgebied zijn toegestaan worden aangeduid met procedurele kennis. Deze kan algoritmisch of heuristisch van aard zijn.

Strategische kennis heeft betrekking op het gehele oplossingsproces en is bovendien de minst vakinhoudelijke van de vier kenniscomponenten.

De organisatie van de vakinhoudelijke kennis is een tweede kernpunt in de theoretische verantwoording en de empirische studies van De Jong. Naast de beschikbaarheid baseert hij zich op 'chunking' en de schematheorie. Onderzoek naar verschillen tussen experts en beginners is in dit theoretisch kader goed te plaatsen, omdat verschillen in procedurele kennis, strategische kennis en in de organisatie van kennis daarin aan te geven zijn. De relativisering van het expert beginner-paradigma ('Wie is een expert en wie is een beginner') neemt De Jong serieus door ook in te gaan op de verschillen tussen beginners.

De onderzoeksvragen vloeien voort uit de kernpunten van het theoretische hoofdstuk. In het eerste experiment, waarbij een experimentele strategie werd aangeleerd ontving de experimentele groep een uur instructie per week gedurende zes weken. Deze instructie begon nadat alle studenten de toets voor het tweede blok uit de cursus hadden afgelegd. De instructie werd gegeven door een ervaren docent van de afdeling Natuurkunde van de TUE, waarbij de nadruk werd gelegd op de systematiek van de oplossingsstrategie en het niet de bedoeling was expliciet specifieke denkhandelingen aan te brengen.

De effectmetingen waren gericht op de vraag: gaan de studenten die de instructie volgen meer conform de aangeboden strategie probleemoplossen? De data werden verkregen door de methode van hardop-denken en het analyseren van protocollen.

De conclusie is dat de studenten er niet toe gebracht zijn volgens de experimentele werkwijze te gaan werken. De oorzaak van dit resultaat zoekt De Jong mijns inziens terecht in het feit dat de instructie kort is geweest en dat de strategie te zeer indruist tegen wat studenten gewend zijn te doen. Ook op grond van de theoretische uitgangspunten kan veilig geconcludeerd worden dat zes uur bijzonder weinig is om een dergelijke strategie aan te leren, zeker als die buiten het regulier onderwijs wordt gegeven. Bovendien bleken een aantal studenten niet alle bijeenkomsten bijgewoond te hebben, werd om experimentele redenen geen schriftelijk studiemateriaal verstrekt en was de instructie ook weinig indringend. Tenslotte werd bij de toets die meetelde voor de studieresultaten het volgen van de strategie niet meegeteld bij de score. Ook dit kan studenten er toe gebracht hebben het zekere voor het onzekere te nemen en die strategie te gebruiken die ze gewend waren.

Het tweede experiment betreft de hypothese, dat goede studenten hun kennis meer volgens (adequate) probleemschemata hebben georganiseerd dan studenten die slecht presteren. Om dit te onderzoeken werd de inhoud van het vak onderverdeeld in 12 probleemttypen. Elk probleemtype bestond uit minstens één kenmerk van een probleemsituatie, één stukje declaratieve kennis en één procedure. De in totaal 65 kenniselementen stonden op kaartjes, die de studenten moesten sorteren in stapeltjes, zodanig dat kaartjes in één stapeltje meer met elkaar te maken hadden dan kaartjes uit verschillende stapeltjes. Er bleek een vrij sterke correlatie tussen de maat voor het georganiseerd zijn van de kennis volgens de probleemschemata en de tentamencijfers van deze studenten voor het vak.

In het derde onderzoek werd de kennis van probleemsituaties van beginnende studenten gemeten. Ze kregen een aantal probleembeschrijvingen uit het vakgebied zeer kort te zien

waarna gevraagd werd een reconstructie te maken van elke probleembeschrijving. In deze situatie moeten studenten voor de reconstructie een beroep doen op hun kennis van probleemsituaties. Belangrijke resultaten uit dit experiment waren dat goede studenten vraagstellingen veel beter reconstrueerden dan zwakke studenten, en dat goede studenten beter dan zwakke studenten hun kennis van probleemsituaties konden koppelen aan declaratieve en procedurele kennis, die voor de oplossing relevant was.

De Jong heeft in zijn onderzoek geprobeerd het beeld van de 'ideale' probleemoplosser te verhelderen, en wel op een cruciaal punt: het verband tussen de inrichting van de kennisbasis en de kwaliteit van het oplossingsproces. Hij heeft daarbij een aantal conclusies kunnen trekken, die zowel van belang zijn voor de theorievorming omtrent het leren oplossen van problemen, als voor de praktische vormgeving van onderwijs. Wat betreft de draagwijdte van de resultaten is vooral van belang dat nu wordt nagegaan of de conclusies ook opgaan voor andere vakinhouden en typen problemen, zoals die waarbij de zoekruimte veel groter is of waarbij het probleemtype een geheel ander is, bijvoorbeeld ontwerpproblemen of verklaringsproblemen.

Wat betreft de praktische implicaties is duidelijk dat voor hoger en voortgezet onderwijs hier ook belangrijke elementen liggen voor het zoeken naar verschillen in bestuderingsprocessen. Een praktische toepassing van de resultaten ligt tenslotte ook op het gebied van de intelligente tutorsystemen. De beschrijving van de cognitieve structuur van goede en zwakke studenten is essentieel voor de ontwikkeling van dit soort computerprogramma's, terwijl tegelijk deze programma's een belangrijke rol kunnen spelen bij verder onderzoek naar het leren oplossen van problemen.

A. Pilot
Universiteit Twente
Onderwijskundig Centrum
Postbus 257
7500 AE Enschede

Johan Zuidema (1988). *Efficiënt spellingonderwijs: een leer-en expertmodel voor het spellen*, 205 p.
(Proefschrift R.U. Utrecht) Leuven: Acco.

Men kan het boek 'Efficiënt Spellingonderwijs' vanuit verschillende gezichtspunten bezien: als een proefschrift ofwel een wetenschappelijke proeve van bekwaamheid; of als een boek waarin Zuidema zijn visie geeft op hoe het spellingonderwijs verbeterd zou moeten worden. Omdat de promotie-commissie de graad van doctor inmiddels heeft verleend aan Johan Zuidema is het niet meer zo relevant het boek als proefschrift te beoordelen. Niet vele promovendi in de Sociale Wetenschappen zal het lukken om te volstaan met één empirisch onderzoek en enkele beschouwende hoofdstukken daarom heen. Vanuit dat gezichtspunt kan Zuidema niet ontevreden zijn en zal de doorsnee lezer een grondiger aanpak in het boek verwachten dan feitelijk aanwezig is.

Op deze plaats is het echter van meer belang te bezien hoe Zuidema zijn idee over efficiënt spellingonderwijs uiteenzet en onderbouwt. Het boek handelt alleen over de werkwoordspelling en alle empirische gegevens en beweringen beperken zich tot de spelling van werkwoordsvormen, een woord dat niet misstaan zou hebben in de titel. Het

boek bevat 5 hoofdstukken en een groot aantal bijlagen (ongeveer 50 pagina's). In het eerste hoofdstuk wordt gewezen op de twee mogelijke manieren om de werkwoordspelling te onderwijzen: de analogie aanpak en de algoritmische aanpak.

In hoofdstuk twee rapporteert Zuidema een onderzoek naar de effectiviteit van deze twee methoden, waarbij hij ook nog de werkvorm varieert: een klassikale en een individuele werkvorm. 20 klassen met in totaal 400 leerlingen werden verdeeld over de 4 condities. Er werd als voortoets een dictee afgenomen (half september '85). Van half september tot half januari volgt de begripsfase. Volgens de analogie methode (vooral Taal voor het leven) en de methode 'De Werkwoordwinkel' wordt wekelijks gedurende anderhalf uur de werkwoordspelling klassikaal geoefend. Vanaf half januari wordt elke groep gesplitst in een groep die klassikaal en een groep die verder individueel onderricht krijgt. Eind april volgt dan de natoets en omstreeks half juni nog een retentie toets, beide in de vorm van een parallel versie van het dictee uit de voortoets. De analyse van de resultaten op het dictee richt zich op de schrijfwijze van de werkwoordsvormen, en wel of het morfologische principe correct is toegepast. Met behulp van covariantie-analyse wordt vervolgens geëvalueerd of de hoofdeffekten (Methode, Werkvorm) significant zijn en of er significante interacties zijn. Uit deze eerste analyse op de 'ruwe data' blijkt dat kinderen die de regelmethode hebben gehad beter presteren, maar dat dit vooral geldt voor de kinderen die klassikaal hebben gewerkt.

Hoofdstuk 3 is getiteld 'verbetering van de diagnostiek' maar betreft eigenlijk een heranalyse van de data uit hoofdstuk 2. Zuidema komt er achter dat leerlingen bij het schrijven van werkwoordsvormen niet 'random' presteren wanneer ze de regels voor werkwoordsvormen niet kennen, maar dat ze daarbij gebruik maken van de spellingsregels die ze al wel kennen. De werkwoordsvormen worden in twee groepen behandeld: de midden-problemen waarbij het erom gaat of een vorm met enkele of dubbele d of t wordt geschreven, en de eind-problemen waarbij het erom gaat of een vorm eindigt op d, t of dt. Voor een kind dat het fonologisch principe beheerst, vormt het enkel-probleem geen moeilijkheid: hij zal die vorm goed schrijven, zij het om de verkeerde reden. Dit betekent dat het nodig is juist de enkel- en dubbel-problemen tegen elkaar af te zetten om te beoordelen of midden-problemen vanwege de juiste redenen goed worden geschreven. Een soortgelijke kwestie doet zich voor bij de eind-problemen. Enerzijds is het interessant te zien hoe Zuidema zich gaat verdiepen in de strategieën van de speller bij het oplossen van spellingsproblemen. Spelling is niet meer een door taalkundigen geformuleerde verzameling van regels en wetmatigheden (met de altijd aanwezige uitzonderingen) maar een conceptuele leertaak geworden, waarbij de nodige psychologie om de hoek komt kijken. Anderzijds is het opmerkelijk dat Zuidema een heel hoofdstuk nodig heeft om duidelijk te maken dat strategie-effecten een rol kunnen spelen en hoe daarvoor met formules gecorrigeerd kan worden: bij reactietijdexperimenten (maar ook bij veel andere taken) is het al lang bekend hoe je met 'hits' en 'false alarms' degelijke strategie-effecten kunt bestuderen. Bovendien vind ik het voor de lezer ook niet elegant eerst de nodige pagina's analyses te beschrijven om vervolgens te constateren dat deze manier van rekenen eigenlijk niet juist is en de berekeningen nog eens opnieuw te doen.

Eigenlijk is hiermee het promotie onderzoek afgelopen. Hoofdstuk 4 beoogt de verschillende didaktische methoden, waarin de werkwoordspelling op een min of meer algoritmische wijze wordt onderricht, te evalueren. Tegelijkertijd wordt echter een expertmodel beschreven, zonder dat er een duidelijke relatie is met de discussie over de didaktische methoden. Bij de evaluatie van deze methoden hanteert Zuidema drie criteria: 1) de complexiteit van het algoritme, 2) de psychologische realiteit, 3) voorbereiding op een goede denkstrategie ofwel: als een leerling volgens een bepaalde methode de spelling heeft geleerd, in hoeverre schrijft zo'n leerling dan volgens een 'expert' (die dan onafhankelijk

van een bepaalde leermethode werkt?). Dan volgt een ingewikkelde beschrijving van een expert, die vol actieve filters blijkt te zitten op alle niveaus van taalverwerking. Helaas wordt nauwelijks aannemelijk gemaakt dat juist dit expertmodel een afspiegeling is van de (automatische en gecontroleerde) cognitive processen die zich afspelen in het hoofd van een ervaren schrijver. Wanneer Zuidema vervolgens de drie criteria gaat toepassen op het werk van zijn collega's maakt hij het zichzelf wel erg gemakkelijk. Hoe bepaal je nu of zo'n methode al of niet complex is, of, nog moeilijker, hoe psychologisch reëel? De vraag of leerlingen bij een bepaalde methode wel of niet volgens een expert zullen denken komt eigenlijk helemaal niet aan de orde. In feite komt het er op neer dat Zuidema redelijk vrijblijvend een oordeel velt. Als expert zal hij zeker geen onredelijk oordeel vellen, maar het blijft een subjectief oordeel.

In hoofdstuk 5 komt Zuidema met een andere oplossing van het didactisch probleem van de werkwoordspelling: een spellinghervorming. Uit allerlei opmerkingen wordt duidelijk dat Zuidema er zich van bewust is dat spellinghervormingen nu eenmaal niet of nauwelijks mogen in dit land en dat zijn voorstel het ook wel niet zal redden. Het zal de lezer langzaam duidelijk zijn geworden: Zuidema is een voorstander van grote regelmaat en het gebruik van regels. In mijn ogen interpreteert Zuidema regelgebruik echter te veel als bewust regelgebruik. Hij is er zich terdege van bewust dat het gebruik van regels geïnternaliseerd en geautomatiseerd kan zijn. Hij is ook op de hoogte van onderzoek waaruit blijkt dat subjecten regels kunnen leren te gebruiken zonder dat ze zich hiervan bewust zijn. Desalniettemin heeft Zuidema weinig vertrouwen in het cognitieve apparaat, dat toch ook in staat is geweest om de talrijke uitzonderingen bij de produktie en perceptie van woordvormen in de *spraak* te leren?

Daarmee heb ik gelijk twee belangrijke problemen van het boek te pakken: hoewel Zuidema wel degelijk oog heeft voor de psychologische vragen die spelen bij het leren van werkwoordsvormen, behandelt hij het thema niet systematisch en onderzoekt hij het niet. De vraag hoe een kind een complex regelsysteem het beste kan leren komt niet voor het eerst naar voren bij het spellingonderwijs: van andere terreinen zoals taalontwikkeling en leren lezen zou relevante informatie kunnen worden ontleend. Dergelijke mogelijkheden worden echter niet onderzocht of vermeld. Zuidema verzuimt zelfs een relatie te leggen tussen het leren van midden-problemen en eind-problemen in zijn eigen onderzoek! Wellicht wijst Zuidema de juiste richting aan voor het spellingonderwijs, hij levert echter weining materiaal aan op basis waarvan wetenschappelijk verantwoord spellingonderwijs gegeven kan worden.

Paul Eling
K. U. Nijmegen
Vakgroep Vergelijkende en
Fysiologische Psychologie
Postbus 9104
6500 HE Nijmegen

W. Gijselaers (1988), *Kwaliteit van het onderwijs gemeten*.
(Proefschrift Rijksuniversiteit Limburg). Maastricht. 215 pagina's.

Kwaliteitszorg en in het verlengde daarvan kwaliteitsbepaling is in het hoger onderwijs een 'hot topic'. Ministeriële ideeën die een grotere verantwoordelijkheid toedenken aan de

instellingen van hoger onderwijs gaan hand in hand met opvattingen over controle van de kwaliteit van het onderwijs door de overheid achteraf. Deze extern geïnspireerde kwaliteitsbewaking heeft geleid tot een toegenomen belangstelling bij de instellingen om ook de interne kwaliteitsbewaking systematisch aan te pakken.

In (vrijwel) alle systemen voor interne kwaliteitsbewaking spelen studentenoordelen een rol. Het proefschrift van Gijselaers over de waarde van studentenoordelen bij de programma-evaluatie aan de Rijksuniversiteit Limburg (RL) komt daarom op een gelukkig moment.

In het proefschrift kunnen drie delen onderscheiden worden. Het eerste deel omvat een uiteenzetting over onderwijsevaluatie in het algemeen en een beschrijving van het evaluatiesysteem bij de medische faculteit van de RL. In het tweede deel wordt ingegaan op de theorieën over en empirisch onderzoek naar de betrouwbaarheid en validiteit van studentenoordelen. In het laatste deel wordt een aantal onderzoeken beschreven die de auteur heeft uitgevoerd. Doel van deze onderzoeken was de constructie van een vragenlijst voor studenten die zou kunnen bijdragen aan optimalisering van de kwaliteit van de onderwijsblokken.

Uitgangspunt bij Gijselaers is, dat de evaluatiebenadering dient samen te hangen met de onderwijsopvatting die gehuldigd wordt en de organisatievorm die daarop gestoeld is. Deze samenhang is terug te vinden bij de medische faculteit van de RL, waar de rationalistische organisatievorm gepaard gaat met een evaluatiesysteem dat met name ten doel heeft de kwaliteit van het onderwijs beheersbaarder te maken, voor de verschillende facultaire bestuursorganen. Mede daarom is een vragenlijst ontwikkeld die aansluit op de specifieke onderwijsopzet aan de RL (Probleem Gestuurd Onderwijs). Deze vragenlijst wordt na elk blok van zes weken bij alle studenten afgenomen.

In hoofdstuk 3 wordt ingegaan op problemen bij en mogelijkheden voor het vaststellen van de betrouwbaarheid en validiteit van beoordelvingsvragenlijsten. Hij sluit daarbij aan bij de klassieke testtheorie. Dit hoofdstuk wordt afgerond met een overzicht van onderzoeksresultaten naar de betrouwbaarheid en validiteit van studentenoordelen. Gijselaers concludeert dat de betrouwbaarheid en validiteit van studentenoordelen van dien aard is dat ze gebruikt kunnen worden voor het beoordelen van de kwaliteit van het onderwijs.

In hoofdstuk 4 wordt de geconstrueerde vragenlijst onderzocht op multidimensionaliteit en schaalhomogeniteit, betrouwbaarheid van de (groeps)oordelen en validiteit. Bij deze onderzoeken wordt aangesloten bij de modellen die in hoofdstuk 3 onderscheiden zijn.

In hoofdstuk 5 tenslotte, wordt beschreven hoe de resultaten gebruikt zijn bij pogingen om het onderwijs te verbeteren en tot welke resultaten dat geleid heeft.

Het sterkste punt van 'Kwaliteit van het onderwijs gemeten' is de systematische, consistente manier waarop gewerkt is aan een instrument dat niet alleen de kwaliteit van het onderwijs kan meten, maar ook gegevens oplevert die bruikbaar zijn voor de verbetering van het onderwijs. Daarmee heeft dit proefschrift ook praktische relevantie. Niet zozeer omdat de vragenlijst zonder meer ook in andere situaties bruikbaar zou zijn, maar omdat in andere situaties dezelfde systematiek gehanteerd kan worden.

Vanwege deze consistentie – en de doorgaans goede leesbaarheid – zijn enkele kleine onnauwkeurigheden in het proefschrift niet echt storend. Zo worden op verschillende plaatsen conclusies getrokken uit (deel)onderzoeken die vervolgens per studie worden samengevat en daarna weer per hoofdstuk. Deze conclusies hebben soms de neiging geleidelijk aan iets positiever of stelliger te worden.

Onduidelijk is de basis waarop criteriumvaliditeit bepaald wordt.¹ De eis die Cohen (1981) aan dergelijk onderzoek stelt, is dat onderwijsinhoud en toets constant blijven, zodat alleen de docent als variabele functioneert. Gijselaers presenteert echter jaargegevens. In die jaargegevens zijn resultaten over verschillende blokken opgenomen. Onduidelijk is nu of deze jaargegevens gebaseerd zijn op gemiddelde correlaties van de verschillende blokken in zo'n

jaar of dat de correlaties over alle groepen, ongeacht de blokken bepaald zijn. In dat laatste geval voldoet het onderzoek niet geheel aan de eisen die Cohen hieraan stelt. Bovendien is vrij eenvoudig na te gaan dat bij analyse per blok totaal andere resultaten verkregen kunnen worden.

Ondanks deze kanttekeningen is het proefschrift alleszins de moeite waard voor degenen die zich bezighouden met kwaliteitsbepaling in het onderwijs.

VERWIJZING

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.

¹ Met dank aan T. Borgesius die mij op dit punt opmerkzaam maakte.

J. Willems, K U Nijmegen, IOWO,

Montessorillan3, Nijmegen

J. F. M. J. Van Hout (1988), *Onderzoekers in opleiding, een verklaringsmodel voor problemen van promotie assistenten en assistenten in opleiding*.

(Proefschrift K.U. Nijmegen). Nijmegen: Instituut voor Onderzoek van het Wetenschappelijk Onderwijs, 265 pp.

Op 15 september 1988 promoveerde J. F. M. J. Van Hout aan de Katholieke Universiteit te Nijmegen op het proefschrift 'ONDERZOEKERS IN OPLEIDING, een verklaringsmodel voor problemen van promotie-assistenten en assistenten in opleiding'.

Deze dissertatie heeft als theoretische doelstelling een bijdrage te leveren aan het onderwijskundig onderzoek van het hoger onderwijs en als praktische doelstelling de problemen, die assistenten bij het maken van een proefschrift ervaren, te verminderen.

De bekommernis met de situatie van promovendi uit de gamma-wetenschappen is het hoofdmotief van de auteur. Hij is met de assistenten begaan. Dit blijkt niet alleen uit de opzet van zijn onderzoek, maar ook uit de bij het proefschrift behorende stellingen, waarin de volgende kwesties worden gehekel: de begeleiding van aio's, de voortgangsregistratie van aio's, de implementatie van het aio-stelsel, de onterecht gelijke behandeling van aio's uit de sociale wetenschappen en letteren enerzijds en de beta-wetenschappen anderzijds en het 'ge'bruik van aio's voor de behoefte van de universiteiten aan wetenschappelijke output.

Dr. Van Hout tracht dit niveau van de bekommernis te overstijgen op drie manieren. Ten eerste, hij maakt een historische studie van het beleid in Nederland ten aanzien van het aio-stelsel. Ten tweede, hij bestudeert de onderzoeksliteratuur met betrekking tot de PhD als één van de postgraduate-opleidingen in Noord-Amerika en Engeland, met het oog op vergelijking. Ten derde, hij doet zelf onderzoek naar de door assistenten ervaren problemen in Nederland en hij tracht hiervoor een verklaringsmodel op te stellen en te testen.

In het Nederlandse beleid van 1968 tot heden worden de voorstellen van Posthumus, de Wet Herstructurering Wetenschappelijk Onderwijs, de Nota Onderwijs voor Velen, de Wet Twee-fasenstructuur, de Beleidsnota Beiaard en de vorming van het aio-stelsel besproken. Uit deze

wetgeving blijkt dat een langdurige discussie aanleiding gaf tot de keuze voor een 'ongedeelde structuur' en een 'opleidingsmodel'. Een ongedeelde in tegenstelling tot een tweedelige structuur, omdat de assistent in opleiding in Nederland zowel student als medewerker-onderzoeker is. Een opleidingsmodel in tegenstelling tot een 'learning by doing'-model, omdat aan de assistent in opleiding een curriculum wordt aangeboden en een opleidings- en begeleidingsplan wordt vastgelegd.

De onderzoeksliteratuur over de 'post-graduate'-opleidingen en meer bepaald de PhD-opleiding van de VS en Engeland wordt bestudeerd, ten eerste omdat onderzoeken in Nederland op dit gebied ontbreken en ten tweede omdat de wetgeving in Nederland voor het postgraduate-PhD-model heeft gekozen, zodat vergelijking mogelijk is. Kenmerken van de PhD-studenten, zoals de leeftijd, de sekse, de motieven tot promoveren en de mate van sociale en intellectuele integratie, alsook problemen bij de PhD-opleiding, zoals de positie, de sociale contacten, de beroepsidentiteit, de financiële situatie, de keuze van het onderzoeksonderwerp, de rapportage, de begeleiding, de werkomgeving en de onderzoeksfaciliteiten, worden uit deze onderzoeksliteratuur gedistilleerd.

Deze Amerikaanse en Engelse voorbeelden dienen als leidraad voor de onderzoeksopzet, waarin problemen van de assistenten-promovendi worden geïnventariseerd en waarin een model ter verklaring van deze problemen wordt opgezet en getest.

Door 166 assistenten uit drie faculteiten (alpha: letteren, beta: wiskunde en natuurkunde, gamma: sociale wetenschappen) in zes Nederlandse universiteiten (waar deze drie faculteiten vertegenwoordigd zijn) is een vragenlijst ingevuld met vragen over de door hen ervaren problemen, over hun individuele kenmerken en over kenmerken van hun werkomgeving. Hieronder volgt een beperkte greep uit de resultaten.

De assistenten variëren sterk in leeftijd. Zij zijn gemiddeld rond de dertig. Zij zijn overwegend van het mannelijk geslacht. De verschillen tussen de drie faculteiten zijn opmerkelijk. De assistenten uit de alpha-richting hebben de sterkste onderwijsoriëntatie en het sterkste universitair wetenschapsmotief. Zij zijn het meest ontevreden over hun positie. Die van de beta-richting zijn het jongst en het meest van het mannelijk geslacht. Zij hebben de zwakste onderwijsoriëntatie, werken het minst zelfstandig en hebben de kleinste onderwijs taak. Zij zijn het meest tevreden met hun positie. Assistenten van de gamma-richting zijn het oudste en de proportie vrouwen is er het hoogst. Zij werken het minst planmatig en hebben de grootste onderwijstaak. Zij hebben de meeste onderwijszorgen en -spanningen.

Alle variabelen worden door dr. Van Hout in één globaal causaal model ondergebracht, dat via de Lisrel-procedure aan een statistische test wordt onderworpen. Zijn twee belangrijkste bevindingen zijn, ten eerste het significante verschil tussen de drie faculteiten en ten tweede, de bevinding dat onderzoeksproblemen en -spanningen aanleiding geven tot onderwijsproblemen en niet omgekeerd.

Ook al is dit onderzoek gedegen van opzet en uitwerking, toch moeten op het punt van de bijdrage aan het onderwijskundig onderzoek en de beleidsrelevantie een aantal kanttekeningen worden geplaatst. Met een response van $166 = 52.2\%$ van 318 vragenlijsten is de steekproef nogal klein uitgevallen. De selectie van promotie-assistenten, die in dienst zijn getreden tussen juli 1979 en juli 1983 en die dus geen assistenten in opleiding of onderzoekers in opleiding kunnen zijn, omdat het aio-stelsel pas in 1986 werd ingevoerd, maakt de beleidsrelevantie geringer dan in de titel van het werk wordt gesuggereerd. Dit geldt ook voor de vergelijkbaarheid met PhD's.

In het proefschrift worden hypothesen en uitgangspunten geformuleerd leidend tot een Lisrel-model dat werd getoetst en niet adequaat bevonden. Daarna is het grote aantal

variabelen samengevoegd, met de bedoeling tot één overzichtelijk causaal model te komen. Dit laatste bleek dan wel adequaat te zijn. Eén en ander heeft veel weg van een ad-hoc strategie ter verkrijging van een geslaagde Lisrel-analyse bij een gering aantal steekprofeenheden. Het is ook jammer maar wellicht onvermijdelijk dat uiteindelijk slechts globaliserende, weinig informatieve uitspraken overblijven, zoals: de invloed van individuele kenmerken loopt via achtergrondkenmerken, de oriëntatie van de assistent heeft invloed op de onderwijsproblemen en de discipline heeft invloed op alle individuele kenmerken.

De auteur besluit dan ook zelf dat zijn verklaringsmodel als voorlopig moet worden gezien en dat nog veel verder onderzoek noodzakelijk is.

J. J. A. Tacq
Erasmus Universiteit Rotterdam RISBO
Postbus 1738
3000 DR Rotterdam

J. A. de Jong, *'Je wordt aan het denken gezet'. Ontwikkeling en evaluatie van een model voor de schoolstage van de PABO.*

Publikatie ISOR - Afdeling Onderwijsonderzoek, diss. Rijksuniversiteit Utrecht 1988 (ISBN 90.5187.002.7).

Zoals de ondertitel aangeeft heeft deze dissertatie betrekking op de ontwikkeling en evaluatie van een schoolstagemodel voor de opleiding van leraren voor het basisonderwijs. Het boek bestaat uit vier delen. In deel I wordt een verantwoording gegeven van het onderzoek, dat uitgevoerd werd op één bepaalde pedagogische academie.

In deel II worden onder de titel Diagnose en Ontwerp naast een literatuurstudie over de schoolstage twee deelonderzoeken besproken: een discrepantie-analyse en een vergelijkende gevalsstudie, die beide uitgevoerd werden t.b.v. de analyse van de beginsituatie. Op grond van documentenanalyse en interviews met een zestal studenten en de coördinator van hun begeleidingsteam werd nagegaan welke discrepanties aan te wijzen waren tussen intenties enerzijds en uitvoering en beleving van het op de betrokken pedagogische academie gehanteerde stagemodel anderzijds. In het tweede deelonderzoek was de vraag op welke wijze de manier van leren van een drietal stagiaires en de manier van begeleiden door hun begeleiders interacteerden. De protocollen die resulteerden uit de interviews en observaties werden geanalyseerd door middel van een vorm van patronenanalyse, daarnaast werd gebruik gemaakt van procesbeschrijvingen.

In deel III wordt allereerst verslag gedaan van de evaluatie van de eerste versie van het ontwikkelde stagemodel, waarbij een kleinschalig discrepantie-onderzoek gedaan werd naar de realisering van de intenties van het model bij een tweetal studenten. Daarbij werd gebruik gemaakt van interviews, observaties en een evaluatielijst.

De auteur baseert zich bij het ontwikkelde stagemodel op de zgn. inquiry-oriented benadering. De schoolstage wordt beschouwd als terrein waar de student het onderwijs en de eigen mogelijkheden kan exploreren. Het ontwikkelde model wordt dan ook aangeduid met de term studiestagemodel. In dit model gaat het om de ontwikkeling van adequate handelingsoriëntaties, d.w.z. 'waarden, aandachtspunten, inzichten en gedragsschema's die ten grondslag liggen aan datgene wat een actor (i.c. een leraar) doet' (p. 69). De stageperiode is in het model ingedeeld in een aantal fasen: oriëntatie- en planningsfase, uitvoeringsfase en evaluatiefase. In de planningsfase formuleert de student leerdoelen en

leervragen (b.v. wat komt er kijken bij het lesgeven aan een combinatieklas) voor zichzelf in overleg met de mentor van de stageschool en de begeleidende opleidingsdocent. In de uitvoeringsfase voert de student naast de lessen die hij geeft, verkennende onderzoekjes uit. Via voorgangsverslagen en supervisiegesprekken wordt de ontwikkeling gevolgd en begeleid. Een belangrijk element daarbij is ook het inzicht in de eigen manier van leren. Het geheel wordt afgesloten door een evaluatiegesprek waarin het gaat om het inzicht in de ontwikkelde handelingsoriëntaties.

In het vervolg van deel III wordt verslag gedaan van een viertal deelonderzoeken die de evaluatie beoogden van de bijgestelde versie van het studiestagemodel. Deze vier deelonderzoeken waren:

- een discrepantie-onderzoek naar de realisering van de intenties van het studiestagemodel bij een klas van zeventien studenten.
- een onderzoek naar de leereffecten van deze studenten in vergelijking met zeventien studenten die volgens het gangbare model stage liepen.
- een correlatieel onderzoek naar het verband tussen de realisering van genoemde intenties en leereffecten.
- een vergelijkende gevalstudie bij zes van de zeventien studenten naar factoren die van invloed waren op de leereffecten.

Naast de toepassing van evaluatielijsten, interviews, observaties en leerverslagen werd voor de effectmeting gebruik gemaakt van een test om de manier van leren in de stage vast te stellen met de bedoeling om een eventuele ontwikkeling naar een meer onafhankelijke en meer onderzoekende leerhouding te kunnen registreren, en een voornemenslijst voor het meten van inzicht in de eigen leerweg. Op deze effectmaten kon geen verschil in effect tussen de twee groepen worden aangetoond. Dat gold ook voor de leerverslagen waarmee nagegaan werd in hoeverre studenten in staat waren om te verwoorden welke handelingsoriëntaties zij verworven hadden. Wel bleek verband aanwezig tussen de *kwaliteit* van het realiseren van het studiestagemodel en een tweetal beoogde leereffecten (toename onderzoekende houding en verwoording handelingsoriëntaties).

In deel IV tenslotte vindt een nadere theoretische verdieping plaats.

De auteur typeert zijn onderzoek als ontwikkelingsonderzoek. Hoewel de aantallen personen waarop het onderzoek betrekking heeft, relatief klein zijn, is toch sprake van een omvangrijk onderzoek gezien het aantal deelonderzoeken - in totaal zeven - dat werd uitgevoerd. Te waarden valt dat deze deelonderzoeken alle uitvoerig verantwoord en gerapporteerd zijn. De onderzoeker heeft getracht de ontwikkeling van het studiestagemodel onderzoeksmatig zorgvuldig te ondersteunen. In het bijzonder is een positief punt dat - al was dat wellicht wat praematuur - gezocht is naar leereffecten die toe te schrijven zouden zijn aan het studiestagemodel, daarbij gebruik makend van een controlegroep en (gedeeltelijk) van voor- en nametingen.

Een probleem bij het soort vergelijkend veldonderzoek dat de auteur heeft uitgevoerd, is de complexiteit van de treatment en het grote aantal factoren dat niet of onvoldoende onder controle te houden is en ruis kan veroorzaken. Het verbaast eigenlijk niet dat geen verschillen gevonden werden. De vraag is of een dergelijk onderzoek met zo'n brede en complexe treatment niet een meer verfijnde methodische opzet vereist.

Ook de onderzoeker zelf levert kritiek op dit deelonderzoek. Dat geldt trouwens algemeen: de auteur stelt zich zeer behoedzaam op en relativeert nadrukkelijk. Dat staat enigszins in contrast met de bijdrage die de auteur meent te leveren aan theorieontwikkeling. Eigenlijk valt het proefschrift in twee parten uiteen. De delen I, II en III hebben betrekking op het uitgevoerde ontwikkelingsonderzoek. Daaraan cq. aan het ontwikkelde studiestagemodel ligt een aantal 'theoretische' uitgangspunten ten grondslag, die vooraf weinig onderbouwd worden. In deel IV wordt hoofdzakelijk verslag gedaan van een

literatuuronderzoek. De verbinding tussen de drie eerste delen en dit - op zichzelf goede - literatuurverslag is zwak, in feite gaat het om twee losse bijdragen. Het ontwikkelingsonderzoek als zodanig vormt naar mijn mening geen duidelijke bijdrage tot theorieontwikkeling. Dat is jammer, m.i. is het zeker mogelijk ontwikkelingsonderzoek zodanig op te zetten dat tevens een theoretische opbrengst verzekerd is.

Deze opmerking doet overigens niets af aan de waardering voor het bewijs van meesterschap dat de schrijver met zijn proefschrift geleverd heeft.

F. K. Kievit
Vakgroep Onderwijsstudies
Wassenaarseweg 52
2333 AK Leiden

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. 13, nr. 5/6

1989

Inhoudsopgave

VOR-Bulletin, Jrg. 13, nr. 5/6

Verslag VOR-ledenvergadering d.d. 30-11-1989	372
VOR-begroting 1990	375
Toelichting op de VOR-begroting 1990	376
Congressen, Symposia, Cursussen, Trainingen	377
Ontvangen publikaties	382

Verslag VOR- ledenvergadering d.d. 30 november 1989

Aanwezig: de leden Creemers, Hoebe, Kooistra, v.d. Vleugel en Stijnen
de bestuursleden Beintema, Claessen, v. Hout, Kanselaar, Roede en v.d. Werf

Afwezig m.k.: Beek, Gerritsen v.d. Hoop, De Glopper, De Gruyter, Simons, Tillema en Warries

1. Mededelingen

Creemers merkt bij de agenda-vaststelling ten aanzien van punt 4 (nieuwe bestuurssamenstelling) op dat nieuwe bestuursleden in functie dienen te worden voorgedragen. De bestuurssamenstelling moet dus in twee stappen verlopen: eerst worden nieuwe leden voorgedragen, daarna volgt de voordracht voor functies. In de convocatie moet voor de leden zichtbaar zijn wie in welke functie wordt voorgesteld. Dat was traditie. Hoebe voegt hieraan toe dat het bestuur jaarlijks aftreedt. De leden moeten jaarlijks het volledige bestuur benoemen.

Het bestuur erkent dat hier een procedurefout is gemaakt en zegt toe volgend jaar weer volgens het huishoudelijk reglement het bestuur voor te dragen.

2. Verslag vorige ledenvergadering

Geen opmerkingen.

3. Begroting 1990 (deze is met toelichting opgenomen in dit VOR-bulletin).

Hoebe merkt op dat in de uitgaven de voorgenomen nieuwe activiteiten van de VOR (zie punt 5 en 6) niet zichtbaar zijn. Het bestuur meent dat de vergadering eerst akkoord moet gaan met de voorstellen. Vervolgens moeten aan de uitgavenkant kosten voor deze activiteiten in de begroting worden opgenomen. Daarvoor moet eerst financiering worden gezocht. In de begroting is geen ruimte meer.

Creemers merkt ten aanzien van de begroting verder op dat gezocht moet worden naar mogelijkheden om het aantal IJER abonnementen te verhogen. Dit zal bij agendapunt 8 worden behandeld. Ten slotte vraagt Creemers zich af of SVO iets terug wil voor de subsidie die zij verstrekt aan TOR. Voorlopig gaat het bestuur ervan uit dat SVO deze subsidie verstrekt enkel onder de voorwaarde dat TOR bereid is beoordelingen van SVO onderzoek op te nemen.

Hierna wordt de begroting vastgesteld.

4. Nieuwe bestuurssamenstelling

Het bestuur heeft voor de aftredende bestuursleden (Claessen en Roede) zelf nieuwe kandidaten gezocht, rekening houdend met werkkring en spreiding over het land.

Voorgedragen worden Dr. S. Stijnen (Open Universiteit) en Dr. K. de Glopper (SCO). Deze voordrachten krijgen de instemming van de leden. De functies van voorzitter en penningmeester zijn vacant. Als kandidaten voor deze functies worden door het bestuur respectievelijk Kanselaar en De Glopper voorgedragen. Ook met deze voordrachten stemt de ledenvergadering in.

Per 1 januari 1990 is de bestuurssamenstelling nu als volgt:

Voorzitter: Kanselaar

Plv. Voorzitter: v. Hout

Secretaris: v.d. Werf

Penningmeester: De Gloppe

Gewone leden: Beintema, Gerritsen v.d. Hoop, Scheerens en Stijnen

5. Gedragscode

Naar aanleiding van de concept-gedragscode die is rondgestuurd naar de instituten en opvraagbaar was door de leden zijn brieven binnengekomen van het RION, het OCTO, het ITS en van Hoeben.

Het RION heeft meegedeeld in te kunnen stemmen met de code, maar verzoekt het bestuur zo snel mogelijk een Kamer van Toezicht en een Commissie van Beroep te regelen. Verder verzoekt het RION het bestuur de mogelijkheden na te gaan om ook de code voor niet VOR-leden binden te verklaren.

Het OCTO deelt mee zich aan te sluiten bij de reactie van het RION.

Het ITS meent eveneens dat het van belang is dat snel duidelijk wordt hoe een Kamer van Toezicht zal worden samengesteld en wat daarvan de taken en bevoegdheden zijn. Verder merkt het ITS op dat het bindend verklaren van de code voor VOR-leden voor haar niet de betekenis heeft die het RION daaraan geeft.

Van Hoeben is een lijst van amendementen op de code ontvangen. Het bestuur heeft nog geen gelegenheid gehad een reactie op deze brieven voor te bereiden.

Het bestuur stelt daarom de leden voor een open discussie te houden over de hoofdlijnen en haar het mandaat te geven om zaken af te handelen.

Van de Vleugel is van mening dat de definitieve tekst, na verschillende commentaar-rondes onder de leden, moet worden voorgedragen aan de ledenvergadering. Het bestuur is van mening dat dit een te langdurige procedure is. De status van de code is daarom voorlopig. Eerst zal enige tijd (bijv. 2 jaar) met deze voorlopige code worden gewerkt, vervolgens wordt nagegaan wat de problemen zijn en daarna wordt de code eventueel aangepast en definitief vastgesteld. Dan kan ook worden bekeken of de code ook dekkend is voor andere dan in academische settings werkende onderzoekers. Creemers wijst er op dat dan ook al vanaf het begin moet worden gewerkt met een Kamer van Toezicht en beroepsprocedures. Juist die leveren ervaringen op.

Na bespreking en verwerking van de lijst van amendementen van Hoeben wordt de code met amenderingen door de ledenvergadering vastgesteld. Ten aanzien van de verdere procedure wordt besloten dat de code in deze vorm nog niet wordt gepubliceerd maar dat eerst een aantal zaken wordt geregeld rond de Kamer van Toezicht, beroepsprocedure en het bindend verklaren van de code voor VOR en niet VOR-leden. Pas als deze zaken geregeld zijn wordt de code met een reglement gepubliceerd. Er wordt naar gestreefd dit alles rond te hebben op de ledenvergadering tijdens de ORD 1990. Dan wordt ook een voordracht gedaan voor de samenstelling van de colleges van Toezicht en Beroep. Intussen wordt de code voorgelegd aan de Registratiekamer voor de privacy wetgeving. Na 2 jaar werken met de code wordt het geheel geëvalueerd.

In verband met de tijd worden de agendapunten 6 en 7 slechts kort behandeld.

6. Relatie SVO/OTS's/VOR

Het bestuur zelf heeft hierover nog geen definitief standpunt ingenomen. Creemers wil benadrukken dat het bestuur goed voor ogen moet blijven houden dat de VOR een zelfstandige beroepsvereniging is en onafhankelijk van SVO moet blijven. Het gevaar bestaat dat de VOR nu de problemen die SVO heeft veroorzaakt met betrekking tot de

OTG's moet oplossen. De VOR moet echter haar eigen verantwoordelijkheid blijven houden.

7. ORD-bundels

Het is de bedoeling dat na de ORD Nijmegen één bundel wordt uitgegeven met een selectie van de beste papers. De cie themaplanning zal als redactie van deze bundel optreden. Namens de cie themaplanning merkt Creemers op dat de advisering tot op heden niet erg succesvol is verlopen. De zelfstandigheid van de organisatiecommissie in de verschillende plaatsen gaat erg ver.

De voorzitter sluit om 17.45 uur de vergadering met dank aan de aanwezige leden.

VOR BEGROTING 1990

Inkomsten

100	Contributie			
	110 leden			
	120 studenten			
300	Intern. Jour. Ed. Res.			
	310 abonneementen			
	320 garantiesubsidie SVO			
500	SOR-bijdrage VOR-prijs			
700	Inkomsten TOR (Royalties)			
800	SVO-subsidie TOR			
900	diversen			

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.		f 37.600.=	
	leden 470 x 80. =		f 1.325.=	
	stud. 25 x 53. =			
				f 38.925.=
	512 redactiekosten TOR		f 4.000.=	
	513 red.k. VOR-bulletin		f 500.=	
				f 43.425.=
	520 IJER			
	150 ab.à 45. =		f 6.750.=	
	linguistic editing		f 2.500.=	
				f 9.250.=
600	bestuur			f 2.000.=
700	VOR-prijs			f 1.000.=
800	Financiële administr.			f 2.875.=
900	diversen			f 450.=
				f 63.000.=
				=====

Toelichting op de VOR-begroting 1990

De begroting voor 1990 is vrijwel identiek aan die voor 1989. Wijzigingen betreffen de inkomsten voor het IJER, en de "financiële" administratie.

Toelichting per post:

Inkomsten

- 100 Voor 1990 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 1989, n.l. fl. 105,- voor gewone leden en fl. 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 is aan de SVO een garantiesubsidie gevraagd voor 2 jaar van 75 abonnementen à fl. 45,- = fl. 3.375,-. Daarnaast subsidieert de SVO maximaal fl. 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 1990 wordt dit geschat op fl. 1.500,-.
- 800 Ten behoeve van het TOR ontvangt de VOR jaarlijks fl. 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 1990 geschat op f 1000,-.
- 420 Net als in 1989 zal het programmaboek voor de ORD 1990 vóór de ORD aan alle VOR-leden worden toegezonden. De hieraan verbonden kosten worden geschat op maximaal fl. 1.500,-.
- 440 Colloquia zijn weer opgevoerd voor fl. 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 1989.
- 600 De uitgaven t.b.v bestuur zijn hetzelfde begroot als voor 1989. Een belangrijk deel hiervan is bestemd voor reiskosten van enkele bestuursleden.
- 700 De jaarlijks tijdens de ORD toe te kennen VOR-prijs bedraagt maximaal fl. 1.000,- (te betalen door de SOR).
- 800 De uitgaven voor de financiële administratie zijn ten opzichte van 1989 iets verhoogd vanwege de uitbesteding van de financiële administratie naar de SCO voor maximaal fl. 2.000,-.

E. Roede, (penningmeester), oktober 1989

Congressen, Symposia, Cursussen, Trainingen

STUDIEDAG ERNSTIGE LEES- EN SPELLINGSPROBLEMEN.

Op 9 mei 1990 wordt onder voorzitterschap van prof. Dr. A. van der Leij door de sectie Speciale Pedagogiek van de Vrije Universiteit te Amsterdam in samenwerking met de Stichting Dyslexie een studiedag georganiseerd over het thema *Ernstige Lees- en Spellingsproblemen*.

De dag zal in het teken staan van behandeling van leerlingen in de leeftijdsperiode van vier tot twaalf jaar. Sprekers zijn uitgenodigd die o.a. werkzaam zijn in de Vakgroepen Kinderstudies (Utrecht), Orthopedagogiek (Nijmegen), de Sectie Speciale pedagogiek (VU Amsterdam) en het Paedologisch Instituut (Amsterdam).

Aan bod komen:

- neuropsychologisch georiënteerde interventie bij kleuters
- de toepassing van managementtechnieken en speciaal materiaal in de derde en vierde groep basisonderwijs
- de evaluatie van speciale methoden voor het Speciaal Onderwijs
- de toepassing computergestuurde lees- en spellingsprogramma's in Basis- en Speciaal Onderwijs.
- individuele behandeling op basis van informatietheoretische en neuropsychologische theorieën.

Behalve voordrachten zullen er marktpresentaties zijn van de ontwikkelde instrumenten en methoden.

De studiedag is bedoeld voor leerkrachten en leesspecialisten in het Basis- en Speciaal Onderwijs en wordt gehouden in de aula van de Vrije Universiteit.

Belangstellenden wordt verzocht zich voor 1 maart schriftelijk te melden bij het secretariaat van de Sectie Speciale Pedagogiek, Vrije Universiteit, van der Boechorststraat 1, 1081 BT Amsterdam, t.a.v. mw. C. Logger. Zij ontvangen dan een inschrijfformulier, een acceptgirokaart, een programma-overzicht en informatie over tijd en plaats. De kosten, inclusief koffie, thee en lunches, zullen ongeveer f. 75,- bedragen.

OPRICHTING VERENIGING RAIN

Recentelijk is de "Reading Association in the Netherlands" (RAIN) opgericht, een Nederlandse afdeling van de "International Reading Association" (IRA). De vereniging stelt zich o.a. als doel: Het stimuleren van onderzoek naar leesproblemen en leeronderwijs, de verspreiding van informatie op het terrein van leesonderzoek, instrument- en leerplanontwikkeling en beleid op het gebied van lezen en leesonderwijs. Nadere inlichtingen en aanmeldingen bij:

drs. J.H.M.W. Clemens
Hogeschool Midden Nederland
Postbus 14007
3508 SB UTRECHT

"CALL FOR PAPERS" SYMPOSIUM BEGRIJPEND LEZEN

RAIN organiseert in juni 1990 een symposium over "Instructie in begrijpend lezen". Een ieder die geïnteresseerd is op dit symposium een bijdrage te leveren, wordt verzocht een voorstel in te dienen vóór 12 februari 1990.

Nadere informatie over de inhoud van het symposium en richtlijnen voor papers te verkrijgen bij:

drs. A.M.A. Walraven

Paedologisch Instituut

Postbus 303

1115 ZG DUIVENDRECHT

POSTDOCTORALE CURSUS KLASSIEKE MULTIVARIATE ANALYSE

Het Interuniversitair Onderzoeksinstituut voor Psychometrie en Sociometrie organiseert de bovenstaande postdoctorale cursus o.l.v. dr. W.H. van Schuur. De cursus wordt gehouden van maandag 5 t/m vrijdag 9 februari 1990 aan de Rijksuniversiteit Groningen. Informatiebrochure en inschrijfformulier schriftelijk aanvragen bij het IOPS-secretariaat, mw. Drs. J.H. Slotboom, Faculteit der Psychologie, Weesperplein 8, 1018 XA Amsterdam.

Kosten: f. 525,- voor (semi) overheid en f. 1025,- voor commerciële instellingen. Voor werklozen is korting mogelijk.

POSTACADEMISCHE CURSUSSEN METHODEN & TECHNIEKEN VAN SOCIAAL-WETENSCHAPPELIJK ONDERZOEK

Het PAOS - Postacademisch Onderwijs Sociale Wetenschappen - organiseert voor het voorjaar van 1990 de volgende cursussen op het gebied van onderzoeksmethoden en -technieken:

Basiscursus Methoden & Technieken van Sociaal-wetenschappelijk Onderzoek

De cursus bestaat uit twee modules waarvoor men apart kan inschrijven:

1. Onderzoeksopzet en dataverzameling: Onderzoeksstrategieën, databronnen, operationaliseren en meten, dataverzamelingstechnieken (vragenlijst, interview, observatie)
Data: 9, 16, 23 en 30 maart, 6 april

2. Data-analyse en statistiek met gebruikmaking van SPSS

Data: 12 en 20 april, 7 en 25 mei, 1 juni

Docenten: dr. B. Emans, drs. J.P. van der Sande (Rijksuniversiteit Groningen),

dr. J. Dessens, dr. W. Jansen, prof. dr. P.G. Swanborn (Rijksuniversiteit Utrecht),

drs. H. Krijnsen (SPSS)

Kosten: f 1250,- per module, f 2050,- voor beide modules

Kruistabellen en gegeneraliseerde lineaire modellen

Theorie en praktische toepassing van loglineaire analyse en logistische regressie.

Data: 16, 17 en 18 mei, 21 en 22 juni.

Docenten: Dr. J. Dessens, Dr. W. Jansen, Prof. dr. A. Verbeek (Rijksuniversiteit Utrecht)

Kosten: f 1450,-

Inlichtingen: PAOS

Postbus 325

2300 AH Leiden

tel. 071-143141

PROBLEM SOLVING WITH CASES AND ITS RELATIONSHIP TO SIMULATION AND GAMING

We are pleased to announce that the University of Twente, Enschede (the Netherlands), will host the Seventh International Conference on Case Method Research and Case Method Application. Previous conference sites were Lausanne (Switzerland), London (England), Boston (USA). Participants in the past included scholars and practitioners from Western and Eastern Europe, North and South America, the Caribbean, Africa, Asia, Indonesia, Australia, New Zealand and Japan.

The program is structured as follows

I Design and Development

Theory and practice of instructional design curriculum development and using cases and its relationship to simulations and games.

II Application and Implementation

Presentations of applications and implementation of cases in university teaching and training programs (private and public sector).

III Research and Evaluation

Reports on ongoing and completed research on learning and teaching problem solving through the use of cases (and its relationship to simulations and games). Reports on ongoing and completed research on designing, applying and implementing cases.

Organization

The conference is organized by the University of Twente (Educational Center, the School of Management Studies and the Department of Education) in cooperation with the World Association for Case Method Research and Application (WACRA), and the European Network for System Simulation and Management Gaming (The Deutsche Planspiel Zentrale).

Registration

Conference site is the *International Conference and Study Center 'Drienerburgh'* on the campus of the University of Twente, located between the cities of Hengelo and Enschede.

The registration fee of \$ 325 (subject to foreign exchange fluctuations) includes board for three days and a copy of the proceedings. On-campus accommodation is available and lodging costs range from \$ 32 to \$ 45 for single bedrooms and from \$ 54 to \$ 63 for double bedrooms (for one night).

University of Twente, OC	Postbox 217	telefax Domestic 053 - 33 66 95
WACRA'90 Conference	7500 AE Enschede	telefax International 31 - 53 - 35 66 95
Joke Oosterhuis-Geers	The Netherlands	telephone International 31 - 53 - 89 26 27

STUDIUM GENERALE ROTTERDAM

28 februari 1990

Senaatszaal

Burgemeester Oudlaan 50

Woudestein (Kralingen)

Erasmus Universiteit Rotterdam

OP DE GRENS VAN TWEE EEUWEN

Nederland in 2000

We zijn kersvers de jaren negentig ingestapt waardoor het eind van deze eeuw dichterbij komt dan in de jaren tachtig leek. Slechts tien jaar scheidt ons nog van de grens van twee eeuwen. De voor de deur staande eeuwwisseling nodigt uit tot bezinning, terugblikken en vooruitkijken. Wat levert een globale verkenning van onze naaste toekomst op?

Aan het eind van de 20e eeuw dreigen we door een toenemende specialisering, verbrokkeling en verkokering van de menselijke kennis het overzicht over de dingen steeds meer te verliezen. Sommigen signaleren decadentie, verloedering, crisisbesef en cultuurpessimisme. Maar er zijn ook politieke ontwikkelingen in Oost-Europa die een keerpunt lijken aan te geven. Studium Generale Rotterdam organiseert een dag waarin sprekers hun gedachten zullen ontvouwen over toekomstige ontwikkelingen.

PROGRAMMA 28 februari

- | | |
|---------------|---|
| 9.00 - 9.30 | Inschrijven, ontvangst met koffie |
| 9.30 - 9.45 | Welkomstwoord
prof. dr. S. W. Couwenberg, Erasmus Universiteit Rotterdam |
| 9.45 - 10.15 | Godsdienst en levensbeschouwing
prof. dr. G. Dekker, Vrije Universiteit Amsterdam |
| 10.15 - 10.45 | Politieke cultuur en politieke verhoudingen
prof. dr. J. van Putten, Vrije Universiteit Amsterdam |
| 10.45 - 11.15 | Gelegenheid tot stellen van vragen, discussie |
| 11.15 - 11.30 | Koffie |
| 11.30 - 12.00 | Onze parlementaire democratie
prof. dr. M.P.C.M. van Schendelen, Erasmus Universiteit Rotterdam |
| 12.00 - 12.30 | Kan technologie het milieuprobleem oplossen?
mw. dr. J. Cramer, STB - TNO Apeldoorn |
| 12.30 - 13.00 | Gelegenheid tot stellen van vragen, discussie |
| 13.00 - 14.00 | lunch |
| 14.00 - 14.30 | Sociale Stratificatie
prof. dr. J. Berting, Erasmus Universiteit Rotterdam |
| 14.30 - 15.00 | Autochtonen en allochtonen in Nederland
prof. Dr. H.B. Entzinger, Wetenschappelijk Raad voor het
Regeringsbeleid Den Haag |

-
- | | |
|---------------|--|
| 15.00 - 15.30 | Gelegenheid tot stellen van vragen, discussie |
| 15.30 - 15.45 | thee |
| 15.45 - 16.15 | Het Nederlandse Onderwijs; actuele ontwikkelingen in relatie tot veranderingen in de Nederlandse samenleving
prof. dr. J.M.G. Leune, Erasmus Universiteit Rotterdam |
| 16.15 - 16.45 | Gelegenheid tot stellen van vragen, discussie |
| 17.00 - 18.00 | borrel |

Deelname: gratis voor scholieren, studenten en medewerkers van de EUR.

Toegangsprijs voor belangstellenden f 10,-. Betaling uitsluitend aan de zaal.

De gratis toegangskaarten kunnen op vertoon van collegekaart van te voren worden afgehaald bij de bureaus van Studium Generale of aan de balie bij de zaal.

Ontvangen publikaties

- Batenburg, Th.A. van (1988). *Een evaluatie van taalmethoden*. Proefschrift RU Groningen. Groningen: RION (Monografieën onderwijsonderzoek 4).
- Bosman, R. & Louwes, W. (1989). *Gezinssituaties en onderwijskansen*. Een toetsing van verklaringen voor de schoolloopbaanverschillen tussen kinderen uit eenoudergezinnen en tweeoudergezinnen. Nijmegen: ITS (OOMO-reeks).
- Breukelen, G.J.P. van (1989). *Concentration, speed and precision in mental tests: a psychometric approach*. Proefschrift KU Nijmegen. Nijmegen: NICI.
- Brink, F.J. van den (1989). *Realistisch rekenonderwijs aan jonge kinderen*. Proefschrift RU Groningen. Utrecht: Vakgroep onderzoek wiskundeonderwijs en onderwijscomputercentrum.
- Brink, H., Wijnstra, J.M. & Janssens, F.J.G. (1989). *Inventarisatie van het leer- en vormingsaanbod bij het bewegingsonderwijs in de basisschool*. PPON-rapport nr. 1. Arnhem: CITO.
- Brink, H., Wijnstra, J.M. & Janssens, F.J.G. (1989). *Inventarisatie van het leer- en vormingsaanbod bij de beeldende vorming in de basisschool*. PPON-rapport nr. 2. Arnhem: CITO.
- CITO (1989). *Meerjarenplan 1990-1993 en Jaarprogramma 1990*.
- Creemers, B., Peters, T. & Reynolds, D. (Eds.) (1989). *School effectiveness and school improvement: proceedings of the Second International Congress, Rotterdam 1989*. Amsterdam/Lisse: Swets & Zeitlinger. f 72,50.
- Dam, G.T.M. ten (1989). *Vrouwen verschillen in de volwasseneneducatie*. Amersfoort/Leuven: Acco. f 34,50.
- Damme, J. van & Dronkers, J. (1989). *Jongeren in school en beroep (Bijdragen aan de Onderwijsresearch no. 25)*. Amsterdam/Lisse: Swets & Zeitlinger. f 39,90.
- Eldering, L. & Klopogge, J. (eds.) (1989). *Different Cultures Same School: ethnic minority children in Europe*. Amsterdam/Lisse: Swets & Zeitlinger.
- Glas, C.A.W. (1989). *Contributions to estimating and testing Rasch models*. Proefschrift Universiteit Twente.
- Haan, M. de, Klem, B., Vedder, J. & Verheul, I. (1989). *Deeltijd lerarenopleidingen MO nader bekeken*. Utrecht: ISOR.
- Hameyer, U. & Loucks-Horsley, S. (eds.) (1989). *New technologies and school improvement. Support policies and practices*. Leuven/Amersfoort: ACCO (ISIP-technical report). f 36,25.
- HRWB (1989). *Vouchers en educatie voor volwassenen*. Advies nr. 56. Boskoop: Macula.
- Jacobs, C. & Vedder, J. (1989). *Eindtermen hogere kaderopleidingen pedagogiek*. Utrecht: ISOR.
- Kanselaar, G. e.a. (1989). *Computerondersteund vreemde talenonderwijs*. Eindverslag van SVO-project 7108. Utrecht: ISOR.
- Keizer, P. (red.) (1989). *Probleemgestuurd onderwijs in de economische wetenschap*. Assen: Van Gorcum. f 28,-.
- Kessel, N. van (1989). *De DOP gevolgd II*. Evaluatie DOP-regeling. Deelrapport 2: een tussenstand. Nijmegen: ITS.
- Kessel, N. van & Kral, M. (1989). *Verlangd basisonderwijs in de gemeente Weert*. Nijmegen: ITS.
- Moen, J. (1989). *Innoveren in universitair onderwijs*. 's-Gravenhage: VUGA.
- Moust, J.H.C., Bouhuijs, P.A.J. & Schmidt, H.G. (1989). *Probleemgestuurd leren*. Groningen: Wolters-Noordhoff (Hoger onderwijs Reeks). f 16,50.
- Nederlands Instituut van Psychologen (1988). *Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen*. Amsterdam: NIP.
- OCTO, Onderzoek Centrum faculteit der Toegepaste Onderwijskunde. *Jaarverslag 1988*. Enschede: Universiteit Twente.
- RION, Instituut voor Onderwijsonderzoek (1989). *Het RION in 1988. Jaarverslag*. Groningen: RION.
- Roede, E. (1989). *Explaining student investment. An investigation of high school students' retrospective causal accounts of their investment in school*. Amsterdam: SCO.
- Santema, M., Wichers, J.E. & Maandag, D.W. (1989). *Middelbare technische scholen en regionaal opleidingsbeleid*. Groningen: Instituut voor Onderwijskunde.

- Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*. Den Haag; SVO (Balans van Onderwijsonderzoek). f 42,50.
- Scheerens, J. & Verhoeven, J.C. (red.) (1989). *Schoolorganisatie, beleid en onderwijskwaliteit*. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, nr. 21).
- Schings, P.L.J.M. (1989). *Hand in Hand: een handleiding voor samenwerking tussen LBO-scholen en bedrijven*. Rotterdam: COA, Stichting Contactcentrum Onderwijs Arbeid Rijndelta.
- Span, P., De Corte, E. & Van Hout Wolters, B. (red.) (1989). *Onderwijsleerprocessen: Strategieën voor de verwerking van informatie*. Amsterdam/Lisse: Swets & Zeitlinger (Bijdragen aan de Onderwijsresearch, nr. 22).
- SVO (1989). *Jaarverslag 1988/Jaarboek: toegankelijkheid van onderwijsonderzoek*. 's-Gravenhage: SVO.
- Tanghe, J.P. & Houte, M. van den (1989). *De glazen school. Doelgroepenbeleid in het onderwijs*. Leuven/Amersfoort: ACCO. f 34,75.
- Tromp, Th.J.M. (1989). *The acquisition of expertise in computer programming*. Amsterdam: Thesis.
- Tulder, M. van, Veenman, S. & Roelefs, E. (1989). *Nascholing leraren basisonderwijs: vormgeving en resultaten*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs (Selecta-reeks).
- Veen, A. & Voorthuis, M. (1989). *PAOW en arbeidsmarkt. Nieuwe mogelijkheden op de arbeidsmarkt voor pedagogen, andragogen en onderwijskundigen*. Amsterdam: SCO.
- Verhellen, E. & Spiesschaert, F. (eds.) (1989). *Ombudwork for children. A way of improving the position of children in society*. Leuven/Amersfoort: ACCO. f 78,75.
- Verhelst, N.D. & Eggen, T.J.H.M. (1989). *Psychometrische en statistische aspecten van peilingsonderzoek*. PPON-rapport nr. 4. Arnhem: CITO.
- Verhoeven, L. & Vermeer, A. (1989). *Diagnose van kindertaal*. Tilburg: Zwijsen. f 49,-.
- Verloop, N. (1989). *Interactive cognitions of student-teachers: an intervention study*. Proefschrift RU Leiden. Arnhem: CITO.
- Vierssen, D.J. van (1989). *Terugblik op het MAVO-project: een evaluatie van een onderwijsvernieuwing*. Proefschrift KU Nijmegen. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.
- Vries, A.Ph. de (1989). *Een kwart eeuw agrarisch leerlingwezen*. Wageningen: Vakgroep Agrarische Onderwijskunde.
- Vries, B. de (1989). *Opleiden in de sector zorg en welzijn*. Nijmegen: ITS.
- Westerhof, K.J. (1989). *Effectiviteit van leerkrachtgedrag*. Proefschrift RU Groningen.
- Willink, B. (1989). *Arbeidsmarkt en curriculum-economie*. Amsterdam: SCO.
- Zwarts, M.A. & Wijnstra, J.M. (1989). *Inventarisatie van het leer- en vormingsaanbod bij het muziekonderwijs in de basisschool* PPON-rapport nr. 3. Arnhem: CITO.