Utrecht University Library
Partner in Science
Universiteit Utrecht

Tijdschrift voor onderwijsresearch 1993

https://hdl.handle.net/1874/208871

OR BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres
Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen
tel. 045-762293

Jaargang 17

1993

VOR:

Verslag VOR-ledenvergadering 27 mei 1993	252
Lezing ORD'93	255
Congressen, Symposia, Cursussen, Trainingen	318
Onderwijsonderzoekers in Europa verenigen zich	384

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591 X

Jaargang 18
1993

Swets & Zeitlinger B.V.
Lisse

Redactie:

Bert P.M. Creemers (secretaris)
Christiaan Hamaker
Bernadette van Hout Wolters
Leo J.Th. van der Kamp
Joost Lowyck
G. Wim Meijnen
Johan M.M. van der Sanden
Jaap Scheerens
Marinus J.M. Voeten
Wynand H.F.W. Wijnen

VERENIGING VOOR ONDERWIJSRESEARCH

Artikelen

De stabiliteit van het schooleffect in de tijd: een analyse op basis van vijf jaar Eindtoets Basisonderwijs van het Cito. *H. Blok en J.B. Hoeksma* 331

Kennisstructuur en hypothesevorming; verschillen tussen beginners en experts in de geneeskunde. *H.P.A. Boshuizen, P.P.M. Hobus, E.J.F.M. Custers en H.G. Schmidt* 163

Opzet en effecten van een preventieve sociale-vaardigheidstraining voor jonge adolescenten. *J.O. Bijstra, H.P. van der Kooi en H.T. van der Molen* 355

Meta-knowledge and the development of expertise in complex domains. *M.G.M. Ferguson-Hessler* 175

Tijdbesteding van docenten in het voortgezet onderwijs op verschillende manieren gemeten: een validatiestudie. *J. Heidelberg, H. van den Bergh, H. Kuhlemeier en J. Hoeks* 306

The influence of parents' expectations on school careers. *A.A. van der Hoeven-van Doornum, M.J.M. Voeten en P. Jungbluth* 369

Probleemoplossen, leren en onderwijzen in exacte vakken: een voorbeeld uit de natuurkunde. *T. de Jong en M.G.M. Ferguson-Hessler* 149

Pygmalion and the effectiveness of 'black' schools; teachers' stereotypes and hidden goal differentations towards ethnic minorities. *P. Jungbluth* 99

Aanpak van kwalificatieproblemen in organisaties: een onderzoek naar strategieën van opleidingskundigen. *F.J. van der Krogt, A.A. Vermulst en I.E.C. ter Woerds* 295

Mijn kruistabel heeft zo veel verwachte waarden kleiner dan 5. Wat moet ik nu doen? *P.M. Kroonenberg en A. Verbeek* 1

De opbrengst van volwasseneneducatie voor vrouwen: economische zelfstandigheid? *S.J. van Leeuwen* 208

Secondary school recommendations in relation to student and school characteristics. *L. Mulder* 111

The quality of the IEA-Reading Activity Inventory pilot version. *M.E. Otter* 261

Cooperation between Educational Priority Policy areas and basic adult education. *G.J. Reezigt* 120

De eerste gehechtheidsrelatie(s) en de cognitieve ontwikkeling van het kind: een literatuuroverzicht. *C. de Ruiter, M.H. van IJzendoorn en J. Goedbloed* 227

De ontwikkeling van expertise in verschillende domeinen: inleiding. *H.G. Schmidt* 133

Kennisencapsulatie. *H.G. Schmidt en H. Boshuizen* 135

Een onderzoek naar betrouwbaarheid en validiteit van studentoordelen over de didactische vaardigheid van universitair docenten. *E. van Schooten, M. Eiting en T.M. Bechger* 11

De invloed van contingente feedback op attributies en prestaties in de klas. *F. Siero en J.P. van Oudenhoven* 343

Multiple correspondentie-analyse van school- en beroepsloopbanen. *T.W. Taris en P.G.M. van der Heijden* 26

Thinking aloud: does it affect regulatory processes in learning? *M.V.J. Veenman, J.J. Elshout and M.G.M. Groen* 322

The effects of teachers' attitudes and practices on the achievements of Dutch low SES pupils. *L.F.J. van der Velden* 85

De veronderstelde bekwaamheid als voorspeller van studieresultaten.. *A. Vrugt, M.Ph. Langereis en Joh. Hoogstraten* 42

Beboeten van spijbelaars: versterking van discipline of verzwakking van motivatie? *F.M. Weerman* 197

Effective provisions for ethnic minorities in primary schools. *M.G. Weide* 74

Improving opportunities for the educationally disadvantaged: introduction to this special issue. *M.P.C. van der Werf* 65

Coördinatie en de effectiviteit van het onderwijs op het vakgebied wiskunde. *B. Witziers* 279

Notities en Commentaren

Perspectief op een beroep? Een commentaar. *A.M.P. Knoers* 242
Hoe soms is vaak?. *Ch.G.J. Michels en M.J.M. Voeten* 58
'Transfer: onderzoek en onderwijs'. Een reactie op de onderliggende denkopvatting. *B. Wolters* 52

Boekbesprekingen

J.H.C. Moust: *De rol van tutoren in probleemgestuurd onderwijs. Contrasten tussen student- en docent-tutoren.* (J. Willems) 245
W.J. Nijhof, H.A.M. Franssen, W.Th.J.G. Hoeben en R.G.M. Wolbert (red.): *Handboek curriculum. Modellen, theorieën, technologieën.* (K.J. Westerhof) 186
W. Veugelers: *Pedagogische opdracht en arbeid. Doelstellingen van docenten in het Algemeen Voortgezet Onderwijs en Beroepsonderwijs.* (C. Klaassen) 248
E. Wagenaar: *Tweetaligheid in het aanvangsonderwijs. Een onderzoek naar de effecten van tweetalig kleuteronderwijs op de schoolloopbaan van Marokkaanse kinderen.* (G.W. Meijnen) 380
H.A.P. Wolfhagen: *Kwaliteit van klinisch onderwijs.* (J.C.M. Metz) 193

Redactionele medewerkers

In 1993 werd redactionele medewerking verleend door:

C.A.J. Aarnoutse	Katholieke Universiteit Nijmegen
T.A. van Batenburg	RION Groningen
M. Beishuizen	Rijksuniversiteit Leiden
M.P.F. Berger	Universiteit Twente
M. Boekaerts	Rijksuniversiteit Leiden
K.P. van den Bos	Rijksuniversiteit Groningen
R.J. Bosker	Universiteit Twente
A.G. Bus	Rijksuniversteit Leiden
J. Cohen-Schotanus	Rijksuniversiteit Groningen
A.H. Corporaal	Inspectie van het Onderwijs De Meern
A.B. Dijkstra	Rijksuniversiteit Groningen
S. Dijkstra	Universiteit Twente
J. Dronkers	SCO-Kohnstamm Instituut Amsterdam
W.H. Gijselaers	Rijksuniversiteit Limburg
D.N.M. de Gruijter	Rijksuniversiteit Leiden
E.G. Harskamp	RION Groningen
R.H. Hofman	RION Groningen
J.G.M. Imants	Rijksuniversiteit Leiden
F.P.C.M. de Jong	Katholieke Universiteit Nijmegen
P. Jungbluth	ITS Nijmegen
M. van der Kamp	Rijksuniversiteit Nijmegen
G. Kanselaar	Rijksuniversiteit Utrecht
K.B. Koster	Rijksuniversiteit Groningen
J.B. Kuhlemeier	CITO Arnhem
H. Kuyper	RION Groningen
R. Lamberigts	Katholieke Universiteit Nijmegen
G. Lang	Rijksuniversiteit Groningen
P.P.M. Leseman	RISBO Rotterdam
C.F.M. van Lieshout	Katholieke Universiteit Nijmegen
J.L. van der Linden	Rijksuniversiteit Utrecht
T. Mooij	ITS Nijmegen
A.F.M. Nieuwenhuis	STOAS Wageningen
M. Overmaat	SCO-Kohnstamm Instituut Amsterdam
W.J. Pelgrum	OCTO Enschede
J.J. Peters	Rijksuniversiteit Groningen
J.M. Pieters	Universiteit Twente
P. Reitsma	Paedologisch Instituut Duivendrecht
J. Rispens	Rijksuniversiteit Utrecht
A.A. Ros	RION Groningen
H.G. Schmidt	Rijksuniversiteit Limburg
P.R.J. Simons	Katholieke Universiteit Nijmegen
T.A.B. Snijders	Rijksuniversiteit Groningen
P.G. Swanborn	Universiteit van Amsterdam
J.G.L. Thijssen	Rabobank Nederland Utrecht
R.K.W. van der Velden	ROA Maastricht
M.P.C. van der Werf	RION Groningen
Th. Wubbels	Rijksuniversiteit Utrecht

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 1-2

Universiteit Utrecht
BIBLIOTHEEK CENTRUM UITHOF

Van de redactie

Voor de Vereniging voor Onderwijsresearch was het jaar 1992 een belangrijk jaar. In 1992 werd immers overgegaan tot een nadere structurering van de vereniging. Voortaan zal de Vereniging voor Onderwijsresearch bestaan uit een aantal divisies. Het Tijdschrift voor Onderwijsresearch zal met deze nadere structurering van de Vereniging voor Onderwijsresearch ongetwijfeld rekening gaan houden. De redactie is er zich van bewust dat het redactionele beleid voortaan ook vanuit de visies van de verschillende divisies zal worden bezien. Uiteraard is het niet mogelijk op voorhand te garanderen dat het werkgebied van alle divisies in de komende jaargangen gelijkelijk aan bod zal komen. Een dergelijke garantie is slechts mogelijk wanneer de redactie een beleid zou kunnen baseren op een zeer ruime hoeveelheid aangeboden kopij. Dit laatste is echter – helaas – niet altijd het geval.

De redactie van het Tijdschrift voor Onderwijsresearch hoopt een beroep te kunnen doen op de besturen van de verschillende divisies om een regelmatige stroom van publiceerbare bijdragen op gang te houden. Op die manier lijkt het mogelijk om de gevarieerdheid die inmiddels in de Vereniging voor Onderwijsresearch aanwezig is ook tot uitdrukking te brengen in het beleid van de redactie van het Tijdschrift voor Onderwijsresearch. Hoewel er natuurlijk bij de redactie een belangrijke verantwoordelijkheid ligt om de inhoud van het tijdschrift zoveel mogelijk in overeenstemming te doen zijn met de stand van zaken binnen de onderwijsresearch dwingt een toenemende differentiatie er als het ware toe het aantal antennes naar buiten toe te vergroten.

Behalve de zorg voor een evenredige aandacht voor de belangstellingsgebieden van de verschillende divisies ziet de redactie geen aanleiding om het geldende redactiebeleid op belangrijke punten te wijzigen. Dit betekent in ieder geval het volgende:

- Voor jaargang 18 zijn opnieuw 1 of 2 themanummers in voorbereiding. Redacteuren voor de themanummers zijn feitelijk aan het werk en de verwachting is dat dit beleidspunt ook in de nieuwe jaargang gestalte kan krijgen.
- De redactie beschikt over een uitvoerige lijst van boeken die besproken kunnen worden of die ter bespreking bij een variatie aan auteurs in behandeling zijn. Een punt van zorg is hier de vraag of een evenwichtige verdeling over de afzonderlijke nummers kan worden gerealiseerd.
- Een aanhoudend punt van zorg voor de redactie is een voldoende voorraad van artikelen die voor publikatie in aanmerking komen. Te veel en te vaak heeft de redactie te maken met pieken en dalen waardoor zowel de evenwichtige samenstelling van de afzonderlijke nummers als een evenwichtige spreiding van nummers over het jaar niet altijd passen bij het beoogde beleid.
- Notities en commentaren worden door de redactie bijzonder op prijs gesteld vooral ook omdat ze nogal eens leiden tot goede discussies.

De redactie wil ook dit jaar weer proberen of auteurs van presentaties tijdens de Onderwijsresearchdagen 1993 bereid zijn hun bijdragen vorm te geven als een artikel dat in het Tijdschrift voor Onderwijsresearch kan worden gepubliceerd. De indruk bestaat dat een te groot aantal goede onderdelen van de Onderwijsresearchdagen ten onrechte een ruimere aandacht middels de publikatie in artikelvorm niet krijgt. Soms lijkt dit een gevolg te zijn van het feit dat men liever een publikatie in een engelstalig tijdschrift nastreeft, in andere gevallen lijkt het erop – o.i. niet terecht – dat de presentatie tijdens de Onderwijsresearchdagen wordt gezien als een afsluiting van de betreffende activiteiten.

De lezers van het Tijdschrift voor Onderwijsresearch mogen ervan overtuigd zijn dat de redactie zich binnen de gegeven mogelijkheden zal inzetten om de inhoud van de afzonderlijke nummers zo aantrekkelijk en zo gevarieerd mogelijk te maken. Een dergelijk voornemen is echter in sterke mate afhankelijk van de gegeven mogelijkheden. Het onder leiding van daartoe uitgenodigde redacteuren werken aan themanummers ziet de redactie in toenemende mate als een goede mogelijkheid. Desalniettemin blijft de redactie in sterke mate afhankelijk van sponta-

RIJKSUNIVERSITEIT TE UTRECHT
2409 975 0

ne bijdragen van auteurs. Opnieuw wil de redactie de lezers aansporen tot een actieve bijdrage aan het Tijdschrift voor Onderwijsresearch. Alleen door een gezamenlijke inspanning van redactie en auteurs kan het tijdschrift de inhoud krijgen die de onderwijsresearch verdient.

In de samenstelling van de redactie zijn met ingang van 1 januari 1993 twee veranderingen opgetreden. Lieven Verschaffel heeft de redactie verlaten. Zijn plaats wordt ingenomen door Joost Lowyck. Namens het bestuur van de VOR heeft Jaap Scheerens in de redactie zitting genomen.

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 3-10

Mijn kruistabel heeft zo veel verwachte waarden kleiner dan 5. Wat moet ik nu doen?

P.M. Kroonenberg* en A. Verbeek[1]
*Vakgroep Algemene Pedagogiek, Rijksuniversiteit Leiden**

ABSTRACT

In an informal way, some dilemmas in connection with hypothesis testing in contingency tables are presented. This discussion is an introduction to a numerical evaluation of Cochran's rule about the minimum expected value in a contingency table.

INLEIDING

Met behulp van een voorbeeldje daagt dit stukje, op een sterk populariserende wijze, de statisticus uit om zich eens te bezinnen op het statistisch geloof en zich af te vragen of statistisch handelen wel in overeenstemming is met statistisch prediken. Dit gedeelte dient als inleiding voor een numerieke beschouwing over Cochrans befaamde vuistregel over de minimum verwachte waarden in een kruistabel. Het geheel wordt afgesloten met een aantal aanbevelingen voor het geval Uw kruistabel zo veel verwachte waarden kleiner dan 5 heeft.

EEN DILEMMA VOOR DE OPRECHTE STATISTICUS (M/V)

Stel U bent op bezoek bij de dokter die U de onaangename mededeling doet, dat U een ernstige zeldzame Ziekte heeft. Zij raadt U aan om een operatie te ondergaan die waarschijnlijk de Ziekte kan verhelpen. Er zijn echter twee soorten operaties (A en B) en als goede statisticus vraagt U natuurlijk of er informatie aanwezig is die het mogelijk maakt om een min of meer rationele beslissing te nemen. "Wel", zegt Uw arts, "veel ervaring met deze ziekte hebben we nog niet en er zijn nog maar 12 eerdere gevallen bekend. Over het verloop van de behandeling weten we wat er in Tabel 1 staat."

De arts voegt er aan toe dat op basis van deze gegevens zij U aanraadt om een operatie te ondergaan en met name Operatie A. Als zorgvuldig statisticus vraagt U zich af of methode A inderdaad beter is.

Het beantwoorden van deze vraag komt neer op de vraag welke behandeling relatief het meeste kans op succes heeft. Er wordt bij het uitsluitend bekijken naar die kans impliciet van uitgegaan dat bij Uw beslissing geen andere argumenten een rol spelen dan of U de behandeling (of geen behandeling) overleeft. Zoals een reviewer terecht opmerkte heeft het voor een ernstig zieke in dat geval weinig zin om tijd te verspillen aan het selecteren van een significantieniveau en om een toets voor de onafhankelijkheid van behandeling en toestand patiënt uit te voeren (wat mijn suggestie in een vroegere versie was).

"Er is immers weinig verloren als de zieke een Type I fout maakt en ten onrechte concludeert dat *soort behandeling* en *toestand van de patiënt* afhankelijk van elkaar zijn. In dit geval zal voor niets een operatie worden ondergaan, maar afgezien van de kosten zal dit de overlevingskansen van de patiënt niet beïnvloeden (de nulhypothese is immers waar). Als de zieke dan toch de toets op onafhankelijkheid wil uitvoeren dan is het ten onrechte concluderen dat

* Adres: Postbus 9555, 2300 RB Leiden.

Tabel 1. Resultaten van de behandeling van patiënten met de Ziekte.

	Operatiemethode			
Toestand Patiënt	Geen	A	B	Totaal
Leeft Nog	0	5	1	6
Overleden	3	1	2	6
Totaal	3	6	3	12

er onafhankelijkheid is een meer fatale fout (Type II). De zieke zou dan immers concluderen dat de keuze van behandelingsmethode niets uitmaakt, terwijl dit wel het geval is" (uit het beoordelingsrapport).

Als U echter helemaal niet houdt van het snijden in Uw lijf, dan ligt het probleem anders. In dat geval zou U eigenlijk alleen een operatie willen ondergaan als die duidelijke voordelen biedt boven niets doen, en dan is het de moeite waarde om uit te zoeken of er een relatie bestaat tussen de toestand van de patiënt en de soort behandeling. Immers in het geval van onafhankelijkheid zult U kiezen voor geen behandeling. Ten minste twee problemen doen zich voor bij het toetsen: Ten eerste wat is de steekproefopzet geweest bij dit onderzoek en ten tweede als we dit weten hoe toetsen we de nulhypothese van onafhankelijkheid tussen soort behandeling en toestand patiënt.

Steekproefopzet

In dit geval zijn er twee (serieuze) mogelijkheden. Ten eerste, de groepen patiënten zijn ontstaan door aselecte toewijzing van de patiënten aan methoden en de ongelijke verdeling over de behandelingscategorieën is toevallig. In dit geval hebben we te maken met een multinomiale proefopzet met een N=12, zodat we met een multinomiale toets aan de gang kunnen gaan. Ten tweede, de verdeling van de patiënten over categorieën is systematisch tot stand gekomen en we hebben te maken met een product-multinomiale proefopzet met het product van drie multinomiale (hier: binomiale) verdelingen met n_1=3, n_2=6 en n_3=3. Er bestaat een hele discussie over hoe men in dit soort situaties hoort te toetsen, dat wil zeggen of men een multinomiale (N vast), product-multinomiale (één marginaal vast), of een (gegeneraliseerde) hypergeometrische (twee marginalen vast) verdeling moet gebruiken (voor referenties en een Nederlandse discussie over dit onderwerp zie bijvoorbeeld Verbeek & Kroonenberg, 1993, Appendix B.2). Uit die discussie blijkt dat wij er de voorkeur aan geven om met twee vaste marginalen te werken, zo ook hier. Er zijn evenwel anderen die niets op hebben met vaste marginalen, en voor hen heeft dit artikeltje weinig te bieden.

Correcte toets

De standaard toets op onafhankelijkheid tegen een niet-gespecificeerd alternatief is Pearson's (1900) X^2-toets [=Σ (geobserveerd – verwacht)^2/verwacht] die *asymptotisch* een χ^2-verdeling heeft. Deze toets wordt meestal, naar onze mening ten onrechte, de "χ^2-toets" genoemd. Ten onrechte, omdat X^2 de toetsingsgrootheid is en er vele toetsen zijn die asymptotisch χ^2 verdeeld zijn (bijv. de loglikelihoodratio-toets, de toets van Kruskal-Wallis, etc.). De praktijk is om in alle steekproefgevallen de χ^2-verdeling te gebruiken. Dat men zich in de praktijk meestal niet druk maakt over de steekproefsituatie heeft er alles mee te maken dat asymptotisch de verschillende steekproefsituaties leiden tot dezelfde toets. Met andere woorden als men X^2 gebruikt als toetsingsgrootheid samen met de χ^2-verdeling ongeacht de steekproefsituatie, dan leidt dit steeds tot hetzelfde resultaat.

Vervelend is wel dat computerpakketten als SPSS (SPSS, 1990) zo vaak waarschuwen dat er een aantal cellen bij zijn die een verwachte waarde hebben kleiner dan 5, zonder aan te geven wat een mens moet met die mededeling. In Tabel 1 hebben bijvoorbeeld alle cellen een verwachte waarde kleiner dan 5. Iedereen herinnert zich dan dat zijn of haar statistiekboek iets zei in de trant van "De X^2-toets mag alleen gebruikt worden als de verwachte waarden groter zijn dan vijf" of iets dergelijks. Opvallend is hoe weinig gezegd wordt over wat er met dat "mag" bedoeld wordt. *Mogen* betekent hier de toets uitvoeren zonder dat men een groot risico loopt een foute beslissing te nemen. Het is namelijk zo dat de $p_{\chi 2}$-waarde [$\Pr_{\chi 2}(X^2 \geq X^2_{obs})$] niet noodzakelijkerwijs dezelfde waarde heeft als de juiste exacte p_{X^2}-waarde [$\Pr_{X^2}(X^2 \geq X^2_{obs})$] dat wil zeggen de p-waarde bepaald op basis van de verdeling van X^2 zelf. Met andere woorden het kan zijn dat de X^2-toets leidt tot verwerpen van de nulhypothese, terwijl de exacte toets tot de andere conclusie zou komen en omgekeerd. Bij exacte toetsen is het overigens zo dat de exacte verdeling in de drie genoemde steekproefsituaties niet hetzelfde is. Besluit men daarentegen te conditioneren op de marginalen zoals wij hier doen dan gebruikt men in alle gevallen de (gegeneraliseerde) hypergeometrische verdeling en is er weer één toets voor alle steekproefsituaties.

Moet U nu Methode A accepteren als de beste of kunnen we net zo goed een driezijdige munt opgooien? De X^2 van Tabel 1 is 6.0 en dit is groter dan de kritieke χ^2-waarde (=5.991, gegeven α=0.05 met df=2; $p_{\chi 2}$ = .0498), dus kunnen we, formalistisch redenerend, concluderen dat gezien de significantie – zelfs bij dit kleine aantal waarnemingen – er sprake is van een afwijking van de nulhypothese van onafhankelijkheid. Als we in Tabel 2 naar de genormaliseerde residuen kijken (Haberman's (1973) adjusted residuals; zie bijvoorbeeld ook Verbeek & Kroonenberg, 1990, 1993, sectie 9.1) dan zien we dat er een groter aantal overlevenden zijn bij behandeling A. Dus het oordeel van de dokter lijkt gezond.

Maar is de $p_{\chi 2}$ wel te vertrouwen in deze situatie? Immers alle verwachte waarden zijn kleiner dan 5. De oplossing is duidelijk, we zullen een beroep moeten doen op de exacte toets. De essentie van die toets is dat we gegeven de marginalen alle mogelijke tabellen bepalen, van elke tabel zowel de X^2 als de hypergeometrische kans uitrekenen op die tabel en vervolgens de kansen optellen van alle tabellen die een X^2-waarde hebben groter of gelijk aan de geobserveerde X^2-waarde. Dit probleem, simpel in concept, heeft velen er toe gebracht om er een algoritme voor te ontwerpen (voor een overzicht, zie Verbeek & Kroonenberg, 1985). De truc is natuurlijk om dit zo efficiënt mogelijk te doen, want met grote tabellen en veel waarnemingen kan het rekenen al gauw uit de hand lopen. Voor zover wij kunnen zien is de exacte p-waarde uitrekenkampioen Mehta en Patel's (1983) netwerkalgoritme, maar ook ons FISHER aftel-algoritme (Verbeek & Kroonenberg, 1985, 1990, 1993) slaat geen gek figuur. Echte wedhondenrennen zijn bij ons weten niet gepubliceerd. Beide algoritmen zijn geïmplementeerd in commercieel verkrijgbare programma's (resp, StatXact – Cytel Software Corporation, 137 Erie Street, Cambridge, Mass.; FISHER – *Progamma*, Postbus 841, 9700 AV Groningen). StatXact kan veel meer situaties aan en heeft een mooiere menustructuur dan FISHER, maar is ook navenant (± 5x) duurder.

Tabel 2. Genormaliseerde Haberman residuen behorend bij Tabel 1.

	Operatiemethode		
Toestand Patiënt	Geen	A	B
Leeft Nog	–2.0	2.3	–0.7
Overleden	2.0	–2.3	0.7

N.B. De genormaliseerde residuen, r_n, zijn gedefinieerd als r_n = (geobserveerde waarde – verwachte waarde)/geschatte standaard afwijking; zij zijn asymptotisch normaal verdeeld, zodat men ze (bij kleine aantallen, ruwweg) kan vergelijken met de kritieke waarden van de standaard normale verdeling.

In Uw precaire gezondheidstoestand ontdekt U dat de exacte toets een p_{X2} oplevert van .12, met andere woorden de toets op onafhankelijkheid wordt niet verworpen. Wat nu? Dan toch maar die driezijdige munt opwerpen of gewoon niet laten opereren? Of misschien Bayesiaan worden? Dit laatste helpt waarschijnlijk niet veel, want juist de benodigde a priori informatie ontbreekt en U moet Uw leven dan aan de Dirichlet verdeling toevertrouwen in plaats van de exacte of χ^2-verdeling. Goede raad is niet zo zeer duur, als wel niet beschikbaar. Zoals een reviewer opmerkte één enkele nieuwe waarneming kan de resultaten weer een geheel ander aanzien geven. Maar ja, U bent nu juist deze nieuwe waarneming. Onmiddelijk komen daarnaast problemen naar voren als hoe zat het onderzoek waaruit Tabel 1 is voorgekomen werkelijk in elkaar. Is er wel randomisatie toegepast? Is er informatie beschikbaar over het risico van de operaties op zich, nog onafhankelijk van de Ziekte, over de kwaliteit van het bestaan met en zonder operatie, etc.?

Om de beslissing nog ingewikkelder te maken, kunt U toch maar de mogelijkheid overwegen om in ieder geval iets te doen, want de Ziekte op zijn beloop laten en wachten tot er andere patiënten al dan niet behandeld vindt U in tweede instantie toch wat griezelig. De tabel horende bij de keuze tussen de twee typen operaties (even veronderstellend dat wij dit onafhankelijk van de vorige toets zouden kunnen doen of gedaan hebben) heeft een X^2 van 2.25 met $p_{X2} = .46$ en $p_{\chi^2} = .13$. In dit geval weet U formeel gezien weer niet welk type operatie U zou moeten kiezen, zij het dat we denken dat de meeste mensen dan maar aan A de voorkeur zouden geven.

Wij laten U verder worstelen met Uw probleem. Alle statistische informatie die wij U kunnen verschaffen hebben we U gegeven. Nog afgezien van deze specifieke contekst confronteert dit voorbeeld U met de gewetensvraag "Hoe sterk gelooft U in hypothesetoetsen wanneer het er echt op aan komt?", laat het tegelijkertijd zien hoeveel keuzes er wel niet gemaakt moeten worden, en bovendien hoeveel er afhangt van de omstandigheden waaronder de data zijn verzameld.

COCHRANS REGEL

Hierboven citeerden wij losjes een willekeurig tekstboek over de toepasbaarheid van de χ^2-toets bij kleine verwachte waarden. Afgezien van enkele voorlopers was het Cochran (1952, 1954) die een preciese formulering gaf over de minimum verwachte waarden in een kruistabel waarbij de asymptotische χ^2-toets nog nauwkeurig genoeg zou zijn. Uit de twee artikelen samen kan de volgende regel worden gedistilleerd:

> ***Voor kruistabellen met meer dan 1 vrijheidsgraad*** *geldt dat als er relatief weinig verwachtingen kleiner zijn dan 5 (zeg 1 van de 5 cellen of meer, of 2 van de 10 cellen of meer), dan is een minimum van 1 toegestaan bij de berekening van* X^2 (Cochran, 1954, p. 420).

In het eerdere artikel gaf Cochran aan wat hij bedoelde met 'toegestaan':

> *Een afwijking wordt als onbelangrijk beschouwd als de exacte P ligt tussen 0.04 en 0.06 terwijl P 0.05 is volgens de* χ^2*-tabel en als the exacte P ligt tussen 0.007 en 0.015 terwijl P 0.01 volgens de* χ^2*-tabel.* (Cochran, 1952, p. 328, 329).

Het is goed om op te merken dat Cochran met deze regel geen uitspraken doet over 2x2-tabellen en dat hij (al staat dit niet in het citaat) alleen uitspraken doet bij conditionering op de marginalen.

Een spannende vraag is natuurlijk, hoe goed die vuistregel van Cochran wel is. Hieronder geven we een indicatie van het antwoord. Een meer volledig verhaal is in voorbereiding. We definiëren eerst een *Cochran-tabel*: Een Cochran-tabel is een kruistabel met twee of meer vrijheidsgraden, alle verwachte waarden groter of gelijk aan één en 80% van de verwachte waarden groter of gelijk aan vijf. Om een indruk te geven van dit soort tabellen: 2x3, 2x4 en 2x5

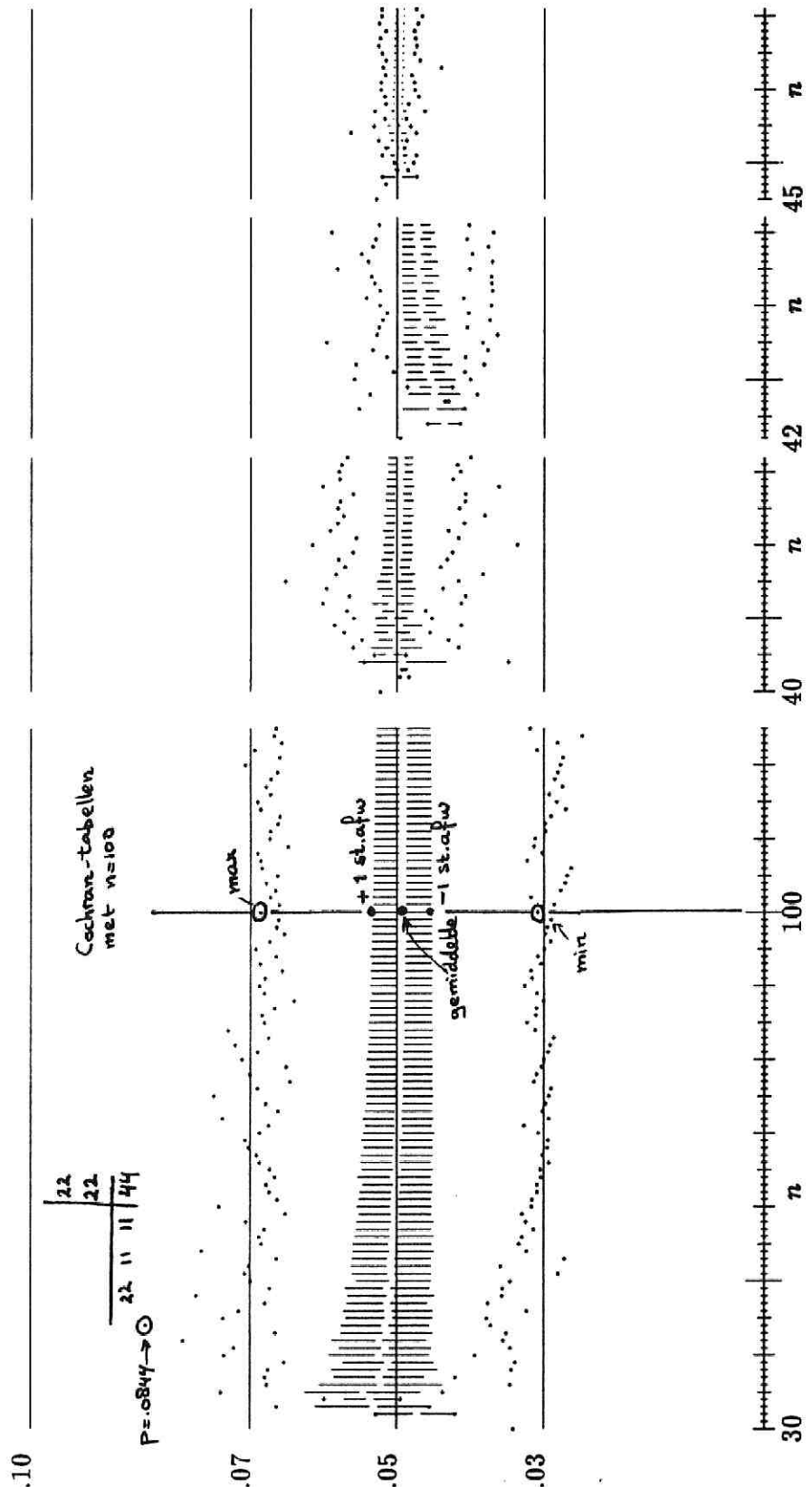


Fig. 1. Voor vaste n, r, en c, wordt in dit figuur het gemiddelde, gemiddelde ±1 st. afw., minimum en maximum afgebeeld van $p = \Pr(X^2 > \chi^2_{df}[.95]$ voor alle r x c Cochran-tabellen met deze n. Voor 2 x 3 tabellen laat de figuur zien dat de p-waarde van de meeste tabellen onder $p = .07$ valt. De tabel met de grootste p heeft een n van 44. Problemen met tabellen met waarde $p < .03$ beginnen pas bij $n = 100$. De kleinste p-waarde vinden we bij $n = 124$. Bij de 1 miljoen tabellen zijn er slechts 73 schendingen.

tabellen hebben nooit meer dan één verwachte waarde kleiner dan vijf. Een 2x3 tabel moet minstens 30 waarnemingen hebben om een Cochran-tabel te zijn, in feite is er voor n=30 maar één, namelijk die met alle cellen gelijk aan 5. Voor een 2x3 tabel heeft men minstens 66 waarnemingen nodig om een Cochran-tabel te krijgen met een verwachte waarde van 1.

Figuur 1 geeft een overzicht van de kwaliteit van de vuistregel van Cochran voor *alle* 2x3 Cochran-tabellen met een n kleiner of gelijk aan 125, *alle* 2x4, 2x5 en 3x3 Cochran-tabellen met $n \leq 72$. Dit zijn er samen een 1,3 miljoen. Op basis van een reeds uit 1979 stammend misverstand hebben we steeds Cochrans regel geïnterpreteerd op basis van het interval [0.03-0.07] in plaats van [0.04-0.06]. Wat hiervan de oorzaak is, is niet meer te achterhalen. Albert Verbeek ontwierp de studie en voerde de berekeningen uit en het kostte zijn 286 machine met coprocessor ongeveer twee weken rekentijd om de gegevens voor Figuur 1 uit te rekenen. Ik [PMK] ben er tot op heden niet toegekomen om Albert Verbeeks bestanden uit te zoeken zodat ik de hele rekenpartij nog een keer over kan doen bij de *echte* Cochran-grenzen. Inspectie van Figuur 1 leert ook dat dit niet echt veel nieuwe inzichten zal verschaffen. Voor de 2x3 tabellen zien we dat de wijdere grenzen de 'correcte' zijn. Er zijn slechts *73 schendingen op ruwweg 1 miljoen tabellen*. Bij de 2x4 en 2x5 tabellen zien we dat de 'echte' grenzen de 'correcte' zijn en bij de 3x3 tabellen zouden de grenzen nog nauwer aangehaald kunnen worden. Opvallend bij elk van de tabellen is dat n maar bar weinig te maken heeft met de grootte van de grenzen, de standaard afwijking en iets meer met het gemiddelde. Al met al is de kwaliteit van Cochrans regel (aangepast voor de 2x3 tabellen) imposant als men weet dat Cochran praktisch geen rekenapparatuur ter beschikking stond. Overigens als we voor een 2x2-tabel dezelfde wijde grenzen aan houden dan moet de minimum verwachte waarde iets zijn in de orde van 10 à 15 (Figuur 2; $\chi^2_{.005,df=1} = 3.8461$). Aan de andere kant kan men bij 2x2 tabellen zelfs op een programmeerbaar zakrekenmachientje de exacte toets heel eenvoudig uitrekenen. BMDP (zie Dixon, 1988, p.261) verschaft in dit geval exacte p-waarden mits de minimum verwachte waarde kleiner is dan 20 en meldt dat

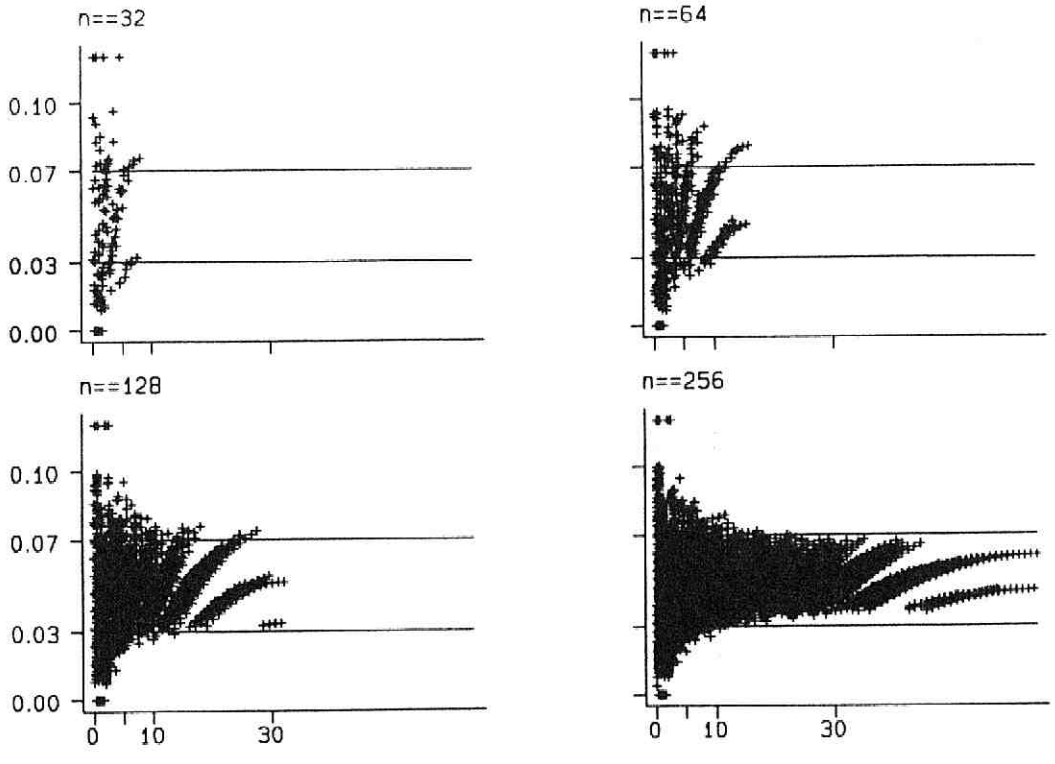


Fig. 2. Alle 2x2 Tabellen met n = 32, 64, 128 of 256 ($Pr(X^2 > \chi^2_1 [.95] = 3.8461)$ tegen Kleinste Verwachte Waarde; p-waarde afgekapt bij .10; maximale waarde was .207).

het bij grotere minimale verwachte waarden niet nodig is. Versie 4.0 van SPSS (SPSS, 1990, p. 104) daarentegen rekent de exacte toets alleen uit als er een cel is met minimum verwachte waarde kleiner dan 5. Dit zou opgetrokken moeten worden naar 15 of 20 om zeker te zijn. De situatie is al verbeterd ten opzichte van Versie 3 toen men de exacte toets alleen uitrekende als $n \leq 20$ (SPSS, 1988).

Een laatste opmerking over de tabellen die de schendingen veroorzaken. Het blijkt dat met name tabellen die gelijke en/of proportionele marginalen hebben, zoals ons voorbeeld in het eerste gedeelte van dit artikeltje, aanleiding geven tot veel grotere p-waarden dan de 0.05 bepaald op basis van de χ^2-verdeling. De voornaamste reden hiervoor lijkt te zijn dat bij gelijke of proportionele marginalen er veel meer tabellen met gelijke X^2-waarden zijn, met andere woorden de verdeling van X^2 wordt discreter en er treden dus grotere sprongen in de kansverdeling op. Het maakt in dergelijke situaties dus veel uit of de geobserveerde X^2-waarde juist groter of kleiner is dan de kritieke waarde bij de gekozen α. Bij 2x3 tabellen is de kritieke waarde, $\chi^2_{\alpha=.05} = 5.991$, net iets kleiner dan een geheel getal. Bij proportionele marginalen komen zulke gehele getallen relatief makkelijk voor en dus ook $X^2_{obs} = 6.0$ en dit maakt de situatie rond het kritieke punt onstabiel. Het hebben van gelijke marginalen hangt uiteraard niet van n af. Dus ook bij grote n kunnen er tabellen voorkomen met grote discrepanties tussen de exacte en asymptotische kritieke waarden. Deze bevinding suggereert dat bij onderzoeken met discrete responsvariabelen het wellicht beter is om niet precies gelijke aantallen personen in de experimentele en controle groep te hebben.

CONCLUSIE

Afgezien van ons voorbeeld, wat kan er nu gezegd worden over de vraag gesteld in de titel van dit artikeltje? Hieronder heb ik een aantal aanbevelingen samengeraapt die niet allemaal uit de tekst voortkomen, maar grotendeels wel berusten op onderzoek dat we gedaan hebben in het verleden. Vuistregels zijn per definite altijd fout en moeten dus ook niet blindelings opgevolgd worden. Een algemeen punt is dat programma's met exacte toetsen niet echt tijdrovend zijn. Voor die gevallen waar dat zou dreigen wordt doorgaans eerst een Monte-Carlo schatting gegeven van de exacte p-waarde, zodat men op grond daarvan als nog kan besluiten om een dan tijdrovende exacte toets te doen.

1. Bij 2x2 tabellen (met vaste marginalen) moet de kleinste verwachte waarde minstens 15 zijn wil men tussen de Cochran-grenzen blijven (zie Figuur 2). De tijd die gemoeid is met een exacte toets is echter dermate minimaal dat men die net zo goed kan uitvoeren. De Yates-correctie is onnodig conservatief.
2. Uit Figuur 1 blijkt dat bij 2x3 Cochran-tabellen (met vaste marginalen) de afwijking ten opzichte van de Cochran-grenzen nogal groot kan zijn, zodat voor alle 2x3 tabellen een exacte toets aan te bevelen is.
3. Voor Cochran-tabellen met meer dan twee vrijheidsgraden is de asymptotische kritieke waarde zeer betrouwbaar en kan men deze in het algemeen met een gerust hart gebruiken. Wel is enige voorzichtigheid geboden bij gelijke of proportionele marginalen. Voor niet Cochran-tabellen met meer dan twee vrijheidsgraden is een exacte toets aan te bevelen.
4. Gezien de kwaliteit van de χ^2-benadering voor de verdeling van X^2 bij twee vaste marginalen, is het bij één of geen vaste marginalen vrijwel nooit nodig om gebruik te maken van exacte toetsen (zie Verbeek & Kroonenberg, 1993, Appendix B.2).

NOOT

1. Dit artikeltje is een uittreksel van een langer gedetailleerder manuscript in wording dat gebaseerd is op aantekeningen van Albert Verbeek, die op 9-9-1990 overleden is. Veel dank is verschuldigd aan de twee beoordelaars die op een aantal onjuiste denkbeelden wezen.

LITERATUUR

Cochran, W.G. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics, 23*, 315-345.

Cochran, W.G. (1954). Some methods for strenghtening the common χ^2 tests. *Biometrics, 10*, 417-451.

Dixon, W.J. (Ed.) (1988). *BMDP statistical software manual* (Vol. 1). Berkeley: CA: University of California Press.

Haberman, S.J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29*, 205-220

Mehta, C.R., & Patel, N.R. (1983). A network algorithm for performing Fisher's exact test in rxc contingency tables. *Journal American Statistical Association, 78*, 427-434.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, Series 5, 50*, 157-175 (Reprinted 1948 in E.S. Pearson (Ed.), *Karl Pearson's early papers*. Cambridge, UK: Cambridge University Press).

SPSS (1988). *SPSS-X user's guide* (3rd edition) . Chicago, IL: SPSS, Inc.

SPSS (1990). *SPSS reference guide*. Chicago, IL: SPSS, Inc.

Verbeek, A. & Kroonenberg, P.M. (1985). A survey of algorithms for exact distributions of test statistics in rxc contingency tables with fixed margins. *Computational Statistics and Data Analysis, 3*, 159-185.

Verbeek, A. & Kroonenberg, P.M. (1990, 1993[2]). *FISHER 3.0: Testing independence in rxc tables.* (2e herziene druk, FISHER 3.10, 1993). Groningen: iec *ProGAMMA*.

Manuscript ontvangen 14-12-1992
Definitieve versie ontvangen 22-3-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 11-25

Een onderzoek naar de betrouwbaarheid en validiteit van studentoordelen over de didactische vaardigheid van universitair docenten

E. van Schooten, M. Eiting en T.M. Bechger
*SCO-Kohnstamm Instituut voor Onderzoek van Opvoeding en Onderwijs**

ABSTRACT

This article describes a study aimed at verifying the reliability and validity of two instruments using student evaluations for the establishment of teaching effectiveness. The construction of the instruments was based on theoretical notions of Vermunt (1989).
Factor analyses of the data show that the factorial validity of the instruments is satisfactory, even though some a priori distinctions are not supported by the factor structure in the data. The reliability of the instruments is satisfactory and the analysis demonstrates that with a certainty of 95% a reliability of at least .80 can be reached with a jury consisting of 23 students. Unfortunately, analysis of variance reveals that for each factor in the instruments there is a considerable amount of bias.

INLEIDING

In Amerika worden al sinds de twintiger jaren oordelen van studenten over de effectiviteit van het onderwijs gebruikt om de kwaliteit van het onderwijs te evalueren en om beslissingen over de promotie of ontslag van individuele docenten te onderbouwen (Marsh, 1987; West, 1988). Barr (1948) rapporteert in een overzicht 138 van dergelijke studies tussen 1905 en 1948 en tussen 1976 en 1984 zijn er 1055 studies over dit onderwerp opgenomen in het ERIC bestand.

Uit het Amerikaanse onderzoek blijkt dat de studentoordelen een redelijk tot goede betrouwbaarheid vertonen. Marsh (1987; 275) rapporteert voor zijn instrument een betrouwbaarheid van .95 als een somscore van 50 studenten wordt gebruikt, .90 voor een somscore van 25 studenten en .74 voor een somscore van 10 studenten. Ook de stabiliteit blijkt goed. De test-hertestbetrouwbaarheid bleek na een periode van enkele jaren nog .83 te bedragen.

Feldman (1977) vond dat onder de assumptie dat studenten elkaars onafhankelijke replicaties zijn, een jury van 20 á 25 studenten voldoende betrouwbaar oordeelt. Bij een oordeel gebaseerd op de beoordelingen van kleinere jury's vond hij partijdigheid of bias. De motivatie van de studenten blijkt een rol te spelen bij het veroorzaken van de partijdigheid en ook blijkt dat de wijze waarop de docent de materie benadert (sterk structurerend, analytisch, enz.), differentieel gewaardeerd kan worden door studenten. Andere factoren zoals geslacht, leeftijd of algemene persoonlijkheidskenmerken van docent of student bleken in Feldmans studie niet of nauwelijks van invloed.

De validiteit van de studentoordelen is nog wel omstreden. Positieve resultaten uit validiteitsonderzoek zijn dat uit multi-trait-multi-method onderzoek blijkt dat het effect van de gegeven cursus niet groot is. De oordelen over twee docenten die beiden dezelfde cursus geven correleren gemiddeld -.05, twee oordelen over één docent, gebaseerd op dezelfde cursus correleren gemiddeld .72 en de gemiddelde correlatie tussen de oordelen over één docent gebaseerd op het lesgeven in twee verschillende cursussen bedraagt .60 (Marsh, 1987; 277). Verder blijkt uit een overzichtsartikel van Cohen (1981) dat de prestaties van studenten positief correleren met studentoordelen over de didactische kwaliteit van de docenten. Howard e.a. (1985) vonden dat de studentoordelen hogere validiteitscoëfficiënten (met verschillende criteria) vertoonden dan oordelen van collega's, zelfbeoordelingen en oordelen van getrainde beoordelaars.

*Adres: Grote Bickersstraat 72, 1013 KS Amsterdam.

Negatieve resultaten met betrekking tot de validiteit van de instrumenten betreffen aanwijzingen voor partijdigheid. In Marsh (1987) worden een onderzoeksoverzicht en resultaten van eigen onderzoek geboden betreffende partijdigheid van studentoordelen over de didactische kwaliteit van docenten. Dat er sprake kan zijn van partijdigheid concludeert Marsh op grond van eigen onderzoek. Hij vond dat de betrouwbaarheid van één beoordelaar gemiddeld .20 bedraagt, terwijl de mediane hertestcorrelatie, berekend over individuele scores die met een jaar tussenpoos zijn bepaald, .56 bedraagt. Hieruit concludeert Marsh dat er systematische variatie in de studentoordelen is, die uniek is voor de oordelen van een individuele student.

Het bias-onderzoek dat Marsh aanhaalt in zijn overzicht, is volgens hem methodologisch zwak. Hij stelt zelfs dat 'The search for potential biases to student ratings has itself been so biased, that it could be called a witch hunt' (Marsh, 1987; 328). Het grootste probleem bij het aangehaalde onderzoek is dat correlaties causaal worden geïnterpreteerd.

Het aangehaalde onderzoek stelt dat er sprake is van partijdigheid als een theoretisch niet met didactische kwaliteit verwante variabele wel met de studentoordelen correleert. Bij deze studies wordt 5 tot 25 % van de variantie in de studentoordelen verklaard door 'interesse voor het vak', 'het verwachte cijfer voor het vak' en 'de moeilijkheid van de cursus of de hoeveelheid werk die de cursus met zich meebrengt voor de student'.

Door een toenemende aandacht voor de kwaliteit van het wetenschappelijk onderwijs komt de beoordeling van de didactische vaardigheid van docenten in het wetenschappelijk onderwijs ook in Nederland steeds meer in zwang. Aan de Faculteit der Geneeskunde van de Universiteit van Amsterdam wordt in samenwerking met de Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam sinds 1990 het project 'Verhoging didactische kwaliteit door beoordeling en scholing' uitgevoerd. Een onderdeel van het project betreft het ontwikkelen van een beoordelingsinstrument waarmee individuele docenten periodiek beoordeeld kunnen worden op hun didactische vaardigheid. Het is de bedoeling om uiteindelijk met behulp van de beoordelingen docenten te selecteren die in aanmerking komen voor didactische cursussen en om, indien mogelijk, de cursusinhoud van deze didactische cursussen aan te passen aan gesignaleerde lacunes. Daarnaast is de docentbeoordeling bedoeld om informatie te leveren aan individuele docenten betreffende hun didactische prestaties en om de mate van vooruitgang na het volgen van didactische cursussen te kunnen bepalen. Ook aan andere universitaire faculteiten zijn instrumenten ontwikkeld voor het beoordelen van de didactische vaardigheid van universitair docenten met behulp van studentoordelen (Vorst, 1991; Os. 1990; Gijselaers, 1988; Jansen & Neve, 1988).

In het onderstaande worden de resultaten gepresenteerd van analyses betreffende de betrouwbaarheid en interne validiteit van twee aan de Faculteit der Geneeskunde van de Universiteit van Amsterdam ontwikkelde instrumenten voor de beoordeling van de vaardigheid van individuele docenten in het uitvoeren van contactonderwijs. Er wordt eveneens aandacht geschonken aan mogelijke partijdigheid van de studentoordelen. Verklaringen van eventueel gevonden partijdigheid kunnen niet worden gegeven, daar er geen achtergrondvariabelen waarvan verwacht werd dat ze partijdigheid zouden veroorzaken, zijn gemeten voor het onderhavige onderzoek.

1. INSTRUMENTCONSTRUCTIE

De twee bovengenoemde instrumenten betreffen een instrument voor het beoordelen van de didactische vaardigheid van docenten in het geven van hoorcolleges, werkcolleges en werkgroepen ('college/werkgroep-instrument') en een soortgelijk instrument voor het practicumonderwijs ('practicuminstrument'). De reden voor het ontwikkelen van twee instrumenten was, dat de relevante aspecten van het te beoordelen didactisch handelen voor het practicumonderwijs enerzijds en het onderwijs dat wordt gegeven middels hoorcolleges, werkcolleges en werkgroepen anderzijds, verschillen. (Vergelijk b.v. de items 'Nagaan of de practicumopdracht goed door de studenten wordt uitgevoerd' versus 'Overzichten en samenvattingen geven'.)

De instrumenten bestaan uit een opsomming van een aantal didactische handelingen waarbij de beoordelaar (student) aan moet geven hoe goed de docent deze handelingen verricht. De waardering werd door de respondenten aangegeven op likert-schalen die lopen van 1 (zeer slecht) tot 5 (zeer goed). De voor beide instrumenten geselecteerde didactische handelingen zijn gebaseerd op theoretische overwegingen beschreven door Vermunt (1989).

Vermunt onderscheidt drie onderwijsstrategieën, te weten 'sterke externe controle', waarbij de docent alles in de hand houdt, 'zwakke externe controle', waarbij studenten verantwoordelijk zijn en de rol van de docent beperkt is, en 'gezamenlijke controle' waarbij de studenten door docenten worden aangespoord de verschillende functies bij het leren zelf uit te voeren. In de instrumenten zijn alleen items opgenomen die vallen onder de onderwijsstrategieën 'sterke externe controle' en 'gezamenlijke controle', omdat de strategie 'zwakke externe controle' handelingen van de student betreft waar de docent geen of zeer weinig invloed op heeft. Dwars op deze indeling onderscheidt Vermunt drie clusters leeractiviteiten; cognitieve en affectieve verwerkingsactiviteiten en metacognitieve regulatieactiviteiten. Binnen elk van de drie clusters leeractiviteiten uit het model van Vermunt worden subcategorieën onderscheiden. Het aantal subcategorieën dat in de instrumenten is geoperationaliseerd, is kleiner dan bij Vermunt om de grootte van het instrument binnen redelijke grenzen te houden en omdat een aantal onderscheidingen bij Vermunt dermate genuanceerd zijn, dat gevreesd werd dat studenten tussen deze items geen onderscheid zouden kunnen maken. Daarnaast bleek uit een proefafname met een eerste versie van de instrumenten dat van enkele categorieën de items niet relevant of beantwoordbaar gevonden werden door de studenten, gegeven het gevolgde onderwijs. Deze categorieën zijn verwijderd. Ook zijn nog wat items toegevoegd die niet op de theorie van Vermunt gebaseerd zijn, maar wel relevant zijn voor het beoordelen van het onderwijs. Deze items betreffen de presentatie door de docent en het gebruik van technische hulpmiddelen.

Het college/werkgroep-instrument bevat 47 items in 13 categorieën. De categorieën vallend onder 'cognitieve verwerkingsactiviteiten' zijn 'relaties aangeven', 'structuur van de stof duidelijk maken' en 'zorgen dat de stof blijft hangen'. Onder affectieve verwerkingsactiviteiten vallen 'motiverend doceren', 'verantwoorden van de stof' en 'studenten serieus bij het vak betrekken'. Onder metacognitieve regulatie-activiteiten tenslotte vallen 'studenten voorbereiden op het studeren', 'controleren of de stof goed overkomt', 'extra aandacht besteden aan de niet goed begrepen onderdelen'. Extra categorieën zijn 'de tijdplanning van het onderwijs', 'spreekvaardigheid en gedrag van de docent', 'het gebruik van technische hulpmiddelen' en 'bevorderen van een actieve studiehouding'.

Het practicuminstrument bevat 31 items in de zes categorieën 'motiverend doceren', 'studenten serieus bij het vak betrekken', 'de studenten voorbereiden op de practicumopdracht', 'controleren of de opdracht goed uitgevoerd wordt', 'extra aandacht besteden aan de niet goed begrepen onderdelen' en 'de presentatie door de docent'.

In beide instrumenten zijn in vrijwel alle categorieën zowel items opgenomen die vallen onder de onderwijsstrategie 'sterke externe controle' als onder de strategie 'gezamenlijke controle'.

Een beschrijving en theoretische verantwoording van de instrumenten wordt gegeven door Verbeek, Van den Bijtel en Ten Cate (1992) en Van den Bijtel & Ten Cate (in voorbereiding).

2. STEEKPROEFTREKKING EN DATA-ANALYSE

Steekproeven van studenten uit de vier verschillende leerjaren van de Faculteit der Geneeskunde hebben hun docenten beoordeeld met behulp van de instrumenten. De steekproeven zijn getrokken door per leerjaar twee practicumgroepen te trekken en alle studenten uit een getrokken practicumgroep om hun medewerking te verzoeken. Aan de studenten uit het tweede en het vierde jaar is gevraagd om voor elke beoordeelde docent na twee weken voor een tweede keer het college- of practicuminstrument in te vullen, teneinde de stabiliteit van de oordelen te bepalen.

De op deze wijze getrokken studenten kregen het verzoek om alle docenten van wie zij in vooraf gespecificeerde blokken of practica les kregen, met behulp van de instrumenten te beoordelen. Gezien de organisatie van het onderwijs aan de medische faculteit was het zowel voor het college-werkgroepinstrument als voor het practicuminstrument helaas onmogelijk alle beoordelaars dezelfde docenten te laten beoordelen.

In totaal hebben 34 studenten één of meer docenten beoordeeld met het college/werkgroepinstrument. Er zijn 74 docenten beoordeeld met dit instrument waarbij iedere beoordelaar gemiddeld 12 docenten beoordeelde. Op deze wijze werden 396 ingevulde vragenlijsten verzameld als eerste beoordeling en 233 ingevulde vragenlijsten voor de test-hertest. Er zijn 28 studenten die beoordelingen met het practicuminstrument hebben uitgevoerd, waarbij 37 docenten werden beoordeeld. Gemiddeld beoordeelde iedere beoordelaar vijf docenten met dit instrument. Van het practicuminstrument zijn 141 exemplaren als eerste beoordeling geretourneerd en 102 vragenlijsten voor de test-hertest.

Er zijn verschillende analyses verricht om de interne validiteit en de betrouwbaarheid van de instrumenten te verifiëren. Ten eerste is door middel van factor- en hiërarchische clusteranalyses nagegaan of de bij de constructie aangebrachte interne structuur van de instrumenten door de empirie bevestigd wordt.

Voor de factoranalyse is de niet-iteratieve FABIN3 schatter van Hägglund (1982) gebruikt. De FABIN3 schatter heeft als voordeel dat de rekentijd geringer is dan voor bij voorbeeld maximum-likelihoodfactoranalyse. Om de passing van alternatieve modellen te vergelijken maken we gebruik van de 'Goodness of Fit Index' (GFI) (Jöreskog en Sörbom, 1986). De GFI meet de relatieve hoeveelheid variantie en covariantie die wordt verklaard door het model. De distributie van GFI is niet bekend maar ervaring en simulatie-experimenten (m.n. Marsh, Balla en MacDonald, 1988, Tabel 3) duiden erop dat een waarde van .97 vereist is om te kunnen spreken van een goede passing. De analyses werden uitgevoerd met een eigen programma.

De clusteranalyses zijn uitgevoerd volgens de methode van Elshout (Elshout, 1969). Bij deze analyse worden clusters van items onderscheiden op basis van de inter-itemcorrelaties. Onderscheiden clusters worden samengevoegd tot grotere clusters op grond van correlaties met items of andere clusters. Op deze wijze wordt een overzicht van de correlationele structuur van het instrument op itemniveau verkregen.

Op grond van de resultaten van de twee bovenbeschreven analyses zijn een aantal items uit de instrumenten geselecteerd die de verschillende in de instrumenten te onderscheiden factoren zo goed mogelijk representeren. Met deze items als afhankelijke variabelen zijn per onderscheiden factor variantie-analyses verricht met als onafhankelijke variabelen de beoordeelde docenten, de beoordelaars en, voor het tweede en het vierde leerjaar, de test-hertest. De onafhankelijke variabelen zijn opgevat als gefixeerde factoren. Bij de analyses zijn de items die een factor representeren als replicaties opgevat.

De variantie-analyses zijn uitgevoerd voor het toetsen van verschillende hypotheses, alle getoetst op vijf-procentsniveau. In de eerste plaats kan door deze analyses nagegaan worden in welke mate de instrumenten discrimineren tussen verschillende docenten. Daar de instrumenten bedoeld zijn om de verschillende docenten te onderscheiden al naar gelang hun didactische vaardigheid, is het essentieel dat er een effect van de docentfactor gevonden wordt.

In de tweede plaats kan door de variantie-analyses bepaald worden of de beoordelingen onafhankelijk zijn van de toevallig gebruikte student-beoordelaars. Daar het in de bedoeling ligt de instrumenten periodiek te gebruiken voor het beoordelen van docenten en er niet altijd van dezelfde groep studenten als beoordelaars gebruik gemaakt zal worden, is een effect van de beoordelaars ongewenst. Een effect van de beoordelaarsfactor hoeft niet persé een onoverkomelijk probleem op te leveren, mits de verschillen tussen beoordelaars kunnen uitmiddelen door het gebruiken van een grote groep beoordelaars per docent.

Een derde reden voor het verrichten van de variantie-analyses is om na te gaan of er een testhertesteffect optreedt. Als dit effect aanwezig blijkt, betekent dat dat beoordelaars niet stabiel zijn in hun oordelen over de test-hertestperiode van twee weken. Het zal duidelijk zijn dat dit, gezien het doel van de instrumenten, onwenselijk is.

Tenslotte kunnen met behulp van de variantie-analyses ook interactie-effecten worden opgespoord tussen de docent-, de beoordelaars- en de test-hertestfactor. Mochten de resultaten dergelijke effecten aan het licht brengen, dan blijkt dat de beoordelaars in statistische zin partijdigheid of 'bias' vertonen. Een interactie tussen docent en beoordelaar betekent bij voorbeeld dat beoordelaars specifieke voor- of afkeur voor bepaalde docenten in hun oordeel meewegen, en dat deze voor- of afkeur per beoordelaar verschilt. Een interactie tussen beoordelaar en test-hertest betekent, dat op een systematische wijze sommige beoordelaars de tweede beoordeling van een docent hoger doen uitkomen dan de eerste keer en andere juist lager. Een interactie tussen docent en test-hertest betekent dat op systematische wijze sommige docenten de eerste keer een lager oordeel krijgen dan de tweede keer en andere docenten juist een hoger oordeel.

Ook is er nog een driewegsinteractie mogelijk, namelijk tussen docent, beoordelaar en test-hertest. De interpretatie van dergelijke effecten is niet eenvoudig. Wel kan gesteld worden dat dit soort effecten eveneens wijst op invaliditeit van de beoordelingen.

Als laatste analyse is bepaald wat de betrouwbaarheid is van de somscores op de instrumenten bij een gegeven aantal beoordelaars en hoeveel beoordelaars er nodig zijn om respectievelijk een jury-alpha van minimaal .95 en een jury-alpha van minimaal .80 te bereiken. Daarnaast is ook nog gekeken hoeveel beoordelaars er nodig zijn voor het bereiken van een bepaalde mate van precisie in de docentbeoordelingen.

3. RESULTATEN

3.1. De structuur van het college- en het practicuminstrument

In het onderstaande worden eerst de resultaten van de factoranalyses besproken en komen vervolgens de resultaten van de clusteranalyses aan de orde. Telkens worden eerst de analyses op de data verkregen met het college/werkgroep-instrument behandeld, en vervolgens de analyses betreffende het practicuminstrument.

3.1.1. De factoranalyses

Het college/werkgroep-instrument

Voor het college/werkgroep-instrument is eerst een éénfactormodel gepast. Hiervoor zijn alle door de studenten uit de vier leerjaren ingevulde college/werkgroepvragenlijsten gebruikt, behalve de vragenlijsten met ontbrekende waarden en de vragenlijsten die een beoordelaar voor een tweede maal over één docent had ingevuld ter bepaling van de stabiliteit van de oordelen (N=354). Het éénfactormodel past matig bij de data, de GFI bedraagt .937. Vervolgens is een tweefactormodel gepast en dit model past goed bij de data. De GFI bedraagt .974 en de betrouwbaarheid van een somscore over alle items gebaseerd op de oordelen van 'de gemiddelde beoordelaar', die het programma ook berekent (Fleishman & Benson, 1987), bedraagt onder het tweefactormodel .968. De correlatie tussen de twee geschatte factoren bedraagt .094. De resultaten van deze analyse staan in Tabel 1.

Daar het tweefactormodel goed past, heeft het geen zin na te gaan hoe goed een model past waarin alle in het instrument onderscheiden factoren gerepresenteerd worden. Een dergelijk model zal zeker perfect passen, doch het spaarzaamheidsprincipe gebiedt ons niet meer factoren te onderscheiden dan nodig zijn voor het verkrijgen van een goede passing.

De meeste items laden positief op beide factoren. Voor het geven van een inhoudelijke interpretatie van de twee gevonden factoren moet gezocht worden naar de zuiverste factoritems. Dit zijn de items die hoog op de éne en laag op de andere factor laden en een lage errorvariantie hebben. Voor de eerste factor zijn dit vooral de items 33 tot en met 40. De meest factorzuivere items van de tweede factor zijn de items 22 tot en met 26, 44, 45 en 46. (Zie voor de inhoud van de items Van den Bijtel & Ten Cate (in voorbereiding)).

Grofweg lijkt de eerste factor items te bevatten die weergeven of de docent in het tempo van doceren en in de presentatie publiek- of studentgericht is (onderwijsstrategie 'sterke externe controle'). De tweede factor lijkt vooral te slaan op het stimuleren van een positieve studiehou-

Tabel 1. Resultaten factoranalyse tweefactormodel college/werkgroep-instrument (N=354). In de tabel staan de gestandaardiseerde factorladingen op de eerste en de tweede factor en de gestandaardiseerde errorvarianties van de items uit de vragenlijst. Daarnaast worden de item-restcorrelaties (Rit) van de items gegeven, evenals de GFI van het tweefactormodel, de volgens Fleishman & Benson (1987) geschatte betrouwbaarheid, de correlatie tussen factor 1 en 2 en het 95%-waarschijnlijkheidsinterval van de alpha.

itemnr:	factor-1:	factor-2:	errorvar:	Rit:	GFI:	betrouwbaarheid:	cor1x2:	alpha-95%-interval:
1	.20	.50	.35	.60	.9735	.9679	.0935	.9574-.9684
2	.31	.38	.31	.66				
3	.26	.51	.38	.63				
4	.39	.22	.47	.55				
5	.50	.17	.31	.65				
6	.48	.24	.27	.70				
7	.46	.18	.27	.66				
8	.40	.31	.33	.66				
9	.30	.45	.34	.65				
10	.20	.44	.51	.51				
11	.47	.15	.40	.58				
12	.57	.22	.47	.63				
13	.59	.30	.55	.65				
14	.56	.27	.49	.65				
15	.19	.53	.33	.62				
16	.18	.45	.52	.50				
17	.05	.54	.42	.47				
18	.36	.24	.33	.61				
19	.36	.49	.34	.70				
20	.47	.19	.37	.62				
21	.42	.41	.44	.66				
22	**.13**	**.51**	**.34**	**.54**				
23	**.18**	**.52**	**.40**	**.57**				
24	**.08**	**.57**	**.30**	**.55**				
25	**−.01**	**.60**	**.39**	**.44**				
26	**.00**	**.55**	**.33**	**.45**				
27	.40	.31	.34	.66				
28	.42	.29	.34	.66				
29	.45	.14	.29	.61				
30	.53	.12	.31	.63				
31	.44	.25	.35	.65				
32	.43	.11	.32	.56				
33	**.53**	**.04**	**.32**	**.55**				
34	**.59**	**.00**	**.31**	**.56**				
35	**.54**	**.07**	**.23**	**.63**				
36	**.43**	**.13**	**.40**	**.54**				
37	**.50**	**.08**	**.27**	**.61**				
38	**.52**	**.05**	**.32**	**.58**				
39	**.50**	**.09**	**.20**	**.66**				
40	**.42**	**.11**	**.30**	**.58**				
41	.46	.17	.48	.56				
42	.32	.15	.73	.39				
43	.42	.07	.34	.52				
44	**.17**	**.54**	**.27**	**.63**				
45	**.05**	**.43**	**.30**	**.47**				
46	**.16**	**.43**	**.34**	**.55**				
47	.41	.26	.35	.63				

ding en het scheppen van de benodigde voorwaarden daarvoor, en met name het verschaffen van duidelijkheid omtrent de functie en de doelstellingen van het te volgen blok, zowel door de relaties met behandelde en te behandelen stof aan te geven als door duidelijk te maken welke eisen er aan de studenten gesteld zullen worden (onderwijsstrategie 'gezamenlijke controle').

Het practicuminstrument

Ook voor het practicuminstrument past een éénfactormodel slecht, de GFI bedraagt .937. Het tweefactormodel past redelijk, de GFI bedraagt .964 en de vanuit het model geschatte betrouwbaarheid van het instrument is .952. De correlatie tussen de twee factoren bedraagt -.003. De resultaten van deze toetsing staan in Tabel 2. Om verwarring te voorkomen worden de gevonden factoren hier x en y genoemd.

Weer zijn de meest factorzuivere items van de twee factoren gezocht. Voor factor x zijn dat de items 13, 14, 28, 29 en 30 en voor factor y de items 21, 22, 24 en 25. Van factor x zijn de meest factorzuivere items afkomstig uit de categorien 'De studenten voorbereiden op de practicumopdracht' (13 en 14) en 'De presentatie door de docent' (28, 29 en 30). De meest factorzuivere items van factor y vallen onder de categorieën 'Extra aandacht besteden aan de niet goed begrepen onderdelen' (21 en 22) en 'De presentatie door de docent' (24 en 25).

Tabel 2. Resultaten factoranalyse tweefactormodel practicuminstrument (N=138). (Zie voor uitleg over de tabel Tabel 1.)

itemnr:	factor-x:	factor-y:	errorvar:	Rit:	GFI:	betrouwbaarheid:	cor1x2:	alpha-95%-interval:
1	.20	.53	.21	.72	.9641	.9520	.0030	.9298-.9568
2	.21	.55	.41	.62				
3	.40	.52	.34	.73				
4	.39	.50	.26	.76				
5	.27	.34	.24	.65				
6	.33	.42	.31	.67				
7	.36	.44	.41	.64				
8	.18	.55	.33	.63				
9	.38	.42	.38	.67				
10	.28	.20	.47	.42				
11	.23	.36	.32	.58				
12	.32	.27	.38	.54				
13	**.47**	**.10**	**.39**	**.46**				
14	**.58**	**.15**	**.29**	**.59**				
15	.39	.27	.47	.52				
16	.26	.54	.37	.68				
17	.20	.42	.30	.62				
18	.26	.40	.41	.58				
19	.28	.45	.33	.66				
20	.19	.48	.37	.61				
21	**.19**	**.63**	**.38**	**.65**				
22	**.00**	**.67**	**.23**	**.60**				
23	.19	.37	.50	.49				
24	**.05**	**.48**	**.71**	**.40**				
25	**.05**	**.53**	**.41**	**.51**				
26	.24	.28	.23	.59				
27	.26	.21	.25	.53				
28	**.64**	**.00**	**.18**	**.47**				
29	**.53**	**.10**	**.24**	**.52**				
30	**.59**	**-.10**	**.48**	**.30**				
31	.26	.43	.27	.68				

Factor x lijkt betrekking te hebben op het bieden van duidelijkheid ten aanzien van wat van de studenten verwacht wordt en de tijdplanning van het onderwijs. Het lijkt hierbij dus te gaan om het organiseren van de activiteiten van studenten tijdens het practicum, enigszins vergelijkbaar met factor 2 (gezamenlijke controle). Factor y lijkt geïnterpreteerd te moeten worden als de mate waarin de docent aandacht geeft aan begripsproblemen bij studenten. Deze factor betreft met name het geven van inhoudelijk goede uitleg (vergelijk factor 1, sterke externe controle).

3.1.2. De clusteranalyses

Uit de factoranalyses blijkt dat de met de beide instrumenten verkregen datasets elk adequaat te beschrijven zijn door een tweefactormodel. Om na te gaan of de in de instrumenten onderscheiden categorieën nog enige empirische realiteit bezitten, is voor elk van de instrumenten een hiërarchische clusteranalyse op itemniveau verricht (Elshout, 1969).

Het college/werkgroep-instrument

Uit de clusteranalyse verricht op de data van het college/werkgroep-instrument blijkt dat de factorzuivere items uit de factoranalyse ook bij de clusteranalyse samenkomen. Item 33 tot en met 40 (factor 1) clusteren, evenals item 44, 45 en 46 (factor 2) en 22 tot en met 26 (ook factor 2). De categorieën zoals onderscheiden bij de constructie van het instrument zijn deels ook herkenbaar in het resultaat van de analyse.

In het instrument zijn bij de constructie 13 categorieën onderscheiden. Twee van deze categorieën blijken niet apart te clusteren maar vormen één groot cluster, één item uit een andere categorie sluit hierbij aan en van één categorie van vier items clusteren er twee keer twee items met andere categorieën samen. De bovengenoemde factorzuivere items voor factor één en twee blijken elk keurig te clusteren.

Concluderend kan gesteld worden dat de factoren uit de factoranalyse herkenbaar uit de clusteranalyse naar voren komen en dat de verschillende in het instrument onderscheiden categorieën over het algemeen clusteren zoals verwacht werd. Bij deze conclusie moet de kanttekening gemaakt worden dat hoewel de items van de in het instrument onderscheiden categorieën over het algemeen in de clusteranalyse clusters vormen, deze clusters onderling zo hoog correleren, dat de data adequaat beschreven kunnen worden met een tweefactormodel.

Het practicuminstrument

Uit de clusteranalyse betreffende het practicuminstrument, blijkt dat ook hier de eerder gevonden tweefactorstructuur in de clusters weerspiegeld wordt. Zowel factor x (item 13, 14, 28, 29 en 30) als factor y (item 21, 22, 24 en 25) vormen een cluster.

Over de overeenkomst tussen de categorieën in het practicuminstrument, die bij de constructie van het instrument zijn onderscheiden, en de clusterstructuur zoals die blijkt uit de analyse, kunnen we kort zijn. De a priori in het instrument onderscheiden categorieën komen niet in de clusterstructuur tot uiting.

Conclusie factor- en clusteranalyses

De conclusie van de factor- en de clusteranalyses luidt dat voor beide instrumenten de data goed met een tweefactormodel te beschrijven zijn. Als we kijken naar de meest factorzuivere items van de verschillende gevonden factoren blijkt dat voor beide instrumenten de twee factoren geïnterpreteerd kunnen worden als factoren die respectievelijk de sterke externe controle dan wel de gezamenlijke controle betreffen. Uit de clusteranalyses blijkt dat de dertien in het college/werkgroep-instrument onderscheiden categorieën grotendeels in de gevonden clusterstructuur terug te vinden zijn. De bij de constructie van het practicuminstrument onderscheiden categorieën zijn niet in de clusterstructuur terug te vinden.

3.2. De variantie-analyses

Op grond van de factor- en de clusteranalyses zijn een aantal voor elke factor representatieve items geselecteerd voor het verrichten van de variantie-analyses. De selectie vond plaats door

de meest factorzuivere items van elke factor te selecteren die ook in de clusteranalyse samen bleken te clusteren. Voor factor 1 zijn de items 33 tot en met 40 gekozen, voor factor 2 de items 22 tot en met 26 en 44, 45 en 46, voor factor x de items 12, 13, 14, 28, 29 en 30 en voor factor y de items 7, 8, 21, 22, 24 en 25. De items 12 voor facor x en 7 en 8 voor factor y zijn toegevoegd, omdat uit de clusteranalyse bleek dat deze items bij de andere factorzuivere items van de factoren clusteren en deze items een redelijk hoge factorlading op hun factor vertonen en een relatief lage lading op de andere factor. Ook de errorvarianties van deze items bleken niet groot. Daarnaast zijn door het toevoegen van deze items per instrument evenveel replicaties per factor beschikbaar voor het verrichten van de variantie-analyses.

Voor ieder leerjaar zijn één of meer analyses verricht, afhankelijk van hoeveel gevulde matrijzen er per factor gevormd konden worden. Een gevulde matrijs houdt in dat alle beoordelaars die op de éne as van de matrijs staan, alle docenten die op de andere as staan hebben beoordeeld. Als er ook naar het effect van de test-hertest gekeken wordt, betekent een gevulde matrijs dat alle beoordelaars alle docenten twee maal hebben beoordeeld. In de cellen die door de matrijs gevormd worden, staan de itemscores van de geselecteerde items (opgevat als 'cases'). Als er meerdere analyses per leerjaar zijn verricht, is erop gelet dat de verschillende matrijzen voor de verschillende variantie-analyses nooit dezelfde docenten of beoordelaars bevatten.

De variantie-analyses zijn verricht per leerjaar èn per factor. De items van verschillende factoren zijn niet in één analyse meegenomen, omdat de items als replicaties opgevat moeten kunnen worden, daar anders de interactie-effecten niet bepaald kunnen worden en omdat anders de residuele variantie zou toenemen, daar de factoren nagenoeg ongecorreleerd zijn. Daarnaast is het denkbaar dat de items van één factor gevoelig zijn voor een bepaald soort bias, en de items van een andere factor niet. Marsh (1987; 309 e.v.) vond dat de verschillende factoren die hij in zijn instrument onderscheidt, niet alle even gevoelig voor bias zijn.

In het onderstaande worden de resultaten gerapporteerd van de variantie-analyses van de data verkregen met het college/werkgroep-instrument en vervolgens de analyses van de data verkregen met het practicuminstrument. De resultaten van deze twee groepen analyses worden respectievelijk in Tabel 3 en 4 schematisch weergegeven.

Het college/werkgroep-instrument

Voor het eerste leerjaar kon voor het college/werkgroep-instrument slechts één matrijs gevormd worden, samengesteld uit drie docenten en drie beoordelaars. Voor het tweede leerjaar zijn twee matrijzen gevormd, éénmaal met vier docenten, drie beoordelaars en test-hertestfactor en éénmaal voor vijf docenten en vijf beoordelaars. Voor het derde leerjaar kon voor het college/werkgroep-instrument één matrijs gevormd worden met vier docenten en drie beoordelaars. Voor het vierde leerjaar is met de data van het college/werkgroep-instrument één matrijs samengesteld, gevormd door de oordelen over drie docenten van vier beoordelaars met een test-hertestbeoordeling. De resultaten van de analyses worden weergegeven in Tabel 3.

Uit de tabel blijkt dat er slechts in de helft van het aantal keren dat er een docenteffect getoetst is, sprake bleek van significantie, terwijl er negen van de tien keer een significant beoordelaarseffect bleek te bestaan, dat soms erg groot uitviel. De test-hertestfactor bleek niet één van de vier keer dat deze factor getoetst kon worden significant en er bleek bij de analyse van tien matrijzen acht keer sprake van één of meer significante interactie-effecten.

Het practicuminstrument

Voor het practicuminstrument konden alleen voor het eerste en het tweede leerjaar volledig gevulde matrijzen gevormd worden. Voor het derde en het vierde leerjaar konden derhalve geen analyses verricht worden. Voor het eerste leerjaar kon er slechts één matrijs gevormd worden met twee docenten en twee beoordelaars. Voor het tweede leerjaar konden drie matrijzen gevormd worden, één met acht docenten, drie beoordelaars en test-hertest, één met twee docenten, zeven beoordelaars en test-hertest en één met twee docenten, drie beoordelaars en test-hertest. De resultaten van deze analyses worden weergegeven in Tabel 4.

Tabel 3. Een overzicht van de resultaten van de variantie-analyses verricht op de data die zijn verzameld met het college/werkgroep-instrument. Een '+' geeft een significant effect aan, een '–' een niet significant effect en als er niets is ingevuld, is het betreffende effect niet getoetst. Onder het kopje 'interacties' staat aangegeven hoeveel getoetste interactie-effecten significant bleken, met tussen haakjes het aantal per analyse getoetste interactie-effecten.

jaar:	factor:	matrijs: doc x beo x t/h	effecten: docent:	beoordelaars:	test-hertest:	interacties:
1	1	3 x 3	+	+		1 (1)
	2	3 x 3	+	+		– (1)
2	1	4 x 3 x 1	–	+	–	2 (4)
	2	4 x 3 x 1	–	+	–	4 (4)
2	1	5 x 5	+	+		1 (1)
	2	5 x 5	+	+		1 (1)
3	1	4 x 3	–	+		1 (1)
	2	4 x 3	+	–		1 (1)
4	1	3 x 4 x 1	–	+	–	2 (4)
	2	3 x 4 x 1	–	+	–	– (4)

Tabel 4. Een overzicht van de resultaten van de variantie-analyses over de data verzameld aan de hand van het practicuminstrument. (Een '+' geeft een significant effect aan, een '–' geeft een niet significant effect aan, als er niets is ingevuld is er geen toetsing van het betreffende effect geweest en onder het kopje 'interacties' staat aangegeven hoeveel interacties significant bleken, met tussen haakjes het aantal per analyse getoetste interactie-effecten.)

jaar:	factor:	matrijs: doc x beo x th	effecten: docent:	beoordelaars:	test-hertest:	interacties:
1	x	2 x 2	–	–		– (1)
	y	2 x 2	–	–		– (1)
2	x	8 x 3 x 1	+	+	–	– (4)
	y	8 x 3 x 1	+	+	–	1 (4)
2	x	2 x 7 x 1	+	+	+	3 (4)
	y	2 x 7 x 1	–	+	–	2 (4)
2	x	2 x 3 x 1	+	+	–	2 (4)
	y	2 x 3 x 1	+	+	–	2 (4)

Uit de tabel blijkt dat er drie keer geen en vijf keer wel een significant docenteffect gevonden is. Vooral beoordelaars uit het eerste jaar geven docentoordelen die niet tussen docenten discrimineren. De docenteffecten lijken niet aan een bepaalde factor gebonden te zijn. De scores gebaseerd op de items uit factor x discrimineren één keer niet en de scores gebaseerd op items uit factor y discrimineren twee keer niet. Verder zijn er van de acht toetsingen van beoordelaarseffecten zes significant. Alleen de beoordelingen van eerstejaars studenten vertonen geen beoordelaarseffect. Bij deze resultaten moet bedacht worden dat de gebruikte matrijs voor de analyse van de oordelen van de eerstejaars erg klein is (2 x 2) en dat het niet vinden van significante effecten voor deze groep wel eens een powerprobleem zou kunnen zijn. Verder blijken interactie-effecten bij het merendeel van de matrijzen aan te tonen.

Conclusie variantie-analyses
De conclusie van de variantie-analyses voor beide factoren van beide instrumenten luidt, dat het aantal docenteffecten eigenlijk te mager is, slechts 10 van de 18 keer bleken de docenteffecten significant, terwijl de instrumenten bedoeld zijn om onderscheid tussen docenten te maken. Aangenomen dat de docenten verschillen in hun didactische vaardigheid, schieten de instrumenten wat dit betreft dus tekort.

De vrijwel altijd optredende beoordelaarseffecten geven aan dat beoordelaars niet vergelijkbaar oordelen over de didactische vaardigheid van docenten. De gemiddelde oordelen van beoordelaars verschillen, er zijn strenge en coulante beoordelaars.

Het slechts sporadisch optreden van significante test-hertesteffecten geeft aan dat de oordelen van individuele beoordelaars stabiel blijven over een periode van twee weken.

Het frequent optreden van significante interactie-effecten legt een duidelijk manco van de instrumenten bloot. Er is sprake van 'bias' of partijdigheid. De interactie-effecten duiden op ernstiger problemen dan de beoordelaarseffecten, omdat deze laatste effecten eventueel door het gebruik van een grote groep beoordelaars per docent zouden kunnen uitmiddelen. De interactie-effecten verraden echter partijdigheid die specifiek is voor bepaalde groepen docenten, terwijl onbekend is hoe deze bias precies werkt. Deze effecten tasten de validiteit van de met behulp van de instrumenten verkregen docentoordelen aan.

3.3 De beoordelaarsbetrouwbaarheid

Gewenste jurygrootte, bepaald door het stellen van een minimaal gewenste betrouwbaarheid
Als uitgangspunt van de analyse moest een matrijs samengesteld worden met het grootst mogelijke aantal beoordelaars en docenten, nadat de test-hertestbeoordelingen waren verwijderd. Voor deze matrijs was het niet noodzakelijk dat alle beoordelaars in de matrijs elke docent in de matrijs beoordeeld hadden. Elke beoordelaar die met één der overige beoordelaars minstens twee docenten gemeenschappelijk had beoordeeld (waardoor een zuivere schatter van de covariatie kan worden berekend), kon worden opgenomen in de matrijs. Op deze wijze kon met de oordelen gegeven via het college/werkgroep-instrument door beoordelaars uit het tweede leerjaar, een jury worden gevormd van 16 beoordelaars die voldoen aan het beschreven criterium. Deze 16 beoordelaars bleken oordelen over 25 verschillende docenten te hebben gegeven, waarbij het gemiddelde aantal door een beoordelaar beoordeelde docenten 8 bedraagt. Uitgangspunt voor de analyse waren de somscores op alle vragen in de collegevragenlijst, mits deze geen ontbrekende itemscores bevatten. Deze somscores zijn dus de beoordelaarsscores. In de matrijs zitten uiteraard missing data. Toch kan de (missing-data-jury-)alpha geschat worden alsof de matrijs geen missing data bevat. Als de uitval aselect is, is deze alpha zuiver ten opzichte van de echte alpha zonder uitval (zie Bijlage). Vervolgens kan met behulp van de Spearman-Brownformule voor parallelle testverlenging (Lord & Novick, 1968) berekend worden hoe groot de jury moet zijn voor het bereiken van een gegeven jury-betrouwbaarheid.

De volgens de bovenbeschreven wijze geschatte alpha voor de somscores van het college/werkgroep-instrument berekend over 16 beoordelaars bedraagt .867 met een 95%-waarschijnlijkheidsondergrens van .7396 en een bovengrens van .9466, uitgaand van het door de 16 beoordelaars gemiddelde aantal beoordeelde docenten van 8. Als we de ondergrens van de wenselijke jury-alpha op .95 stellen, is duidelijk dat in het gunstigste geval deze grens bereikt wordt met 16 beoordelaars. In het ongunstigste geval moeten we uitgaan van een alpha van .7396 bij een jury van 16 beoordelaars. De Spearman-Brownformule voor parallelle testverlenging (zie Bijlage) laat zien dat de verlengingsfactor in dat geval 6.68 bedraagt. In het ongunstigste geval zijn er voor het bereiken van een jury-alpha van .95 dus 6.68 x 16 = 107 beoordelaars per docent nodig.

Daar een ondergrens van .95 voor de betrouwbaarheid vrij hoog is, is dezelfde berekening nogmaals uitgevoerd met een iets lagere ondergrens voor de betrouwbaarheid. De Groot & Van Naerssen (1969; 262-263) stellen als absolute ondergrens voor de betrouwbaarheid van een instrument waarmee personen gerangschikt worden .80. Deze betrouwbaarheid wordt in het ongunstigste geval bereikt met 23 en in het gunstigste geval met 4 beoordelaars.

Gewenste jurygrootte, bepaald door het stellen van een minimaal gewenste precisie van de somscores

Uitgaande van de datamatrix, met ontbrekende waarnemingen, van 16 beoordelaars en 25 docenten, is een variantie-analyse verricht met uitsluitend de docenten als gefixeerde factor. De residuele variantie die in deze analyse is geschat, kan worden opgevat als de beoordelaarsvariatie (systematisch en fouten) bij constante docentscores. In dit geval volstaan we met de schatter en laten we het 95%-waarschijnlijkheidsinterval buiten beschouwing. De beoordelaarsvariatie kan beschouwd worden als een schatter van de populatievariantie. Als we stellen dat een docentscore (47 items met 5-puntsschalen, dus minimaal score 47 en maximaal score 235) met een waarschijnlijkheid van 95% moet liggen plus of min 5 punten rond het gevonden gemiddelde, dan kan uitgerekend worden hoe groot de jury moet zijn om een dergelijke precisie te bereiken. De standaardfout van een gemiddelde is de wortel uit de verhouding van populatievariantie en jurygrootte. De bij de analyse gevonden variantieschatting bedraagt 435.633, zodat om een standaardfout van 2.5 te bereiken een jury van afgerond 70 beoordelaars nodig is. Het bovenstaande houdt dus in dat, gegeven dat de populatievariantie 435.633 bedraagt, een docent die is beoordeeld door 70 beoordelaars en een gemiddelde score van b.v. 100 heeft, in werkelijkheid voor 95% zeker een score heeft die ligt tussen 95 en 105.

4. DISCUSSIE

Uit het bovenstaande blijkt dat de instrumenten voldoende betrouwbaar zijn. Een betrouwbaarheid van .80 wordt in het slechtste geval bereikt met een jury van 23 beoordelaars. Hierbij moet wel bedacht worden dat de respondenten uit slechts enkele practicumgroepen afkomstig zijn. De verkregen oordelen zouden hierdoor in statistische zin afhankelijk kunnen zijn. Daar komt bij dat, zoals ook Gijselaers (1988, 115) aanhaalt, de oordelen op retrospectieve observaties berusten en dat, gezien de mogelijkheid die studenten hebben over de oordelen van gedachten te wisselen, de afhankelijkheid van de oordelen hierdoor zou kunnen toenemen. De betrouwbaarheid zou hierdoor overschat kunnen zijn. Marsh (1987) rapporteert, zoals in de inleiding is aangehaald, vergelijkbare betrouwbaarheden voor studentoordelen over de didactische vaardigheid van docenten.

Ook de interne structuur van de instrumenten lijkt redelijk. Het college/werkgroep-instrument komt er wat dat betreft beter af dan het practicuminstrument, maar de analyses van het college/werkgroep-instrument zijn uitgevoerd op een groter aantal waarnemingen en daardoor betrouwbaarder, dan de analyses van het practicuminstrument. De problemen zitten in de resultaten van de variantie-analyses.

Uit de variantie-analyses blijkt dat de instrumenten onvoldoende discrimineren tussen verschillende docenten, al lijkt het practicuminstrument wat dit betreft iets beter dan het college/werkgroep-instrument. Een ander probleem dat uit de variantie-analyses naar voren komt, is dat er vrijwel altijd significante beoordelaarseffecten gevonden worden. Beoordelaars blijken te verschillen in hun gemiddelde oordelen, oftewel er is verschil in strengheid. Dit hoeft geen onoverkomelijk probleem te zijn, daar dit soort verschillen tussen beoordelaars kunnen uitmiddelen door het gebruik van een grote, aselect getrokken jury.

Het ernstigste probleem dat blijkt uit de resultaten van de variantie-analyses, is dat de oordelen partijdig blijken te zijn. Deze partijdigheid blijkt uit het veelvuldig optreden van interacties tussen de docent-, de beoordelaars- en de test-hertestfactor. Deze interactie-effecten kunnen niet uitmiddelen door het gebruik van grotere jury's. Om het probleem van de interactie-effecten aan te pakken is verdergaand onderzoek nodig. Er zou gekeken kunnen worden of de door Marsh en Feldman gerapporteerde oorzaken (zie inleiding) van invloed zijn.

Duidelijk is dat de scores verkregen met de instrumenten op de wijze zoals voor dit onderzoek is gebeurd, niet als 'spijkerhard' beschouwd kunnen worden, en om die reden kunnen er geen al te verstrekkende consequenties, b.v. selectie, aan bepaalde docentscores verbonden worden. Voor het geven van informatie aan docenten over hoe studenten het didactisch hande-

len van de docent waarderen, voor het selecteren van docenten voor het volgen van didactische cursussen en voor het bepalen van het effect van didactische cursussen lijken de instrumenten een redelijke basis te vormen.

LITERATUUR

Barr, S.A. (1948). The measurement and prediction of teaching efficiency: A summary of investigations. *Journal of Experimental Education. 16*, 203-283.

Benton, S.E. (1982). *Rating College Teaching: Criterion Validity Studies of Student Evaluation-of-Instruction Instruments*. AAHE-ERIC/Higher Education Research Report No. 1 (ED 221 147).

Bijtel, J.A.L. van den & Ten Cate, Th.J. ten (1992). *Docentbeoordeling bij de Faculteit Geneeskunde aan de Universiteit van Amsterdam*. (in voorbereiding).

Cohen, P.A. (1981). Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies. *Review of Educational Research, 51*, 3, 281-309.

Elshout, J.J., Elshout-Mohr, M. & Wijgaart C. v.d. (1969). *Marimaxcor: hiërarchische clusteranalyse uitgaande van een inter-correlatie-matrix.* Bulletin Technisch Centrum, Universiteit van Amsterdam.

Feldman, K.A. (1977). Consistency and variability among college students in rating their teachers and courses: a review and analysis. *Research in Higher Education, 6*, 233-274.

Feldt, L.S., Woodruff, D.J., & Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93-103.

Fleishman, I., & Benson, J. (1987). Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurement, 47*, 925-939.

Gijselaers, W.H. (1988). *Kwaliteit van het onderwijs gemeten. Studies naar de betrouwbaarheid, validiteit en bruikbaarheid van studentoordelen*. Maastricht, R.L. (Proefschrift).

Groot, A.D. de, & Naerssen, R.F. van (1969). *Studietoetsen. Construeren, afnemen, analyseren*. dl. 2, Den Haag, Mouton.

Hägglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika, 47*, 209-222.

Howard, G.S., Conway, C.G., & Maxwell, S.E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology, 77*, 187-196.

Jansen, P.J. & Neve, H. de (1988). *Studeren en doceren aan het hoger onderwijs. Vakmanschap als leeropdracht*. Leuven, ACCO.

Jöreskog, K.G., & Sörbom, D.(1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood and least squares methods*. Mooresville, IN: Sientific Software, Inc.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.

Marsh, H.W. (1987). Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H.W., Balla, J.R., & MacDonald, R.P. (1988). Goodness-of-Fit Indexes in Confirmatory Fatcor Analysis: The Effect of Sample Size., *Psychological Bulletin, 103*, 391-410

Os, W. van (1990). De evaluatie van cursussen binnen het analyse model voor het onderwijs in studierichtingen. In: Heijnen, G.W.H., Joostens, T.H. & Vroeijenstijn, A.I. *Kwaliteitszorg. Waarborg voor kwaliteit in het hoger onderwijs*. Groningen, COWOG. 189-198.

Vermunt, D.H.M. (1989). The interplay between internal and external regulation of learning, and the design of process-oriented instruction. Paper presented at the third Conference of the European Association of Research on Learning and Instruction, Madrid, September 4-7. In Lowyck, J. *Learning models in self-guided instruction*. Madrid.

Verbeek, A.E., Bijtel, J.A.L. van den, & Cate, Th.J. ten (1992). *Instrumentontwikkeling voor de beoordeling van docenten. Eerste verslag project beoordeling en scholing van de Faculteit Geneeskunde UvA*. Amsterdam, Stichting Centrum voor Onderwijsonderzoek (SCO-rapport 281).

Vorst, H.C.M. (1991). *Advies ter behoeve van de verdere ontwikkeling van AMOS-UvA II*, Amsterdam, UvA.

West R.F. (1988). The short-term stability of student ratings of instruction in medical school. *Medical Education, 22*, 104-112.

Manuscript ontvangen 4-12-1992
Definitieve versie ontvangen 29-3-1993

BIJLAGE

De missing-data-jury-alpha:
Als de uitval aselekt is is deze schatter zuiver ten opzichte van de alpha zonder uitval. Bewijs: stel aantal beoordelaars is k.

$\alpha = (k/(k-1))\,(1-\Sigma_{j=1,k}\,\sigma^2_j/\sigma^2_{tot})$, met σ^2_j als variantie van de j-de beoordelaar en σ^2_{tot} als variantie van de somscores van alle beoordelaars. Nu is σ^2_{tot} de som van alle varianties/covarianties in de variantie-covariantiematrix en $\Sigma_{j=1,k}\,\sigma^2_j$ is de som van alleen de varianties, zeg 'spoor'. Zodat

$\alpha = ((k/(k-1))(1-(\text{spoor/som}))$.

Uitleg:
De gangbare formule voor coëfficiënt alpha luidt

(1) $\alpha = (k/(k-1))\,(1-\Sigma_{j=1,k}\,\sigma_{jj}/\sigma^2)$

met k het aantal items of beoordelaars, σ_{jj} de variantie van het j-de item of de j-de beoordelaar en σ^2 de variantie van de som van item- of beoordelaarsscores. Als men niet beschikt over veel rekenfaciliteiten is (1) een makkelijke manier om de betrouwbaarheid te schatten. Coëfficiënt alpha onderschat de betrouwbaarheid en is de betrouwbaarheid wanneer de ware itemscores perfect correleren en gelijke varianties hebben (Lord & Novick, 1968). Omdat de variantie van de som gelijk is aan de som van alle varianties/covarianties der samenstellende deelscores, ofwel

(2) $S = \Sigma_{i=1,k}\,\Sigma_{j=1,k}\,\sigma_{ij}$,

terwijl

(3) $TR = \Sigma_{j=1,k}\,\sigma_{jj}$,

of het spoor van de variantie/covariantiematrix van de items of beoordelaars, volgt

(4) $\alpha = (k/(k-1))\,(1-TR/S)$

De schatter voor alpha vindt men door TR en S te berekenen uit de zuivere of onzuivere schatters van de variantie/covariantiematrix. Als alle schatters zijn gebaseerd op dezelfde steekproef met grootte N, maakt het wegens de breuk in (4) niet uit of men de onzuivere schatters vermenigvuldigt met N/(N-1). Is er sprake van ontbrekende gegevens en berekent men op basis van alle beschikbare gegevens (onder paarsgewijze uitsluiting) de zuivere schatters van de variantie/covariantiematrix, dan zal de schatter van α, mits de variabele N niet al te klein is (Feldt, Woodruff & Salih, 1987), zuiver zijn indien de uitval willekeurig is. Over de standaardfout van de aldus berekende alpha is weinig bekend, maar ze zal afnemen bij toenemende N en k (Feldt, Woodruff & Salih, 1987). In het onderhavige geval is de variabele N klein, maar de omvang van de jury (k) groot. Elke beoordelaar die met de overige beoordelaars minstens twee docenten gemeenschappelijk heeft beoordeeld, waardoor een zuivere schatter van de covariatie kan worden berekend, kan worden opgenomen in de jury. Op deze wijze kon een jury gevormd worden van 16 beoordelaars. uitgangspunt waren de somscores op alle vragen in de collegevragenlijst, mits deze geen ontbrekende itemscores bevatten. Deze somscores zijn dan de beoordelaarsscores. Via (4) schatten we vervolgens de jury-alpha. Onder aanname van homogene juryverkleining of -uitbreiding kan vervolgens via de Spearman-Brownformule (Lord & Novick, 1968) geschat worden hoe groot de jury moet zijn om een bepaalde betrouwbaarheid te realiseren.

De Spearman-Brownformule luidt

$Rxx' = ((N * Ryy') / (1 + (N-1) Ryy'))$,

waarbij Rxx' de gewenste betrouwbaarheid is en Ryy' de gevonden betrouwbaarheid. N is het aantal keren dat de test verlengd moet worden om Rxx' te bereiken. Als we N willen berekenen, kunnen we uitgaan van de onderstaande vorm van dezelfde formule;

$N = (Rxx'(1-Ryy')) / (Ryy'(1-Rxx'))$.

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 26-41

Multiple correspondentie-analyse van school- en beroepsloopbanen

Toon W. Taris en Peter G.M. van der Heijden*

ABSTRACT

This article addresses the use of multiple correspondence analysis (MCA) in the analysis of sequential data, such as event histories. First the principles of MCA are discussed. Subsequently a typical data set concerning the educational and employment careers of 265 young adults is analyzed. Then a simple four-class career typology is constructed. Graphical representations of the careers within each group indicate that these indeed differ strongly, providing evidence of the validity of the approach taken here. The paper concludes that MCA is a tool that enables the scientist to construct viable classifications of careers, which yields rather much understanding in the reasons why careers are grouped in a specific way.

1. INLEIDING

Het verloop van school- en beroepsloopbanen is voor onderwijsonderzoekers al geruime tijd een dankbaar object van studie. Longitudinaal onderzoek op dit terrein levert vaak data in de vorm van *event histories* op: sequenties van kwalitatieve toestanden die een subject gedurende de geobserveerde periode bezette, alsmede de tijdstippen waarop overgangen van de ene naar de andere toestand plaatsvonden. Deze toestanden zijn over het algemeen de categorieën van een nominale variabele, bijvoorbeeld 'werkend', 'werkloos' en 'niet beschikbaar voor de arbeidsmarkt'.

In de beschrijving en verklaring van het verloop van dergelijke loopbanen zijn grofweg twee benaderingen te onderscheiden. De eerste benadering richt zich op de *afzonderlijke overgangen* van de ene toestand naar de andere; bijvoorbeeld overgangen van werkloosheid naar werk, van alleenstaand naar samenwonend, of van werkzaam in een bepaalde baan naar niet meer werkend in die baan. De sequentie van toestandsovergangen van een bepaald subject wordt aldus opgedeeld in de afzonderlijke episoden, waarin het subject in één bepaalde toestand verkeerde (bijvoorbeeld de tijd dat men werkloos was). Op dit moment lijkt dit de dominante benadering, wat wellicht samenhangt met het feit dat er geavanceerde statistische technieken beschikbaar zijn voor de analyse van dit type gegevens (Allison, 1984; Blossfeld, Hamerle & Mayer, 1989; Tuma & Hannan, 1984, geven overzichten van deze technieken, die nauw gerelateerd zijn aan overlevingsduuranalyses uit de biologie). Deze technieken zijn bij uitstek geschikt voor de identificatie van factoren die het plaatsvinden van een bepaalde overgang beïnvloeden.

De tweede benadering richt zich op de beschrijving en verklaring van het verloop van de loopbanen gedurende een bepaald *interval*. In dit interval kunnen meerdere overgangen plaatsvinden, die in principe allemaal in de analyse betrokken worden. Centraal in deze benadering staat het vergelijken en classificeren van deze loopbanen, en het gebruiken van deze classificaties in vervolg-analyses (Abbott & Hrycak, 1990). Daarbij wordt aangenomen dat het resultaat niet goed is te modelleren als het resultaat van een of ander onderliggend stochastisch mecha-

*Toon Taris is verbonden aan de Vrije Universiteit, Vakgroep Arbeids- en Organisatiepsychologie, De Boelelaan 1081, 1081 HV Amsterdam, e-mail V70UTOON@HASARA11.
Peter van der Heijden is werkzaam bij de werkgroep Methoden en de Vakgroep Sociologie van de FSW van de Rijksuniversiteit Utrecht.
Adres: T.W. Taris, Vrije Universiteit, Vakgroep Arbeids- en Organisatiepsychologie, De Boelelaan 1081, 1081 HV Amsterdam.

nisme (bijvoorbeeld een eerste-orde Markovproces, Van de Pol, 1989). Een belangrijk probleem hierbij is dat het aantal theoretisch mogelijke loopbaanpatronen over het algemeen zeer groot is: als s het aantal mogelijke toestanden (bijvoorbeeld beroepen, schooltypen, enzovoort) is, en t het aantal tijdstippen waarop deze toestanden zijn waargenomen gedurende een bepaalde periode, dan bedraagt het aantal theoretisch mogelijke loopbanen s^t. Als er gedurende 6 tijdstippen geobserveerd wordt en er 4 toestanden onderscheiden worden, dan zijn er in principe $4^6 = 4096$ verschillende loopbanen denkbaar. Het aantal empirisch in de steekproef voorkomende sequenties zal meestal kleiner zijn, maar toch al snel het bevattingsvermogen van de onderzoeker overschrijden. Daarmee is het een belangrijke vraag hoe het grote aantal mogelijke patronen gereduceerd kan worden tot een beperkt aantal interpreteerbare en analytisch bruikbare patronen. In dit artikel wordt hiervoor een methode gepresenteerd. Deze techniek, multiple correspondentie-analyse, is geschikt voor de analyse van twee-dimensionale tabellen met rijen en kolommen, en niet-negatieve getallen in de cellen. Deze twee-dimensionale matrix wordt zo goed mogelijk benaderd door een matrix van een lagere rang (dat wil zeggen, met minder rijen en/of kolommen), waardoor de belangrijkste aspecten van de data in een laag-dimensionele ruimte kunnen worden bestudeerd en een beter inzicht in de gegevens verkregen kan worden. De rijen en kolommen kunnen worden afgebeeld in een grafische representatie van deze laag-dimensionele ruimte. Indien in een dergelijke representatie twee rijen (kolommen) dicht bij elkaar liggen, dan lijken deze sterk op elkaar; in het andere geval zijn de verschillen juist erg groot. Multiple correspondentie-analyse is geschikt voor de analyse van veel verschillende typen data (vergelijk bijvoorbeeld Gifi, 1990); doel van dit artikel is te laten zien hoe MCA specifiek bij de analyse van loopbanen van nut kan zijn. Ook deze kunnen immers worden weergegeven in de vorm van een twee-dimensionele tabel, waarbij elke rij een sequentie van bijvoorbeeld een persoon gedurende een bepaalde periode representeert, en een kolom een combinatie van een toestand waargenomen op een tijdstip – bijvoorbeeld een bepaald beroep waargenomen op tijdstip 2 – weergeeft.

In de volgende paragraaf worden eerst de principes van MCA kort besproken. Daarna wordt een toepassing van deze techniek gepresenteerd op de werk- en schoolloopbanen van jongeren over een periode van 2 jaar; aan de hand hiervan zal dieper worden ingegaan op multiple correspondentie-analyse van loopbanen.

2. CORRESPONDENTIE-ANALYSE VAN LOOPBANEN

Op de rijen van Tabel 1a zijn de loopbanen van 5 personen gegeven. De kolommen geven de 4 mogelijke toestanden weer waarin elk persoon zich kan bevinden, respectievelijk 'onderwijsvolgend', 'werkloos', 'verricht niet bij opleiding passende arbeid', en 'verricht wel bij de opleiding passende arbeid'. De cellen van Tabel 1a geven het aantal weken dat iemand zich in een bepaalde toestand heeft bevonden. De eerste persoon heeft bijvoorbeeld gedurende de twee jaar dat deze is geobserveerd geen onderwijs gevolgd, is in totaal 40 weken werkloos geweest, en heeft achtereenvolgens 12 weken niet en 52 weken wel bij de opleiding passende arbeid verricht.

Om de loopbanen van deze vijf personen in deze twee jaar te vergelijken kan gekeken worden naar de proportie tijd die zij ieder in elk van de vier toestanden hebben doorgebracht. Deze proporties zijn gegeven in Tabel 1b. Te zien is bijvoorbeeld dat de loopbaan van de eerste persoon enigszins lijkt op die van de derde persoon, maar sterk verschilt van die van de vierde persoon.

Door de vijf personen werden in totaal 126 weken in de eerste toestand zijn doorgebracht, 77 in de tweede toestand, 103 in de derde, en 214 in de vierde. De proporties in Tabel 1c geven aan in hoeverre elk van de personen heeft bijgedragen aan elk van de toestanden. Te zien is bijvoorbeeld dat de vierde persoon voor 56% heeft bijgedragen aan de 126 weken van de eerste toestand; de eerste persoon droeg voor 52% bij aan de tweede toestand, enzovoort.

Kleine tabellen zoals Tabel 1a zijn gemakkelijk met het oog te bestuderen, maar dat wordt

Tabel 1. De loopbanen van 5 hypothetische respondenten gedurende 2 jaar.

	Tabel 1a: aantal weken gespendeerd in diverse toestanden.					Tabel 1b: proportie tijd gespendeerd in diverse toestanden.					Tabel 1c: bijdrage van elk van de respondenten aan elke toestand.				
toestand	f	g	h	i	tot	f	g	h	i	tot	f	g	h	i	tot
resp a	0	40	12	52	104	.00	.38	.12	.50	1.00	.00	.52	.12	.24	.20
resp b	14	10	40	40	104	.13	.10	.38	.38	1.00	.11	.13	.39	.19	.20
resp c	0	0	24	80	104	.00	.00	.23	.77	1.00	.00	.00	.23	.37	.20
resp d	70	17	17	0	104	.67	.16	.16	.00	1.00	.56	.22	.17	.00	.20
resp e	42	10	10	42	104	.40	.10	.10	.40	1.00	.33	.13	.10	.20	.20
tot	126	77	103	214		.24	.15	.20	.41	1.00	1.00	1.00	1.00	1.00	1.00

Nb. f = onderwijsvolgend, g = werkloos, h = verricht niet bij de opleiding passende arbeid, i = verricht wel bij de opleiding passende arbeid.

moeilijker als het aantal personen of het aantal toestanden groter wordt; in dat geval wordt correspondentie-analyse behulpzaam. Stel dat de loopbanen van de vijf personen vergeleken worden. De proporties in Tabel 1b geven dan duidelijk weer hoe een loopbaan van een persoon door de tijd heen is opgebouwd. In correspondentie-analyse worden de proporties van Tabel 1b als coördinaten in een hoogdimensionele ruimte gebruikt. De vijf personen worden elk door 1 punt weergegeven in een in dit geval vier-dimensionele ruimte waarbij de proporties als coördinaten dienen. Zo heeft de eerste persoon de coördinaten (.00, .38, .12, .50), de tweede persoon (.13, .10, .38, .38), enzovoort. De afstanden tussen de punten geven nu een indicatie van de overeenkomst in de desbetreffende loopbanen: personen die dicht bij elkaar liggen volgden loopbanen die op elkaar lijken, en personen die ver uiteen liggen hebben loopbanen die erg verschillen.

In principe kunnen loopbaanproporties het meest verschillen op de toestand waarin in totaal de meeste weken terecht zijn gekomen. In dit voorbeeld is dit toestand 4, waarin 41% van alle weken valt. Hiervoor wordt in correspondentie-analyse gecorrigeerd door de assen van de vier-dimensionele ruimte 'op te rekken', door de eenheidslengte van elke as te vermenigvuldigen met $(1/p_j)^{1/2}$, waarbij p_j de proportie weken is die in deze toestand zitten. Voor de vier toestanden in het voorbeeld zijn deze oprekfactoren respectievelijk (2.03, 2.60, 2.25, 1.56).

Voor een meer formele definitie van de afstanden tussen de loopbanen moet eerst wat notatie ingevoerd worden. De elementen van de matrix in Tabel 1a worden genoteerd als x_{ij}, waarbij i (i=1, ..., I) de rijen indiceert en j (j=1, ..., J) de kolommen. De elementen van de matrix in Tabel 1b worden aangegeven met $p_{j|i} \equiv x_{ij}/x_{i+}$, waarbij $x_{i+}=\Sigma_j x_{ij}$. Analoog hieraan worden de elementen van de matrix in Tabel 1c als $p_{i|j} \equiv x_{ij}/x_{+j}$ genoteerd. De *chi-kwadraatafstand* $\delta(i,i^*)$ tussen de rij i en i^* in de vier-dimensionele ruimte is gedefinieerd als

$$\delta^2(i,i^*) = \sum_{j=1}^{J} \frac{(p_{j|i} - p_{j|i^*})^2}{p_j} \qquad (1)$$

Voor elke toestand wordt dus het verschil gekwadrateerd, en gedeeld door p_j. Vervolgens worden al deze verschillen opgeteld. Doordat voor elk paar loopbanen deze chi-kwadraatafstand berekend wordt ontstaat aldus een afstandsmatrix.

Het *zwaartepunt* representeert de 'gemiddelde' loopbaan, en heeft in het voorbeeld de coördinaten (.24, .15, .20, .41); deze zijn gelijk aan de proporties p_j. De afstand tussen rij i en de

oorsprong geeft dan aan hoezeer rij i afwijkt van het gemiddelde. Indien rij i en rij i^* in dezelfde richting afwijken van de oorsprong, dan wijkt hun loopbaan op vergelijkbare wijze af van de gemiddelde loopbaan.

De afstanden tussen de punten in de vier-dimensionele ruimte zijn nu helder gedefinieerd. Maar hoe moet deze ruimte bestudeerd worden? Het is niet gemakkelijk om afstanden te bekijken in ruimten met meer dan twee dimensies. Er is dus alle reden toe om te zoeken naar een manier om deze bestudering te vergemakkelijken. In correspondentie-analyse gebeurt dit door de ruimte zo te roteren dat achtereenvolgende dimensies in de geroteerde ruimte zoveel mogelijk laten zien van de afstanden in de vier-dimensionele ruimte. Voor elke volgende dimensie α is het te maximaliseren criterium

$$\lambda_\alpha^2 \equiv \Sigma_i p_i r_{i\alpha}^2 \tag{2}$$

waarbij r_i de coördinaat van rij i op de nieuwe (d.w.z. geroteerde) dimensie α is. De rijproportie p_i in deze vergelijking geeft aan dat de betergevulde rijen een grotere rol spelen in de bepaling van de uiteindelijke oplossing. Omdat voor het voorbeeld alle punten in een drie-dimensionele deelruimte van de vier-dimensionele ruimte liggen, zijn alleen λ_1^2, λ_2^2 en λ_3^2 groter dan nul. Hun som $\lambda_1^2 + \lambda_2^2 + \lambda_3^2$ wordt *inertie* genoemd. De inertie is een maat voor de totale (gewogen) afstand in de ruimte; $\lambda_1^2 / (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)$ is bijvoorbeeld de proportie van de inertie die is afgebeeld in de eerste dimensie. Voor het voorbeeld zijn de eerste drie waarden λ_α^2, achtereenvolgens .41, .13 en .05, en deze drie dimensies laten respectievelijk 67%, 23% en 9% van de totale inertie zien. Als alleen de eerste twee dimensies bestudeerd worden, dan wordt 91% van de totale inertie bestudeerd.

Een twee-dimensionele afbeelding van de rijpunten is te vinden in figuur 1a. Daar is te zien dat de loopbanen b, c, d en e op een lijn liggen die ongeveer evenwijdig loopt aan de horizontale as van de eerste dimensie, en dat vooral loopbaan a hoog scoort op de tweede dimensie. Loopbaan d lijkt het meest op loopbaan e, en het minst op loopbaan a en c. Om te weten waarom dit zo is, moet de relatie tussen de loopbanen en de toestanden onderzocht worden. Dit gebeurt tegelijkertijd in dezelfde correspondentie-analyse.

Voor de kolommen van de matrix wordt een vergelijkbare procedure als voor de kolommen gevolgd; uitgangspunt zijn nu echter de proporties in Tabel 1c. Elke toestand is als een punt af te beelden in een vijf-dimensionele ruimte, waarbij de proporties per kolom als coördinaten gebruikt worden. De vier toestanden liggen gezamenlijk in een drie-dimensionele deelruimte van de vijf-dimensionele ruimte. Het zwaartepunt heeft als coördinaten het gewogen gemiddelde van de vier toestanden, namelijk (.20, .20, .20, .20, .20). De assen van de vijf-dimensionele ruimte worden opgerekt met een factor $(1/p_i)^{1/2}$. Voor het voorbeeld betekent dit dat elke dimensie op dezelfde wijze wordt opgerekt, omdat $p_i = .20$ voor elke i. De chi-kwadraatafstand tussen toestand j en j^* wordt analoog aan die voor de afstanden tussen de rijen berekend, en ook de manier van roteren is gelijk aan die voor de rijen.

De oplossing voor de kolompunten is gegeven in figuur 1b. Daar is te zien dat de personen vergelijkbaar scoren op toestand h en i, en nogal verschillend op f en g; deze liggen immers ver van de oorsprong. Een beter begrip van deze oplossing ontstaat door deze oplossing in figuur 1b te relateren aan de oplossing voor de rijen in figuur 1a. De rijcoördinaten $r_{i\alpha}$ en de kolomcoördinaten $c_{j\alpha}$ zijn als volgt gerelateerd:

$$c_{j\alpha} = \lambda_\alpha^{-1} \sum_{i=1}^{I} p_{j|i} r_{i\alpha} \tag{3a}$$

en

$$r_{i\alpha} = \lambda_\alpha^{-1} \sum_{j=1}^{I} p_{i|j} c_{j\alpha} \tag{3b}$$

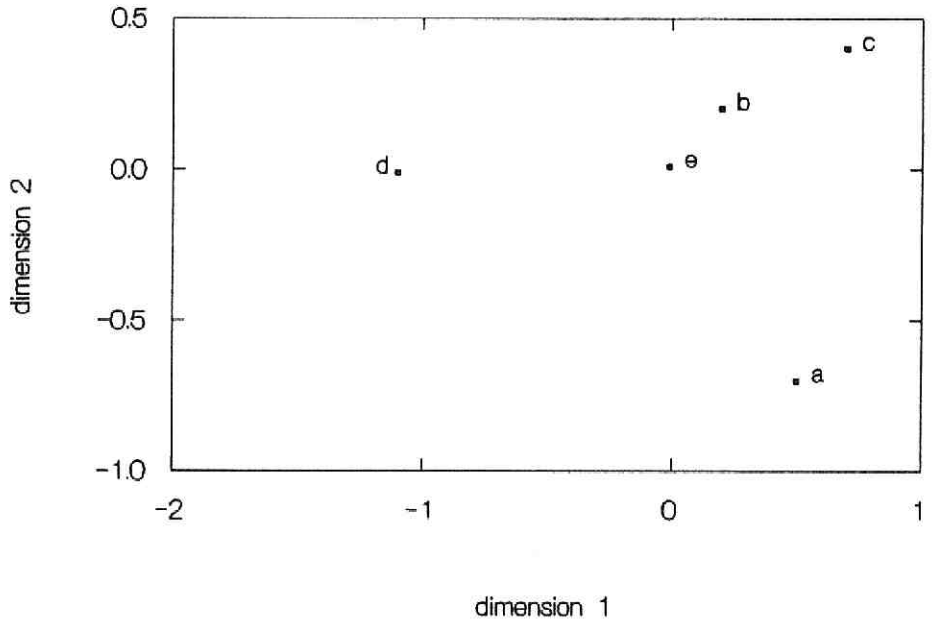


Fig. 1a. Weergave van de rijen in een twee-dimensionele ruimte.

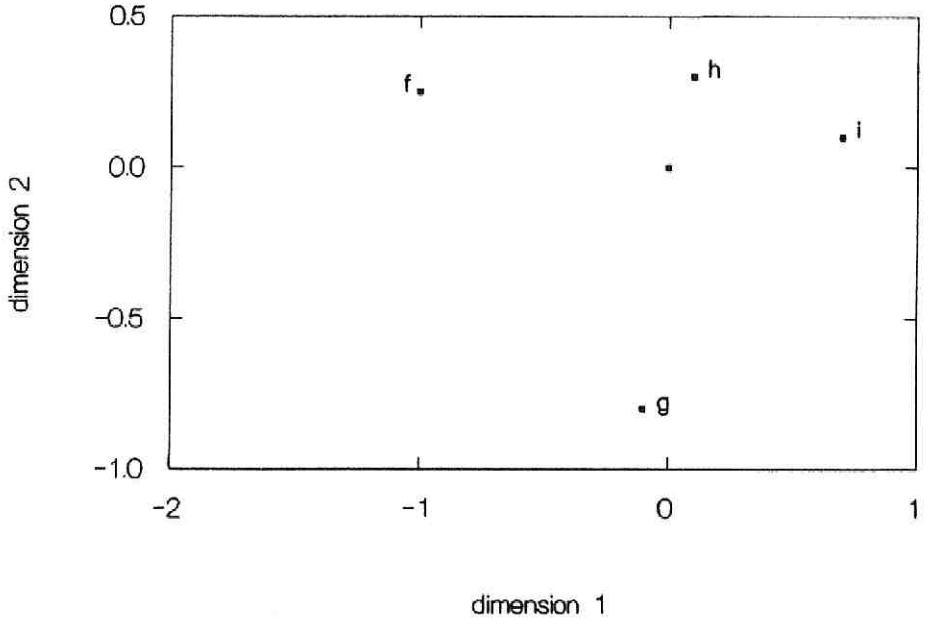


Fig. 1b. Weergave van de kolompunten in een twee-dimensionele ruimte.

Fig. 1. Afbeeldingen van respectievelijk de rij-punten (loopbanen) en kolom-punten (categorieën) van de gegevens in Tabel 1.

De eerste vergelijking laat zien dat, op een schaalfactor λ_α^{-1} na, de kolomcoördinaat $c_{j\alpha}$ voor kolom j op dimensie α te berekenen is als het gewogen gemiddelde van alle rijcoördinaten op dimensie α, waarbij de proporties p_{jli} als gewicht worden gebruikt. Op een schaalfactor na liggen de kolompunten dus in het gewogen gemiddelde van de rijpunten. De tweede vergelijking laat zien dat de rijpunten tegelijkertijd in het gewogen gemiddelde van de kolompunten liggen. Verder is het van belang te weten dat $\Sigma_i\, p_i r_{i\alpha} = 0$ en $\Sigma_j\, p_j c_{j\alpha} = 0$; indien het gemiddelde profiel gebruikt wordt voor het berekenen van een gewogen gemiddelde, dan geeft dit de coördinaten van de oorsprong. De interpretatie van de relatie tussen de oplossing voor de rijpunten

en die voor de kolompunten is nu eenvoudig. Ruwweg geldt dat, indien $p_{j|i} > p_j$ (en dan geldt tegelijkertijd dat $p_{i|j} > p_i$, want uit $p_{j|i} = p_{ij}/p_i > p_j$ volgt dat $p_{i|j} = p_{ij}/p_j > p_i$), dat rijpunt i dicht bij kolompunt j ligt. Dit is na te gaan in de Figuren 1a en 1b, gecombineerd met de Tabellen 1b en 1c. In Tabel 1b is te zien dat loopbaan a als speciaal kenmerk heeft dat de tweede toestand een veel grotere proportie heeft dan gemiddeld, namelijk .38 versus .15 (en in Tabel 1c is te zien dat .52 veel groter is dan .20). Daarom wijken loopbaan a en b op dezelfde wijze af van de oorsprong voor zover het de tweede dimensie betreft. Op dezelfde manier geldt dat de vierde loopbaan d een veel hogere proportie heeft op toestand 1, namelijk .67 versus .24.

Tot zover deze inleiding in correspondentie-analyse van loopbanen. Resumerend: correspondentie-analyse geeft een grafische representatie van loopbanen zoals die zijn samengevat zoals in Tabel 1a. Deze representatie laat zien in hoeverre loopbanen op elkaar lijken, en op welke punten ze overeenstemmen of juist verschillen. Alhoewel op deze manier al veel inzicht in de overeenkomsten en verschillen tussen loopbanen kan worden verkregen, zullen nu enkele manieren besproken worden om de informatie die aanwezig is in loopbaangegevens nog beter te benutten.

Verfijningen

In Tabel 1a zijn de gegevens van vijf loopbanen in twee jaar tijd samengevat. Indien er een duidelijke volgorde is in de toestanden en vrijwel iedereen dezelfde volgorde doorloopt, dan gaat hier geen informatie verloren. Als de volgorde in de tijd bijvoorbeeld altijd 'onderwijsvolgend', 'werkloos', 'niet bij opleiding passende arbeid', en 'wel bij de opleiding passende arbeid' is, dan kan uit de gegevens van Tabel 1a de loopbaan van elk van de vijf personen exact gereconstrueerd worden. In de praktijk is het verloop van loopbanen echter ingewikkelder; de volgorde van toestanden ligt niet vast, en men kan bijvoorbeeld korte perioden van werkloosheid afwisselen met perioden van passende en niet-passende arbeid. Dit kan niet gereconstrueerd worden aan de hand van Tabel 1a. Deze informatie is echter vaak wel beschikbaar, en in dat geval kan een meer verfijnde tabel dan Tabel 1a worden samengesteld. In Tabel 2 zijn bijvoorbeeld de gegevens van de vijf loopbanen uitgesplitst in twee jaren, die elk sommeren tot 52 weken. Als deze twee deeltabellen opgeteld worden, dan geeft dit Tabel 1a. Op de data in Tabel 2 is een correspondentie-analyse uit te voeren. Dit zal een oplossing opleveren met 5 loopbanen en 2x4 jaar-toestand combinaties.

Dit opsplitsen van de tabel is steeds verder door te voeren; uiteindelijk is het mogelijk een tabel van 5 loopbanen en 104x4 week-toestand combinaties te construeren. Deze matrix van 5x(104x4) wordt wel een super-indicatormatrix genoemd (Gifi, 1990), en correspondentie-analyse van een super-indicatormatrix staat ook wel bekend als *multiple* correspondentie-analy-

Tabel 2. Uitsplitsing van de gegevens in Tabel 1a naar jaar van observatie.

	Jaar 1				jaar 2				
toestand	f	g	h	i	f	g	h	i	totaal
resp a	0	30	12	10	0	10	0	42	104
resp b	14	4	30	4	0	6	10	36	104
resp c	0	0	24	28	0	0	0	52	104
resp d	52	0	0	0	18	17	17	0	104
resp e	42	10	0	0	0	0	10	42	104
	108	44	66	42	18	33	37	172	

Nb. f = onderwijsvolgend, g = werkloos, h = verricht niet bij de opleiding passende arbeid, i = verricht wel bij de opleiding passende arbeid.

se of HOMALS (HOMogeniteitsanalyse via Alternating Least Squares). Correspondentie-analyse van de 5x(104x4)-matrix zal echter nogal instabiele resultaten te zien geven. Het minder verfijnd representeren van de data, door bijvoorbeeld gebruik te maken van de 5x(2x4)-matrix zoals gegeven in Tabel 2, is een manier om tot een stabielere oplossing te komen. In dit licht is het van belang dat de correspondentie-analyse van de matrix van 5x(2x4) gelijk is aan de correspondentie-analyse van de matrix van 5x(104x4) met gelijkheidsrestricties op de toestandsscores, namelijk de restrictie dat toestandsscores voor een bepaalde toestand *binnen* een jaar aan elkaar gelijk zijn (zie Van der Heijden en De Leeuw, 1989).

Ontbrekende gegevens

In de praktijk heeft men vaak te maken met ontbrekende gegevens. De observatie van een loopbaan kan bijvoorbeeld later beginnen of eerder ophouden dan de observatie van andere loopbanen. In panel-onderzoek is onder andere vaak sprake van uitval (sommige personen werken wel aan de eerste meting mee, maar niet aan de volgende metingen); in dat geval is voor deze observaties sprake van een kortere waarnemingsperiode dan voor de overigen. Er zijn verschillende manieren om om te gaan met ontbrekende gegevens bij correspondentie-analyse (overzichten worden gegeven in Meulman, 1982, en Van der Heijden en Escofier, 1988). Hier zullen er twee besproken worden aan de hand van Tabel 1a.

Ten eerste, als bijvoorbeeld van de eerste loopbaan voor minder weken bekend is wat de toestand is - zodat bijvoorbeeld de gegevens respectievelijk 0, 40, 6, 40 zijn - dan analyseert men de matrix waarbij voor deze rij dan ook 0, 40, 6, 40 is ingevuld; er wordt dus geen enkele correctie toegepast. De analyse wordt verder op geen enkele manier aangepast, hoewel deze loopbaan een wat kleiner gewicht p_i zal hebben bij het bepalen van de assenstand. Ook bij het verder verfijnen van de analyse wordt op dezelfde wijze te werk gegaan. Zo wordt bij de superindicatormatrix voor die weken waarvoor geen gegevens bekend zijn (0,0,0,0) ingevuld.

Ten tweede, uitgaande van de gegevens 0, 40, 6, 40 voor de eerste loopbaan zijn er ontbrekende observaties voor 18 weken. Het is dan mogelijk om voor alle loopbanen een extra kolom die het aantal missende observaties aangeeft te definiëren, en vervolgens voor de eerste loopbaan in deze kolom '18' in te vullen. Dit is een goed alternatief indien de ontbrekende gegevens van verschillende loopbanen op dezelfde manier werden veroorzaakt, bijvoorbeeld, zij zijn allen later begonnen of eerder geëindigd. Het is echter duidelijk dat op dit gebied meer onderzoek gewenst is, bijvoorbeeld om te onderzoeken hoe 'linkse censurering', 'rechtse censurering' en andere vormen van ontbrekende gegevens het best behandeld kunnen worden.

Loopbanen als sequenties van toestanden

Bij de hier voorgestelde procedure om loopbanen te kwantificeren moet worden opgemerkt dat de informatie dat bepaalde toestanden volgen op of voorafgaan aan andere niet expliciet in de analyse wordt meegenomen; de kolommen van de te analyseren tabel kunnen met elkaar verwisseld worden, zonder dat hierdoor de uiteindelijke resultaten veranderen. Indien twee individuen zich op alle geobserveerde tijdstippen in dezelfde toestand bevinden - dus: voor alle kolommen dezelfde score hebben - zijn zij identiek. Naarmate er meer verschillen ontstaan zullen de twee loopbanen meer verschillend gekwantificeerd worden. Neem het volgende, tamelijk extreme, voorbeeld:

a b c d e f g h	(1)
b c d e f g h i	(2)
b c d e e f g h	(3)
j k l m n o p q	(4)

De eerste twee sequenties hebben op geen enkel tijdstip een element gemeen. De twee sequenties zijn echter bijna identiek; ze bevatten vrijwel dezelfde elementen in exact dezelfde volgorde. 'Zelfde tijd, zelfde toestand'-technieken zullen deze twee sequenties als verschillend beschouwen; het feit dat de volgorde van de gemeenschappelijke elementen gelijk is telt niet mee.

Deze twee sequenties hebben echter allebei 4 elementen gemeenschappelijk met sequentie 3, en zullen daarom beiden dicht bij 3 – en dus ook dicht bij elkaar – komen te liggen. Sequentie 4 heeft geen enkel element met de 3 voorgaande gemeenschappelijk, zelfs niet indien de volgorde van de elementen meegenomen wordt, en zal daarom ver van de andere 3 elementen worden geplaatst.

Dit voorbeeld toont aan dat het loutere feit alleen dat twee sequenties geen elementen op dezelfde tijdstippen gemeenschappelijk hebben niet hoeft te betekenen dat deze sequenties als zeer verschillend worden gezien; dat hangt in sterke mate af van de andere loopbanen in de data. Als er een grote samenhang is tussen de opeenvolgende categorieën van sequenties (zoals in de eerste drie sequenties van ons voorbeeld), heeft dit tot gevolg dat deze sequenties als ongeveer gelijk zullen worden gekwantificeerd. Met andere woorden, het feit dat er een bepaalde volgorde in de elementen van de sequenties zit heeft wel degelijk invloed op de resultaten.

Tenslotte

Correspondentie-analyse van loopbanen, of meer in het algemeen van tijdbudgetten en event-histories werd voor het eerste als zodanig voorgesteld door Saporta (1981, 1985) en Deville en Saporta (1980, 1983). Zij behandelen dit onderwerp als een vorm van harmonische analyse voor nominale tijdreeksen, en benadrukken geometrische representaties van dit type gegevens. De Leeuw, Van der Heijden en Kreft (1985) en de Leeuw en Kreft (1985) benadrukken eerder de interpretatie van homogeniteitsanalyse en de mogelijkheid die deze interpretatie biedt om te komen tot een kwantificatie van loopbanen. Dit is vooral zinvol indien de afstanden tussen de loopbanen zonder al te veel verlies van informatie in 1 dimensie afgebeeld kunnen worden.

3. TOEPASSING: MULTIPLE CORRESPONDENTIE-ANALYSE VAN SCHOOL/WERKLOOPBANEN

Vraagstelling en opzet

In het tweede deel van dit paper zal correspondentie-analyse gebruikt worden voor de analyse van de school/werkloopbanen van 265 Nederlandse jongeren gedurende de periode 1987-1989. De gegevens werden eerder geanalyseerd door Claes, Quintanilla & Whitely (1992), wier onderzoek als doel had te beschrijven via welke paden mensen in een bepaalde toestand (een vaste baan in hetzij de administratieve, hetzij de metaalsector) terechtkomen.

De rest van deze sectie is gesplitst in drie delen. In het eerste deel worden kort de gebruikte gegevens en de gevolgde procedure beschreven. In het tweede deel komen de mogelijkheden van correspondentie-analyse bij de analyse van loopbanen aan de orde. Getoond zal worden hoe, op basis van afbeeldingen van categorieën van variabelen en personen, inzicht in de data verkregen kan worden. Toetsingen van hypothesen over de structuur van de data blijven hier echter achterwege. In het derde deel wordt eerst op basis van de resultaten van de correspondentie-analyse een classificatie van loopbanen geconstrueerd. Daarna wordt nagegaan hoe binnen elk van deze clusters de loopbaan verloopt.

Methode: data en procedure

De gegevens werden gedurende de herfst van 1989 verzameld in het kader van een internationaal onderzoek betreffende de werksocialisatie van jongeren (WOSY International Research Group, 1989; Whitely, Peiró & Sarchielli, 1992). In dit artikel worden alleen de Nederlandse data gebruikt. De te analyseren steekproef bestaat uit 265 mensen die ten tijde van het interview ongeveer 6 maanden in hun toenmalige (eerste) baan werkzaam waren. De gemiddelde leeftijd van de respondenten bedroeg 19,7 jaar. De helft van de onderzoeksgroep was werkzaam in beroepen waarbij in principe intensief met computers omgegaan wordt (administratief medewerkers, leerling-programmeurs en dergelijke). De andere helft bestond uit personen werkzaam in de metaal (lassers, bankwerkers en dergelijke).

Gedurende het interview werd informatie verzameld over de meningen, waarden en het gedrag ten aanzien van werk, alsmede over de school/werkloopbaan over de 2 jaar voorafgaande aan het interview. Voor ieder kwartaal moest worden aangegeven wat men hoofdzakelijk had gedaan, wat in totaal dus 8 tijdspunten oplevert. De mogelijke categorieën op deze tijdspunten waren: 'tijdelijke baan', 'vaste baan', 'in opleiding in verband met huidige baan', 'in opleiding, niet in verband met huidige baan', 'werkloos', 'in militaire dienst', en 'anders' (bijvoorbeeld ziek). Van de 265 personen hadden er 23 missende waarden op één of meer tijdspunten. Hiervoor werd op de betreffende variabelen de waarde 0 ('ontbrekend') ingevuld. De 8 tijdstippen waren daarmee te coderen in een super-indicatormatrix, die met behulp van multiple correspondentie-analyse (via het SPSS-programma HOMALS) werd geanalyseerd. De eerste 5 eigenwaarden van de oplossing bedroegen respectievelijk .60, .52, .41, .28, en .25. Op basis van het 'knik'-criterium (Cattell, 1966) werd de analyse tot de eerste 3 dimensies beperkt.

Resultaten

Vervolgens werd bekeken of de categorieën van de variabelen (in de kolommen) verschillend gerelateerd zijn aan de rijen (de loopbanen). Een afbeelding, omwille van de interpreteerbaarheid beperkt tot slechts 2 dimensies, van de lokatie van de categorieën van de variabelen is gegeven in figuur 2.

Uit deze figuur blijkt dat enkele categorieën sterk onderscheiden tussen de kolommen, terwijl andere categorieën aanzienlijk minder belangrijk zijn. Globaal kan gezegd worden dat de latere categorieën (gemeten op tijdstip 7 en 8) dichter bij het nulpunt liggen dan de overige categorieën. Dit geeft aan dat deze twee tijdstippen minder belangrijk zijn in het maken van onderscheid tussen de rijen. Omdat de steekproef bestaat uit mensen die in principe 6 maanden in de huidige baan werkzaam waren, is dit geen verrassend resultaat.

Verder zijn er ook enkele clusters van categorieën te zien. Het hebben van een vaste baan op de tijdstippen 1 tot en met 6 ligt bijvoorbeeld links onder in Figuur 2, vlak bij elkaar; dit geeft aan dat diegenen die op tijdstip 1 een vaste baan hebben, op tijdstip 2 tot en met 6 zeer waarschijnlijk ook nog een vaste baan zullen hebben. Het in opleiding zijn, maar niet in verband met de huidige baan, ligt voor alle tijdstippen ongeveer bovenin Figuur 2; het in opleiding zijn in verband met huidige baan ligt rechts van het nulpunt; en het hebben van een tijdelijke baan ligt ruwweg in het midden van Figuur 2. Het in militaire dienst zijn, het werkloos zijn en de categorie 'anders' liggen voor de verschillende tijdstippen relatief ongestructureerd door elkaar, in de rechterbovenhoek van de figuur. Hieruit kan geconcludeerd worden dat men relatief gemakkelijk wisselt tussen deze drie toestanden.

In Figuur 3 is de configuratie van de rij-punten (loopbanen) gegeven. Voor onze doeleinden is dit de belangrijkste figuur, omdat hieruit afgelezen kan worden welke personen een vergelijkbare loopbaan hebben gevolgd.

Figuur 3 laat zien dat er enkele clusters van rij-punten kunnen worden onderscheiden. Een groot aantal rij-punten is geconcentreerd in de rechterhelft van Figuur 3, wat aangeeft dat deze loopbanen sterk op elkaar lijken. Daarnaast kan een klein cluster linksonder in Figuur 3 onderscheiden worden; deze punten liggen zeer dicht bij elkaar, maar ver uit de buurt van het nulpunt en de meeste andere punten. Bovenin is ook een cluster zichtbaar, dat weliswaar intern niet zo homogeen is als het cluster rechts, maar ook niet bij het grote cluster lijkt te horen. Tussen deze relatief sterke clusters zweven nog enkele andere punten. De bij deze punten behorende mensen hebben blijkbaar weinig voorkomende en onderling sterk verschillende loopbanen gevolgd.

Classificatie van loopbanen

Globaal geeft Figuur 3 dus de indruk dat het zinvol zou kunnen zijn om zo'n 4 clusters te onderscheiden. Waar precies de grenzen van deze clusters liggen is echter niet helemaal duidelijk; het is niet mogelijk om deze clusters nu reeds te onderscheiden. Daarom is eerst op basis van de via correspondentie-analyse verkregen waarden voor de personen op de eerste drie dimensies een cluster-analyse uitgevoerd, zodat er besloten kan worden welke punten nog wel, en welke punten niet in een bepaald cluster thuishoren.[1] Uiteindelijk werden er op basis van een

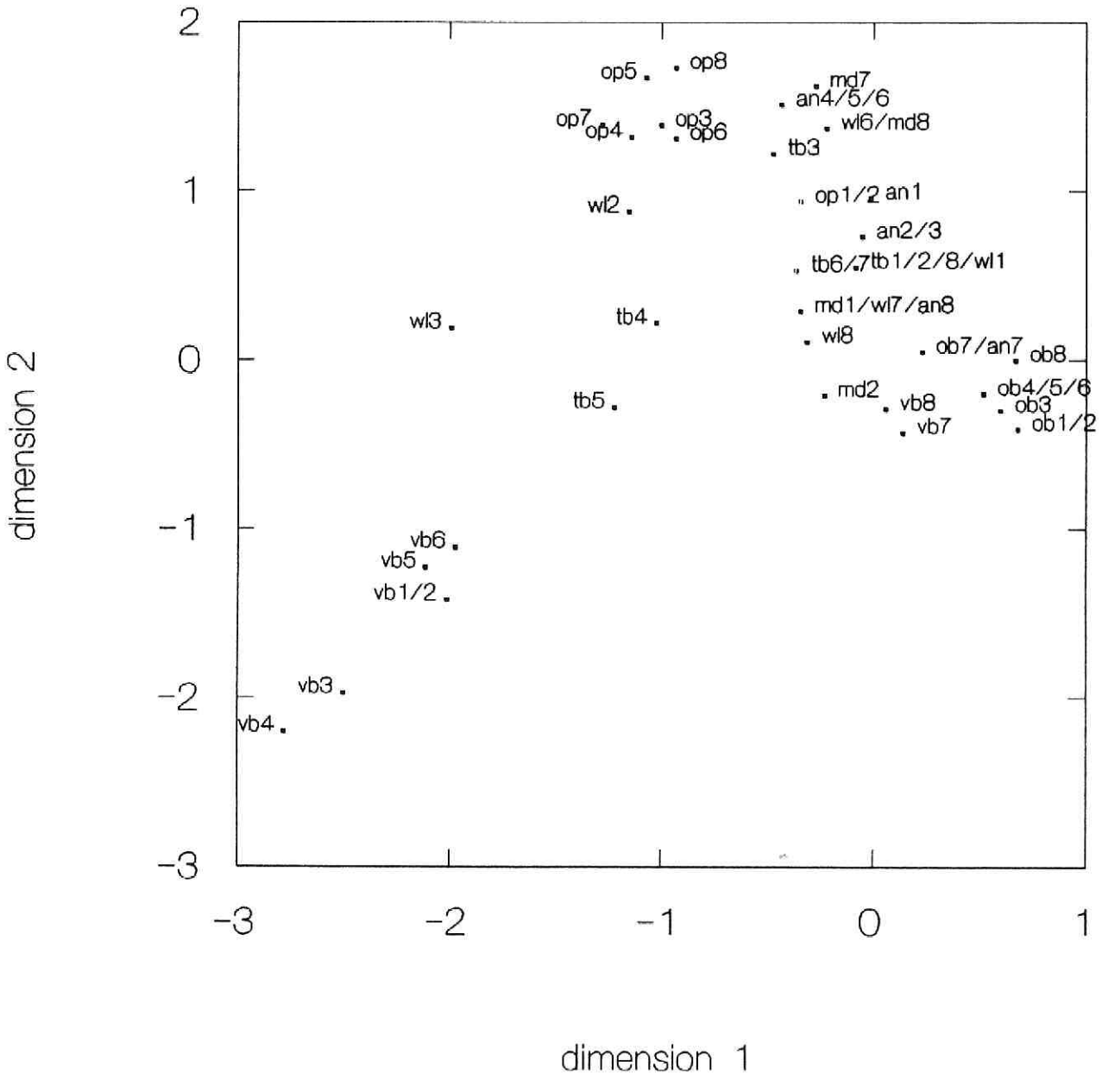


Fig. 2. Afbeelding van de categorieën van de variabelen in een twee-dimensionale ruimte, op basis van via MCA verkregen schaalwaarden.
Nb. vbi = vaste baan op tijdstip i, op = in opleiding (niet in verband met baan), ob = in opleiding (in verband met baan), md = in militaire dienst, wl = werkloos, vb = vaste baan, tb = tijdelijke baan, an = anders. Zeer dicht bij elkaar liggende punten zijn ter wille van de leesbaarheid samengevoegd.

nearest-neighbour cluster-analyse van de gekwadrateerde euclidische afstanden zoals berekend op basis van de eerste drie HOMALS-dimensies 4 verschillende clusters van loopbanen geconstrueerd, die zowel empirisch – op basis van de toename van de error sum of squares in de cluster-analyse – als inhoudelijk van elkaar te onderscheiden leken te kunnen worden. In Figuur 3 is ruwweg aangegeven welke mensen uiteindelijk in een bepaald cluster ingedeeld werden. In Figuur 4 is de indeling in clusters inzichtelijk gemaakt, door in diagrammen voor elk van de clusters aan te geven wat de personen op de verschillende tijdstippen deden. Tabel 3 tenslotte geeft de gemiddelden en standaardafwijkingen van de 4 clusters op elk van de 3 MCA-dimensies.

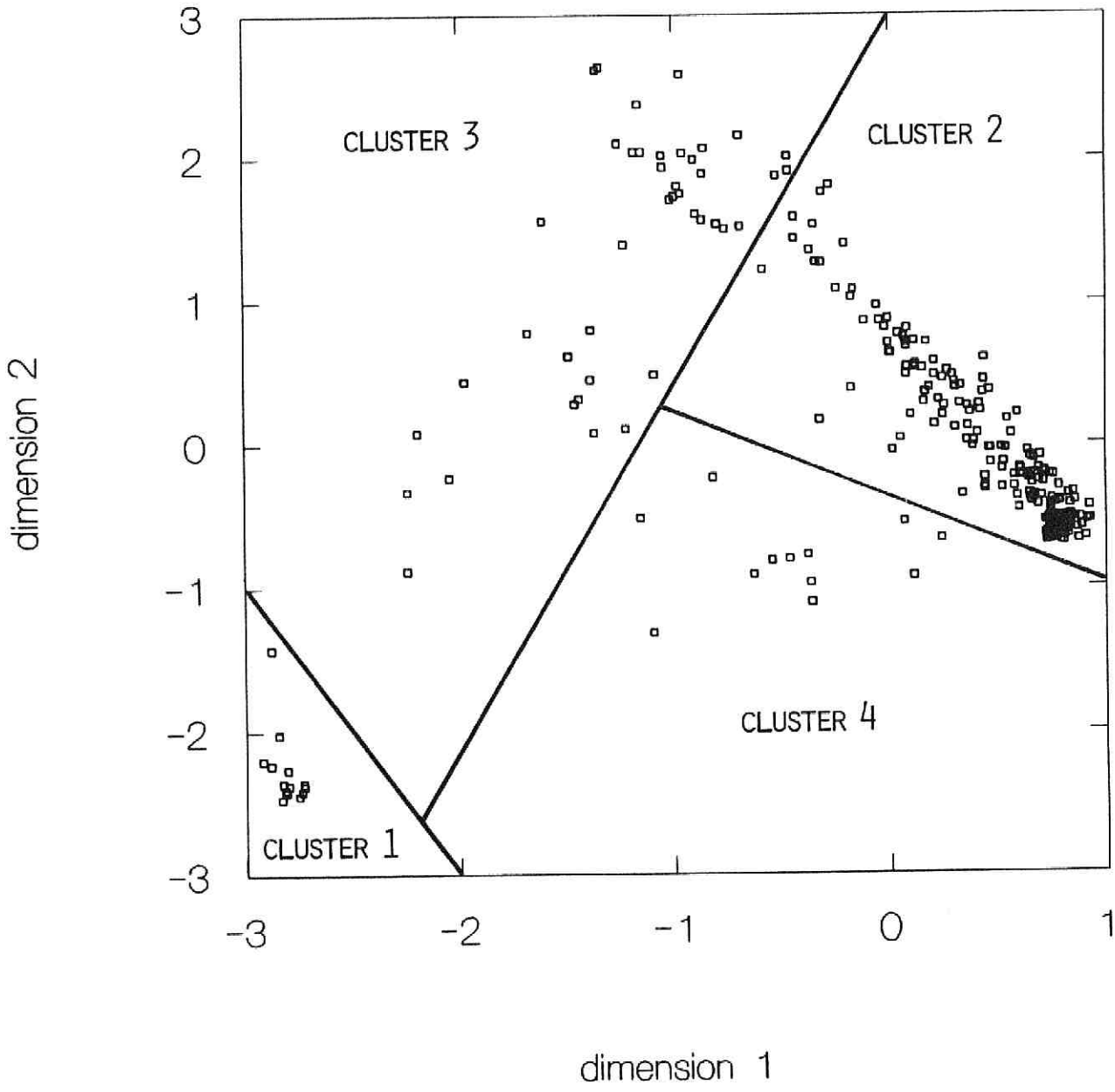


Fig. 3: Afbeelding van de 265 rijpunten (loopbanen) in een twee-dimensionale ruimte, op basis van de via MCA verkregen schaalwaarden.

Tabel 3 bevestigt globaal de indruk van figuur 3. Het is te zien dat de clusters wat betreft hun gemiddelden op de eerste drie MCA-dimensies vooral verschillen op de eerste twee dimensies, terwijl de derde dimensie eigenlijk weinig onderschied maakt tussen de groepen. Met betrekking tot de standaardafwijkingen kan opgemerkt worden dat cluster 1 een relatief zeer homogeen cluster betreft. Met name cluster 3 is intern tamelijk los, zoals blijkt uit de standaardafwijkingen die op alle drie de dimensies relatief erg groot zijn.

In Figuur 4a zijn de marginale verdelingen van alle personen op de verschillende tijdstippen weergegeven. Op tijdstip 1 (T1) zijn bijvoorbeeld 38 mensen werkzaam in een vaste baan; 125 mensen bevinden zich in opleiding, in verband met hun huidige baan; een andere groot deel (83 personen) is wel in opleiding, maar niet in verband met hun huidige baan; de overige categorieën zijn klein en onbetekenend. Op tijdstip 8 (T8) heeft het grootste deel van de mensen (156 personen) een vaste baan, een ander deel (46 personen) een tijdelijke baan; de andere categorieën zijn weer relatief klein. De globale indruk van deze figuur is dat de steekproef overgaat van

Tabel 3. Gemiddelden en standaardafwijkingen van de vier onderscheiden clusters op de eerste drie MCA-dimensies.

	dimensie 1		dimensie 2		dimensie 3	
	gem.	SD	gem.	SD	gem.	SD
cluster 1	-2.925	.019	-2.299	.115	.068	.065
cluster 2	.508	.257	-.120	.432	-.010	.522
cluster 3	-1.199	.584	1.509	1.011	-.025	2.230
cluster 4	-.438	.367	-.801	.252	-.006	.249

'in opleiding' naar 'werkend'. Figuur 4a geeft een weergave van de toestand van de totale groep op elk van de verschillende tijdstippen; dit is in feite te beschouwen als een representatie van de marginaal van de te analyseren kruistabel (zie de paragraaf 'verfijningen').

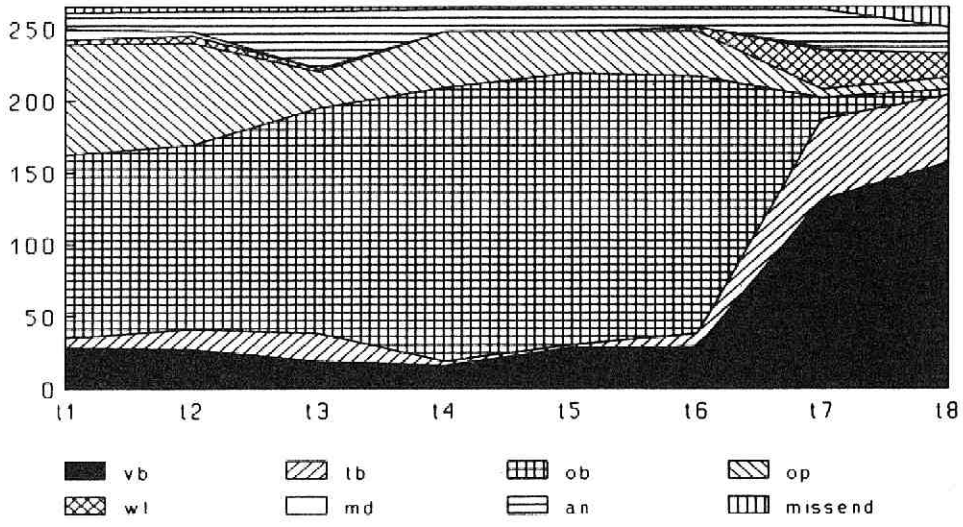
Deze grote groep personen werd op basis van de clusteranalyse opgedeeld in 4 verschillende groepen. Er is een kleine groep die gedurende de geobserveerde periode vrijwel constant een vaste baan had (figuur 4b); een tweede, zeer grote groep (N = 195, figuur 4c) blijkt verantwoordelijk voor de in het totaalbeeld optredende indruk van overgang van opleiding (in verband met de huidige baan) naar werk. Dit is het patroon dat gezien de opzet van het onderzoek verwacht mocht worden. Veel van de personen in deze steekproef komen immers via het leerlingstelsel aan hun baan. Er is een derde groep van redelijke omvang (N = 44, figuur 4d), die van groep 2 onderscheiden kan worden doordat ze in de middelste perioden in opleiding, *niet* in verband met de huidige baan zijn. Tenslotte is er nog een klein groepje (N = 12, figuur 4e) dat zich van de rest onderscheidt doordat zij relatief vaak in de categorie 'anders' vallen. Vermoed wordt dat het hier gaat om mensen die bijvoorbeeld langdurig ziek zijn geweest. De totale groep van 265 personen is dus op te delen in een grote groep, die van opleiding (in verband met de huidige baan) naar een - al dan niet vaste - baan gaat; daarnaast zijn er 3 andere, kwantitatief minder belangrijke patronen te onderscheiden.

4. SAMENVATTING, CONCLUSIES EN DISCUSSIE: CORRESPONDENTIE-ANALYSE VOOR DE ANALYSE VAN LOOPBANEN?

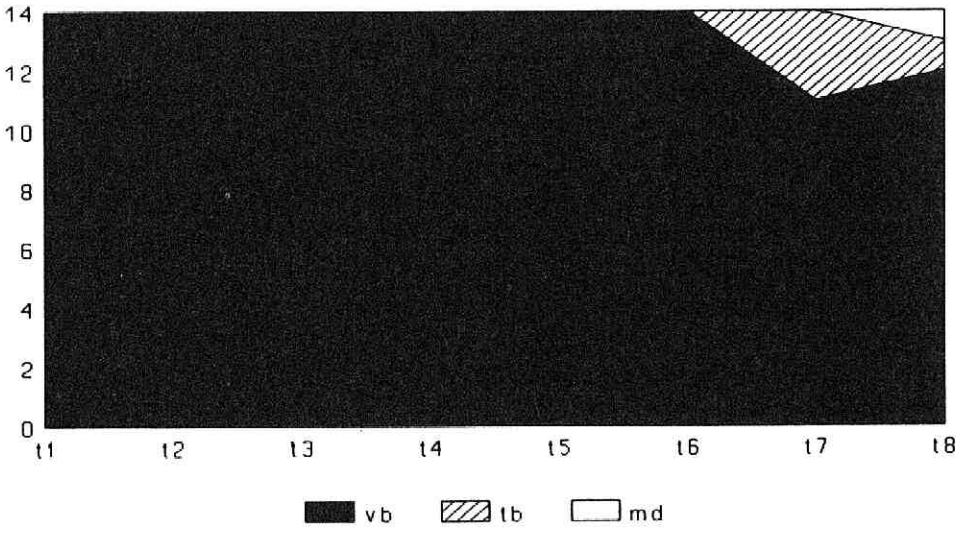
In dit artikel werd een techniek beschreven die gebruikt kan worden voor de analyse van loopbanen. Allereerst werden de principes van deze techniek, multiple correspondentie-analyse, uiteen gezet. Multiple correspondentie-analyse is geschikt voor de analyse van nominale variabelen die weergegeven zijn in een tweedimensionale kruistabel. Sequenties van toestandsovergangen (loopbanen) kunnen ook in een dergelijke matrix gecodeerd worden. Vervolgens werd, met behulp van gegevens over de opleiding- en werkloopbaan van 265 Nederlandse jongeren, getoond hoe correspondentie-analyse gebruikt kan worden om inzicht te verkrijgen in het verloop van loopbanen. Allereerst werden op basis van via correspondentie-analyse verkregen afstanden tussen categorieën en personen twee-dimensionale afbeeldingen geconstrueerd. Daarna werd, op basis van de via correspondentie-analyse verkregen afstanden tussen personen, een clusteranalyse uitgevoerd, die leidde tot een variabele 'loopbaantype'. Deze kan in vervolganalyses gebruikt worden als te verklaren, dan wel als verklarende variabele.

Uit deze exercities bleek dat via correspondentie-analyse inzicht verkregen kan worden in de structuur van sequentiële data. Duidelijk werd welke variabelen en categorieën het meest bijdroegen aan het onderscheid tussen de loopbanen van de personen. Geconcludeerd kan worden dat correspondentie-analyse ook voor de in loopbanen geïnteresseerde onderzoeker een interes-

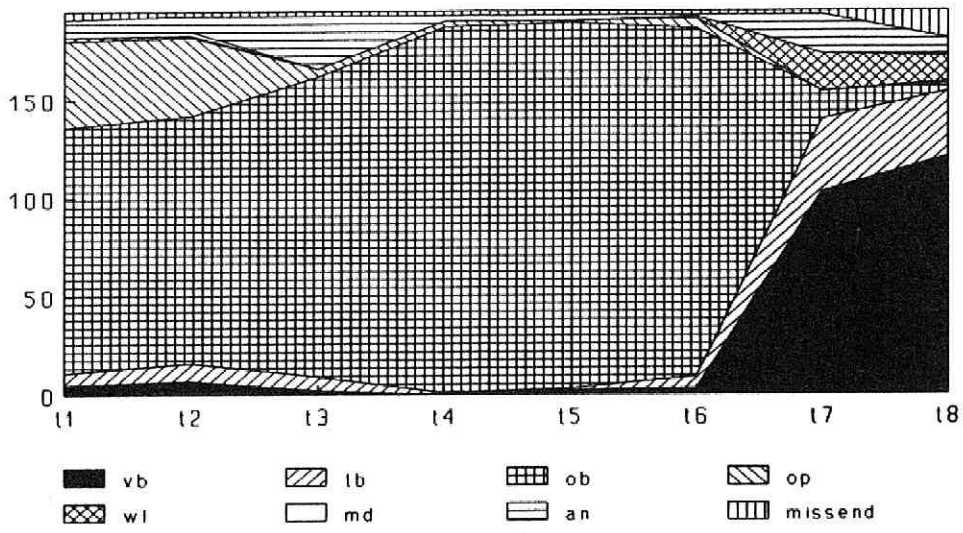
4a: totale groep (N = 265)



4b: groep 1 (N = 14)



4c: groep 2 (N = 195)



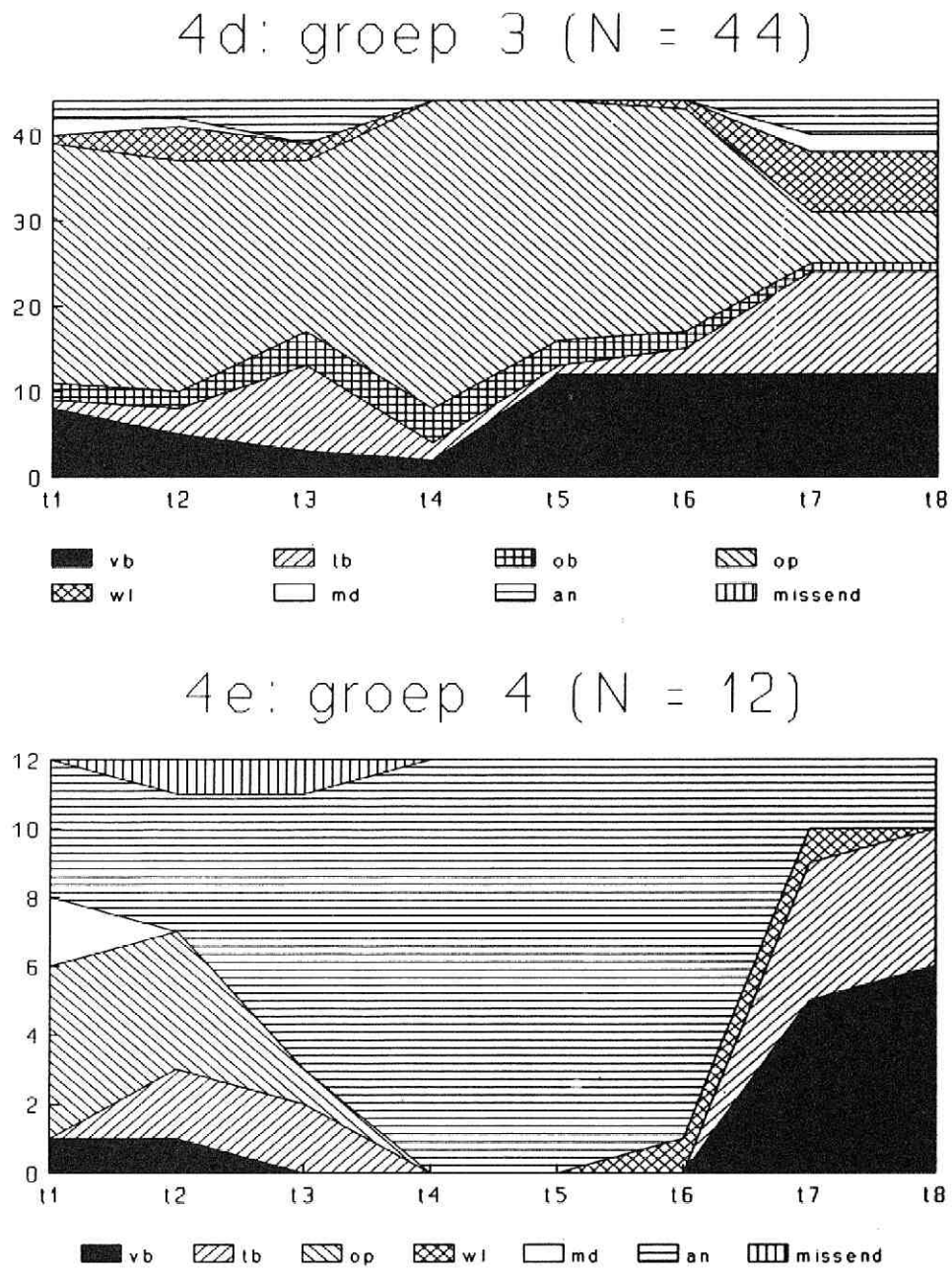


Fig. 4. De loopbaanpatronen van respectievelijk de totale steekproef (4a) en de 4 via MCA daaruit afgesplitste deelgroepen (4b-e). Nb. vb = vaste baan, tb = tijdelijke baan, wl = werkloos, op = opleiding, niet in verband met baan, ob = opleiding in verband met baan, md = in militaire dienst, an = overig, missend = geen score voor desbetreffende periode bekend.

sante techniek kan zijn, waarbij vooral veel inzicht verkregen kan worden in de mate waarin bepaalde variabelen/categorieën van variabelen bijdragen aan het onderscheid tussen personen.

Er zijn ook enkele nadelen te noemen van de correspondentie-analyse van loopbaangegevens. Allereerst is er het probleem van de wijze waarop er in MCA met missende gegevens omgegaan wordt. Als de score op een bepaald tijdspunt voor een bepaalde persoon ontbreekt, kan er een aantal oplossingen getroffen worden; het is echter niet duidelijk in hoeverre deze de resultaten systematisch beïnvloeden. Dit probleem is voor het besproken type data tamelijk belangrijk omdat de gegevens vaak verzameld worden via een panel-design, waarbij de infor-

matie over tussenliggende perioden meestal via retrospectieve vragen verzameld zal worden. Bij een grotere periode tussen de metingen van het panel zal het voor deze persoon echter steeds moeilijker worden voor alle tijdspunten aan te geven, wat hij/zij op het betreffende punt deed. Dat kan als gevolg hebben dat er relatief veel missende waarden in zulke datasets zullen voorkomen. Er zijn meerdere oplossingen voor dit probleem denkbaar, maar het is nog niet duidelijk in hoeverre deze de analyses systematisch beïnvloeden. Hierbij moet opgemerkt worden dat dit type problemen ook voor andere technieken speelt, en zeker niet uniek voor MCA is.

Ten tweede het ordeningsprincipe: er worden niet zozeer *loopbanen* met elkaar vergeleken worden, maar alleen *toestanden* op specifieke tijdspunten. De kolommen van de te analyseren tabel kunnen met elkaar verwisseld worden zonder dat de resultaten veranderen; de informatie over de volgorde van toestanden wordt dus strikt genomen niet meegenomen. Bij de bespreking van de toepassing bleek overigens dat het sequentiële karakter van de gegevens wel degelijk de resultaten zal beïnvloeden: over het algemeen zal gelden dat dezelfde categorieën op verschillende tijdstippen dichter bij elkaar zullen liggen dan verschillende categorieën op dezelfde tijdstippen, en zullen gelijksoortige sequenties – zelfs zonder dat ze elementen op bepaalde tijdstippen gemeenschappelijk hebben – als ongeveer gelijk worden gekwantificeerd.

Het is de vraag in hoeverre deze kwestie potentiële gebruikers ervan moet weerhouden om MCA te gebruiken bij de analyse van hun gegevens. Zoals gezegd is MCA zeer wel in staat verschillende loopbaantypen te onderscheiden; de in dit paper gegeven toepassing illustreert deze uitspraak. Daarnaast zijn er nauwelijks praktisch toepasbare alternatieven die dit nadeel niet hebben. Een andere 'zelfde tijd, zelfde toestand'-technieken als directe clustering van de toestanden via dummy-variabelen (voor het voorbeeld hierboven zouden dat 56 variabelen worden: 7 dimensies voor de toestanden * 8 tijdstippen) betrekt evenmin de volgorde van toestanden expliciet in de analyse, en geeft aanzienlijk minder inzicht in de structuur van de gegevens. De door Abbott & Hrycak (1990) voorgestelde 'optimal matching'-procedure (die met name de volgorde van toestanden vergelijkt om tot afstanden tussen loopbanen te komen) lijkt vooralsnog niet veel meer dan een aansprekend idee. Hier ontbreekt echter de ruimte om in te gaan op de voor- en nadelen van deze alternatieve benaderingen; voor een uitgebreid overzicht van deze en andere benaderingen van de analyse van loopbaangegevens wordt verwezen naar Taris (in voorbereiding).

Als laatste nadeel moet nog genoemd worden dat (multiple) correspondentie-analyse een tamelijk exploratieve methode is; er worden geen betrouwbaarheidsintervallen met betrekking tot de locatie van de rij- dan wel kolompunten geconstrueerd. Inzicht in de stabiliteit van de oplossing kan echter verkregen worden via bootstrap- of jackknife-analyses (Efron, 1981: Gifi, 1990).

DANKWOORD

Deze publicatie is gebaseerd op data die verzameld werden in het kader van een internationaal onderzoek betreffende de werksocialisatie van jongeren (WOSY). Het Nederlandse WOSY-project maakt deel uit van het onderzoeksprogramma 'Het proces van socialisatie van jong-volwassenen'. Dit onderzoeksprogramma wordt uitgevoerd aan de Vrije Universiteit te Amsterdam door de vakgroepen Methoden en Technieken van Sociaal-Wetenschappelijk Onderzoek (faculteit SCW) en de vakgroep Arbeids- en Organisatiepsychologie (faculteit FPP). De auteurs danken de leden van de internationale WOSY-projectgroep voor het mogen gebruiken van hun data, Jan van Gastel voor zijn aandeel in de verzameling van de Nederlandse data, en twee anonieme beoordelaars van dit tijdschrift voor hun opbouwende kritiek op een eerdere versie van dit stuk.

NOOT

1. In Figuur 2 en 3 werden de lokatie van de rij- en kolompunten op de eerste 2 dimensies getoond. Omdat de derde dimensie gezien diens eigenwaarde wel degelijk van belang is werd besloten ook deze bij de

clusteranalyse te betrekken. Opgemerkt kan nog worden dat het empirisch weinig bleek uit te maken of de oplossing op basis van 2, dan wel 3 dimensies werd uitgevoerd. Dit is niet zo heel verwonderlijk gezien het feit dat elke volgende dimensie minder informatie bevat dan de voorgaande; zie ook Tabel 3.

LITERATUUR

Abbott, A. & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology, 96,* 144-185.

Allison, P.D. (1984). *Event history analysis: Regression for the social sciences.* Beverly Hills: Sage.

Blossfeld, H.P., Hamerle, A. & Mayer, K.U. (1989). *Event history analysis: Statistical theory and application in the social sciences.* New York: Lawrence Erlbaum.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioural Research, 1, 245-276.*

Claes, R., Quintanilla, A.R. & Whitely, W. (1992). Career preparation patterns. *Revue Internationale du Psychologie Sociale, 5, 1,* 37-60.

Deville, J.-C. & Saporta, G. (1980). Analyse harmonique qualitative. In E. Diday (ed.), *Data analysis and informatics.* Amsterdam: North-Holland.

Deville, J.-C. & Saporta, G. (1983). Correspondence analysis, with an extension towards nominal time series. *Journal of Econometrics, 22,* 169-189.

Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other methods. *Biometrika, 68,* 589-599.

Gifi, A. (1990). *Non-linear multivariate analysis.* New York: John Wiley & Sons.

Heijden, P.G.M. van der & Escofier, B. (1988). *Multiple corresponding analysis with missing data.* Rennes: Institut National de Recherche en Informatique et en Automatique.

Heijden, P.G.M. van der & Leeuw, J. de (1989). Correspondence analysis, with special attention to the analysis of panel data and event history data. In C.C. Clogg (ed.), *Sociological Methodology 1989.* Oxford: Basil Blackwell.

Leeuw, J. de, Heijden, P.G.M. van der, & Kreft, I. (1985). Homogeneity analysis of event history data. *Methods of Operations Research, 50,* 299-316.

Leeuw, J. de & Kreft, I. (1985). *De definitie en kwantifikatie van de schoolloopbaan.* Leiden: Vakgroep Datatheorie, RR-85-20.

Meulman, J. (1982). *Homogeneity analysis of incomplete data.* Leiden: DSWO-Press.

Van de Pol, F.J.R. (1989). *Issues of design and analysis of panels.* Amsterdam: Sociometric Research Foundation.

Saporta, G. (1981). *Méthodes exploratoires d'analyse de données temporelles.* Ongepubliceerde doctoraalscriptie. Parijs: Université P. et M. Curie.

Saporta, G. (1985). Data analysis for numerical and categorical individual time-series. *Applied Stochastic Models and Data Analysis, 1,* 109-119.

Taris, T.W. (in voorbereiding). *Analysis of career data from a socialization perspective.* Amsterdam, vakgroep Arbeids- en Organisatiepsychologie (academisch proefschrift).

Tuma, N.B. & Hannan, M.T. (1984). *Social dynamics.* New York: Lawrence Erlbaum.

Whitely, W., Peiró, J.M. & Sarchielli, G. (1992). WOSY theoretical framework, research methodology and potential implications. *Revue Internationale du Psychologie Sociale, 5, 1,* 9-36.

WOSY International Research Group (1989). Socialización del joven: Un estudio transnational (Work socialization of youth: A transnational study). Papeles del Psicológico, 39/40, 32-35.

Manuscript ontvangen 27-1-1993
Definitieve versie ontvangen 30-3-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 42-51

De veronderstelde bekwaamheid als voorspeller van studieresultaten

A. Vrugt, M.Ph. Langereis en Joh. Hoogstraten
*Faculteit der Psychologie, Universiteit van Amsterdam**

ABSTRACT

In this article a study is described which was conducted to examine the relationship between academic perceived self-efficacy, conceptions about the learnability of capabilities required for the task, and performance. We measured academic perceived self-efficacy, grade goals, attributions of possible failure to lack of talent, intelligence and course grades in three domains: Introductory psychology, Personality psychology, and Social psychology. Further, the conceptions about the learnability of the capabilities required for the study of psychology were measured. The results showed that self-efficacy magnitude and grade goals for Introductory psychology contributed significantly to the grades obtained for this course. Moreover, self-efficacy magnitude and the conceptions about the learnability of required capabilities simultaneously contributed to the attributions to lack of talent. For Introductory psychology these factors also contributed to course grades.

INLEIDING

Bandura's theorie (1986, 1989) over waargenomen eigen competentie ("perceived self-efficacy") beschrijft op welke wijze iemands denkbeelden over zijn competentie zijn prestaties beïnvloeden, en door welke factoren deze denkbeelden beïnvloed kunnen worden. Het begrip "waargenomen eigen competentie" wordt door Bandura (1986) als volgt omschreven: "... people's judgement of their capabilities to organize and execute courses of action required to attain designated types of performances" (p. 395). Uit deze omschrijving komt naar voren dat het subjectieve oordeel over de eigen capaciteiten domeinspecifiek is. Het gaat hier dus niet om een persoonlijkheidskenmerk, zoals eigenwaarde. Dit betekent ook dat iemand zich op een bepaald domein, zoals het studeren voor tentamens, heel competent, en op een ander domein, zoals het schrijven van een paper, veel minder competent kan achten. Uit Bandura's omschrijving kan ook afgeleid worden dat het oordeel over de eigen competentie tot stand komt in relatie tot een bepaald prestatieniveau. Iemand kan zichzelf bij voorbeeld heel goed in staat achten een taak op gemiddeld niveau uit te voeren, maar menen dat hij minder goed in staat is een excellente prestatie op dezelfde taak te leveren.

Onderzoek heeft uitgewezen dat de denkbeelden over de eigen competentie op een bepaald terrein van invloed zijn op de prestaties die behaald worden. Bij Wood en Locke (1987) blijkt dat de waargenomen eigen competentie de studieprestaties beïnvloedt, en Taylor, Locke, Lee en Gist (1984) hebben aangetoond dat de waargenomen eigen competentie samenhangt met de wetenschappelijke productiviteit. Uit studies van Hill, Smith en Mann (1987) en Gist, Schwoerer en Rosen (1989) komt naar voren dat de waargenomen eigen competentie eveneens van invloed is op het werken met computers en op het goed gebruiken van computerprogramma's.

De wijze waarop de waargenomen eigen competentie de prestaties beïnvloedt, laat zich als volgt beschrijven. Als een persoon er gunstige denkbeelden over de eigen competentie op nahoudt, zal hij een relatief hoog prestatieniveau of uitdagend doel nastreven (Locke & Latham, 1990). Hij zal zich niet makkelijk uit het veld laten slaan, zich inspannen, zoeken naar nieuwe oplossingen en ook bij een moeilijke opgave volharden (Cervone & Peak, 1986). Deze inspanning, strategie-ontwikkeling en volharding brengen de feitelijke competentie tot ontwikkeling,

*Adres: Universiteit van Amsterdam, Faculteit der Psychologie, Roetersstraat 15, 1018 WB Amsterdam.

en leiden tot goede prestaties. Als iemand er ongunstige denkbeelden over de eigen competentie op nahoudt, zal hij een minder hoog prestatieniveau nastreven. Twijfel en onzekerheid tijdens de taakuitvoering tasten de concentratie aan, en bij een moeilijke taak wordt de moed vrij snel opgegeven. Al met al komt de feitelijke competentie dan niet erg tot ontwikkeling, en dat reflecteert zich in minder goede prestaties (Bandura, 1989).

Om de bijdrage van de waargenomen eigen competentie aan prestaties te kunnen bepalen, moet in onderzoek worden gecontroleerd voor de feitelijke bekwaamheid. Zo heeft Collins (1982, geciteerd in Bandura, 1986, p. 391) de feitelijke bekwaamheid en de waargenomen eigen competentie van leerlingen op het gebied van wiskunde gemeten. Vervolgens hebben de leerlingen een wiskundetaak uitgevoerd. Bij eenzelfde niveau van feitelijke bekwaamheid leverden leerlingen die zichzelf competenter achtten betere prestaties dan leerlingen die zichzelf minder competent achtten. Dit bleek uit de inzet waarmee problemen te lijf werden gegaan, uit de strategieën die werden toegepast en uit het aantal goede oplossingen. Wood en Locke (1987) hebben het begrip "academic self-efficacy" geoperationaliseerd. Zij onderscheiden een aantal deelgebieden, zoals het begrijpen van de stof, die tezamen inhoud geven aan de academische eigen competentie. De waargenomen eigen academische competentie werd gemeten, evenals de doelen die men nastreefde. De tentamenresultaten werden gebruikt als prestatiemaat, en als maat voor de feitelijke bekwaamheid werd gebruik gemaakt van een gangbare test van academische vaardigheden. De resultaten van Wood en Locke (1987) wijzen uit dat de waargenomen eigen academische competentie, na controle voor de feitelijke bekwaamheid een duidelijke bijdrage levert aan de tentamenresultaten. Ook komt naar voren dat de waargenomen eigen academische competentie de doelen die nagestreeft worden, beïnvloedt, en deze doelen oefenen op hun beurt invloed uit op de prestaties.

Het onderhavige onderzoek repliceert de studie van Wood en Locke (1987). In overeenstemming met de resultaten van deze auteurs verwachten wij dat ook bij onze onderzoeksgroep (eerstejaarsstudenten in de psychologie) de waargenomen eigen academische competentie en de nagestreefde doelen van invloed zijn op de studieresultaten.

Bandura (1989) heeft in navolging van Elliot en Dweck (1988) gesuggereerd dat de denkbeelden die iemand erop nahoudt over de maakbaarheid van intelligentie van invloed kunnen zijn op de prestatiemotivatie. Er worden twee perspectieven onderscheiden. Volgens het eerste perspectief is intelligentie een bekwaamheid die ontwikkeld kan worden door kennis te verwerven en door zich vaardigheden eigen te maken en die te verbeteren. Personen die zich door deze opvatting laten leiden, zullen zich bij voorkeur bezig houden met uitdagende taken die hen de mogelijkheid bieden hun bekwaamheid te ontwikkelen. Dit perspectief activeert een beheersingsgerichte benadering van tegenslagen. Als deze personen falen, zullen zij niet bij de pakken neerzitten, maar ernaar streven nieuwe strategieën te ontwikkelen en met behulp van inzet en volharding tot prestatieverbetering te komen. Vanuit het tweede perspectief is intelligentie een min of meer stabiel gegeven. Intelligentie wordt beschouwd als een talent dat een noodzakelijke voorwaarde vormt voor het leveren van goede prestaties. Alleen als dit talent aanwezig is, kan inspanning vruchten afwerpen. Vanuit deze gedachtengang verschaft de kwaliteit van de taakprestatie aanwijzingen over de mate waarin de betrokkene over het benodigde talent beschikt. Falen wordt dan ook opgevat als een indicatie van te geringe intellectuele capaciteiten. Een dergelijke attributie kan gevoelens van hulpeloosheid activeren; zij zet in elk geval niet aan tot een beheersingsgerichte benadering.

Deze visies op de maakbaarheid van intelligentie zouden, volgens Bandura, in aanvulling op de denkbeelden over de eigen competentie, van invloed zijn op de prestaties. Om deze hypothese adequaat te kunnen toetsen, dient zij naar ons idee gespecificeerd te worden. De denkbeelden over de eigen competentie worden bestudeerd met betrekking tot verschillende onderdelen van de eerstejaarsstudie in de psychologie. Daarom ligt het in de rede om de visies op de maakbaarheid van intelligentie ook te specificeren en te omschrijven in relatie tot de eerstejaarsstudie in de psychologie. Dan gaat het om de mate waarin men meent dat de bekwaamheid op dit gebied maakbaar of leerbaar is.

De hypothese luidt als volgt: studenten met relatief gunstige denkbeelden over de eigen

competentie en met een sterk geloof in de maakbaarheid van de bekwaamheid op het gebied van de studie zullen eventueel falen in geringere mate toeschrijven aan gebrek aan talent, en betere resultaten behalen dan studenten met minder gunstige denkbeelden over de eigen competentie en met minder geloof in de maakbaarheid van deze bekwaamheid. Ook verwachten wij, in aansluiting bij Wood en Locke, dat de denkbeelden over de eigen competentie in combinatie met de nagestreefde doelen de tentamenresultaten zullen beïnvloeden. Omdat Bandura de nadruk legt op de domeinspecificiteit van de denkbeelden over de eigen competentie, hebben wij ook deze hypothese getoetst in relatie tot verschillende studie-onderdelen.

METHODE

Aan het onderzoek is meegewerkt door 438 eerstejaarsstudenten in de psychologie die in het kader van studieverplichtingen hebben deelgenomen aan een collectieve testzitting. Deze zitting vond plaats enkele weken na het begin van de relevante studie-onderdelen en enkele weken voorafgaande aan de tentamens. De deelnemers hebben tijdens deze zitting de Academic-Self-Efficacy-vragenlijst (ASE) voorgelegd gekregen, evenals vragen over doelen, attributies aan "gebrek aan talent" en vragen over de maakbaarheid van de bekwaamheid voor hun studie. Alle vragen, behalve de laatste, zijn beantwoord in relatie tot drie studie-onderdelen: Inleiding in de psychologie, Persoonlijkheidsleer en Sociale psychologie.

ASE

De "Academic Self-Efficacy"-vragenlijst van Wood en Locke (1987) is vertaald en integraal afgenomen. Deze auteurs onderscheiden de eigen academische competentie in zeven taakdomeinen: begrijpen van de stof, onderscheiden van begrippen, uitleggen van begrippen, herinneren, concentratie tijdens de les, aantekeningen maken tijdens de les en concentratie tijdens het tentamen. Voor elk taakgebied zijn vier of vijf items (vijf bij herinneren) met de volgende intervallen gebruikt. Bij voorbeeld: "Ik begrijp 50% (70%, 90% of 100%) van de begrippen die aangeboden worden." De deelnemers werd verzocht aan te geven of ze het beschreven niveau konden realiseren (ja of nee). Deze vragen representeren de omvang van de eigen competentie (Self-Efficacy Magnitude: SEM). De deelnemers werd eveneens verzocht bij elk niveau van competentie aan te geven hoe zeker ze waren over hun mogelijkheden om dit niveau te realiseren (op een schaal van 0 tot 10). Deze vragen representeren de sterkte van de eigen competentie (Self-Efficacy Strength: SES). Evenals Wood en Locke hebben we de betrouwbaarheid van de SES bepaald. Bij elk vak en elk taakgebied bleken de (laatste) items die betrekking hebben op de hoogste moeilijkheidsgraad laag te correleren met de overige items. Bij de vakken Sociale psychologie en Persoonlijkheidsleer correleerde ook het voorlaatste item bij "Herinneren" laag met de rest; bij Inleiding in de Psychologie was dit het geval voor het eerste item met betrekking tot "Herinneren". Na verwijdering van deze items is de Cronbachs alpha per vak berekend (21 items per vak). De resultaten zijn als volgt: Persoonlijkheidsleer a = .87, Sociale psychologie α = .90, Inleiding in de psychologie α = .91. In navolging van Wood en Locke hebben we de overeenkomstige items uit de SEM eveneens verwijderd. Om te bepalen in hoeverre de eigen competentie in combinatie met de denkbeelden over de maakbaarheid de studieresultaten heeft beïnvloed is gebruik gemaakt van de som van de SEM-scores per vak. De correlaties tussen de som van de SEM-scores voor Persoonlijkheidsleer, Sociale psychologie en Inleiding in de psychologie zijn als volgt: Persoonlijkheidsleer/Sociale psychologie r =.69; Persoonlijkheidsleer/Inleiding in de psychologie r = .61; Sociale psychologie/ Inleiding in de psychologie r = .62. Met behulp van de mediaan zijn per vak twee groepen onderscheiden die hoge of lage competentie representeren. (Persoonlijkheid: hoge competentie m = 19.5, sd = 1.2, lage competentie m = 14.7, sd = 2.5; Sociale psychologie: hoge competentie m = 20.1, sd = .85, lage competentie m = 15.3, sd = 2.7; Inleiding in de psychologie: hoge competentie m = 20, 2, sd = .79, lage competentie m = 15.3, sd = 2.4).

Wood en Locke hebben alle geselecteerde SES-items gesommeerd en daarover de verdere

analyses uitgevoerd. Dit betekent dat de totale SES-score gebaseerd is op de zekerheid over het wél en over het níet realiseren van de verschillende moeilijkheidsniveau's. Als twee personen dezelfde hoge SES-score hebben, kan dat inhouden dat de een heel zeker is over het wél realiseren van de verschillende niveau's, terwijl de ander heel zeker is over het níet realiseren daarvan. Omdat wij vooral geinteresseerd zijn in de zekerheid over de competentie, en niet over de incompetentie, hebben wij alleen de zekerheidsscores gebruikt over het niveau dat men meende te kunnen realiseren.De correlaties tussen de som van de zekerheidsscores over de competentie per vak zijn als volgt: Persoonlijkheidsleer/ Sociale psychologie r = .78; Persoonlijkheidsleer/ Inleiding in de psychologie r = .76; Sociale psychologie/ Inleiding in de psychologie r = .81.

Doelen

De vragen van Wood en Locke (1987) over de nagestreefde doelen zijn eveneens overgenomen. Een voorbeeld van zo'n vraag is: "Welke score probeert u voor het tentamen Sociale Psychologie te behalen". Het verzoek was om de relevante score te omcirkelen. Evenals Wood en Locke hebben wij voor elk vak vier items gebruikt. Cronbachs alpha voor de doelenschaal voor elk vak is als volgt: Persoonlijkheidsleer α = .72, Sociale psychologie α = .77, Inleiding in de psychologie α = .68. Bij het vak 'Inleiding' is een item verwijderd dat laag correleerde met de rest.

Studieresultaten

Omdat voor de tentamenresultaten van Sociale psychologie een andere schaal is gebruikt dan bij de andere studie-onderdelen, hebben we, met het oog op de vergelijkbaarheid, bij elk studie-onderdeel bepaald of de proefpersoon gezakt (0) dan wel geslaagd (1) is. Deze scores zijn gebruikt als prestatiemaat. De slaag-percentages per studie-onderdeel zijn: Persoonlijkheidsleer 55%, Sociale psychologie 33%, en Inleiding in de psychologie 60%. De correlaties tussen de prestatiematen zijn als volgt: Persoonlijkheidsleer/Sociale psychologie r = .80; Persoonlijkheidsleer/Inleiding in de psychologie r = .43; Sociale psychologie/ Inleiding in de psychologie r = .46.

Attributies aan gebrek aan talent

Wij hebben vijf vragen geformuleerd over gebrek aan talent als verklaring voor falen. Deze zijn afgenomen in relatie tot elk van de drie studie-onderdelen. Een voorbeeld is: "Als ik voor het tentamen Persoonlijkheidsleer zou zakken, betekent dit dat ik op dit gebied onvoldoende kwaliteiten bezit". De antwoorden zijn gegeven op zespuntsschalen (extremen: duidelijk eens en duidelijk oneens). Cronbachs alpha is: Persoonlijkheidsleer α = .77, Sociale psychologie α = .82, Inleiding in de psychologie α = .81.

Maakbaarheid van de eigen bekwaamheid

Er zijn vier vragen gesteld over de maakbaarheid van de eigen bekwaamheid op het gebied van de psychologiestudie. Een voorbeeld is: "Door inzet en volharding kun je je bekwaamheden in de psychologie ontwikkelen". De antwoorden zijn gegeven op zespuntsschalen (extremen duidelijk eens en duidelijk oneens). Cronbach's alpha is .66. Op basis van de mediaan zijn de deelnemers onderverdeeld in twee groepen die een relatief grote of geringe maakbaarheid representeren. (Grote maakbaarheid m = 5.2, sd = .31; geringe maakbaarheid m = 4.3, sd = .39.)

Intelligentie

Wood en Locke hebben voor feitelijke bekwaamheid gecontroleerd door gebruik te maken van een bekwaamheidstest. Omdat een dergelijke test van academisch bekwaamheid, voor zover wij weten, in Nederland niet voorhanden is, hebben wij gecontroleerd voor intelligentie. Er is gebruik gemaakt van de gemiddelde score op een aantal intelligentietests die jaarlijks worden afgenomen bij eerstejaarsstudenten in de psychologie. Het betreft zes Guilfordtests, namelijk: Vocabulaire, Reeksen, Verborgen Figuren, Conclusies, Verbale analogieën en Rekensnelheid. Deze tests zijn afgenomen tijdens enkele collectieve testzittingen; dat waren andere zittingen

dan de zitting waarin de overige vragen zijn afgenomen. Ook deze variabele is in verband met de hypothese over de rol van competentie en maakbaarheid gedichotomiseerd met behulp van de mediaan. (Intelligentie hoog: m = 22.6, sd = 2.83; intelligentie laag: m = 15.5, sd = 2.52.)

RESULTATEN

In aansluiting bij de resultaten van Wood en Locke werd in de eerste plaats verwacht dat de academische eigen competentie en de nagestreefde doelen van invloed zijn op de tentamenresultaten. Wij hebben dezelfde analyses toegepast als Wood en Locke. Eerst zijn, voor elk vak afzonderlijk, de product-moment-correlatie- coëfficiënten berekend. De resultaten staan in Tabel 1.

Het meest opvallende is dat SEM bij elk vak correleert met Doelen. Bij Sociale psychologie zien wij verder dat alleen de variabele Doelen met het tentamenresultaat correleert. Bij Persoonlijkheidsleer hebben zich geen significante correlaties met het tentamenresultaat voorgedaan. Bij Inleiding in de psychologie zien wij daarentegen dat SEM, SES en Doelen met het tentamenresultaat correleren.

Er zijn tevens enkele hiërarchische regressie-analyses uitgevoerd. Bij de eerste analyses is in de eerste stap de bijdrage van Intelligentie aan het tentamenresultaat bepaald, daarna is nage-

Tabel 1. Correlatiematrices van variabelen per studie-onderdeel.

SOCIALE PSYCHOLOGIE

	Intell	SEM	SES	Doelen
Tentamen	.04	.06	.08	.13**
Doelen	.03	.32**	.12*	.—
SES	- .04	.13**	.—	.—
SEM	- .01	.—	.—	.—

* p < .05
** p < .01

PERSOONLIJKHEIDSLEER

	Intell	SEM	SES	Doelen
Tentamen	.04	.06	.06	.09
Doelen	- .02	.34**	.23**	.—
SES	- .03	.25**	.—	.—
SEM	- .03	.—	.—	.—

* p < .05
** p < .01

INLEIDING IN DE PSYCHOLOGIE

	Intell	SEM	SES	Doelen
Tentamen	.06	.17**	.10*	.23**
Doelen	.02	.31**	.21**	.—
SES	.00	.15**	.—	.—
SEM	- .02	.—	.—	.—

* p < .05
** p < .01

gaan wat de bijdrage is van respectievelijk SEM, SES en Doelen. De resultaten voor elk van de drie vakken staan bovenin Tabel 2.

Uit deze tabel blijkt dat Intelligentie het tentamenresultaat niet noemenswaardig beïnvloed heeft. Verder zien wij bij Sociale psychologie en Persoonlijkheidsleer dat SEM, SES en Doelen heel weinig variantie in het tentamenresultaat verklaren. Bij Inleiding in de psychologie wordt 7% van de variantie verklaard. Verder blijkt dat SEM en Doelen elk een significante bijdrage geleverd hebben aan het tentamenresultaat. Bij Wood en Locke werd omstreeks 24% van de variantie verklaard.

Om te bepalen of zich sekseverschillen hebben voorgedaan, zijn de bovenbeschreven analyses eveneens uitgevoerd voor vrouwen en mannen afzonderlijk. De resultaten wijzen uit dat de verschillen tussen vrouwen en mannen te verwaarlozen zijn. Daarom wordt deze variabele bij de bespreking van de regressie-analyses verder buiten beschouwing gelaten.

Om enig zicht te krijgen op het effect van de volgorde waarin de variabelen zijn ingevoerd, is deze volgorde bij de tweede serie regressie-analyses gewijzigd. Na Intelligentie is de variabele Doelen ingevoerd, gevolgd door respectievelijk de SEM en SES. Deze resultaten staan in het middelste deel van Tabel 2. Bij Inleiding in de psychologie zien wij de meest ingrijpende wijzigingen: als de variabele Doelen eerst wordt ingevoerd, wordt de bijdrage van deze variabele aan de tentamenresultaten groter, en dat gaat ten koste van de bijdrage van SEM.

In de laatste serie regressie-analyses fungeerde de variabele Doelen als criterium. De resultaten van deze analyses staan onderin Tabel 2. Bij elk van de drie vakken levert SEM een significante bijdrage aan Doelen. Alleen bij Inleiding in de psychologie is ook de bijdrage van

Tabel 2. Regressie-analyses voor tentamenresultaten en doelen.

I. Afhankelijke variabele - Tentamenresultaten

	Sociale psychologie		Persoonlijkheidsleer		Inleiding	
Volgorde van invoer	mult R	t	mult R	t	mult R	t
1. Intelligentie	.04	.90	.04	.87	.07	1.45
2. SEM	.08	1.26	.08	1.38	.18	3.51**
3. SES	.11	1.52	.09	.98	.20	1.75
4. Doelen	.16	2.40*	.12	1.65	.27	3.78**
Totaal	R^2=.026	F=2.64 p=.03	R^2=.015	F=1.60 p=.17	R^2=.07	F=8.14 p<.001
1. Intelligentie	.04	.90	.04	.87	.07	1.45
2. Doelen	.14	2.76*	.11	2.15*	.25	4.94**
3. SEM	.14	.45	.12	.73	.27	2.12
4. SES	.16	1.37	.12	.73	.27	1.05
Totaal	R^2=.026	F=2.64 p=.03	R^2=.015	F=1.60 p=.17	R^2=.07	F=8.14 p <.001

II. Afhankelijke variabelen - Doelen

Volgorde van invoer	mult R	t	mult R	t	mult R	t
1. Intelligentie	.03	.52	.02	.47	.005	.10
2. SEM	.31	6.49**	.33	7.02**	.32	6.75**
3. SES	.31	1.29	.36	3.01*	.36	3.78**
Totaal	R^2=.10	F=14.73 p <.001	R^2=.13	F=19.88 p <.001	R^2=.13	F=20.46 p <.001

* $p < .05$
** $p < .001$

SES significant. De verklaarde variantie varieert van 10 tot 13 %, hetgeen overeenkomt met de resultaten van Wood en Locke.

Ook werd verwacht dat personen die relatief gunstige opvattingen over hun eigen competentie combineren met een sterk geloof in de maakbaarheid van de bekwaamheid voor de studie eventueel falen in geringere mate zouden toeschrijven aan gebrek aan talent dan personen met minder gunstige opvattingen over hun competentie en een gering geloof in de maakbaarheid.

Er zijn, voor elk vak afzonderlijk, variantie-analyses uitgevoerd met SEM (laag/hoog), Maakbaarheid (laag/hoog), Intelligentie (laag/hoog) en Sekse (vrouw/man) als factoren en gebrek aan talent bij falen als afhankelijke variabele. Bij elk vak heeft zich een hoofdeffect voor SEM voorgedaan (Sociale psychologie $F_{(1, 436)} = 11.59$, $p < .001$; Persoonlijkheidsleer $F_{(1, 436)} = 12.45$, $p < .001$; Inleiding in de psychologie $F_{(1, 436)} = 4.22$, $p < .05$). Ook zijn er hoofdeffecten voor Maakbaarheid opgetreden (Sociale psychologie $F_{(1, 436)} = 8.32$, $p < .005$; Persoonlijkheidsleer $F_{(1, 436)} = 8.92$, $p < .005$; Inleiding in de psychologie $F_{(1, 436)} = 8.08$, $p < .005$). Verder hebben zich geen effecten voorgedaan. De relevante gemiddelden staan in Tabel 3.

Onze hypotheses hebben betrekking op verschillen tussen specifieke groepen. Daarom is voor de toetsing gebruik gemaakt van t-toetsen. De resultaten wijzen uit dat bij lage intelligentie personen met relatief gunstige denkbeelden over de eigen competentie en de maakbaarheid mogelijk falen in mindere mate hebben toegeschreven aan gebrek aan talent, dan personen met relatief ongunstige denkbeelden over de eigen competentie en de maakbaarheid. Er is eenzijdig getoetst. (Sociale psychologie: $t = 2.47$, $df = 104$, $p = .02$; Persoonlijkheidsleer: $t = 2.98$, $df = 99$,

Tabel 3. Gemiddelden voor gebrek aan talent.

SOCIALE PSYCHOLOGIE

		INTELLIGENTIE LAAG		INTELLIGENTIE HOOG	
		SEM		SEM	
		laag	hoog	laag	hoog
		m	m	m	m
MAAKBAARHEID	laag	2.1	2.0	2.2	2.1
	hoog	2.1	1.9	2.1	1.7

PERSOONLIJKHEIDSLEER

		INTELLIGENTIE LAAG		INTELLIGENTIE HOOG	
		SEM		SEM	
		laag	hoog	laag	hoog
		m	m	m	m
MAAKBAARHEID	laag	2.2	1.9	2.3	2.0
	hoog	2.1	1.8	1.9	1.7

INLEIDING IN DE PSYCHOLOGIE

		INTELLIGENTIE LAAG		INTELLIGENTIE HOOG	
		SEM		SEM	
		laag	hoog	laag	hoog
		m	m	m	m
MAAKBAARHEID	laag	2.1	2.0	2.2	2.1
	hoog	2.1	1.9	1.9	1.7

NB. Een lagere score representeert een geringere toeschrijving van mogelijk falen aan gebrek aan talent.

$p = .002$; Inleiding in de psychologie: $t = 1.84$, $df = 101$, $p = .03$). De verwachte verschillen zijn eveneens opgetreden bij hoge intelligentie (Sociale psychologie: $t = 4.36$, $df = 98$, $p < .001$; Persoonlijkheidsleer: $t = 4.12$, $df = 94$, $p < .001$; Inleiding in de psychologie: $t = 3.35$, $df = 94$, $p = .001$).

Tevens verwachtten wij dat personen met relatief gunstige denkbeelden over de eigen competentie en een groot vertrouwen in de maakbaarheid van de relevante bekwaamheid betere studieresultaten zouden behalen dan personen met minder gunstige denkbeelden en minder vertrouwen in de maakbaarheid.

Ook in dit geval zijn variantie-analyses uitgevoerd met SEM, Maakbaarheid, Intelligentie en Sekse als factoren. De tentamenscore fungeerde als afhankelijke variabele. Er hebben zich geen significante hoofd- of interactie-effecten voorgedaan. Zoals op grond van de eerder beschreven resultaten te verwachten viel, is bij toetsing met t-toetsen geen steun gevonden voor deze hypothese bij de studieonderdelen Sociale psychologie en Persoonlijkheidsleer. Bij Inleiding in de psychologie bleken personen met hoge intelligentie, die er gunstige denkbeelden over hun eigen competentie en de maakbaarheid op na hielden, betere studieresultaten te behalen dan personen met hoge intelligentie en met weinig gunstige denkbeelden over hun competentie en de maakbaarheid ($t = 1.85$, $df = 101$, $p = .03$ bij eenzijdige toetsing). Bij lage intelligentie hebben deze verschillen zich niet voorgedaan.

DISCUSSIE

Uitgangspunt van het onderhavige onderzoek was dat de denkbeelden over de eigen competentie van invloed zijn op de prestaties die behaald worden, en dat deze invloed domeinspecifiek is. Onze resultaten illustreren de domeinspecifieke werkzaamheid van de eigen competentie. Immers bij de vakken Sociale psychologie en Persoonlijkheidsleer bleken de eigen competentie en de nagestreefde doelen geen noemenswaardige bijdrage te leveren aan de tentamenresultaten. Bij Inleiding in de psychologie leverden de omvang van de eigen competentie (SEM) en de nagestreefde doelen een significante bijdrage aan de tentamenresultaten, terwijl over het geheel genomen 7% van de variantie in de tentamenresultaten verklaard werd. Deze domeinspecifieke effecten bij dezelfde studenten zijn in overeenstemming met de theoretische uitgangspunten. De denkbeelden van studenten over hun eigen competentie blijken immers in het ene domein wel en in het andere géén bijdrage te leveren aan de tentamenresultaten. Nader onderzoek zal moeten uitwijzen welke factoren hierbij een rol spelen

Ook bleek bij het vak Inleiding in de psychologie dat de omvang (SEM) en de sterkte (SES) van de eigen competentie een significante bijdrage leverden aan de nagestreefde doelen, en dus via deze doelen eveneens hebben bijgedragen aan de studieresultaten. Deze resultaten komen overeen met die van Wood en Locke. Er is ook een duidelijk verschil: in het onderzoek van Wood en Locke verklaren de academische waargenomen eigen competentie en de doelen tezamen omstreeks 24% van de variantie in de studieresultaten, bij ons is dat voor Inleiding in de psychologie slechts 7%.

Het is de vraag waaraan dit verschil toegeschreven kan worden. Er zijn tenminste twee factoren die een rol gespeeld kunnen hebben. Wood en Locke hebben tijdens de betrouwbaarheidsanalyse van de SES-scores, ter verhoging van de betrouwbaarheid de items die betrekking hebben op het makkelijkste niveau van competentie verwijderd. In onze studie droegen juist de items die het moeilijkste niveau representeren bij aan de onbetrouwbaarhed van de SES. Dit doet vermoeden dat de cursussen waarin Wood en Locke de eigen competentie bestudeerd hebben aanzienlijk makkelijker waren dan de studie-onderdelen in ons onderzoek. Waarschijnlijk is het bij cursussen die relatief moeilijk zijn ook moeilijker om de eigen competentie goed te taxeren, of, met andere woorden, om de eigen competentie zodanig te taxeren dat zij in lijn is met de resultaten die behaald gaan worden.

De tweede factor heeft betrekking op het tijdstip waarop de waargenomen eigen academische competentie is gemeten. Bij Wood en Locke was dat direct na het eerste tentamen dat in de

cursus werd afgenomen, maar wel voorafgaande aan de bekendmaking van de tentamenscores. In ons onderzoek hebben de studenten hun competentie beoordeeld voordat ze enige ervaring konden opdoen met het desbetreffende tentamen. Deze tentamenervaring heeft de studenten in het onderzoek van Wood en Locke waarschijnlijk een duidelijker beeld gegeven van de vereisten voor het eindtentamen (waarvan de resultaten als criterium gebruikt zijn). Door "enactive learning" (Bandura, 1989) of zelfgegenereerde "feedback" (Locke & Latham, 1990) hebben zij hun denkbeelden over de eigen competentie waarschijnlijk beter kunnen afstemmen op de vereisten van het eindtentamen. Wel hebben 'onze studenten' voorafgaande aan Inleiding in de psychologie deelgenomen aan een enigszins vergelijkbaar onderdeel waarover zij ook tentamen hebben gedaan. Bij Inleiding in de psychologie hadden de studenten dus ervaring met een soortgelijk tentamen. Dit verklaart misschien ook waarom juist bij Inleiding in de psychologie de waargenomen eigen competentie en de nagestreefde doelen de tentamenresultaten relatief goed voorspellen.

Het is zeer wel mogelijk dat veel van de studenten in ons onderzoek er weinig accurate denkbeelden over de eigen competentie op nahielden, niet alleen omdat zij niet op de hoogte waren van de tentamenvereisten, maar ook omdat zij nog geen ervaring hadden met een academische studie in de psychologie. Als ervaring op een bepaald terrein ontbreekt, worden de denkbeelden over de eigen competentie vooral ontleend aan eerdere ervaringen die vergelijkbaar lijken te zijn. In dit geval zullen dat vooral de studie-ervaringen op de middelbare school zijn. Omdat dergelijke ervaringen slechts ten dele overeen komen met de psychologiestudie, zal het beeld over de eigen competentie op het gebied van deze studie vaak niet zo accuraat zijn. Volgens Gist en Mitchell (1990) zou dit gebrek aan accuratesse vooral bij de grote middenmoot voorkomen. Daarom hebben wij de studenten geselecteerd die voor alle drie de tentamens gezakt of geslaagd zijn. Bij deze studenten voorspelden de denkbeelden over de eigen competentie en de nagestreefde doelen ruim 20% van de variantie in de studieresultaten. Het kan niet uitgesloten worden dat de eerdere onderwijservaringen van de studenten in het onderzoek van Wood en Locke beter aansloten bij het cursorisch onderwijs dat zij in hun onderzoek hebben opgenomen.

Ook is bestudeerd in hoeverre het geloof in de maakbaarheid van de bekwaamheid voor de psychologiestudie in combinatie met de denkbeelden over de eigen competentie van invloed zijn op de attributies aan gebrek aan talent bij mogelijk falen, en op de studieresultaten. De resultaten die betrekking hebben op de attributies aan gebrek aan talent bieden steun aan de verwachtingen. Bij controle voor het intelligentieniveau schreven personen met gunstige denkbeelden over hun competentie en geloof in de maakbaarheid van bekwaamheid eventueel falen in mindere mate toe aan gebrek aan talent dan personen met minder gunstige denkbeelden en minder geloof in die maakbaarheid. De verschillen in studieresultaten bleken ook in dit geval domeinspecifiek te zijn. Alleen bij het vak Inleiding in de psychologie en alleen bij een relatief hoog intelligentieniveau behaalden personen met gunstige denkbeelden over de eigen competentie en de maakbaarheid betere resultaten dan personen met minder gunstige denkbeelden.

De niet-domeinspecifieke resultaten met betrekking tot gebrek aan talent en de domeinspecifieke resultaten met betrekkking tot de studieprestaties vestigden er weer eens de aandacht op dat personen verschillende verklaringen voor hun gedrag kunnen hanteren zonder dat die doorwerken in het observeerbare gedrag. Verder onderzoek zal moeten uitwijzen of er, naast de eerder beschreven factoren die de studieresultaten differentiëel beïnvloed hebben, nog andere psychologische factoren werkzaam zijn die mediëren tussen attributies en gedrag.

LITERATUUR

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, N.J.: Prentice Hall.

Bandura, A. (1989). Self-regulation of motivation and action through internal standards and goal systems. In L.A. Pervin (Ed.), *Goal Concepts in personality and social psychology*. Hillsdale, N.J.: Lawrence Erlbaum.

Cervone, D., & Peake, P.K. (1986). Anchoring efficacy and action: The influence of judgemental heuristics on self-efficacy judgements and behavior. *Journal of Personality and Social Psychology, 50*, 492-500.

Elliot, E.S., & Dweck, C.S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*, 5-12.

Gist, M.E., & Mitchell, T.R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17*, 183-211.

Gist, M.E., Schwoerer, V., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology, 74*, 884-891.

Hill, T, Smith, N.D., & Mann, M.F. (1987). Role of efficacy expectations in predicting the decision to use advanced technologies: The case of computers. *Journal of Applied Psychology, 72*, 307-313.

Locke, E.A., & Latham, G.P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, N.J.: Prentice Hall.

Taylor, M.S., Locke, E.A., Lee, L., & Gist, M. (1984). Type A behavior and faculty research productivity: What are the mechanisms? *Organisational Behavior and Human Performance, 34*, 402-418.

Wood, R.E., & Locke, E.A. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measurement, 47*, 1013-1024.

Manuscript ontvangen 6-1-1993
Definitieve versie ontvangen 31-3-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 52-57

Notities en Commentaren

'TRANSFER: ONDERZOEK EN ONDERWIJS'. EEN REACTIE OP DE ONDERLIGGENDE DENKOPVATTING

Ben Wolters
*Vakgroep Onderwijskunde, Katholieke Universiteit Nijmegen**

1. INLEIDING

In hun overzichtsartikel refereren Simons en Verschaffel (1992) (onderzoeks)literatuur die in belangrijke mate de opvatting weerspiegelt dat transfer tot stand komt als gevolg van een actieve en bewuste oriëntatie op de taak. Pregnant komt dat naar voren in het volgende citaat - m.b.t. de, vooral voor het onderwijs relevante, 'high road'[1] waarlangs transfer kan worden bewerkstelligd: 'Essentieel ... is het *bewust en opzettelijk expliciteren en abstraheren van kennis-elementen of cognitieve vaardigheden* vanuit een bepaalde context' (p.7). Hermans (1983) stelt dat deze actieve, bewuste houding een beroep doet op 'de halve persoonlijkheid van de onderwijspsychologie'. Hij meent dat de andere helft van de persoonlijkheid van de ontwikkelingspsychologie wordt gekenmerkt door receptiviteit in plaats van activiteit. Is het mogelijk eveneens transfer tot stand te brengen door een beroep te doen op de receptieve helft van de persoonlijkheid?

Een positief antwoord op deze vraag wordt kansrijker wanneer er zoiets als receptieve informatieverwerking bestaat. Actieve informatieverwerking vormt immers ook de basis voor de beschrijving die de auteurs van transfer geven.
In dit commentaar zal daarom achtereenvolgens worden ingegaan op twee vragen:
- bestaat er zoiets als receptieve informatieverwerking, naast actieve informatieverwerking?
- kunnen uit receptieve informatieverwerking nader aan te duiden, onderwijsrelevante transfer-condities worden afgeleid?

2. INTUÏTIE

De leerling die een probleem oplost, kan dat op een zodanige manier doen dat hij[2] in de toekomst zo goed mogelijk in staat zal zijn min of meer verwante problemen ook met succes op te lossen. De condities waaronder deze zogenoemde transfer zo groot mogelijk is, hebben niet slechts betrekking op de bovengenoemde actieve informatieverwerking, maar mogelijk ook op receptieve informatieverwerking, d.w.z. intuïtie.

In de denk- en leerpsychologie wordt vanouds, d.w.z. al vóór ongeveer 1970, onderscheid gemaakt tussen twee aspecten van denken: 'reasoning vs. imagination', 'Regelbewusztsein vs. Beziehungsbewusztsein', 'reality-adjusted thinking vs. autistic thinking', 'association vs. bisociation', 'conscious work vs. unconscious inspiration', 'reproductive thinking vs. productive thinking', 'vertical thinking vs. lateral thinking', 'logical reasoning vs. original thinking', 'rational knowing vs. intuition', analytisch denken vs. intuïtief denken (Wolters, 1980, noot 1).

In de moderne cognitieve opvatting van probleemoplossen is het intuïtieve denkaspect ook steeds aanwezig als een soort nevenstroom (vgl Wolters,1987). In een recent artikel m.b.t. de didactiek van de wiskunde, toont Fischbein (1990) aan dat een informatieverwerkings-model dat uitsluitend is gebaseerd op de computermetafoor, niet geheel voldoet om het oplossen van (wiskundige) problemen te verklaren. Hij komt tot de conclusie dat er twee manieren van informatieverwerking bestaan:

* Postbus 9103, 6500 HD Nijmegen.

- de logische, sequentiële, analytische, formele manier;
- de globale, onbemiddelde manier.

Een vergelijking van de kenmerken van receptieve informatieverwerking: *intuïtief denken* en actieve informatieverwerking: *analytisch denken* bij verschillende auteurs, levert het volgende overzicht op:

Intuïtief denken	Analytisch denken
1. parallelle informatieverwerking	1. lineaire informatieverwerking
2. niet bewust	2. bewust
3. niet in woorden of tekens gevat	3. wel in woorden of tekens gevat
4. kinaesthetisch, visueel	4. verbaal
5. 'Gestalt'	5. logisch
6. gebonden aan concrete situatie (Wolters,1987)	6. abstract

Fischbein (op.cit.) verwijst voor zijn beschrijving van analytisch denken naar het model van informatieverwerking. Voor zijn beschrijving van intuïtief denken heeft hij geen theorie voorhanden. Hij beperkt zich tot de vrome wens dat het model van informatieverwerking ook dit type denken zal gaan omvatten. Een theoretische onderbouwing is echter noodzakelijk om condities te formuleren waaronder transfer kan optreden.

Een theorie van intuïtie is wel geformuleerd door Bastick (1982; vgl. Wolters, 1987; 1991a; 1991b). Daarin speelt het *beeld dat van de* (probleem)situatie wordt gevormd, een centrale rol.

Beeld van de situatie

In het *beeld* van de situatie worden aspecten van de situatie niet afzonderlijk waargenomen, maar als een totaal, een geheel, een 'Gestalt'. Het beeld dat van de situatie wordt gevormd, wordt gekenmerkt door een *woordeloze,* maar toch *betekenisvolle gelijktijdigheid* die *niet bewust* is, maar ook niet volledig onbewust. Deze compacte formulering zal verder worden uitgewerkt en verduidelijkt aan de hand van de volgende alledaagse situatie. Stel U loopt in een drukke straat in een vreemde stad. Allerlei onbekende mensen lopen voorbij, maar plotseling ontdekt U in de zaterdagmiddagdrukte een bekend gezicht: uw oom die in een nabijgelegen dorp woont. Deze herkenning vindt niet plaats op grond van een logische gevolgtrekking uit bepaalde kenmerken: kleur ogen, stand wenkbrauwen, inplantingslijn van het hoofdhaar, vorm van de mond, enz. Er is een plotselinge herkenning op grond van een patroon, ook wel 'Gestalt' genoemd, d.w.z. op grond van een *gelijktijdige* aanwezigheid van informatie. Vandaar het oude chinese spreekwoord: 'a picture is worth a thousand words' (in Bastick,1982).

Het woordeloze, doch betekenisvolle en half onbewuste karakter van het beeld kan worden verduidelijkt aan de hand van de volgende situatie. Een keuringsarts voor dienstplichtigen in de V.S. moest bij binnenkomst van iedere recruut een tweetal standaardvragen stellen. Op een gegeven moment viel het hem op dat hij eigenlijk de vragen niet eens zou hoeven te stellen, omdat hij het antwoord al wel zelf kon geven; en hij constateerde tot zijn verrassing dat het antwoord dat hij in gedachten had, klopte met het feitelijke antwoord van de recruut. Dat vormde voor hem aanleiding een systematisch onderzoek te doen naar dit verschijnsel. Hij nam zich voor het beroep van de recruut te bepalen zonder het hem eerst te vragen. Hij verkreeg zo van iedere recruut een opgave van feitelijk uitgeoefend beroep en geschat beroep. Hij constateerde dat voor een tweetal beroepen - boeren en mecaniciens - er een duidelijk verband bestond tussen feit en schatting. Dit verband was dermate sterk dat het niet aan het toeval kon worden toegeschreven (Berne,1949).

Hoe komt dat? Vanuit de vraag: wat zou zijn beroep zijn? herkent de keuringsarts in het *beeld* of de verschijningsvorm - houding, beweging, oogopslag, e.d. en manier van doen - van de recruut/beroepsbeoefenaar, de boer (vgl. het gezicht van de oom). De keuringsarts moet

natuurlijk wel bekend zijn met mensen die een dergelijk beroep uitoefenen. De vorming van het beeld van de recruut en de herkenning van dit beeld als van een boer, is een onderbewust, gelijktijdig proces. De idee: dat is een boer, komt tot stand *op basis van* een onderbewuste, betekenisvolle, woordeloze gelijktijdigheid. Daarbij redeneert de keuringsarts niet; hij gaat niet bewust argumenten op een rijtje zetten. Een zodanig tot stand gekomen idee wordt een *intuïtieve* idee genoemd. Deze intuïtieve idee springt in het bewustzijn, min of meer vergezeld van een (klein) 'Aha-Erlebnis'.

Het is wel duidelijk dat kennis van zaken een belangrijke functie vervult in de beeld-vorming. Kennis maakt wel deel uit van het beeld van de situatie, maar niet op een discursieve, doch op een onderbewuste, ongearticuleerde manier. Bijvoorbeeld, iemand die geen arts is, zal van zijn vriend zeggen: 'hij is de laatste tijd zo lusteloos'. De arts zal deze lusteloosheid vanzelf, zonder bewuste activiteit, 'invullen' met allerlei medisch-psychologische kennis van zaken die adequaat is aan het beeld dat de patiënt oproept. Naarmate de arts over minder medische kennis beschikt, des te meer kans loopt hij om een foute of onvolledige diagnose te stellen. De kans dat de idee ook de toets van analytisch denken en/of empirische verificatie kan doorstaan, is afhankelijk van de mate waarin het beeld wordt gedifferentieerd door de onderbewust aanwezige kennis van zaken.

Vervolgens, het is alleen maar mogelijk om zich een beeld van de situatie te vormen door zich in te leven in de situatie. Wanneer een leerling die een probleem moet oplossen alleen maar geïnteresseerd is in de oplossing, maar niet in het probleem, komt er van een identificatie met het probleem niet veel terecht. De leerling laat het probleem niet op zich inwerken, onbevooroordeeld en vrij, om de vorming van een beeld een kans te geven.

Tenslotte, het beeld zoals hier bedoeld, is afgeleid van het woord *fysionomie*. De leerling die zich een beeld vormt, neemt fysionomisch waar. Fysionomisch waarnemen is al een oud begrip (Werner, 1959), maar er is weinig gebruik van gemaakt in de opbouw van de psychologie. Met betrekking tot intuïtie als een vorm van kennisverwerving kan het echter een duidelijke functie krijgen (vgl Wolters, 1987), zoals uit de nu volgende, welbekende baksteenopdracht naar voren komt.

Wanneer kinderen van ongeveer 10 jaar worden geconfronteerd met de open opdracht: 'noem zoveel gebruiksmogelijkheden van een baksteen als je kunt', conformeren sommigen zich sterk aan hetgeen gebruikelijk is: 'huis bouwen, toren bouwen' enz. Andere kinderen geven ook wel deze variaties, maar noemen bovendien: grafsteen voor een vogel, (in een) wanddecoratie, schuurmiddel, grondstof voor rode verf, hoofdkussen voor een robot, enz. Het is mogelijk dat de leerling deze ongebruikelijke ideeën genereert als gevolg van een bewust en systematisch aftasten van de zoekruimte door zich bijvoorbeeld af te vragen: waarvoor zou een bakker zoal een baksteen kunnen gebruiken, een kunstenaar, een soldaat, een landbouwer, een matroos, ... (Frijda & Elshout,1976,425-426). Deze verklaring in de 'analytische' traditie sluit principieel een andere verklaring niet uit.

De leerling die ongewone gebruiksmogelijkheden opsomt, beperkt zich niet tot de geleerde kennis van bakstenen, maar vormt zich een *beeld* van de baksteen. Dat wil hier zeggen dat hij de baksteen on-middel-lijk waarneemt, niet be-middel-d door conventies. Hij laat de baksteen onbevangen, als een kind dat nog niet weet waarvoor bakstenen worden gebruikt, op zich inwerken. Deze doorleefde onbevangenheid heeft tot gevolg dat de leerling als het ware teruggaat naar het zintuigelijke, naar de van belevingen doortrokken kwaliteiten van de baksteen: grof, rood, vierkant, enz. Hij is in staat allerlei eigenschappen van de baksteen te beleven in hun oorspronkelijkheid: rood, grof, rechthoekig, hard. De eerder genoemde term fysionomisch waarnemen is hier volop van toepassing. De variaties op bouwen zijn in het algemeen geen gevolg van een dergelijk fysionomisch waarnemen van de baksteen, maar hebben betrekking op louter formele kennis.

Deze baksteenopdracht toont aan dat beeld-vorming ook een goede manier is om een impasse in het probleemoplossen te doorbreken die langs bewust-systematische weg niet valt te bereiken. De leerling die een probleem moet oplossen zal niet gefixeerd mogen zijn op de impasse, maar erop vertrouwen dat een gedistantieerde, maar positieve betrokkenheid en onbevangen-

heid de goede voedingsbodem vormen voor de vorming van een beeld waarin het vruchtbare idee is vervat. De leerling zal de vanzelfsprekende gedachte achter zich moeten laten dat de enige weg naar de oplossing van het probleem is gelegen in een bewuste, logische, systematische en analytische gedachtengang. Er is een andere, even legitieme weg die haar rationaliteit kan bewijzen via de gedachtengang achteraf: het idee lijkt de impasse te doorbreken omdat het idee ....(en dan volgt de redenering).

Wanneer de leerling zich bewust geworden is van *intuïtief denken*, kan hij ook bepalen dat hij als beheerser van het denkproces moet terugtreden en een receptieve houding ten aanzien van de probleemsituatie moet innnemen (Hermans, 1983).

Hiermee is de eerste vraag beantwoord. Er bestaat een receptieve vorm van informatieverwerking (intuïtie) naast de actieve vorm van informatieverwerking (analyse). Intuïtief denken vormt een aanvulling op analytisch denken, omdat intuïtief denken betrekking heeft op de vorming van ideeën en niet op de uitwerking en toetsing ervan. In analytisch denken wordt relevante voorkennis geactualiseerd als gevolg van een bewuste, systematische zoekprocedure[3].

Nu het bestaan van receptieve informatieverwerking is aangetoond, kan worden overgegaan tot de beantwoording van de tweede, in de Inleiding gestelde, vraag: kunnen transfercondities worden afgeleid uit het receptieve, intuïtieve denken?

3. TRANSFER VIA INTUÏTIE

In hun overzichtsartikel stellen Simons & Verschaffel (op.cit.) onder meer als transferconditie dat integratie van de leerstof in het persoonlijke kennisbezit de kans vergroot dat in nieuwe situaties de benodigde kennis en vaardigheden opgespoord kunnen worden (op. cit., p.11). Evenzeer geldt dat de kans dat de leerling via intuïtief denken op een idee komt kan worden vergroot door het scheppen van een *onderwijs- en opvoedingsklimaat* waarin leerlingen op een vruchtbaar idee kunnen komen wanneer het systematisch afzoeken van de cognitieve structuur niets oplevert.

Het voert te ver om hier uitvoerig op het betreffende klimaat en de hierna te bespreken brainstormingprocedure en alles wat daarmee samenhangt, in te gaan. Dat is elders (Wolters, 1991b; 1991c) reeds gebeurd. Daarom zal hier worden volstaan met een samenvatting.

Uit empirisch onderzoek naar de opvoeding van creatieve geleerden en kunstenaars, en van kinderen die blijk geven van veel ideeënproduktiviteit op open opdrachten, blijkt namelijk dat het pedagogische en onderwijsklimaat waarin deze mensen ideeën produceren, wordt gekenmerkt door stimulansen voor de opbouw van een eigen waarden- en normenstelsel. Ouders van ideeënproducerende kinderen wekken in het hanteren van regels niet de indruk dat die regels heilig zijn. Regels zijn alleen maar nodig voor een goed reilen en zeilen van de samenleving. Zij vertegenwoordigen geen absolute norm, die niet mag worden overtreden op straffe van liefdesonthouding, aantasting van het gevoel van eigenwaarde en zelfrespect van het kind. De straf op de overtreding van de regel is overeenkomstig het vergrijp en niet zodanig overtrokken dat de persoonlijke integriteit van het kind wordt aangetast. Het kind voelt zich aanvaard en zal zijn gedrag niet noodgedwongen conformeren aan het ouderlijke waarden- en normenpatroon om zo zijn gevoel van eigenwaarde niet te verliezen (Drevdahl, 1964; Krause, 1977; 1979; MacKinnon, 1966; Weisberg & Springer, 1961; vgl. Wolters, 1991b; 1991c). Kinderen (leerlingen) die in een dergelijk klimaat leven, kunnen de taak op zich laten inwerken, zonder bezig te hoeven zijn met zichzelf. Daardoor wordt fysionomisch waarnemen, de vorming van een beeld van de situatie, en het daaruit voortvloeiende idee, onbelemmerd mogelijk. De leerling kan onbemiddeld betrokken zijn op de taak. Tot zover het onderwijs- en opvoedingsklimaat als transferconditie van intuïtief denken.

De auteurs merken vervolgens in hun artikel op dat ontbrekend transfervermogen kan worden gecompenseerd door gebruik te maken van instructietechnieken die het proces van decontextualisering ondersteunen of oproepen. 'De docent heeft daarbij tot taak lerenden aan te zetten

tot reflectie, o.a. door vragen te stellen, cognitieve conflicten te induceren en leer- en denkprocessen zichtbaar te maken. ... in dit verband (werden) drie technieken met succes gehanteerd:
a. het expliciet laten verbaliseren van kennis en strategieën;
b. het laten opschrijven van ideeën en
c. het laten discussiëren in kleine groepen' (op. cit., p.13).

Een belangrijke transferconditie voor intuïtief denken op het vlak van instructie vormt de *brainstormingprocedure*. Deze procedure stelt bepaalde voorwaarden aan de wijze van discussiëren, waardoor een bepaalde vorm van decontextualisering, namelijk produceren van ideeën, wordt bevorderd. De brainstormingprocedure is erop gericht ideeën te genereren om oplossingen te vinden voor moeilijk oplosbare problemen of om bestaande oplossingen te verbeteren. Voor het welslagen van die procedure is wezenlijk dat onder de deelnemers een permissief klimaat heerst. Dat wil zeggen, ideeën die worden gelanceerd, worden niet meteen bekritiseerd, maar gezien als een mogelijke (gedeeltelijke) bijdrage aan de oplossing. Dat heeft tot gevolg dat de druk tot aanpassing aan de gevestigde normen in opvoeding, onderwijs en maatschappij, steeds minder werkzaam wordt. De zogenaamde censuur die ideeën niet eens tot bewustzijn laat komen, wordt verminderd of zelfs opgeheven, waardoor fysionomisch waarnemen mogelijk wordt en verrassende ideeën kunnen opkomen. Tot zover de brainstormingprocedure.

Tot nu toe is in het algemeen over transfer gesproken. De auteurs voeren ook het begrip transferafstand in. Deze term heeft betrekking op de mate van gelijkenis tussen de leer- en de transfersituatie. Is deze gelijkenis groot, dan wordt gesproken van 'near transfer'; is de gelijkenis klein dan is sprake van 'far transfer'. Deze gelijkenis is echter moeilijk objectief te bepalen onder meer als gevolg van het verschil in voorkennis tussen subjecten, vinden de auteurs.

De transferafstand wordt vermoedelijk niet alleen bepaald door de voorkennis, maar ook door de mentale set (Luchins, 1969). Deze mentale set kan worden doorbroken door leerlingen een receptieve, ontvankelijke houding ten aanzien van taken bij te brengen - onder de hiervoor genoemde condities. Het feit dat de objectieve gelijkenis niet hoeft samen te vallen met de subjectieve gelijkenis, vormt in het algemeen een pleidooi voor het ontwikkelen van een receptieve houding ten opzichte van taken, om daardoor de toegankelijkheid van relevante informatie zo groot mogelijk te doen zijn. Anders gezegd, om te bereiken dat de fysionomie zo goed mogelijk wordt gedifferentieerd door relevante voorkennis - voorzover aanwezig in het lange duur geheugen.

NOTEN

1. In navolging van Salomon & Perkins (1989) onderscheiden de auteurs twee verschillende wegen naar transfer, namelijk de 'low road' en de 'high road'. 'Deze "high road" is bovenal aangewezen wanneer men transfer nastreeft van wetenschappelijke denkschema's, heuristische procedures en metacognitieve vaardigheden, leermethoden, enz.' (loc. cit.).
2. Waar hij/zij/hem is geschreven, wordt ook zij/haar bedoeld
3. Dat wil overigens niet zeggen dat denken uit twee achtereenvolgende fasen bestaat: intuïtie en analyse. Tijdens de uitvoering van een probleem komt het vaak voor dat de denker op een idee moet komen.

LITERATUUR

Bastick, T. (1982). Intuition. How we think and act. Chichester: John Wiley & Sons.

Berne, E. (1949). The nature of intuition. *Psychiatric Quarterly, 23,* 203-226.

Drevdahl, J.E. (1964). Some developmental and environmental factors in creativity. In C.W. Taylor (Ed.), *Widening horizons in creativity* (pp. 170-186). New York: Wiley

Fischbein, E. (1990). Intuition and information processing in mathematical activity. *International Journal of Educational Research, 14,* 31-51

Frijda, N.H. & Elshout, J.J. (1976). Probleemoplossen en denken. In Michon, J.A., Eijkman. E.G.J. & Klerk, L.F.W. de (Red.) *Handboek der Psychonomie* (413-441). Deventer: Van Loghum Slaterus

Hermans, H.J.M. (1983). De halve persoonlijkheid van de onderwijspsychologie. In Dijkstra, S., Dudink, A.C.M. & Takens, R.J. (Red.) *Psychologie en onderwijs* (185-194). Lisse: Swets & Zeitlinger.

Krause, R. (1977). *Produktives Denken bei Kindern. Untersuchungen über Kreativität.* Weinheim: Beltz Verlag.

Luchins, A.S. (1967). Mechanization in problem solving: the effect of Einstellung. In Ray,W.S. *The experimental psychology of original thinking* (173-176). New York: Macmillan & Co.

MacKinnon, D.W. (1966). Instructional media in the nurturing of creativity. In C.W. Taylor & F.E. Williams (Eds.), *Instructional media and creativity* (pp. 179-216). London: Wiley & Sons.

Salomon, G. & Perkins, D.N. (1989). Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*, 113-142

Simons, P.R.J. & Verschaffel, L. (1992). Transfer: onderzoek en onderwijs. *Tijdschrift voor Onderwijsresearch, 17*, 3-17

Weisberg, P.S. & Springer, K.J. (1961). Environmental factors in creative functioning. *Archives in general Psychiatry*, 5, 554-564.

Werner, H. (1959). *Einführung in die Entwicklungspsychologie*. München: Barth.

Wolters, B.J. (1980). De functie en de meetbaarheid van het verbeeldende denkaspect in probleemoplossen. *Pedagogisch Tijdschrift, 5*, 495-506

Wolters, B.J.M. (1987). Kan de impasse in het creativiteitsonderzoek doorbroken worden? *Pedagogische Studiën, 64*, 485-498.

Wolters, B. (1991). Docentbekwaamheden en produktieve denkvaardigheden. *Nederlands tijdschrift voor Opvoeding, Vorming en Onderwijs, 7*, 261-284 (a)

Wolters, B. (1991). De funktie van intuïtief handelen bij de opbouw van onderwijsbekwaamheden. *Pedagogisch Tijdschrift, 68*, 449-461 (b)

Wolters, B. (1991). Leren: terugkoppelen en reflectie. *Gids voor de opleidingspraktijk*, nr.8 rubriek 6.12, pp. 1-32. (c).

Manuscript ontvangen 13-10-1992
Definitieve versie ontvangen 17-3-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 1, pp. 58-62

Notities en Commentaren

HOE SOMS IS VAAK?

Chris G.J. Michels en Marinus J.M. Voeten[1]
Vakgroep Onderwijskunde, K.U. Nijmegen

Voor de meting van persoonlijkheidskenmerken door middel van vragenlijsten wordt vaak gebruik gemaakt van antwoordalternatieven die peilen naar de frequentie van een bepaald gedrag. Hierbij worden termen gebruikt zoals 'bijna nooit', 'soms', 'vaak', 'bijna altijd'. Bekende voorbeelden in het Nederlandstalige gebied zijn vragenlijsten voor de meting van emotiedisposities als angst en woede (cf. Van der Ploeg, Defares, & Spielberger, 1981; 1982) of voor de meting van disposities op het terrein van stress-hantering (Hendriksen, 1990) en de expressie en controle van woede (Maes, Van Elderen, Van der Ploeg, & Spielberger, 1985). Bij dit type meetinstrumenten wordt ervan uitgegaan, dat dergelijke disposities om bijvoorbeeld gevoelens van angst of woede te ervaren adequaat kunnen worden gemeten door aan de respondent middels frequentie-aanduidingen te vragen hoe deze zich in het algemeen in bepaalde situaties gedraagt.

In de genoemde voorbeelden wordt een angst- of woede-dispositie gedefinieerd in termen van individuele verschillen in de frequentie waarmee een emotie doorgaans wordt ervaren. De onderzoeker dwingt de respondent zich uit te drukken in termen die in het dagelijkse spraakgebruik gehanteerd worden om een frequentie aan te geven. Deze termen zijn echter allesbehalve exact in hun verwijzing naar de frequentie van gedrag. Frequentiebegrippen zijn frequent ambigu: hoe vaak is 'soms'? En bovendien: hoe verhoudt het 'vaak' van de één zich tot het 'bijna altijd' van een ander?

In het kader van een groter onderzoek naar de emotie 'boosheid' werd bij leerlingen van 4-havo/vwo een pilot-studie verricht naar de numerieke betekenis die zij geven aan 17 verschillende frequentiebegrippen. Allereerst moesten de begrippen worden gerangschikt van 'absoluut nooit' tot 'absoluut altijd'. Vervolgens dienden de leerlingen aan de begrippen een percentage toe te kennen op een schaal van nul tot honderd. De bedoeling van deze studie was enerzijds meer inzicht te verkrijgen in de frequenties die leerlingen 'überhaupt' associëren met veel in vragenlijsten gebruikte antwoordalternatieven als 'bijna nooit', 'soms', 'vaak' en 'bijna altijd'. Anderzijds was het de bedoeling te evalueren in welke mate het frequentiecontinuüm door deze begrippen wordt gedekt. De begrippen werden contextonafhankelijk aangeboden.

Uit onderzoek is bekend dat de numerieke invulling bij frequentiebegrippen afhankelijk is van de context waarin de begrippen worden gehanteerd. Zo zal de frequentie die wordt geassocieerd met het begrip 'vaak' hoger liggen in de context van diefstallen van autoradio's in Amsterdam dan in de context van aardbevingen in Limburg. Er treedt dus bij de numerieke invulling van een frequentiebegrip een verschuiving op in de richting van de geschatte frequentie van de context (cf. Pepper & Prytulak, 1974; Newstead & Collis, 1987). Onafhankelijk van enige context frequenties toekennen, leidt tot informatie over de omvang van individuele verschillen in frequenties die geassocieerd worden met frequentiebegrippen. Hoogstraten (1979) rapporteert, in een nog steeds zeer lezenswaardig boek over valkuilen in sociaal-wetenschappelijk onderzoek, over onderzoek naar antwoordtendenties waarbij aan respondenten puur antwoordcategorieën werden aangeboden zonder bijbehorende vragen.

1. Met dank aan E. Croes, K. Kiesling, B. Koopman en M. van Mierlo voor het verzamelen van de data.
Adres: Vakgroep Onderwijskunde, Postbus 9103, 6500 HD Nijmegen.

METHODE

Aan het onderzoek werkten 96 leerlingen mee (van wie 66 meisjes en 30 jongens, in leeftijd variërend van 15-19 jaar), afkomstig van vier havo/vwo-klassen van vier aselect gekozen scholen in Oost-Nederland, in grootte variërend van 700 tot 1450 leerlingen.

De leerlingen kregen tijdens de reguliere schooltijd een lijst van 17 frequentiebegrippen in willekeurige volgorde voorgelegd met de opdracht de begrippen te rangschikken van 'absoluut nooit' (rangnummer 1) tot 'absoluut altijd' (rangnummer 17), waarbij ex aequo's niet waren toegestaan. Bij deze opdracht werd de leerlingen de hint gegeven dat het tot hulp zou kunnen zijn bij het rangschikken eerst groepjes te maken van begrippen die qua frequentie dicht bij elkaar liggen. Ná de rangschikking moesten de leerlingen als tweede opdracht bij elk van de gerangschikte begrippen middels een getal tussen 0 (= absoluut nooit) en 100 (= aboluut altijd) aangeven welke relatieve frequentie bij het begrip hoort. Bij deze opdracht was het wél toegestaan zo nodig meerdere keren hetzelfde percentage toe te kennen. De bij de twee opdrachten horende instructie stond boven aan de begrippenlijst en werd door de testleider zo nodig individueel toegelicht.

De frequentiebegrippen werden voor het merendeel ontleend aan een pilot-studie van Bonke en Verhoeven (1978). Van die begrippen werd het begrip 'helemaal niet' vervangen door 'absoluut nooit', en werden vier begrippen toegevoegd, namelijk 'altijd', 'absoluut altijd', 'af en toe' en 'niet zo vaak'. Tabel 1 geeft een overzicht van de 17 gehanteerde begrippen.

RESULTATEN

Rangschikking frequentiebegrippen

Na inspectie van de data werden 10 leerlingen uit het bestand verwijderd op grond van ongeldige rangschikkingen en/of onwaarschijnlijke gegevens. Over de gegevens van de resterende 86 leerlingen werd van de rangnummers per frequentiebegrip de modus bepaald, het percentage leerlingen in de modus, het gemiddelde en de standaardafwijking van de rangnummers en het laagste en hoogste toegekende rangnummer ('bereik'). Tabel 1 geeft een overzicht vań deze beschrijvende gegevens, waarbij de begrippen zijn geordend in volgorde van de modus en bij gelijke modus in volgorde van het gemiddelde.

Zoals uit Tabel 1 blijkt lopen de modi vrij regelmatig op, met uitzondering van de begrippen 'soms' en 'niet zo vaak' enerzijds en 'meestal' en 'zeer vaak' anderzijds, die een gelijke modus bezitten. Ook de gemiddelde rangnummers lopen – parallel aan de modi – relatief gelijkmatig op, waarbij de begrippen 'soms', 'niet zo vaak', 'af en toe' en 'zo nu en dan' op geringe afstand van elkaar rond de 6.5 liggen. Gelet op het percentage leerlingen in de modus en op de standaardafwijking en het bereik van de toegekende rangnummers, blijken de respondenten het veel meer eens te zijn over de rangschikking van de begrippen die een lage óf een hoge frequentie aanduiden dan over de begrippen die minder extreme frequenties aanduiden. Over de plaats die de begrippen '(bijna) nooit', 'zelden' en '(bijna) altijd' in de rangschikking innemen, is minstens 80 procent van de leerling het eens. De overeenstemming ten aanzien van de overige begrippen varieert tussen de 30 – 40 procent, met uitzondering van de middelste categorie in de rangschikking 'nogal eens', waarover nog een behoorlijke overeenstemming (61%) bestaat.

Percentages toegekend aan de frequentiebegrippen

Tabel 2 geeft de gemiddelden, standaardafwijkingen, mimima en maxima weer van de aan de afzonderlijke frequentiebegrippen toegekende percentages. De rangordening van de begrippen in deze tabel op grond van de gemiddelde percentages blijkt overeen te komen met die op grond van de modale rangnummers in Tabel 1.

Gelet op de verdeling van de gemiddelde percentages over het continuüm van 0 tot 100 blijken de (relatieve) frequenties van de begrippen 'soms', 'niet zo vaak', 'af en toe' en 'zo nu en dan' dicht bij elkaar te liggen (tussen 26-31%). Ook de frequenties van de begrippen 'dik-

Tabel 1. Rangschikking van frequentiebegrippen: Modus, gemiddelde en spreiding van de toegekende rangnummers (N=86).

	Modus	%	Gem.	St.dev.	Bereik
absoluut nooit	1	100.0	1.0	0.00	1- 1
nooit	2	100.0	2.0	0.00	2- 2
bijna nooit	3	80.2	3.2	0.44	3- 5
zelden	4	77.9	3.9	0.49	3- 6
soms	5	40.7	6.1	1.14	5- 8
niet zo vaak	5	37.2	6.3	1.46	3-10
af en toe	7	37.2	6.6	0.91	5- 9
zo nu en dan	8	38.4	7.0	0.95	5- 8
nogal eens	9	60.5	9.4	0.94	7-14
dikwijls	10	33.7	10.8	1.21	9-14
tamelijk vaak	11	27.9	10.9	1.28	9-14
vaak	13	34.9	12.1	1.05	9-15
meestal	14	31.4	12.2	1.73	9-15
zeer vaak	14	46.5	13.7	0.85	12-16
bijna altijd	15	81.4	14.7	0.75	10-15
altijd	16	98.8	16.0	0.11	15-16
absoluut altijd	17	100.0	17.0	0.00	17-17

% = het percentage respondenten in de modus

Tabel 2. Percentages toegekend aan de frequentiebegrippen: Gemiddelde, spreiding en bereik van de percentages (N=86).

	Gem.	St.dev.	Bereik
absoluut nooit	0.0	0.00	0- 0
nooit	3.0	3.07	0- 10
bijna nooit	9.2	5.77	1- 25
zelden	11.7	5.81	2- 30
soms	26.1	12.00	5- 62
niet zo vaak	26.7	14.23	5- 67
af en toe	28.5	11.71	7- 60
zo nu en dan	31.0	12.69	4- 60
nogal eens	50.1	14.56	20- 80
dikwijls	61.9	13.40	30- 85
tamelijk vaak	62.4	15.01	20- 92
vaak	71.2	13.36	30- 90
meestal	71.8	15.90	25- 97
zeer vaak	81.5	9.43	55- 98
bijna altijd	87.6	8.86	45- 99
altijd	96.0	4.90	75-100
absoluut altijd	100.0	0.00	100-100

wijls' en 'tamelijk vaak' (gemiddeld 62%) en de begrippen 'vaak' en 'meestal' (gemiddeld 72%) blijken nauwelijks van elkaar te verschillen. Voorts valt op dat geen begrip voorkomt in het frequentiegebied van 31 tot 50 procent en ook bevindt zich een hiaat in het gebied tussen 12 en 26 procent. Het begrip 'nogal eens' blijkt qua numerieke invulling precies in het midden van de schaal te liggen (50%). Evenals bij de rangschikking het geval was blijkt ook hier een

behoorlijke overeenstemming tussen de respondenten te bestaan ten aanzien van de frequenties die worden geassocieerd met begrippen die een zeer lage frequentie aanduiden. De spreidingen van de begrippen 'nooit', 'bijna nooit' en 'zelden' zijn klein. Deze overeenstemming blijkt veel minder aanwezig bij de begrippen die een hoge frequentie aanduiden: enkel voor de categorie 'altijd' blijkt een overeenkomstige spreiding voor te komen. Voor de overige begrippen zijn de spreidingen hoog tot zeer hoog en blijkt het bereik minstens 50 procent van de schaal te omvatten. Aldus ontstaat een grote overlap in de relatieve frequenties van de onderscheiden frequentiebegrippen. Door deze grote individuele verschillen tussen respondenten in de relatieve frequenties die aan deze begrippen worden toegekend, komt het voor dat de ene respondent 'soms' associeert met een frequentie van 62 procent, terwijl een andere respondent 'vaak' een percentage toekent van 30 procent.

DISCUSSIE

In de inleiding van deze notitie werd de vraag gesteld welke relatieve frequenties door leerlingen – onafhankelijk van enige inhoud of context van vragen – worden geassocieerd met de frequentiebegrippen 'bijna nooit', 'soms', 'vaak' en 'bijna altijd' en in welke mate het frequentiecontinuum door deze begrippen wordt gedekt. Op grond van de gegevens in deze pilot-studie kan worden gesteld dat deze begrippen qua gemiddelde percentage vrij gelijkmatig over het continuum verspreid liggen (resp. 9, 26, 71 en 88 procent). In deze reeks begrippen ontbreekt een term voor het midden van het frequentiebereik. In sommige gevallen is het opnemen van een middencategorie bij de antwoordmogelijkheden op een vragenlijst niet noodzakelijk of zelfs minder wenselijk. Een dergelijke situatie kan zich bijvoorbeeld voordoen als het er om gaat opinies van personen te meten. Indien de vragen in het meetinstrument qua frequentie echter legitiem in een '50%-respons' moeten kunnen resulteren, zou op basis van onze resultaten het begrip 'nogal eens' (50 procent op de schaal) aan de antwoordalternatieven kunnen worden toegevoegd. Uit de resultaten blijkt echter tevens dat slechts die frequentiebegrippen redelijk eenduidig qua frequentie worden geïnterpreteerd, die zich bevinden aan de beide extremen van het continuum zoals 'bijna nooit' of 'bijna altijd'. De begrippen 'soms' en 'vaak' blijken een dermate groot bereik te bezitten, dat zij niet alleen elkaar overlappen, maar ook de nabij gelegen begrippen 'bijna nooit' en 'bijna altijd'. Door deze ambiguïteit blijkt voor de ene respondent 'soms' frequenter te zijn dan 'vaak' (of zelfs frequenter dan 'bijna altijd'!) voor een andere respondent (en omgekeerd). Weliswaar zijn voor 'soms' en 'vaak' synoniemen voorhanden onder de onderzochte begrippen, zoals 'niet zo vaak' of 'meestal', maar deze frequentiebegrippen lijden aan hetzelfde euvel. 'In view of the magnificent imprecision of frequency words, it is amazing that we can communicate at all. Do you know what I mean?' (Hakel, 1968, p. 534).

Het is niet onmogelijk dat de in deze studie gevolgde werkwijze – eerst de rangordening van de begrippen, vervolgens de toekenning van een relatieve frequentie – invloed heeft gehad op de toekenning van percentages aan de begrippen. Voor de respondenten wordt de taak van toekenning van percentages vergemakkelijkt, wanneer men eerst een rangorde heeft gemaakt. Verder kan men zich afvragen, of gemiddeld genomen dezelfde percentages zouden zijn toegekend aan de 'centrale' begrippen ('bijna nooit', 'soms', 'vaak' en 'bijna altijd') indien zou zijn gevraagd naar de relatieve frequentie van deze begrippen buiten de context van de overige 13 begrippen om. Uit onderzoek van Newstead en Collis (1987) is gebleken, dat de interpretatie van frequentiebegrippen niet afhankelijk is van de aanwezigheid van andere begrippen, indien er voldoende begrippen aanwezig zijn om het gehele frequentiecontinuum te bestrijken. Aan deze conditie lijkt in de voorliggende studie ruimschoots te zijn voldaan. Ook bleek in de genoemde studie de interpretatie van de begrippen onafhankelijk te zijn van de volgorde van aanbieding. In dat licht is het niet waarschijnlijk dat de hier gevolgde procedure tot een artefact heeft geleid.

Groter is het probleem met de evaluatie van de resultaten in het licht van het feit, dat de begrippen zijn aangeboden zonder inhoud of context. Dit niet zozeer voor wat betreft de rangschikking van de begrippen. Uit onderzoek is ampel gebleken, dat rangordeningen van frequen-

tiebegrippen in hoge mate consistent zijn, onafhankelijk van (al dan geen) verschillende contexten, of inhouden (cf. Hakel, 1968; Pepper & Prytulak, 1974; Newstead & Collis, 1987). Het gebruik van vage frequentie-aanduidingen zal dus in het algemeen tot adequate informatie leiden op een ordinale schaal. Bovendien kan in dit verband worden opgemerkt, dat respondenten zich in hun antwoordgedrag zullen aanpassen bij de antwoordmogellijkheden die de onderzoeker hen suggereert, bijv. door van links naar rechts enkele termen aan te bieden die een toenemende frequentie van voorkomen aanduiden. Wanneer de onderzoeker de pretentie heeft om tot meting op een intervalschaal te komen, zijn er meer problemen om een meting met behulp van vage frequentietermen te rechtvaardigen. De hier gevonden gemiddelde percentages zullen vermoedelijk eerder van toepassing zijn voor een context waarmee een betrekkelijk hoge frequentie wordt geassocieerd dan voor contexten met lage frequenties (cf. Pepper & Prytulak, 1974).

Niet alleen de context waarover de vraag gaat is relevant. Ook het referentiekader van de respondent zelf is van belang. Wanneer aan een aan alcohol verslaafde wordt gevraagd hoe vaak deze drinkt, dan zal het antwoord 'vaak' waarschijnlijk een andere frequentie inhouden dan datzelfde antwoord van een persoon die niet aan alcolhol is verslaafd. Afgezien van extreme frequentiebegrippen als 'altijd' of 'nooit' – welke niet zo goed bruikbaar zijn omdat ze te absoluut zijn (cf. Janssen 1982) – kunnen frequentiebegrippen tamelijk onbruikbare indicatoren zijn voor de feitelijke frequentie van het gedrag van individuen, als dat gedrag veel versus weinig voorkomt. Voor onderzoek naar gedragsdisposities op het terrein van de emoties ligt hier een duidelijk probleem. Dit probleem als gevolg van individuele verschillen in de interpretatie van frequentiebegrippen lijkt niet eenvoudig op te lossen. Het aangeven van een 'normfrequentie' of 'normpercentage' in de stam van de vraag blijft stuiten op het probleem, dat elk individu waarschijnlijk een verschillende 'verwachte' frequentie hanteert ten aanzien van de emotie waarover de vraag handelt (cf. Newstead en Collis (1987). Hoe schaalt een respondent een frequentiebegrip in? Op grond van eigen emotionele ervaringen? Of op grond van de waarneming van het individu van emotionele ervaringen van een of andere 'normgroep'? Deze problematiek lijkt zeker het onderzoek naar emoties parten te spelen.

LITERATUUR

Bonke, P.M.M. & Verhoeven, A.F.M. (1978). *Angstmeting in onderwijssituaties*. Nijmegen, Katholieke Universiteit, Nijmeegs Instituut voor Onderwijsresearch.

Hakel, M.D. (1968). How often is often? *American Psychologist, 23*, 533-534.

Hendriksen, J.G.M. (1990). *Stress bij basisschoolleerlingen*. Academisch proefschrift, Katholieke Universiteit Nijmegen.

Hoogstraten, J. (1979). *De machteloze onderzoeker*. Meppel: Boom.

Janssens, J.M.A.M. (1982). *'Ogen' doen onderzoek: Een inleiding in de methoden van sociaal wetenschappelijk onderzoek*. Lisse: Swets & Zeitlinger.

Maes, S., Elderen, T. van, Ploeg, H.M. van der & Spielberger, C.D. (1987). *De woede-expressie en controlelijst*. Lisse: Swets & Zeitlinger.

Newstead, S.E. & Collis, J.M. (1987). Context and the interpretation of quantifiers of frequency. *Ergonomics, 30*, 1447-1462.

Pepper, S. & Prytulak, L.S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. *Journal of Research in Personality, 8*, 95-101.

Ploeg, H.M. van der, Defares, P.B. & Spielberger, C.D. (1981). *Handleiding bij de Zelf-Beoordelings Vragenlijst, Z.B.V. Een Nederlandstalige bewerking van de Spielberger State-Trait Anxiety Inventory, S.T.A.I.-DY*. Lisse: Swets & Zeitlinger.

Ploeg, H.M. van der, Defares, P.B. & Spielberger, C.D. (1982). *Handleiding bij de Zelf-Analyse Vragenlijst, Z.A.V. Een Nederlandstalige bewerking van de Spielberger State-Trait Anger Scale*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 9-4-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 2, pp. 65-73

Improving opportunities for the educationally disadvantaged: introduction to this special issue

Greetje van der Werf
RION, Institute for Educational Research, University of Groningen *

ABSTRACT

The aim of this special issue is to analyze what educational characteristics improve the school achievement level of disadvantaged pupils. The issue contains five articles in which results of research into effective school and instruction characteristics for specific disadvantaged groups are presented. The data in three of the articles are derived from the evaluation of the Dutch Educational Priority Policy. After describing the background to the EPP, the present educational disadvantages, the recent Priority Policy and the design of the EPP evaluation, school and instruction characteristics are reviewed, which are theoretically relevant for pupils in general and for disadvantaged groups in particular. Finally the results of five separate articles are also incorporated. It is concluded that only a few theoretically important characteristics are empirically supported. In general, classroom-level characteristics would seem to be the most important.

INTRODUCTION

Almost everywhere in Europe and the United States the educational opportunities of pupils from low income families and from ethnic minorities are less favourable when compared to those of other children. Educational research has established time and again a firm relationship between the socioeconomic and cultural background of children and educational achievement.

The aim of this special issue is to find out which educational provisions could improve the achievements in school of Dutch and immigrant children from low income families. The focus will be on research results in primary schools, where particular policies for the educationally disadvantaged have appeared to be both effective and ineffective.

From the provisions aimed at improving achievement level, most of the variables under study in the different articles of this special issue are derived as predictor variables to explain differences between schools and teachers concerning the educational performance of different disadvantaged groups. The design of the research on which data set most of the articles are based – the data set of the national evaluation of the Educational Priority Policy – is also described. Finally the results of the articles are drawn together and conclusions drawn on what works and does not work for the educationally disadvantaged.

HISTORICAL BACKGROUND

Although educational disadvantage in the United States already drew political attention in the 1950s, this did not happen in the Netherlands before the late 1960s. The acknowledgement of educational disadvantage as a political and social problem, also within Dutch society, was strongly affected by the research findings of educational sociologists and psychologists (e.g. Van Heek, 1968; Rupp, 1969; Van Calcar, 1969). As a consequence of their research, different projects and programs were started to improve the educational opportunities for disadvantaged children. In the beginning these were carried out mainly in large cities such as Amsterdam, Rotterdam and Utrecht. Later, in 1974, a national, so-called Stimulation policy was launched.

* RION, P.O. Box 1286, 9701 BG Groningen.

Initially all activities were directed at Dutch low SES children, as in the early 1970s these children formed the most educationally disadvantaged group. However, soon after the start of the Stimulation policy, the number of foreign residents in the Netherlands increased dramatically. Today over a million of the 14,5 million inhabitants of the Netherlands is of foreign origin. That is some 7 per cent of the population and the number is still growing.

The Dutch government did not really pay attention to ethnic minorities in society until the early 1980s, as until then policy was based on the idea that the Netherlands was not an immigration country and that in due time most ethnic minorities would return to their country of origin.

However, it became increasingly clear that this was not going to happen and in 1982 a policy document on minorities was accepted. The objective of this policy was to give ethnic minority groups the same opportunities within society as the autochtonous Dutch, especially concerning the education of immigrant children. It was acknowledged that the educational opportunities of these children were just as poor as those of Dutch children from low income families (low SES children). It was also recognised that although second generation immigrant children would not have the same language problems as their parents, they still might be educationally disadvantaged due to most, especially Turkish and Moroccan children, coming from low SES families. Thus in 1986 the new Educational Priority Policy was launched, which drew together the former policies for Dutch low SES children and ethnic minorities.

THE EDUCATIONAL PRIORITY POLICY

The Educational Priority Policy (EPP) aims to improve educational opportunities for children who are disadvantaged because of social, economic or cultural circumstances. The emphasis is on improving achievements in arithmetic and the Dutch language at primary school level. The most important target groups are Dutch low SES children and ethnic minorities.

The EPP consists of both general and area facilities. The general facilities mean that all primary schools with children from the target groups have extra teachers, in proportion to the number of disadvantaged pupils within their schools. This is done by giving particular groups a higher weight when counting the number of children in a school. Dutch low SES pupils count for 1.25, immigrant children for 1.90, while bargees' children and caravan children are weighted with respectively 1.40 and 1.70. Per school the weighted number of children is divided by the actual number and multiplied by 100. On the basis of the resulting score (the school score), which can vary from 100 to 190, schools are provided with additional teaching staff over and above the number of teachers they receive on the basis of school size.

The second part of the program consists of providing priority areas with extra facilities. Priority areas comprise primary and secondary schools working together with educational and welfare bodies to carry out particular projects and programs for disadvantaged pupils. Such activities are only possible where the mean school score of the primary schools within the area is at least 170. To receive facilities, priority areas have to submit a plan of goals and activities. The facilities are then introduced to the area as a whole and divided among the participants by the area authority.

An important characteristic of the Priority Policy is that schools and priority areas are free to choose for themselves the activities by which they aim to improve the educational opportunities of target groups. This has resulted in many differences between schools and areas regarding provisions, and probably as a consequence of this, also in pupils' educational achievements. Until now, however, the EPP has not shown any significant effect; the educational achievement of disadvantaged pupils is still much lower than those of other children, as the next paragraph illustrates.

EDUCATIONAL DISADVANTAGE NOWADAYS

In Table 1 the most recent figures on test scores for arithmetic and the Dutch language in the first grade of secondary school as well as on first grade pupils in higher general secondary schools (HAVO) and pre-university schools (VWO), the two highest levels of secondary education, are presented respectively for Dutch low SES children, immigrant children and Dutch children from middle and higher income families. In Table 2 the same figures are presented for the different ethnic minority groups. In both tables the achievement scores are standardized and normalized scores with a mean of 50 and a standard deviation of 10.

Table 1. Mean arithmetic and language scores of first grade pupils and percentages of these pupils attending the two highest levels of secondary schooling (HAVO/VWO), per socioeconomic group and ethnic origin.

	Arithmetic	Language	% HAVO/VWO
Dutch low SES	47	48	8
Immigrant	44	44	4
Dutch middle/higher	53	53	30

Source: Driessen & Van der Werf (1992)

From Table 1 it becomes clear that the difference in achievement between Dutch low SES and immigrant children is almost as large as the difference between Dutch low SES and Dutch children from middle and higher income families. Both differences are about half a standard deviation. The difference between immigrant and Dutch middle-class children is almost one complete standard deviation.

Compared to the differences in achievement, the differences in level of secondary schooling attained shows even larger discrepancies. The percentage of children from Dutch middle and higher income families attending the highest levels of secondary education (30%) is more than three times higher than the corresponding number of Dutch low SES children (10%); for immigrant children the situation is even worse. Only 5% attend HAVO and VWO schools.

From Table 2 it appears that Turkish and Moroccan children are low achievers when compared with other ethnic minority groups, though children from Surinam and the Antilles do not fare much better. Turkish and Moroccan children also attend the two highest levels of secondary schooling less often.

Table 2. Mean arithmetic and language scores of first grade pupils and percentages of these pupils attending the two highest levels of secondary schooling (HAVO/VWO), per ethnic group.

	Arithmetic	Language	% HAVO/VWO
Aruba	46	47	10
Antilles	48	49	10
Surinam	45	46	10
Morocco	43	42	3
Turkey	44	43	3
Others 1[1]	48	48	14
Others 2[2]	51	51	25

1. Others 1 are children from countries that in general are associated with educational disadvantage (Greece, Italy, Spain, Yugoslavia). Children from these countries receive a weight factor of 1.90.
2. Others 2 are children from countries that are not associated with disadvantage (Germany, USA, etc). They do not receive a weight factor.

Source: Driessen & Van der Werf (1992).

Comparing ethnic minorities with Dutch low SES children, it appears that the latter achieve almost as well as Aruban and Antillian pupils, but lower than children from countries denoted as 'others'. Concerning secondary education level, Dutch low SES pupils score even lower than all ethnic minority groups apart from Turkish and Moroccan children.

The figures presented did not take into account the fact that there are also differences in the initial cognitive abilities of the individual groups. Especially Turkish and Moroccan children score much lower on these abilities than the other groups of children (Driessen & Van der Werf, 1992), probably because of ethnic bias in the intelligence test and family background (80% of their parents were not educated beyond primary school level). Other ethnic minority groups as well as Dutch low SES children also score less well on intelligence tests than Dutch middle-class pupils. So it might be possible that differences in educational achievement is partly due to differences in initial abilities. Further analyses, however, have shown that this was not the case. The differences in mean achievement scores remain almost the same after correction for differences in intelligence; the mean score of Dutch middle and higher SES children is one point lower, while the mean score of immigrant children is increased by one point, for arithmetic only. There is no change in the mean score of Dutch low SES children.

The conclusion from the figures presented above is that the educational needs of ethnic minorities as well as Dutch low SES pupil still deserve attention from policymakers, schools and teachers. This has already been acknowledged by the fact that the Educational Priority Policy is to be continued. More important however is what exactly schools and teachers might do to improve the educational achievement of disadvantaged pupils. In schools and priority areas a large amount of energy and time is spent on all kinds of activities, the effects of which on achievement still remain unclear. Therefore in evaluating EPP, all school and instructional characteristics that seem to be relevant for improving pupil achievement are studied. These characteristics are derived from theoretical knowledge, frequently empirically supported, on effective schools and effective instruction on the one hand, and theoretical notions, most of them not empirically supported, about specific measures for disadvantaged pupils on the other.

EFFECTIVE EDUCATION THEORY FOR THE EDUCATIONALLY DISADVANTAGED

At the end of the 1970s and early 1980s reviews appeared in educational literature on school characteristics that were empirically shown to be effective for the educational achievement of children, especially disadvantaged children (e.g. Edmonds, 1979; Purkey & Smith, 1983). Although the research the reviews are based on was not always methodologically sound, a rather consistent picture emerges about five characteristics that seem to be important in explaining differences in achievement between schools. These are: 1) strong educational leadership; 2) high expectations of learning results; 3) stressing basic skills; 4) an orderly and safe climate in school; and 5) regular and systematic evaluation of pupil results.

Until now these five characteristics were still not confirmed in Dutch research. More and more it is concluded that instructional characteristics are more important than characteristics at school level because instruction at classroom level is more closely related to pupils' learning processes than school level characteristics. School characteristics must be considered as conditions for optimal instructional processes in classrooms (Scheerens, 1989; Creemers, 1991). At school level the conditions considered to be the most important are: a policy that stresses the basic skills; educational leadership; competent teachers; financial and material resources; organizational and structural characteristics; and school climate.

From reviews on effective instruction the most important classroom level characteristics appear to be the amount of time a teacher spends on activities that are directed at improving learning results and the handling of management and instruction strategies that support educational results. These include: 1) maximizing learning time; 2) keeping order and setting clear rules; 3) giving appropriate subject-matter; 4) structuring lessons; 5) regular evaluation and

reinforcement; 6) setting clear goals; 7) giving content directed instruction (see e.g. Walberg, 1986; Bloom, 1984; Madaus, Airasian & Kellaghan, 1980; Brophy & Good, 1986; Peterson and Walberg, 1979; Brophy, 1986; Creemers, 1991). Other characteristics of effective teachers are: 1) stressing traditional learning content and 2) having high expectations of pupils' results and confidence in their learning potential (see e.g. Rutter et al., 1979).

These classroom level characteristics appear to be effective for all categories of children within schools. Besides these characteristics, expectations about effective characteristics for specific categories of students (differential effectiveness) can be found in literature. Most of these characteristics take as a starting point an assumed cultural gap between home and school. To close this gap, cooperation between schools and neighbourhood and social bodies is proposed as well as close links with parents, including parental participation in school in the form, for instance, of family activation programs. Recent research shows that such programs have little effect on the degree of parent participation. Van Tilborg (1987) however concludes that a specific form of family activation could have a positive effect on the learning results of low SES children, including ethnic minorities. Such an example is the home intervention program HIPPY (in the Netherlands called Op-Stap), in which parents learn to stimulate the cognitive development of their children. Another way is to stimulate parents, especially immigrant mothers, to learn basic Dutch language and arithmetic skills themselves. This could lead to a more favourable learning climate for their children. Cooperation with adult education bodies could optimize the conditions for such stimulation.

Also the link between learning content and pupils' experiences at home is an important issue. Activities for this include classroom talks, projects, intercultural education, cooperative learning and flexible working methods. Sometimes measures are proposed that are directed explicitly at improving learning results of target groups, like remedial teaching and differentiation of learning content.

In compensatory education much attention was traditionally given to language development, and this can still be seen in schools participating in the priority program, where much attention is given to improving oral and written expression.

The activities directed at specific categories of children are relevant both for Dutch low SES children and ethnic minority groups. For the latter additionally there is Education in Own Language and Culture (OETC), the content and form of which, whether or not in combination with intercultural education, might have a positive effect on the attitudes and learning results of ethnic minorities.

The provisions directed at the target groups of the priority program is by no means exhaustive, but the above-mentioned measures and activities are the most common ones in schools with high numbers of low SES children and ethnic minority groups. Along with the characteristics of effective schools and effective instruction, they are the most relevant characteristics to be analyzed for their effectiveness concerning the educationally disadvantaged. Most of them are measured in the evaluation of the EPP, from which the data in three of the articles are derived (Weide; Van der Velden; Mulder). Some of them are measured in additional studies, also reported in this issue (Jungbluth; Reezigt).Together the five articles offer a comprehensive picture about which characteristics are both effective and ineffective for specific disadvantaged groups. Before drawing conclusions from the articles, the design of the evaluation of the EPP is described, mainly for the sake of avoiding overlap between the different articles. For information on the first results of the evaluation itself refer to Tesser, Mulder & Van der Werf (1991) and to Van der Werf, Weide & Tesser (1991).

THE EVALUATION OF THE PRIORITY POLICY

The core part of the evaluation is a longitudinal study of successive cohorts of children in primary schools. The first measurement took place at the end of September 1988. Pupils from grades 4, 6 and 8 were tested in Dutch language, arithmetic, social competence and intelligence.

For grade 8 pupils information was also collected concerning their recommended entry level for secondary education. School level and classroom level variables were measured by means of written questionnaires for principals and teachers. Background features of the pupils and data on parental educational interest and stimulation were collected from parents.

Data collection takes place once in two years in some 700 schools, and each time in grades 4, 6 and 8. Thus every two years the measurement of a new grade 4 cohort is started, while former grade 8 pupils are followed-up in secondary schools. The sample consists of a 3.5 per cent random sample of all primary schools. Stratification criteria were the area factor and the school score factor. The school score factor has three levels, which represent the social composition of a school's pupil population. These three levels are schools with a score less than 105, schools with a score between 105 and 115 and schools with a score higher than 115. The area factor has two levels, schools that belong and schools that do not belong to a priority area. This stratification of the random sample resulted in six strata. Strata with less than 100 schools were added up until 100. For information on the representativeness of the sample, response rate and reasons for non-response, refer to Tesser et al (1991).

The evaluation design makes it possible to analyse the effects of the policy over a period of time, by comparing the achievements of the target groups with the random sample of primary school children. On the other hand the design offers the possibility to test differences in effectiveness between schools participating in a priority area and schools which do not, or between schools with different priority facilities.

At this moment the results of the second measurement in primary education taken in 1990 are available. In 1989 a cohort study in secondary education was also started. The results presented in Tables 1 and 2 are derived from that study.

Besides these longitudinal studies that form the core of the evaluation, some smaller additional studies were carried out. One example is the study on the cooperation of priority areas with adult education bodies, about which an article is here presented (Reezigt).

CONTENT AND MAIN CONCLUSIONS OF THIS SPECIAL ISSUE

Not every article deals with all predictor variables summarized in the above and each one uses different outcome measures for different disadvantaged groups, but together they all lead to some conclusions about what works and does not work for disadvantaged groups.

The articles by Weide and Van der Velden describe the effects of virtually all theoretically relevant school, teacher and instructional characteristics. Weide presents the effects of these variables on the language and arithmetic achievement of ethnic minorities. Van der Velden deals with the effects on language and arithmetics scores of Dutch low SES children. Jungbluth's article focuses on teacher characteristics only. He presents results about teachers' stereotypes and hidden goal differentiation towards immigrant children, and the effects of these characteristics on their language and arithmetic development. Mulder's article concentrates on differences between school principals' recommendations given to immigrant children for entry level for secondary education. To explain these differences she uses one of the structural variables at school level, namely the composition of the pupil population in terms of their average achievement. In Reezigt's article existing relationships between schools and priority areas on the one hand and adult education bodies on the other are described.

Overall, it appears that differences in effectiveness do exist between schools in terms of mean arithmetic and language achievements, even when taking into account individual pupil background characteristics, as well as in terms of principals' recommendations for entry level for secondary education. Although in every analysis the percentages of between schools variance are not the same and sometimes they are rather small, it must be concluded that schools and teachers do matter. Mulder's results however show that, at least with regard to recommendation for entry level for secondary education, a rather substantial part of between schools differences is explained by context factors like a school's mean achievement, which have much to do with

its pupil population. Such context effects were found earlier with regard to language and arithmetic achievements (e.g. Tesser et al, 1991; Van der Werf et al (1991). But even when comparing schools which are very much alike with respect to the pupil population, remarkable differences in effectiveness remain (Van der Velden; Weide; Jungbluth), which has important consequences for the achievements of both immigrant and Dutch low SES children.

So, differences in effectiveness between schools are partly explained by context factors like percentage of immigrant children within schools and schools' mean achievement, but after taking this into account, still differences in effectiveness remain, which might be explained by educational characteristics. Unfortunately the articles in this special issue do not give a clear and consistent picture about which characteristics are effective and which are not. Overall, however, it seems that classroom level characteristics are the most important ones. Van der Velden's results show that teachers' use of minimum goals have a positive effect on arithmetic achievements of Dutch low SES children. This effect was also found earlier for all children by Van der Werf et al (1991). The negative effect of ability grouping on arithmetic and language achievement, also found by Van den Velden and earlier shown by Reezigt & Weide (1989), is rather consistent with this finding. Ability grouping mostly goes hand-in-hand with different goals set for groups of pupils with different abilities, and is frequently related to socioeconomic or ethnic background. The negative effects of teachers' biased expectations and hidden goals differentiation, showed by Jungbluth, completes this picture.

Other effective characteristics that emerge from Weide's article and that are in agreement with earlier research into school and instruction effectiveness are time availability, the use of monitoring and frequently discussing pupils with learning difficulties. These variables all indicate the emphasis of schools on pupil achievement. For the rest, none of the characteristics mentioned in effective schooling literature was supported empirically in the separate articles. Some of them even seemed to have a negative effect (e.g. testing children and the principal's stimulation of teachers to emphasize high academic standards), as Weide's article shows.

The specific measures, considered theoretically to be important for disadvantaged pupils, seemed to have no effect at all. Two of them even have a negative effect: providing additional language lessons for immigrant children (Weide) and ability grouping (Van der Velden). The negative effect of the first one was also found earlier by Van der Werf & Weide (1991).

Rather surprising is that Weide found parental involvement with school and multicultural education had a positive effect. She explains this by suggesting that these variables are indicators of an open mind by teachers towards ethnic minorities and their parents. How this open mind explains the positive effect on achievement, however, is unclear. Until now in Dutch research a substantial effect of activities directed at pupils' parents was not found. In this respect Reezigt's article offers some new suggestions: instead of stimulating parents to be involved with the educational progress of their children, it seems more important that schools stimulate partents to educate themselves. Therefore schools should cooperate with basic adult education centres, firstly to make it easier for these institutions to reach parents, and secondly to leave adult learning to the professionals. Such co-operation to educate parents of disadvantaged children could be a useful addition to home intervention programs in order to give children a better start to primary education on the one hand and to a school's own activities to improve directly a pupil's achievement on the other.

Summarizing what works and what does not work for disadvantaged groups, it is clear that we know more about school activities and characteristics that are not effective than about those that are consistently effective. The effectiveness of home intervention programs also seems to be doubtful (Eldering & Vedder, 1992). The effectiveness of educating parents of the educationally disadvantaged is still unknown. Despite this, there are differences in opportunities in primary schools, which means there is still reason to be optimistic that opportunities for disadvantaged groups can be improved.

REFERENCES

Amico, J. d' (1982). Using effective school studies to create effective schools: no recipes yet. *Educational Leadership, 40,* 60-62.

Appelhof, P. (1986). *Een pleidooi voor effectief stimuleringsonderwijs (A plea for effective stimulative schooling).* In J.C. Voogt & A. Reints (Eds.). Naar beter onderwijs. Tilburg: Zwijssen.

Bloom, B.S. (1984). The 2 Sigma problem: The search for methods of group instruction as one-to-one tutoring. *Educational Researcher, 4,* 4-16.

Brophy, J. (1986). *Synthesizing the results of research linking teacher behaviour to student achievement.* San Francisco: AERA-paper.

Brophy, J.E., & Good, T.L. (1986). Teacher behaviour and student achievement. In M.C. Wittrock (Ed.). *Handbook of research on teaching* (p. 328-376). New York: MacMillan Publishing Company.

Calcar, C. van (1967). *Enschedese onderzoekingen* (Research in Enschede). Groningen: Wolters-Noordhoff.

Creemers, B.P.M. (1991). *Effectieve instructie. Een empirische bijdrage aan de verbetering van het onderwijs in de klas* (Effective instruction: An empirical contribution to improvement of schooling in the classroom). The Hague: SVO Balansreeks.

Driessen, G., & Werf, G. van der (1992). *Het functioneren van het voortgezet onderwijs. De positie van leerlingen in het eerste leerjaar* (The functioning of secondary education. The position of pupils in the first grade). Nijmegen/Groningen: ITS/RION.

Edmonds, R.R. (1979). *A discussion of the literature and issues related to effective schooling.* Cambridge: Center for Urban Studies, Harvard Graduate School of Education.

Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership, 37,* 15-24.

Eldering, L. & Vedder, P. (1992). *Op stap. Een opstap naar meer schoolsucces?* (On the road. A step to more succes in school ?). Lisse: Swets & Zeitlinger.

Heek, F. van, et al. (1968). *Het verborgen talent. Milieu, schoolkeuze en schoolgeschiktheid.* (The hidden talent. Socioeconomic background, school choice and school aptitude). Meppel: Boom.

Levine, D.U., & Lezotte, L.W. (1990). *Unusually effective schools. A Review and Analysis of Research and Practice.* The National Center for Effective Schools Research and Development.

Mackenzie, D.E. (1983). Research for school improvement: an appraisal of some recent trends. *Educational Researcher, 4,* 5-17.

Madaus, G.F., Airasian, P.W., & Kellaghan, T. (1980). *School effectiveness. A reassessment of the evidence.* New York: Mc Graw-Hill Book Company.

Meijnen, W. (1984). *Van zes tot twaalf* (From six to twelve). Harlingen: Flevodruk (SVO-reeks no. 79).

Peterson, P., & Walberg, H. (1979). *Research on teaching: concepts, findings and implications.* Berkeley, CA.: McCutcheon.

Purkey, S.C., & Smith, M.S. (1983). Effective schools: a review. *The Elementary School Journal, 83;* 427-452.

Reezigt, G.J. & Weide, M.G. (1989). *Effecten van differentiatie. Resultaten survey onderzoek* (Effects of differentiation. Results survey research). Groningen: RION.

Rupp, J.J.C. (1969). *Opvoeding tot schoolweerbaarheid. Een Utrechtse kinderpsychologische studie* (Education in social competence in school. A child psychology study in Utrecht). Groningen: Wolters-Noordhoff.

Rutter, M., Maugham, B., Mortimore, J., Austin, J., & Smith, A. (1979). *Fifteen thousand hours. Secondary schools and their effects on children.* London: Open Books, Publishers Ltd.

Rutter, M. (1983). School effects on pupil progress: research findings and policy implications. *Child Development, 54,* 1-29.

Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten* (What makes schools effective? Summary and analysis of research results). The Hague: SVO Balansreeks.

Slavin, R., & Madden, N.A. (1989). What works for students at risk: A Research Synthesis. *Educational Leadership, 46,* 4-13.

Tesser, P., Mulder, L., & Werf, M.P.C. van der (1991). *De eerste fase van de longitudinale OVB-cohortonderzoeken* (The first measurement of the longitudinal EPP cohort studies). Nijmegen/Groningen: ITS/RION.

Tilborg, L. van (1987). *De betekenis van het arbeidersgezin voor het leerniveau en de schoolloopbaan van het kind* (The effect of low SES family background on the child's achievement level and educational career). The Hague: SVO-reeks.

Walberg, H. (1986). *Een synthese van onderzoek naar onderwijsproduktiviteit* (A synthesis of research on educational productivity). In W.J. Nijhof & E. Warries (Eds.). *De opbrengst van onderwijs en opleiding* (The results of education and schooling). Lisse: Swets & Zeitlinger.

Werf, M.P.C. van der, Weide, M.G., & Tesser, P. (1991). *Het OVB in de school en in de klas. De eerste meting in het basisonderwijs* (The EPP in the school and in the classroom. The first measurement in primary schools). Groningen/Nijmegen: RION/ITS.

Werf, M.P.C. van der & Weide, M.G, (1991). Effectief onderwijs voor allochtone leerlingen (Effective education for ethnic minority pupils). *Tijdschrift voor Onderwijsresearch, 16,* 231-243.

Wolf, J.C. van der (1984). *Schooluitval: een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het reguliere onderwijs* (Drop out: an empirical study on the relationship between school level factors and drop-out in regular education). Lisse: Swets & Zeitlinger.

Manuscript ontvangen 17-9-1992
Definitieve versie ontvangen 30-3-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 2, pp. 74-84

Effective provisions for ethnic minorities in primary schools

Marga G. Weide
*COWOG, University of Groningen**

ABSTRACT

To establish the effectiveness of different educational provisions in schools for ethnic minorities a secondary analysis was carried out on a sample of 551 pupils in 45 primary schools having between 30 and 70 per cent of children of non-Dutch origin. The results show that schools with high levels of ethnic minorities differ in quality but not in equity. The differences in quality are small, and larger for Dutch language than for mathematics. Although these differences can be explained by educational provisions, it is difficult to pinpoint separate educational provisions that are effective for ethnic minority groups. There are indications that extra language lessons or exercises for ethnic minorities are not very effective provisions, at least not in the way they are usually handled. On the other hand, availability of time for a school subject, monitoring, regularly discussing pupils with learning and behavior problems and a teacher who pays attention to parents and background of ethnic minority pupils all seem to have a positive effect on mean achievement.

OBJECTIVES

In the Netherlands ethnic minorities and their performance in school have been a matter of concern for policymakers as well as researchers since the early eighties, when the Dutch government recognized that many of them would be staying permanently. From that time on many studies and projects were carried out describing the state of affairs concerning ethnic minorities and trying to find explanations for their relatively low performance in school. Although it is clear that much of what a child achieves is related to individual characteristics, research has shown that schools can make a difference, and that some schools are more effective than others (Edmonds, 1979; Brookover et al. 1979; Mortimore et al., 1988; Brandsma & Knuver, 1989).

In this article the effectiveness of schools and teachers for ethnic minorities is explored. To establish the effect of a school (or teacher) two dimensions of effectiveness are considered: quality and equity. The quality of a school is defined as the mean achievement level of its pupils, after taking individual pupil characteristics into account. When the quality of a school is high, however, the mean achievement of a specific group of pupils (e.g. ethnic minorities) can still be relatively low. Thus it is important to look at the equity dimension as well, i.e. the difference in achievement between children 'at risk' and other children. A school where the difference in achievement between these two groups is small, is considered to be effective. Here equity is defined as the difference in achievement between Turkish and Moroccan children on the one hand and Dutch working-class children on the other, although the latter group itself is considered 'at risk' when compared to middle-class children. This choice was made because Turkish and Moroccan children (who invariably belong to the working class) are the most disadvantaged and perform much worse than their Dutch counterparts (Tesser, Mulder & Van der Werf, 1991).

The research questions to be answered are:

do schools/teachers differ in quality, and if so, what educational provisions account for these differences?

* COWOG, University of Groningen, P.O. Box 335, 9700 AH Groningen.

- do schools/teachers differ in equity, and if so what educational provisions account for these differences?

The educational provisions that are taken into account are taken from literature reviews about effective education for all children (Brophy & Good, 1986; Scheerens, 1989; Levine & Lezotte, 1990; Creemers, 1991) and about effective measures and compensation programs for children at risk (Edmonds, 1979; Scheerens, 1987; Stedman, 1987; Slavenburg, Peters & Van Galen, 1989; Slavin & Madden, 1989; Levin, 1990). For a more extensive description of relevant educational provisions see the article by Van der Werf in this special issue.

METHOD

Sample

In order to answer the research questions, secondary analyses were performed on data collected for the national evaluation of the Educational Priority Policy (see also Van der Werf in this issue). Data were collected between September and December in 1988 and 1990 in grades 4, 6 and 8 of the same schools. Children were tested for intelligence, language and mathematics, while principals, teachers and parents had to fill in written questionnaires.

Schools having a minimum of 30 per cent and a maximum of 70 per cent of children of ethnic minority groups were selected for the analyses. The purpose of this was to select schools that would have enough ethnic minorities to take separate actions for this group, but not so many as to make a comparison with Dutch children impossible. In these schools children were selected who were tested in 1988 in grade 6 (age group 9 to 10) and in 1990 in grade 8 (age group 11 to 12). Moreover the children had to be no older than two years of age when they came to the Netherlands, a criterion that of course mainly concerns ethnic minority groups.

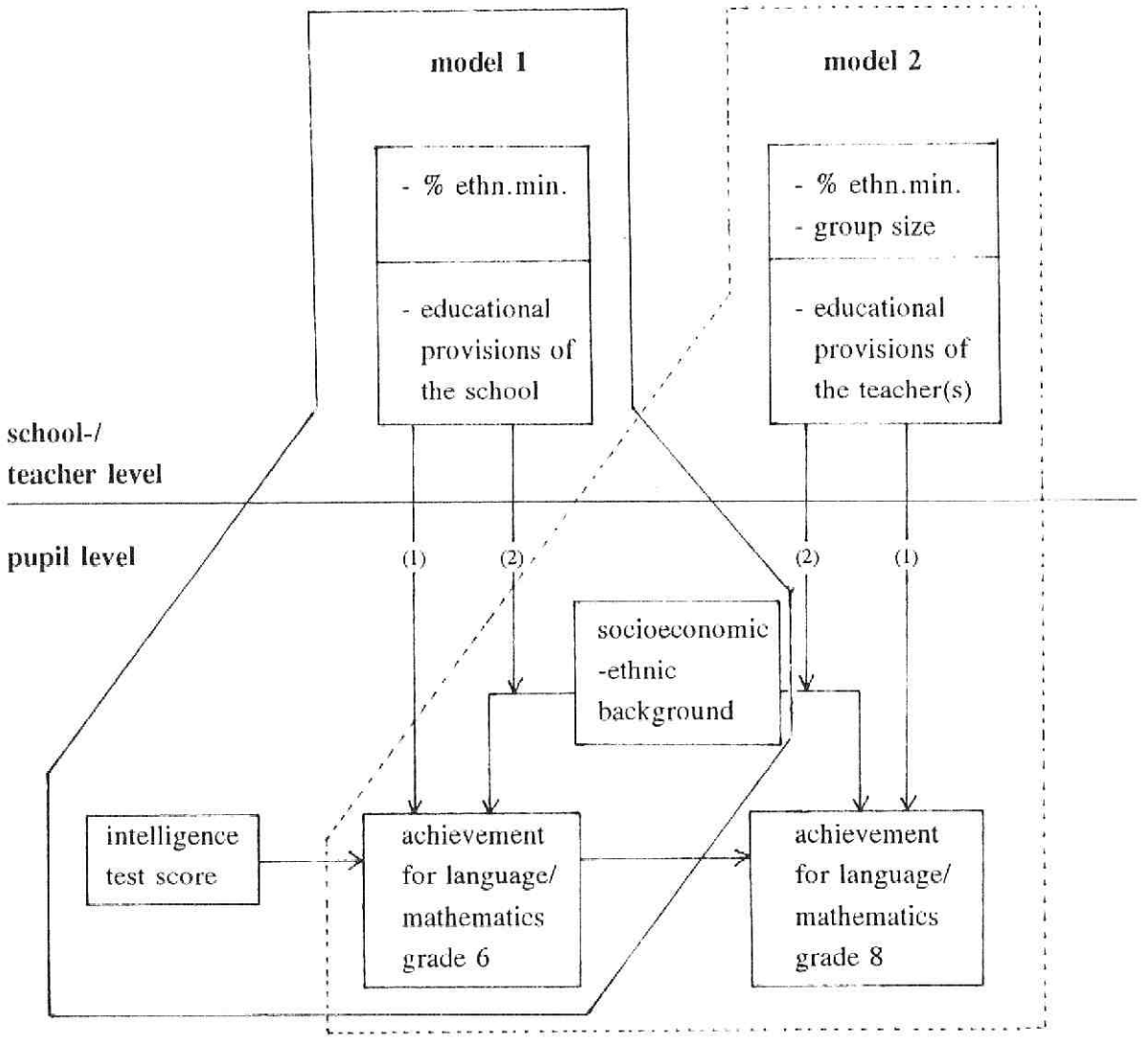
Schools were only selected if:
- they had five or more children with a complete set of data, of whom at least one child was Dutch and one child was Turkish or Moroccan;
- the principal had filled in a questionnaire both in 1988 and in 1990;
- data were available of both teachers the children had in grade 6 and grade 7;
- the period between the two testing dates was not less than one year and ten months and no more two years and two months.

There were 45 schools left that met these criteria and they are considered to be representative for 'average' schools with ethnic minority groups. Finally, 551 children were included in the analyses, divided into 51 'teacher'groups on the basis of the combination of teachers they had in grades 6 and 7.

Research model

In order to distinguish the effects of educational provisions at school level from the effects of educational provisions at teacher level, two multilevel models were used, a school model and a teacher model (see Figure 1).

The effect of the school is established at the first moment of data collection (model 1). At that moment the school and its educational provisions should already have had an impact on pupil achievement. At pupil level, language and mathematics achievement in grade 6 are the dependent variables and intelligence and the socioeconomic-ethnic background of a child are the covariates. The effect of the teacher(s) is established at the beginning of grade 8 (model 2). Language and mathematics achievement in grade 8 are the dependent variables and achievement in grade 6 and socioeconomic-ethnic background are covariates in the model. The percentage of ethnic minorities and group size are included in the model as covariates at school level, because in earlier research these variables were found to have effects on the quality of schools



model 1: school- effectiveness model
model 2: teacher-effectiveness model

Fig. 1. Multilevel research models.

(Glass & Smith, 1979; Tesser & Mulder, 1990; Dors et al., 1991). The educational provisions are the variables of main interest. These are supposed to have an impact both on the quality and the equity dimension; that is, to influence the achievement level of all pupils (arrow 1), as well as to affect the relationship between socioeconomic-ethnic background of pupils and their achievement (arrow 2).

Variables

The intelligence and achievement scores are obtained by means of standardized tests (Doddema-Winsemius & Van der Werf, 1989; Van Bergen, 1989); the socioeconomic-ethnic background is a combination of a pupil's weight factor (an indicator of a pupil's social class; see also Van der Werf in this issue) and ethnic background (based on the country of origin of both parents). On the basis of the weight factor children are divided into two socioeconomic groups:

high/middle and low. The latter are also subdivided into different ethnic groups. The 551 pupils are thus classified into seven socioeconomic-ethnic groups: 111 pupils from high/middle income families (98 Dutch and 13 of non-Dutch origin), 214 Dutch, 22 Surinamese, 76 Turkish, 57 Moroccan, 38 other ethnic minority and 33 'mixed' pupils from low income families (children with one parent born in the Netherlands and one parent born abroad).

The different provisions have been identified as:
- compensatory measures (e.g. remedial teaching, extra language lessons for ethnic minorities);
- policies emphasizing academic standards (e.g. set times for Dutch language/reading and mathematics, efficient organization of lessons, evaluation, monitoring);
- effective instruction (e.g. maximizing learning time, setting minimum goals, grouping);
- home-school relation (e.g. activities/information for parents, parental involvement);
- adapting education to the social and cultural background of pupils (e.g. special teaching methods for ethnic minorities, multicultural education).

'Effective instruction' refers exclusively to the educational provisions provided by teachers, while the other terms refer to educational provisions both at school and teacher level. At both levels 25 different educational provisions have been specified, which are partly operationalised by means of single items, partly by means of scales (reliabilities between .60 and .83). For a further specification the reader is referred to Van der Werf, Weide and Tesser (1991) and Weide (1992).

Analyses

The data were analysed by means of the VARCL multilevel program (Longford, 1988). The statistical model takes into account the fact that data have been collected using a two-stage sampling design. The total variance in the dependent variable (e.g. achievement) is divided into variance at pupil level (within-schools variance) and variance at school/teacher level (between-schools variance).

To establish which variables at pupil and school/teacher level are of importance, use is made of three different indices: the proportion of variance accounted for, the change in model fit, and the t-test for the regression coefficient of separate independent variables. To establish the effect of pupil variables, a significance level of one per cent is used; for school and teacher variables the significance level is set at five per cent. The effect of a socioeconomic-ethnic category is established by comparison with the Dutch lower income category. The socioeconomic-ethnic category is included in the model after having established the effect of individual differences in 'capacity', i.e. intelligence in the school model and achievement in grade 6 in the teacher model.

Separate analyses were executed for Dutch language and mathematics for both the school and teacher model.

RESULTS

As has been said before, two models were specified to establish the effect of the school and the teacher. Firstly the results of the analyses concerning the school model will be discussed, then the results concerning the teacher model.

The school-effectiveness model

In the school model the effect of the school is established on the achievement of pupils in grade 6. Firstly the relations between the intelligence and socioeconomic-ethnic background of pupils and their achievement are considered, and the extent in which schools differ in quality is established. Then the effectiveness of the different educational provisions at school level is discussed. Finally the results of the analyses concerning the effectiveness of schools on the equity dimension are presented.

Table 1. Proportions of variance at the pupil and school level for language and mathematics achievement in grade 6.

	language			mathematics		
variance	mod.0	mod.1	mod.2	mod.0	mod.1	mod.2
pupil level	85 %	73 %	64 %	89 %	68 %	65 %
school level	15 %	8 %	7 %	11 %	4 %	3 %
total	100 %	81 %	71 %	100 %	72 %	68 %
accounted for	–	19 %	29 %	–	28 %	32 %

model 0: empty model
model 1: intelligence included
model 2: intelligence and socioeconomic-ethnic background included

Table 1 shows that in the model without independent variables (the empty model) the initial variance at school level is 15 per cent and 11 per cent for language and mathematics, respectively.

The two pupil covariates account for nearly one third of the total variance. Intelligence and socioeconomic-ethnic background together explain about one quarter of the variance at pupil level; for language achievement both covariates seem equally important, but for mathematics achievement, intelligence has much more impact than socioeconomic-ethnic background. Turkish and Moroccan children perform less well than Dutch children of low income families on both language and mathematics tests. As for the differences between schools concerning language achievement, intelligence alone accounts for almost half of the initial variance at school level; for mathematics this is over sixty per cent. Socioeconomic-ethnic background accounts for a further six to nine per cent of the between-schools variance. At school level a seven and three per cent variance is left unexplained for language and mathematics, respectively, which means that the differences in quality between schools are not very large.

The covariate at school level, the percentage of ethnic minorities in a school, has no significant effect on either language or mathematics achievement. In order to discover what educational provisions have the most effect, the four categories of educational provisions that can be distinguished at school level were separately included in the model. Table 2 shows the proportion of variance accounted for by each category of educational provisions (the variance at school level that was left unexplained in the model with pupil covariates is set to 100 per cent).

The variables identified as 'policies emphasizing academic standards' account for the largest proportion of variance: 51 per cent for language and 100 per cent for mathematics. In the latter case the fit of the model also improves. The variables identified as 'home-school relation' and 'adapting education to the social and cultural background of pupils' account for the smallest proportion of variance: between 0 and 18 per cent; 'compensatory measures' do not account for much more: about a quarter of the variance.

For language only three variables have a significant relation with achievement, two from the category 'policies emphasizing academic standards' (efficient use of time by teachers and monitoring) and one 'compensatory measure' (extra language lessons). When only these three significant variables are included in the model, however, the regression coefficient of the variable 'efficient use of time by teachers' no longer differs significantly from zero. So in the final model for language, only two variables with significant effects are left: providing extra language lessons for ethnic minorities, which is negatively related to language achievement, and the use of a monitoring system, which has a positive impact on achievement.

Table 2. Proportions of variance at the school level for language and mathematics achievement in grade 6.

			language			
variance	mod.2	mod.3	mod.4	mod.5	mod.6	mod.7
school level	100 %	77 %	49 %	100 %	90 %	68 %
accounted for	–	23 %	51 %	0 %	10 %	32 %
number of school variables included		4	11	3	7	2
number of significant variables (5 %-level)		1	2	0	0	2
improved model fit?		no	no	no	no	yes
			mathematics			
variance	mod.2	mod.3	mod.4	mod.5	mod.6	mod.7
school level	100 %	74 %	0 %	82 %	91 %	0 %
accounted for	–	26 %	100 %	18 %	9 %	100 %
number of school variables included		4	11	3	7	4
number of significant variables (5 %-level)		0	4	0	0	4
improved model fit?		no	yes	no	no	yes

model 2: only pupil covariates included
model 3: pupil covariates and 'compensatory measures'
model 4: pupil covariates and 'policies emphasizing academic standards'
model 5: pupil covariates and 'home-school relation'
model 6: pupil covariates and 'adapting education to the social and cultural background of pupils'
model 7: pupil covariates and significant school variables

For mathematics all significant variables belong to the category 'policies emphasizing academic standards'. These concern availability of time for mathematics, discussing pupils with learning and behavior problems on a regular basis (both positively related to achievement), specially testing ethnic minorities and the principal stimulating teachers to emphasize high academic standards (both negatively related to achievement). The significant educational provisions in the final model (model 7 in table 2) account for 32 and 100 per cent of the variance at school level for language and mathematics, respectively.

In order to see whether schools differ at the equity dimension, socioeconomic-ethnic background was included in the model as a covariate with a random regression coefficient. For each socioeconomic-ethnic category it was tested whether the mean difference in achievement between children belonging to that particular category and Dutch children of low income families is equal for all schools. Here only the difference in achievement between Turkish/Moroccan and Dutch children of low income families is considered. In the model intelligence was included as a covariate at pupil level; no variables at school level were specified. The results of the analyses

show that schools do not differ in the difference in achievement between Turkish/Moroccan children and Dutch children of low income families, so there are no equity differences between schools. Thus the second part of the second research question can be ignored.

The teacher-effectiveness model

In the teacher model the effect of the teachers of grades 6 and 7 on pupil achievement, measured in the beginning of grade 8, is established. In order to discover to what extent teachers differ in quality, firstly the influence of the former achievement of pupils in grade 6 and their socioeconomic-ethnic background on their achievement in grade 8 is considered. Secondly the results of the analyses on the effectiveness of educational provisions are presented. Finally the differences between teachers on the equity dimension are discussed.

Table 3 shows that in the empty model the initial variance at teacher level is 11 and 10 per cent for language and mathematics, respectively.

Former achievement in grade 6 accounts for 40 to 50 per cent of the variance at pupil level and half of the variance at teacher level. Socioeconomic-ethnic background does not add much to this. Taking into account former achievement in grade 6, only Turkish children perform lower on language in grade 8 than Dutch working-class children. For mathematics, Turkish as well as Moroccan children achieve at the same level as Dutch working-class children. The covariates together account for 42 and 52 per cent of the total variance for language and mathematics, respectively. At teacher level only a six and four per cent variance is left unexplained for language and mathematics, respectively, so teachers do not differ very much in 'quality'.

The covariates at teacher level, the percentage of ethnic minorities in a group and group size, have no effect on language or mathematics achievement. Thus in the analyses where the different categories of educational provisions were, separately, included in the model, these covariates were excluded. The results of the analyses are presented in Table 4. In the table the variance at teacher level that was left unexplained in the model with pupil covariates is set to 100 per cent.

Now the picture for language and mathematics is rather different from the results of the school model. For the first school subject the category 'home-school relation' accounts for the highest proportion of variance (47 per cent), while for mathematics 'adapting education to the social and cultural background of pupils' has the greater impact, accounting for 64 per cent of the variance at teacher level. In both cases the fit of the model improves and one of the four, respectively three, variables has a significant positive relation with achievement. For language two other variables have a significant relation with achievement; one from the category 'com-

Table 3. Proportions of variance at the pupil and teacher level for language and mathematics achievement in grade 8.

	language			**mathematics**		
variance	mod.0	mod.1	mod.2	mod.0	mod.1	mod.2
pupil level	89 %	53 %	52 %	90 %	46 %	44 %
teacher level	11 %	6 %	6 %	10 %	5 %	4 %
total	100 %	59 %	58 %	100 %	51 %	48 %
accounted for	–	41 %	42 %	–	49 %	52 %

model 0: empty model
model 1: achievement grade 6 included
model 2: achievement grade 6 and socioeconomic-ethnic background included

Table 4. Proportions of variance at the teacher level for language and mathematics achievement in grade 8.

				language			
variance	mod.2	mod.3	mod.4	mod.5	mod.6	mod.7	mod.8
teacher level	100 %	76 %	75 %	95 %	53 %	81 %	40 %
accounted for	–	24 %	25 %	5 %	47 %	19 %	60 %
number of teacher variables included		5	6	7	4	3	2
number of significant variables (5 %-level)		1	1	0	1	0	2
improved model fit?		no	no	no	yes	no	yes
				mathematics			
variance	mod.2	mod.3	mod.4	mod.5	mod.6	mod.7	mod.8
teacher level	100 %	87 %	82 %	82 %	87 %	36 %	77 %
accounted for	–	13 %	18 %	18 %	13 %	64 %	23 %
number of teacher variables included		5	6	7	4	3	1
number of significant variables (5 %-level)		0	0	0	0	1	1
improved model fit?		no	no	no	no	yes	yes

model 2: only pupil covariates included
model 3: pupil covariates and 'compensatory measures'
model 4: pupil covariates and 'policies emphasizing academic standards'
model 5: pupil covariates and 'effective instruction'
model 6: pupil covariates and 'home-school relation'
model 7: pupil covariates and 'adapting education to the social and cultural background of pupils'
model 8: pupil covariates and significant teacher variables

pensatory measures' and one from the category 'policies emphasizing academic standards'. The significant (negative) effect of the latter variable (availability of time for language) disappears, however, when only the significant variables are included in the model. Thus for language only two variables have a significant effect on achievement in grade 8, and for mathematics only one variable, accounting for respectively 60 per cent, and 23 per cent of the variance at teacher level.

Of the significant variables, providing ethnic minorities with extra language exercises is negatively related to language achievement. A similar negative relation was already found in the school model between extra language lessons and language achievement. Positive relations are found between organizing information-evenings for parents and language achievement and between multicultural education and mathematics achievement.

As with the school model, in the teacher model socioeconomic-ethnic background was also specified as a covariate with a random regression coefficient, in order to see whether teachers differ in equity. This time only achievement in grade 6 was included as a covariate in the model.

Again no equity differences were found. This result is not surprising, since the differences in mean achievement in grade 8 between Turkish/Moroccan children and Dutch children of low income families are very small or insignificant, at least when differences in former achievement in grade 6 are taken into account. Here, too, the second part of the second research question can be ignored.

DISCUSSION

In this article an attempt was made to answer the question which educational provisions are effective for ethnic minorities. To establish this, two dimensions of effectiveness have been considered: the mean achievement of the pupils of a school (quality) and the difference in achievement between Turkish/Moroccan children on the one hand and Dutch working-class children on the other (equity). In order to distinguish the effect of the school from that of the teacher, two models of effectiveness were analyzed: a school model and a teacher model. In the first model the effect of the school on pupil achievement in grade 6 (taking into account the differences in individual intelligence scores) was established. In the second model the effect of the teachers of grade 6 and grade 7 on the achievement of the same pupils in grade 8 was assessed, taking into account differences in achievement in grade 6. Both models were analyzed separately for Dutch language and mathematics.

The results show that schools with a large number of ethnic minorities differ in quality, but not in equity, as it was defined here. Initially between 10 and 15 per cent of the total variance is between schools or teachers, but about half (sometimes even more) of these differences can be ascribed to differences at individual pupil level. Between three and seven per cent between-schools variance was left unexplained, after the pupil covariates had been included in the model. So the differences in quality are not very large, and they are larger for language than for mathematics. In other research, usually the opposite is found, i.e. more between-schools variance for mathematics than for language (see e.g. Meijnen, 1984; Brandsma & Knuver, 1989; Reezigt & Weide, 1989). This is generally explained by the fact that language is learned at home as well as at school, while mathematics is almost exclusively learned at school. However, for ethnic minorities this is not always the case; often their knowledge of the Dutch language is very limited when they start school, and they largely learn the language at school. So the fact that schools with large numbers of ethnic minorities differ more in Dutch language than mathematics achievement level might be explained by the different ways schools deal with this second-language challenge.

Of the five categories of educational provisions distinguished here, 'policies emphasizing academic standards' accounts for most of the between-schools variance in the school model, both for language and for mathematics. In the teacher model the categories 'home-school relation' and 'adapting education to the social and cultural background of pupils', account for most of the between-teacher variance in language achievement and mathematics achievement respectively. It is remarkable that in the teacher model the variables identified as 'policies emphasizing academic standards' and 'effective instruction' have hardly any impact on achievement.

There are only a few significant relations between educational provisions and achievement, and about half of these are negative. The significant variables differ per school subject, and although for some variables it stands to reason they are related to language rather than mathematics achievement, or vice versa (e.g. extra language lessons for ethnic minorities), for most variables this is not so (e.g. monitoring, homework, efficient managing of time, activities for parents).

Both in the school and teacher model, providing extra language lessons/exercises for ethnic minorities, has a negative relation with language achievement. In the school model it can in fact be argued that the relation is reversed and that extra language lessons are a consequence of low achieving ethnic minority groups. However, this argument does not hold for the teacher model, where extra language exercises were given in the period between the two testing dates. An

outlier study concerning the same dataset, but a different sample of schools, confirmed this finding: negative outliers (for language achievement) provided ethnic minorities more often with extra language lessons (Van der Werf & Weide, 1991). So this kind of compensation does not seem to be very effective, and in fact yields negative results. This is in accordance with other findings of the negative effects of compensating measures, which cause pupils to miss regular lessons, or are unrelated with the standard timetable (see e.g. Slavin & Madden, 1989). The other negative relations between specially testing ethnic minorities and a principal who stimulates teachers to emphasize high academic standards on the one hand, and mathematics achievement on the other, can also be interpreted as 'reverse' relations. These practices are possibly triggered by low achievements. Negative relations between certain activities of a schoolprincipal and pupil achievement were also found in other research (e.g. Brookover et al., 1979; Mortimore et al., 1988; Brandsma & Knuver, 1989; Van der Grift & Akkermans, 1991).

The positive relations in the school model seem more easy to interpret: having a monitoring system and regularly discussing pupils with problems make it possible to detect problems at an early stage and to act upon these in a quick and efficient manner. So it is plausible that schools that incorporate these policies have pupils who perform better. Also it is easy to understand why a school that gives much time to mathematics has pupils who do relatively well in this subject. In the teacher model, the positive effects of organizing information-evenings for parents and multicultural education on achievements are likely to be indirect. These provisions can be seen as indicators of an atmosphere in which the teacher pays attention to parents and has an open approach towards teaching ethnic minorities. In other research the importance of such an atmosphere has been confirmed (Stedman, 1987; Levin, 1990). Nevertheless, it is rather surprising that these variables have a significant effect on achievement, while variables that are indicators of the actual behavior of teachers, such as efficient use of time in the classroom, frequency of testing and providing pupils with direct feedback in the classroom, do not.

The main conclusion is that it is difficult to pinpoint individual educational provisions that are effective for ethnic minorities. Only a few educational provisions have a significant relation with achievement and sometimes the relations are negative. Moreover the significant relations that were found are not always obvious or easy to interpret, while more obvious relations are insignificant. Of course we must take into consideration that in this study teacher behavior was not observed, so these conclusions are based on what teachers say they do. It is also important to emphasize that the significant relations found should be interpreted with caution because of the large number of relations tested. In this article only the possible effects of individual educational provisions is explored. It is possible, however, that certain combinations of provisions are effective, while these on their own have no significant effect. Other researchers have warned against focusing on one 'effective' provison and argued that a combination of different provisions is likely to be most effective (Stedman, 1987; Levin, 1990; Levine & Lezotte, 1990; Creemers, 1991). For future research in this area, a research approach that focuses more on finding effective combinations of educational provisions rather than individual ones seems a logical follow-up.

REFERENCES

Bergen, J.B.A.M. van (1989). *Verantwoording constructie toetsen voor de evaluatie van het onderwijsvoorrangsbeleid. (Accountability for the construction of tests for the evaluation of the educational priority policy.)* Arnhem: CITO.

Brandsma, H.P. & Knuver, J.W.M. (1989). *Basisschoolkenmerken als determinanten van het functioneren van leerlingen. Deelrapport II: Resultaten. (Primary school features as decisive factors of pupil functioning. Report II. Results.)* Groningen: RION.

Brophy, J & Good, T. (1986). Teacher behavior and student achievement. In M.C. Wittrock (ed.). *Third Handbook of Research on Teaching, (pp. 328-375).* New York: McMillan, Inc.

Brookover, W.B., Beady, C., Flood, P., Schweitzer, J. & Wisenbaker, J. (1979). *School social systems and student achievement. Schools can make a difference.* New York: Praeger.

Creemers, B.P.M. (1991). *Effectieve instructie. Een empirische bijdrage aan de verbetering van het onderwijs in de klas. (Effective instruction. An empirical contribution to the improvement of classroom education.)* The Hague: SVO, Balansreeks.

Doddema-Winsemius, H. & Werf, M.P.C. van der (1989). *Selectie/constructie van toetsen voor het meten van sociale redzaamheid en intelligentie ten behoeve van de evaluatie OVB. (Selection/construction of tests for measuring social cognition and intelligence on behalf of the EPP evaluation.)* Groningen: RION.

Dors, H.G., Karsten, S., Ledoux, G., Steen, A.H.M. & Meijer, P.G. (1991). *Etnische segregatie in het onderwijs: beleidsaspecten. (Ethnic segragation in education: policy issues.)* Amsterdam: SCO.

Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership, 1979, 37,* 15-24.

Glass, G.V. & Smith, M.L. (1979). Meta-analysis of research on class-size and achievement. *Educational Evaluation and Policy Analysis, 1,* 2-16.

Grift, W. van der & Akkermans, W. (1991). Onderwijskundig leiderschap en leerlingprestaties in het basisonderwijs. (Educational leadership and pupil achievement in primary education.) *Tijdschrift voor Onderwijsresearch, 16, 4,* pp. 244-257.

Levin, H.M. (1990). *Accelerating the progress of at risk students.* Paper presented at the 8th Rockefeller Institute Policy Seminar, Nov. 8, 1990. New York.

Levine, D.U. & Lezotte, L.W. (1990). *Unusually effective schools. A review and analysis of research and practice.* The National Center for Effective Schools Research and Development.

Longford, N. (1988). *VARCL-manual.* Princeton, NJ: Educational Testing Service.

Meijnen, G.W. (1984). *Van zes tot twaalf. (From six to twelve.)* Harlingen: Flevo-druk.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988). *School Matters: the junior years.* Somerset: Open Books.

Reezigt, G.J. & Weide, M.G. (1989). *Effecten van differentiatie. Resultaten survey-onderzoek. Deelrapport IV. (Effects of adaptive instruction. Results survey. Report IV.)* Groningen: RION.

Scheerens, J. (1987). *Enhancing educational opportunities for disadvantaged learners. A review of Dutch research on compensatory education and educational development policy.* Amsterdam: North-Holland Publishing Company.

Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten. (What makes schools effective? Summary and analysis of research results.)* The Hague: SVO. Balansreeks.

Slavenburg, J.H., Peters, T.A. & Galen, W.L.M. van (1989). *The Dutch Education and Social Environment Project: closing report.* Rotterdam: School Advies Dienst.

Slavin, R.E. & Madden, N.A. (1989). What works for students at risk: a research synthesis. *Educational Leadership, 46, 5,* 4-13.

Stedman, L.C. (1987). It's time we changed the effective schools formula. *Phi Delta Kappan, 69, 3,* 215-224.

Tesser, P. & Mulder, L. (1990). Concentratie en prestaties van allochtone leerlingen in het basisonderwijs. (Concentration and achievement of ethnic minorities in primary education.) *Migrantenstudies, 6, 2,* 31-44.

Tesser, P., Mulder, L. & Werf, G. van der (1991). *De eerste fase van de longitudinale OVB-onderzoeken. Het leerlingenonderzoek. (The first phase of the longitudinal EPP-surveys. The pupil survey.)* Nijmegen/Groningen: ITS/RION.

Teunissen, J. (1988). *Etnische relaties in het basisonderwijs. 'Witte' en 'zwarte' scholen in de grote steden. (Ethnic relations in primary education. 'White' and 'black' schools in large cities.)* Dissertation University of Utrecht.

Weide, M.G. (1992). *Verslag schaalanalyses tweede meting OVB. (Report schale analyses second measurement EPP.)* Groningen: RION.

Werf, M.P.C. van der & Weide, M.G. (1991). Effectief onderwijs voor allochtone leerlingen. (Effective education for ethnic minorities.) *Tijdschrift voor Onderwijsresearch, 16, 4,* 231-243.

Werf, M.P.C. van der, Weide, M.G. & Tesser, P. (1991). *Het onderwijsvoorrangsbeleid in de school en in de klas. De eerste meting in het basisonderwijs. (The educational priority policy in the school and classroom. The first measurement in primary education.)* Groningen/Nijmegen: RION/ITS.

Manuscript ontvangen 17-9-1992
Definitieve versie ontvangen 30-3-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 2, pp. 85-98

The effects of teachers' attitudes and practices on the achievements of dutch low SES pupils

Lud van der Velden
*OCTO, University of Twente**

ABSTRACT

Results are reported from a study into the effects of teacher attitudes and practices on the cognitive functioning of Dutch low SES pupils.
Multilevel models were used to establish different effectiveness measures within a subset of the data from the national evaluation of the Educational Priority Policy. This was selected with respect to the reliability of pupil data and the comparability of the social composition of the schools.
Within this data set, adherents of two distinct ideologies (a curriculum-centered equal opportunity vision and a pupil-centered reform-pedagogical vision) could be identified. However, they differed only slightly with respect to approach and no effectiveness differences were found for either. Only two approach variables showed an effect on the cognitive functioning of Dutch disadvantaged pupils: the use of class ability grouping was negatively related to both language and mathematics scores, while the use of minimum goals for mathematics was positively related to mathematics scores.

INTRODUCTION

In Dutch literature on educational opportunities for disadvantaged children (Mathijssen, 1975; Meijnen, 1979; Hofstee, Koster & Meijnen, 1984; APS, 1986; Van der Werf & Hoeben, 1987), three diverse schools of thought can be found.

The first school, commonly known as the 'equal opportunity vision', believes that teachers and schools should use a curriculum-centered or cognitive oriented approach in which effort is directed at improving the cognitive development of children. The second, known as the 'reform-pedagogical vision', believes that teachers and schools should use a pupil-centered or non-cognitive approach in which effort is directed at the social-emotional well-being of children. The third, the 'collective emancipation vision', states that teachers and schools should use a group oriented approach, in which effort is directed at improving the social/economic position of children.

Within the 1988 data set of the national evaluation of the Educational Priority Policy (EPP), teachers' opinions on how to improve the educational opportunities of disadvantaged children could be scaled down to the 'equal opportunity' versus 'reform-pedagogical' schools. Teachers adhering to the 'collective emancipation vision' could not be identified. The distinction between the two groups of teachers is mainly in the degree to which they emphasized the cognitive goals of education. Therefore, in the following only the terms 'cognitive' and 'non-cognitive' will be used.

The difference in emphasis teachers place on cognitive goals to improve the educational opportunities for disadvantaged children is supposed to lead to different educational practices. Teachers with a cognitive attitude are supposed to pay more attention to learning basic skills in mathematics and Dutch language, to spending more teaching hours on these subjects and to emphasising achievements, etc. On the other hand, teachers with a non-cognitive attitude will allocate less educational resources on these subjects in order to spend these resources on non-

* OCTO, P.O. Box 217, 7500 AE Enschede.

cognitive goals. In other words, teachers with a cognitive attitude are supposed to teach more in line with so-called 'effective school and instruction characteristics' compared to others (see also Van der Werf & Hoeben, 1987).

Arguments for this hypothesis have already been presented by Van der Velden (1991). In the data set of the first measurement of the evaluation of the EPP in 1988, teachers with a cognitive attitude indeed scored higher on 'effective school and instruction characteristics' than those with a non-cognitive attitude. Until now, however, it is still unknown whether a cognitive attitude towards education for disadvantaged pupils also has an effect on their achievement.

In school and instruction effectiveness studies cognitive attitude often appears to be one of the effective characteristics that is found among other effective school and instruction characteristics (see e.g. Edmonds, 1979; Scheerens, 1989; Levine & Lezotte, 1990; Creemers, 1991). It is unclear whether this is an independent effect or is mediated by other school and instruction variables.

The question to be answered here is whether the cognitive attitudes of teachers effects the achievements of disadvantaged pupils and if so, whether this effect is additional to the one concerning the instructional practice of teachers or is mediated by this. For practical reasons – to avoid overlap with Weide's article – the focus here will be on Dutch low SES children.

The following three research questions will be answered:

1. Does a teacher's attitude (cognitive versus non-cognitive) affect the achievements of Dutch low SES children?
2. Does a teacher's educational practice affect the achievements of Dutch low SES children?
3. If a teacher's attitude affects achievement is this an independent affect or is it mediated by the educational practice of teachers?

In the EPP evaluation data set a large set of variables measuring the educational practice of teachers was available. Several preliminary analyses on these data revealed that only two approach variables were related to the achievements of Dutch low SES children: the use of 'class ability grouping' showed negative effects on both language and mathematics, and the use of 'minimum goals for mathematics' showed a positive effect on mathematics. Therefore this study will be limited to these two educational practice variables.

METHOD

Sample

The study uses data collected in 1988 and 1990 for the national evaluation of the EPP. A detailed description of the data gathered in 1988 can be found in Tesser, Mulder & Van der Werf (1991). A report on data gathered in 1990 is forthcoming. A stratified sample of schools was made (with a large overlap for the two years), with an overrepresentation of schools with relatively high numbers of disadvantaged pupils (see also Van der Werf).

From all the data available, only the measurements for grade 6 in 1988 and grade 8 in 1990 for Dutch low SES pupils were used for this article. The following criteria were used to select schools and pupils from these samples:

1. Only schools who participated twice and were tested within the September-December period in both years.
2. Schools should have at least 33 per cent Dutch low SES pupils (at school level) and at most 85 per cent. Schools should have at least 1 per cent non-disadvantaged pupils and 1 per cent immigrant pupils and at most 50 per cent non-disadvantaged or 50 per cent immigrant pupils.
3. At least 80 per cent of grade 6 pupils tested in 1988 should be traceable in grade 8.
4. Only Dutch low SES pupils whose gender was known and who had valid test scores on aptitude, language and mathematics in grade 6, and language and mathematics in grade 8, were selected.

5. At least 5 Dutch low SES pupils per school should be present for the final analysis.
6. Schools with missing values on the three teacher variables were excluded.

The final data set consisted of 74 schools and 756 Dutch low SES pupils, with one group and one teacher (the one who taught the pupils in grade 6 during the 1988/89 school year) for every school. So the group, teacher and school levels are the same within this data set. For every group we have three characteristics of the grade 6 teacher: the 'attitude' and two 'practice' variables.

Variables at pupil level

At pupil level test scores specially developed for the evaluation of the EPP (Van Bergen, 1989; Doddema-Winsemius & Van der Werf, 1988) are available. The grade 6 data are scores on aptitude, language and mathematics tests while grade 8 data consists of language and mathematics scores. The tests are reliable, with Cronbach's alpha about .80 for each one. The tests were normalized using a nationally representative weighted combination of the strata. All effects regarding these tests can therefore be interpreted with respect to a nationally representative mean of zero and a standard deviation of one. Within the selected, unweighted data set of the 756 Dutch low SES pupils, the test scores have means between -.15 and -.33 and have a typical standard deviation of .9 at pupil level. Thus, the final data set consists of pupils scoring a little under the national mean and with a somewhat smaller distribution relative to the spread of the total pupil population. At school level, the typical standard deviation of the test scores is .4. In the appendix, the mean, standard deviation, minimum and maximum of the group averages of the test scores, are reported.

Variables at teacher level

An 18 item attitude-scale (Cronbach's alpha= .78) measures the importance teachers attach to cognitive goals and to a cognitive approach. This scale was used to differentiate between teachers with a 'cognitive' attitude and teachers with a 'non-cognitive' attitude. The scale includes items like "the importance of preparing for secondary education versus the importance of developing according to one's own interest", "the importance of scholastic achievements" and "the importance of relating to pupils' perception of the environment". For more information regarding the content and psychometric characteristics of this scale, see Van der Velden (1991). An overview of the 18 items with their item-rest correlations is presented in Table 1.

As mentioned before, in this article only the effects of two variables measuring the daily educational practices of grade 6 teachers are analysed. The use of class ability grouping was measured by a 7 item scale (Cronbach's alpha = .74). This scale consists of items like "all pupils work at their own pace" and "fast learners are confronted with elaborate curriculum matters". The 7 items with their item-rest correlations are presented in Table 2.

The use of minimum goals for mathematics was measured using one item with three response categories: (1) no use of minimum goals, (2) the use of individual goals and (3) the use of common goals.

In the appendix, some descriptive results of the attitude and practice variables are reported. It should be noted that for all the grade 4, 6 and 8 teachers available in the 1988 data set of the national evaluation (±1900 teachers), the correlations between the attitude and practice variables is -.23 for class ability grouping and .17 for minimum goals for mathematics (Van der Velden, 1992). Within the selected data set, however, the correlations between the attitude of grade 6 teachers and their practices is -.17 for class ability grouping and .11 for minimum goals for mathematics. So, although the relationship between attitude and the two selected practice variables was not high in the total data set, it is even smaller in the selected data set.

Variables at school level

Although no variables regarding the attitudes or practices of a school are analysed, there are some school-level variables present in the analysis. First of all, the dates on which the schools

Table 1. Items for measuring 'cognitive attitude'.

The importance of cognitive goals and a cognitive approach *Cronbach's alpha = .78*	Item-rest correlation
The importance of:	
making demands on achievement	.11
more time for language and arithmetic	.26
scholastic achievement	.43
preparing for secondary education vs. developing individual interests	.44
knowledge-transfer vs. general training	.56
mastering basic skills vs. developing individual abilities	.30
knowledge-transfer vs. learning to co-operate	.54
preparing for secondary education	.47
intellectual vs. social skills	.55
cognitive development	.46
relating to pupils' perception of the environment*	.38
improving social skills*	.25
parental involvement*	.20
pupils feeling comfortable*	.34
giving pupils extra attention*	.04
giving attention to non-cognitive goals*	.40
working at one's own speed and level*	.44
flexibly evaluating achievement*	.26

*NB: Most items had five responses ranging from (1) strongly disagree, (2) disagree, (3) neutral to (4) agree (5) strongly agree. Some items had three responses which were recoded as 1, 3 and 5. High scores indicate an emphasis on cognitive goals, while low ones indicate an emphasis on non-cognitive goals. Items marked with a * are reversed.*

Table 2. Items for measuring 'class ability grouping'.

Approaches to class ability differences *Cronbach's alpha = .74*	Item-rest correlation
All pupils work at their own pace	.39
Slow learners are not confronted with certain parts of the curriculum	.35
Fast learners are paced at a higher level	.56
All pupils receive material that is for their particular level	.49
Slow learners are subjected to lower demands	.29
Fast learners are confronted with advanced parts of the curriculum	.54
Slow learners are paced at a lower level	.60

NB: For all items responses ranged from (1) never, (2) sometimes, (3) regularly to (4) mostly, (5) always.

were tested in 1988 and 1990 are available. Within the final data set, schools were tested in the period between mid-September and mid-December, both in 1988 and 1990. This spread in test dates is supposed to be relevant for pupil test scores: pupils in schools tested later should have higher test scores than those tested earlier. Differences in test scores related to differences in test date cannot be considered to reflect 'real' differences in achievements and corrections for these test date effects should be made before analysing the effects of other variables.

Furthermore, the social composition of a school's pupil population, indicated by the percent-

age of non-disadvantaged pupils, Dutch low SES pupils and ethnic minorities is taken into account when comparing the group averages of Dutch low SES pupils. As with the test date, the effects of differences in social composition on language and mathematics scores should not be considered reflecting differences in effectiveness between teachers. Descriptive results of school level variables are also reported in the appendix.

Data analysis

In general, four kinds of effectiveness criteria can be distinguished (see Bosker, 1991): 1. Gross achievement (mean achievement with no adjustment at all); 2. Net achievement (with an adjustment for differences in aptitude and other variables, but not for prior achievement); 3. Gross achievement gain (adjusted for prior achievement only); and 4. Net achievement gain (with adjustments for prior achievement, aptitude and other variables).

In this study, net achievement gain is considered to be the most relevant criterium for school effectiveness. Grade 8 output scores are first related to all relevant grade 6 input scores and than the effect of 'treatment' during the grade 6 period is estimated. However, using net achievement gain as an effectiveness criterium can lead to a certain 'overadjusting'. Grade 6 input scores are output scores themselves for the grade 1-5 period and are possibly related to teacher and school characteristics during this period. As far as teacher and school characteristics during the grade 1-5 period are related to both grade 6 achievement scores and grade 6 teacher characteristics, adjusting for prior achievement could lead to underestimating the influence of teacher characteristics. So, whether grade 8 achievement scores should be adjusted for grade 6 achievement scores remains open to discussion. Therefore, net achievement is also included as a relevant effectiveness criterium. Having language and mathematics scores each modeled in two ways (net achievement gain and net achievement), there are four different effectiveness indicators per group, teacher or school.

At first, so-called 'empty models' are analysed using individual grade 8 test scores for Dutch low SES pupils on both language and mathematics. As a second step, the individual grade 8 test scores are linked to differences in grade 6 test scores for these pupils on aptitude (for net achievement), or aptitude, language and mathematics (for net achievement gain), to gender and to between-school differences in background variables (the social composition of the school population and test dates). This step leads to the four effectiveness scores: net achievement and net achievement gain for both language and mathematics. The analyses needed to build these models are not reported here; only the resulting models are presented.

As a third step, the attitude variable is added to the four effectiveness models. This will give an answer to the first research question: the affect of attitude on the achievements of Dutch low SES children.

As a fourth step, the educational practice variables are added to the four effectiveness models. This will give an answer to the second research question: the affect of educational practice on the achievements of Dutch low SES children.

Finally, the attitude and practice variables are included. The difference in fit between the models in the fifth and the fourth step will give an answer to the third research question: whether attitude affects effectiveness or is mediated by practice. If, in the third step, it is concluded that attitude has no direct relation with 'effectiveness', the fifth step is not made.

Besides the 'explained variance' percentages per level, being the familiar way of expressing how well a specific model describes the data, the 'modeled variance' percentage of the individual and group scores is reported. In Snijders & Bosker (1993), it is shown that the 'modeled variance' percentage is theoretically a more sound measurement of how well the model fits the data. It gives an estimate of the proportional reduction in error when predicting the individual or group scores of the different models[1].

RESULTS

Net achievement gain for language

From the results in Table 3 it appears that the language pretest in grade 6 has a strong effect on the language posttest in grade 8. After correcting for pretest scores, gender (boys outperform

Table 3. Multilevel analysis for grade 8 language (756 Dutch low SES pupils from 74 schools): Net achievement gain.

Step: model	1: Empty model	2: Net achievement gain	3: Net ach. gain + Attitude	4: Net ach. gain + Approach
Fixed part				
Grand mean	-.333	-.125	-.321	.191
Pupil level				
- Gender *(girls vs. boys)*		-.154 *(.048)*	-.152 *(.048)*	-.151 *(.048)*
- Aptitude grade 6		.227 *(.030)*	.226 *(.030)*	.224 *(.030)*
- Language grade 6		.451 *(.031)*	.450 *(.031)*	.451 *(.031)*
Group level				
Experimental design				
- Test date 1988 *(per 10 days)*		-.025 *(.014)*	-.028 *(.014)*	-.024 *(.013)*
- Test date 1990 *(per 10 days)*		.059 *(.012)*	.060 *(.012)*	.052 *(.012)*
Social composition at school level				
- Perc. ethnic minorities *(per 10%)*		-.053 *(.026)*	-.064 *(.027)*	-.045 *(.025)*
Attitude grade 6 teacher				
- Cognitive attitude *(per 10 points)*			.047 *(.032)*	
Approach grade 6 teacher				
- Class ability grouping *(per 10 points)*				-.166 *(.059)*
Random part	*Initial var.*	*Remaining variance per level*		
Pupil level	.654	.414	.414	.414
Group level	.062	.017	.015	.012
Deviance	1873.2	1504.9	1502.8	1497.4
	Proportional distribution of the initial variance per level	*Remaining percentage variance per level (compared to the total initial variance)*		
Within groups	91.3%	57.8%	57.9%	57.8%
Between groups	8.7%	2.4%	2.2%	1.7%
	100.0%			
		Percentage 'explained' variance per level (compared to the initial variance per level)		
Within groups		36.7%	36.6%	36.7%
Between groups		72.0%	75.2%	80.9%
		Percentage 'modeled' variance per level		
Pupil level		39.7%	40.0%	40.5%
Group level		53.0%	54.4%	57.1%

NB: For the fixed part, regression coefficients are reported, with standard errors within parentheses.

girls) and aptitude still have an effect. Moreover, the test date variables (for which the effect sizes and standard errors are multiplied by 10) have the expected effects. Their inclusion makes sure that artificial effects due to differences in time points when the tests were taken are controlled for in the sequel. There is also a compositional effect concerning the percentage of immigrant pupils at school level: the higher the number of ethnic minorities, the lower the achievement scores of Dutch low SES pupils. The other covariates (mean gender and the percentage of non-disadvantaged and Dutch low SES pupils) did not appear to have any effect after the afore-mentioned variables were included in the model. It was also unnecessary to model the pretest or aptitude variables as centred variables at individual level and their means at group level. The introduction of all the significant covariates in the model shows that 39.7 per cent of individual level variation is accounted for and 53 per cent of group level variation.

After taking individual and group level covariates into account, teachers' attitude is included in the prediction model. The results indicate that attitude has no significant effect on net achievement gain in language performance for low SES pupils of grades 6 to 8.

When adding 'class ability grouping' to the net achievement gain model for language, a significant effect is found: the more teachers use this, the lower the scores of low SES pupils. After taking the covariates and this approach variable into account, the model shows that 40.5 per cent of the individual level variation is accounted for and 57.1 per cent of the group level variation.

Net achievement gain for mathematics

The results for net achievement gain on mathematics are comparable to the results for net achievement gain on language. The mathematics pretest in grade 6 has a strong effect on the mathematics posttest in grade 8. After correction for the pretest scores, gender (boys outperform girls) and aptitude still have an effect.

The test date variables again have the expected effects, although the effect for the 1990 test date is considerably smaller than that for 1990 on language. There is also a compositional effect concerning the percentage of non-disadvantaged pupils at the school level: the more non-disadvantaged pupils, the lower the achievement scores of Dutch low SES pupils. This result was not expected. The other covariates (mean gender and the percentages of Dutch low SES pupils and ethnic minorities) did not appear to have any effect after the afore mentioned variables were included in the model. It was also unnecessary to model the pretest or aptitude variables as centred variables at individual level and their means at group level. The introduction of all the significant covariates in the model shows that 50.1 per cent of individual level variation is accounted for and 61.6 per cent of group level variation. So, differences in net achievement gain on mathematics can be predicted more precisely than differences in net achievement gain on language.

After taking the individual and group level covariates into account, teachers' attitude is included in the prediction model. The results indicate that attitude has no significant effect on net achievement gain in mathematics performance. So, as for language, attitude has no significant effect on net achievement gain in mathematics scores for low SES pupils of grades 6 to 8.

When adding 'minimum goals for mathematics' and 'class ability grouping' to the net achievement gain model for mathematics, only one significant effect is found: the more teachers use class ability grouping, the lower the scores of low SES pupils. The use of minimum goals for mathematics does not have a significant effect and the fit of the model does not improve when this approach variable is taken into account (along with 'class ability grouping'). When entered on its own, however, the use of minimum goals does show a significant effect in that low SES pupils of teachers who use minimum goals have better mathematics scores than those of teachers who do not. After taking the covariates and the two approach variables into account, the model shows that 51.3 per cent of the individual level variation is accounted for and 67 per cent of the group level variation. When minimum goals is not entered in the model, almost the same amount of modeled variance is reached.

Table 4. Multilevel analysis for grade 8 mathematics (756 Dutch low SES pupils from 74 schools): Net achievement gain.

Step: model	1: Empty model	2: Net achievement gain		3: Net ach. gain + Attitude		4: Net ach. gain + Approach	
Fixed part							
Grand mean	-.203	.308		-.335		.461	
Pupil level							
- Gender *(girls vs. boys)*		-.246	*(.045)*	-.246	*(.045)*	-.246	*(.045)*
- Aptitude grade 6		.182	*(.030)*	.182	*(.030)*	.185	*(.030)*
- Mathematics grade 6		.543	*(.031)*	.543	*(.031)*	.533	*(.031)*
Group level							
Experimental design							
- Test date 1988 *(per 10 days)*		-.026	*(.014)*	-.026	*(.015)*	-.022	*(.013)*
- Test date 1990 *(per 10 days)*		.042	*(.013)*	.042	*(.013)*	.035	*(.012)*
Social composition at school level							
- Perc. non-disadvantaged pupils *(per 10%)*		-.051	*(.029)*	-.053	*(.030)*	-.054	*(.027)*
Attitude grade 6 teacher							
- Cognitive attitude *(per 10 points)*				-.005	*(.035)*		
Approach grade 6 teacher							
- Minimum goals for mathmetics						.058	*(.040)*
- Class ability grouping *(per 10 points)*						-.155	*(.064)*
Random part	*Initial var.*	*Remaining variance per level*					
Pupil level	.665	.355		.355		.356	
Group level	.103	.028		.028		.019	
Deviance	1905.4	1406.0		1406.0		1394.8	
	Proportional distribution of the initial variance per level	*Remaining percentage variance per level (compared to the total intitial variance)*					
Within groups	86.7%	46.3%		46.3%		46.3%	
Between groups	13.3%	3.7%		3.7%		2.4%	
	100.0%						
		Percentage 'explained' variance per level (compared to the initial variance per level)					
Within groups		46.6%		46.6%		46.6%	
Between groups		72.4%		72.4%		81.8%	
		Percentage 'modeled' variance per level					
Pupil level		50.1%		50.1%		51.3%	
Group level		61.6%		61.6%		67.0%	

NB: For the fixed part, regression coefficients are reported, with standard errors within parentheses.

Net achievement for language

The results for net achievement on language show that aptitude has a strong effect on the grade 8 language posttest when the grade 6 language pretest is *not* taken into account. Gender (boys outperform girls) still has an effect, but this effect is somewhat smaller than the effect of gender

Table. 5. Multilevel analysis for grade 8 language (756 Dutch low SES pupils from 74 schools): Net achievement.

Step: model	1: Empty model	2: Net achievement		3: Net achievem. + Attitude		4: Net achievem. + Approach	
Fixed part							
Grand mean	-.333	-.441		-.665		-.132	
Pupil level							
- Gender *(girls vs. boys)*		-.105	*(.054)*	-.103	*(.054)*	-.102	*(.054)*
- Aptitude grade 6		.412	*(.030)*	.412	*(.030)*	.410	*(.030)*
Group level							
Experimental design							
- Test date 1990 *(per 10 days)*		.054	*(.013)*	.055	*(.013)*	.048	*(.013)*
Social composition at school level							
- Perc. ethnic minorities *(per 10%)*		-.065	*(.029)*	-.076	*(.030)*	-.057	*(.028)*
Attitude grade 6 teacher							
- Cognitive attitude *(per 10 points)*				-.047	*(.035)*		
Approach grade 6 teacher							
- Class ability grouping *(per 10 points)*						-.156	*(.066)*
Random part	*Initial var.*	*Remaining variance per level*					
Pupil level	.654	.535		.535		.535	
Group level	.062	.019		.017		.014	
Deviance	1873.2	1695.2		1693.4		1689.8	
	Proportional distribution of the initial variance per level	*Remaining percentage variance per level (compared to the total initial variance)*					
Within groups	91.3%	74.8%		74.8%		74.7%	
Between groups	8.7%	2.6%		2.3%		1.9%	
	100.0%						
		Percentage 'explained' variance per level (compared to the initial variance per level)					
Within groups		18.1%		18.1%		18.2%	
Between groups		69.9%		73.0%		78.1%	
		Percentage 'modeled' variance per level					
Pupil level		22.6%		22.9%		23.4%	
Group level		42.0%		43.4%		45.8%	

NB: For the fixed part, regression coefficients are reported, with standard errors within parentheses.

in the net achievement gain model. The gender effect is barely significant, but the fit of the model is better when this is included.

The 1990 test date variable has almost the same effect in the net achievement model as for the gain model. This also applies to the compositional effect of the percentage of ethnic minorities at school level. The other covariates (mean gender and the percentages of Dutch low SES and non-disadvantaged pupils) again do not appear to have any effect after the afore-mentioned variables were included in the model. It was also unnecessary to model the pretest or aptitude

Table 6. Multilevel analysis for grade 8 mathematics (756 Dutch low SES pupils from 74 schools): Net achievement.

Step: model	1: Empty model	2: Net achievement		3: Net achievem. + Attitude		4: Net achievem. + Approach	
Fixed part							
Grand mean	-.203	.126		.207		.432	
Pupil level							
- Gender *(girls vs. boys)*		-.359	*(.052)*	-.360	*(.052)*	-.357	*(.052)*
- Aptitude grade 6		.462	*(.030)*	.462	*(.030)*	.459	*(.029)*
Group level							
Experimental design							
- Test date 1990 *(per 10 days)*		.058	*(.016)*	.058	*(.016)*	.046	*(.014)*
Social composition at school level							
- Perc. non-disadvantaged pupils *(per 10%)*		-.068	*(.037)*	-.072	*(.038)*	-.071	*(.032)*
Attitude grade 6 teacher							
- Cognitive attitude *(per 10 points)*				-.015	*(.044)*		
Approach grade 6 teacher							
- Minimum goals for mathematics						.086	*(.048)*
- Class ability grouping *(per 10 points)*						-.243	*(.077)*
Random part	*Initial var.*	*Remaining variance per level*					
Pupil level	.665	.492		.492		.493	
Group level	.103	.053		.053		.030	
Deviance	1905.4	1663.3		1663.2		1645.9	
	Proportional distribution of the initial variance per level	*Remaining percentage variance per level (compared to the total initial variance)*					
Within groups	86.7%	64.1%		64.1%		64.2%	
Between groups	13.3%	6.9%		6.9%		3.9%	
	100.0%						
		Percentage 'explained' variance per level (compared to the initial variance per level)					
Within groups		26.0%		26.0%		25.9%	
Between groups		48.1%		48.2%		70.7%	
		Percentage 'modeled' variance per level					
Pupil level		28.9%		29.0%		31.9%	
Group level		38.8%		38.9%		51.9%	

NB: For the fixed part, regression coefficients are reported, with standard errors within parentheses.

variables as centred variables at individual level and their means at group level. The introduction of all the significant covariates in the model shows that 22.6 per cent of individual level variation is accounted for and 42 per cent of group level variation. So, differences in net achievement on language (especially differences at the individual level) can be predicted less precisely than differences in net achievement gain on language.

After taking individual and group level covariates into account, teachers' attitude is included

in the prediction model. The results indicate that attitude has no significant effect on net achievement in language performance. So, again, it does not matter for low SES pupils whether their teachers have a cognitive attitude towards education or not.

When adding 'class ability grouping' to the net achievement model for language, a significant effect is found, with an effect size almost the same as for the net achievement gain model: the more teachers use class ability grouping, the lower the language scores of low SES pupils. After taking the covariates and approach variable into account, the model shows that 23.4 per cent of individual level variation is accounted for and 45.8 per cent of group level variation.

Net achievement for mathematics

For net achievement on mathematics the effect of aptitude is roughly the same as that on net achievement on language. The effect of gender (boys outperform girls), however, is much bigger for net achievement on mathematics than for any of the other models already discussed. The effect of the 1990 test date variable and the compositional effect of the percentage of non-disadvantaged pupils at school level are somewhat larger than the effects these variables had on net achievement gain on mathematics. The other covariates (mean gender and the percentages of Dutch low SES-pupil and ethnic minorities) again do not appear to have any effect after the afore-mentioned variables were included in the model. It is also unnecessary to model the pretest or aptitude variables as centred variables at individual level and their means at group level. The introduction of all the significant covariates in the model shows that 28.9 per cent of individual level variation is accounted for and 38.8 per cent of group level variation. So, as for language, differences in net achievement on mathematics can be predicted less precisely than differences in net achievement on mathematics.

After taking individual and group level covariates into account, teachers' attitude is included in the prediction model. The results indicate that attitude has no significant effect on net achievement on mathematics.

When adding 'minimum goals for mathematics' and 'class ability grouping' to the net achievement model for mathematics, again only one significant effect is found: the more teachers use class ability grouping, the lower the scores of the low SES pupils. The use of minimum goals for mathematics does not have a significant effect, but the fit of the model does improve when this approach variable is taken into account (along with 'class ability grouping'). The effect of this variable indicates that low SES pupils of teachers who use minimum goals have better mathematics scores than those of teachers who do not. Both approach effects are larger in the net achievement model than in the net achievement gain model. After taking the covariates and the two approach variables into account, the model shows that 31.9 per cent of individual level variation is accounted for and 51.9 per cent of group level variation.

DISCUSSION

From the results described above, it can be concluded that teacher attitude (cognitive versus non-cognitive) towards educating disadvantaged pupils has nothing to do with the actual achievement of Dutch low SES children. Neither for language nor mathematics scores of low SES pupils does it make a difference whether their teachers have a cognitive or a non-cognitive attitude. So the first research question of this article (does teacher attitude affect the achievements of low SES pupils) has to be answered in the negative.

Effects of two educational practice variables have been found on the achievements of Dutch low SES children. The use of class ability grouping has appeared to be related to achievement: effects were found for net achievement and net achievement gain both for language and mathematics. The use of minimum goals for mathematics has appeared to be related to net achievement for mathematics. Thus, the second research question (does a school's educational practices affect the achievements of low SES pupils) has to be answered positively.

Due to the lack of a relation between attitude and achievement, the third research question

(which practice variables mediate the relation between attitude and achievements) becomes irrelevant.

Although attitudes do seem to represent a 'real' empirical division between teachers (given the estimated reliability of the scale), and although attitudes do seem to be at the heart of discussions on educational inequality, there is no relation between attitude and achievements. This can be explained by pointing at the fact that, especially in the selected data set, the relation between attitude and practices is also very weak.

Regarding the relation between practices and achievements, one should realise that only two out of several dozen practice variables were used in this article. These were not selected on theoretical grounds but mainly because in preliminary analyses and in other research these appeared to be the most influential practice variables (see also Van der Werf in this special issue).

NOTE

1. Snijders & Bosker (1993) have proposed an alternative way of expressing 'modeled' variance in two-level models. Explaining variation at individual level could be computed from the reduction of the total variance, whereas explaining variation at group level could be computed from the reduction of the variance between the groups.

   For the individual scores the 'modeled variance' in a specific model is defined as the reduction in the total amount of variance compared to the empty model:

   $100\% * (1 - (VAR/Y_{ij}-X_{ij}\beta)/VAR(T_{ij}))$, where $VAR(Y_{ij}-X_{ij}\beta) = \hat{\delta}^2+\hat{\tau}^2$ for the specific model and $VAR(Yij) = \hat{\delta}^2 + \hat{\tau}^2$ for the empty model. (N.b.: $\hat{\delta}^2$ is the estimated amount of within-school variation and $\hat{\tau}^2$ is the estimated amount of between-school variation within the specific or empty model).

   For the group scores, the 'modeled variance' in a specific model is defined as the reduction in the amount of variance considered to be related to the group level, including the amount of variance on the individual level that is transfered from the group to the individual level within the multilevel analysis, and that is related to the uncertainty caused by the sampling procedure:

   $100\% * (1-(VAR(\overline{Y}_{.j}-\overline{X}_{.j}\beta)/VAR(\overline{Y}_{.j}))$, where $VAR(\overline{Y}_{.j}-\overline{X}_{.j}\beta) = (\hat{\delta}^2/n) + \hat{\tau}^2$ for the specific model and $VAR(Y_{.j}) = (\hat{\delta}^2/n) + \hat{\tau}^2$ for the empty model and n is the harmonic mean of the group sizes. For the analyses in this article, the harmonic mean of the group sizes is approximately 8.

This research was financed by the Commission for Program Evaluation. The data are from SVO-project 8075.

REFERENCES

APS (1986). *Actieplannen in voorrangsgebieden. Deel 1.* Amsterdam: APS.

Bergen, J.B.A.M. van (1989). *Verantwoording constructie toetsen voor de evaluatie van het onderwijsvoorrangsbeleid.* Arnhem: CITO.

Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch, 16. 4, pp. 206-218.*

Creemers, B.P.M. (1991). *Effectieve instructie. Een empirische bijdrage aan de verbetering van het onderwijs in de klas.* The Hague: SVO.

Doddema-Winsemius, H. & Werf, M.P.C. van der (1988). *Selectie/constructie van toetsen voor sociale redzaamheid en intelligentie ten behoeve van de evaluatie van het OVB.* Groningen: RION.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37, pp. 15-24.*

Hofstee, W.K.B., Koster, K.B. & Meijnen, G.W. (1984). *Programmavoorstel voor evaluatie-onderzoek naar de effecten van onderwijsvoorrangsbeleid.* The Hague: Ministry of Education and Science.

Levine, D.U. & Lezotte, L.W. (1990). *Unusually Effective Schools. A review and analysis of research and practice.* Madison: National Center for Effective Schools Research and Development.

Mathijssen, M.A.J.M. (1975). Sociale ongelijkheid van onderwijskansen; vijf interpretaties van sociale werkelijkheid. *Sociologische Gids, 6, pp. 427-445.*

Meijnen, G.W. (1979). *Maatschappelijke achtergronden van intellektuele ontwikkeling.* Groningen: Wolters-Noordhoff bv.

Meijnen, G.W. (1980). Schooltypen in het lager onderwijs en milieuspecifieke leerprestaties. *Mens & Maatschappij, 55, pp. 385-409.*

Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten.* The Hague: SVO.

Snijders, T.A.B. & Bosker, R.J. (1993): Modeled variance in two-level models. To appear in: *Social Research and Methods.*

Tesser, P., Mulder, L. & Werf, G. van der (1991). *De eerste fase van de longitudinale OVB-onderzoeken. Het leerlingenonderzoek.* Nijmegen/Groningen: ITS/RION.

Velden, L.F.J. van der (1991). De samenhang tussen de 'Denkrichting' van leerkrachten en de 'Vormgeving' van het onderwijs. In: Abram, I.B.H.; Creemers, B.P.M. & Leij, A. van der (Eds.) (1991). *OnderwijsResearchDagen 1991. Curriculum.* Amsterdam: SCO.

Werf, M.P.C. van der & Hoeben, W.Th.J.G. (1987). *Voorstel voor aanvullend onderzoek in het kader van de evaluatie onderwijsvoorrangsbeleid.* Groningen: RION.

Werf, M.P.C. van der, Weide, M.G. & Tesser, P. (1991). *Het onderwijsvoorrangsbeleid in de school en in de klas. De eerst meting in het basisonderwijs.* Groningen/Nijmegen: RION/ITS.

Manuscript ontvangen 17-9-1992
Definitieve versie ontvangen 5-4-1993

APPENDIX

Description of groep level variables; the first six are individual level variables aggregated to the groep level. There are 74 valid cases on all variables.

	Mean	Std. dev.	Minimum	Maximum
Mean normalized Z-score on language in 1990 (Dutch low SES pupils in grade 8)	-.345	.392	-1.154	.601
Mean normalized Z-score on mathematics in 1990 (Dutch low SES pupils in grade 8)	-.206	.411	-1.398	.737
Mean normalized Z-score on aptitude in 1988 (Dutch low SES pupils in grade 6)	-.220	.379	-.837	.638
Mean normalized Z-score on language in 1988 (Dutch low SES pupils in grade 6)	-.258	.411	-1.278	.602
Mean normalized Z-score on mathematics in 1988 (Dutch low SES pupils in grade 6)	-.168	.380	-.987	.551
Mean gender (1: 100% Boys; 2: 100% Girls) (Dutch low SES pupils in grade 6 & 8)	1.51	.16	1.00	1.91
Test date 1988 (grade 6) (number of days since Aug. 1, 1988)	92.2	21.3	56	142
Test date 1990 (grade 8) (number of days since Aug. 1, 1990)	84.1	24.1	42	133
Percentage Dutch low SES pupils in 1988 (% Dutch disadvantaged pupils at school level)	59.5	11.5	36	86
Percentage ethnic minorities in 1988 (% foreign disadvantaged pupils at school level)	15.4	11.5	2	47
Percentage non-disadvantaged pupils in 1988 (% non-disadvantaged pupils at school level)	24.2	10.5	4	47
Cognitive attitude, teacher grade 6 (low: strongly non-cognitive high: strongly cognitive)	48.6	8.9	32	70
Minimum goals for mathematics, teacher grade 6 (1. No minimum goals (11 teachers) 2. Individual minimum goals (20 teachers) 3. Common minimum goals (43 teachers))	2.4	.7	1	3
Class ability grouping, teacher grade 6 (low: no ability grouping high: much use of ability grouping)	17.2	4.9	7	32

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 2, pp. 99-110

Pygmalion and the effectiveness of 'black' schools; teachers' stereotypes and hidden goal differentiations towards ethnic minorities

Paul Jungbluth
*Institute for Applied Social Sciences (ITS), University of Nijmegen**

ABSTRACT

The aim of this article is to identify teacher-related determinants on the educational achievement of ethnic minorities. It focusses especially on teachers' perception of pupil's abilities and on covert goal differentiation towards individual pupils. Both kinds of teacher perspectives are found to contribute to ethnic differences in achievement. The ethnic minority status of pupils appears to be relevant mainly because of its accompanying social-class identity. A teacher's ethnic stereotyping apparently originates from, or coincides with, social-class bias.
Social bias in teacher perspectives also contributes to explaining school effectiveness. At individual level it explains the social-class related part of 'ethnic' differences in achievement. At school level it also explains some of the differences in achievement of ethnic minorities. It is concluded that research on education inequality should pay more attention to social bias in teacher perspectives as should Educational Priority Policies.

INTRODUCTION AND RESEARCH QUESTIONS

The research described consists of a series of four studies.[1] Central to these studies is the transition of ethnic minority pupils from primary to secondary school. It is common knowledge that most ethnic minorities in the Netherlands stay far behind the average educational level attained by the indigenous population. Turkish and Moroccan pupils, in particular, more often attend lower levels of secondary schooling, than their Dutch counterparts (cf. Tesser, Mulder & Van der Werf, 1991; Driessen, 1991). When looking for an explanation for this the external characteristics of ethnic minority groups are emphasised: different language, different culture, the duration of their stay in the Netherlands, sometimes their socioeconomic background and the educational support at home (cf. Van Langen & Jungbluth, 1991; 1992).

This article focusses on one possibly school-related determinant of educational effectiveness for ethnic minorities: the Pygmalion-phenomenon, which relates to the idea that teacher expectations about individual pupils influence their actual achievement (Rosenthal & Jacobson, 1968; Dusek, 1985). This concept of teacher expectancies working as self-fulfilling prophecies is combined with ideas on the educational reproduction of social inequality (Zinnecker, 1973; Hausser, 1980; Jungbluth, 1984). On the one hand, there is the idea that teacher expectations of pupils are affected by certain gender and class stereotypes. On the other, there is the idea of different teaching approaches, formal or informal, affecting different groups of pupils, in different schools or even in one school or class setting. The research questions posed are a follow-up to findings of former surveys on both sex- and class-related educational inequality.

One survey on the sex bias of secondary school teachers not only revealed sex biased teacher stereotypes but also a form of gender related definitions of school success in line with the traditional forms of separate education for both girls and boys. The idea of inequal opportunity was apparently blocked out as a result of sex stereotypes concerning both pupils' abilities and

* Institute for Applied Social Sciences, University of Nijmegen, P.O. Box 9048, 6500 KJ Nijmegen.

the respective forms of educational success for boys and girls (Zinnecker, 1973; Jungbluth, 1982).

Several surveys on class-related teacher stereotypes and, in particular, class-related forms of goal setting, even made it possible in a longitudinal design to explain in part class inequality in educational achievement in primary schools (Jungbluth, 1984; Van der Hoeven et al, 1990). Again, both in terms of selective perception of aptitude and the different levels of attainment defined as adequate for pupils from certain social classes, the reproduction of inequality is not thought of by teachers as illegitimate or even problematic.

The theoretical framework *in which social inequality in achievement is explained by different goal levels set by teachers as a result of social bias in expectations* has partly been checked before for both gender and socioeconomic status (SES) (Jungbluth, 1984; Van der Hoeven – van Doornum, Voeten & Jungbluth, 1990). Now, additional evidence is sought for ethnic minority groups. If found, it could lead to modifying the Dutch Educational Priority Policy, which up till now has only focused on extra educational programs, denying that the educational system reproduced social inequality by open and formal streaming based on social class. Do these variable goals still exist – possibly hidden behind individual differentiation and now gaining an ethnic connotation?

Three major questions will be answered:

- *are teachers' perceptions of pupils' abilities ethnically biased and, if so, does this affect the implicit goal levels teachers set for individual pupils?*
- *to what extent can ethnic bias explain ethnic inequality in the educational achievement of pupils?*
- *what is the relation between teachers' ethnic- or class-biased perceptions and educational effectiveness at school level (mean achievement, corrected for pupil's IQ)?*

METHOD

Sample

The data were collected during spring 1990 at 44 primary schools originally selected because they offered programs in Turkish and Arabic on the timetable (for more details on the collection of data see Driessen, 1991). At the time the data were collected, all 44 schools had at least four immi-grants in grade eight (age eleven/twelve). Data were collected at different levels. Grade eight pupils sat language, arithmetic and intelligence tests, while their teachers completed a questionnaire on their teaching methods and attitudes, and answered questions about each individual pupil. Parents also received a short questionnaire, which included questions on their own educational and occupational level.

The database contained information on 807 students, of which 434 were Dutch, 160 Turkish, 97 Moroccan, 45 Surinamese, 5 Antillean, 3 Mollucan and 63 others (varying from southern and northern European to Africans and Asians). The percentage of immigrant pupils on the total school population of the 44 schools ranged from 8 to 97 per cent, with an average of 48 per cent, which clearly shows that this is not a national representative sample survey. Not only are immigrants over-represented in 'black' schools, as they are called today, but so are Dutch working-class children among the remaining pupils (Verhaak & Jungbluth, 1990).

Variables

Three categories of variables were used in the analysis:

A. Variables to indicate the educational achievement and aptitude of individual pupils (measured by standardized tests). Both educational achievement and non-verbal reasoning have been indicated over standardized tests, which were shortened for this occasion. Reliability as measured by Cronbach's alpha is still above 0.90, both in indigenous and in ethnic minority groups; arithmetic and language items were summarized to a general achievement score (per cent items correct).

B. Variables to indicate the socioeconomic background (SES) of all pupils (measured by a questionnaire for parents). To indicate SES, both occupation and (if available) educational level of both parents was used, combined through HOMALS-procedures (cf. Van Langen & Jungbluth, 1992).
C. Variables to indicate both the abilities that teachers believe individual pupils to have and the hidden differences in educational goals that teachers find adequate for each individual pupil (cf. Hausser, 1980), and indicated by Likert scales, which were used when interviewing teachers about individual pupils. These scales were indicated for the following items:
perceived abilities ('COGPAS') (Cronbach's alpha: 0.87)
needs a lot of help
is good capable of thinking well in the abstract
sees easily beyond things
level of goals set ('COGACT') (alpha: 0.73)
I'm very moderate in what I demand from this pupil
I just offer subjects necessary for this child's future

All three categories of variables are defined at the individual pupil level, including the information from teachers.

Results

Educational achievement ethnic minority groups

To start with, the original relation between achievement and SES and the ethnic factor are presented, ignoring teacher variables. In Figure 1 educational achievement is presented after correction for non-verbal reasoning. Turkish and Morrocan pupils in particular lag far behind in terms of achievement in all categories of non-verbal reasoning. Apparently there is an additional ethnic factor or a factor related to an ethnic one, which could account for these differences. In the ANOVA analysis presented in Table 1, SES is used as a concurrent factor with the 'ethnic' background. Different from the presentation in the graph, in the ANOVA analysis IQ is used in its full metric range.

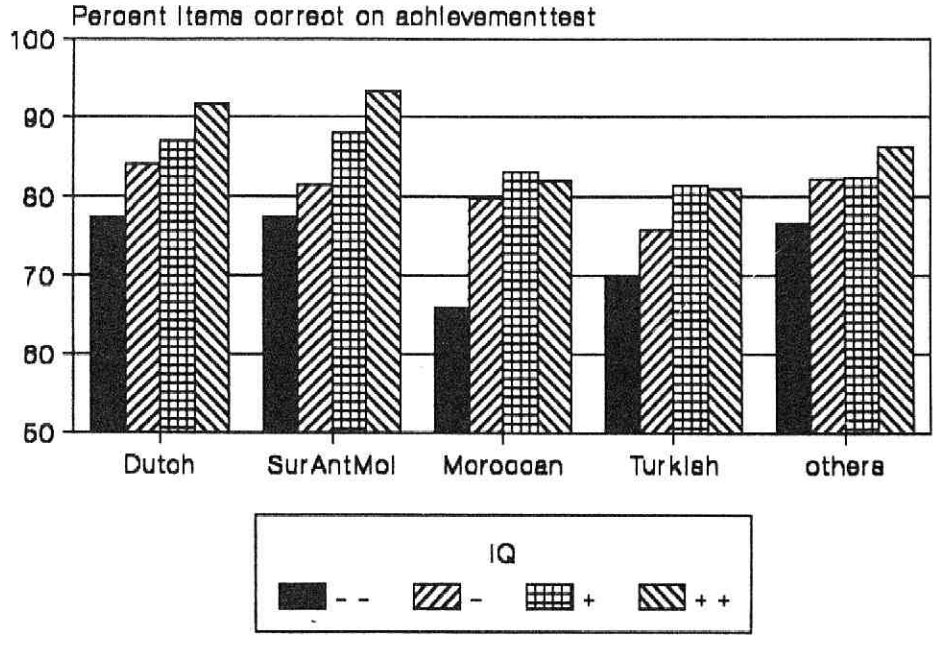


Fig. 1. Educational achievement of different ethnic minority groups, corrected for non-verbal reasoning (quartiles; all bars representing at least 10 pupils) (The 'SurAntMol' category consists of children from the former Dutch colonies (Surinam, Antilles, and the Molucas).

Table 1. Analysis of variance: relations between SES and ethnic background with achievement, controlled for non-verbal reasoning.

Source of Variation	Sum of Squares	DF	Mean Square	F	P
Covariates					
NON-VERB. IQ	32604.673	1	32604.673	271.417	.000
Main Effects	10381.771	10	1038.177	8.642	.000
IMMIGRANT	4079.246	4	1019.812	8.489	.000
SES	2193.873	6	365.646	3.044	.006
2-way Interactions	3027.738	22	137.624	1.146	.291
IMMIGR*SES	3027.738	22	137.624	1.146	.291
Explained	46014.183	33	1394.369	11.607	.000
Residual	91177.011	759	120.128		
Total	137191.193	792	173.221		

The analysis presented shows that after correction for non-verbal IQ (which accounts for 23.8 per cent of variance in achievement), the 'ethnic' factor accounts for 3 per cent, the SES factor for 1.6 per cent and there is an additional overlap of 3 per cent of variance. Without SES the 'ethnic' factor explains 6.0 per cent, so about half of the immigrant-related differences in what pupils with equal aptitude learn could also be accounted for by socioeconomic differences. (Figure 4).

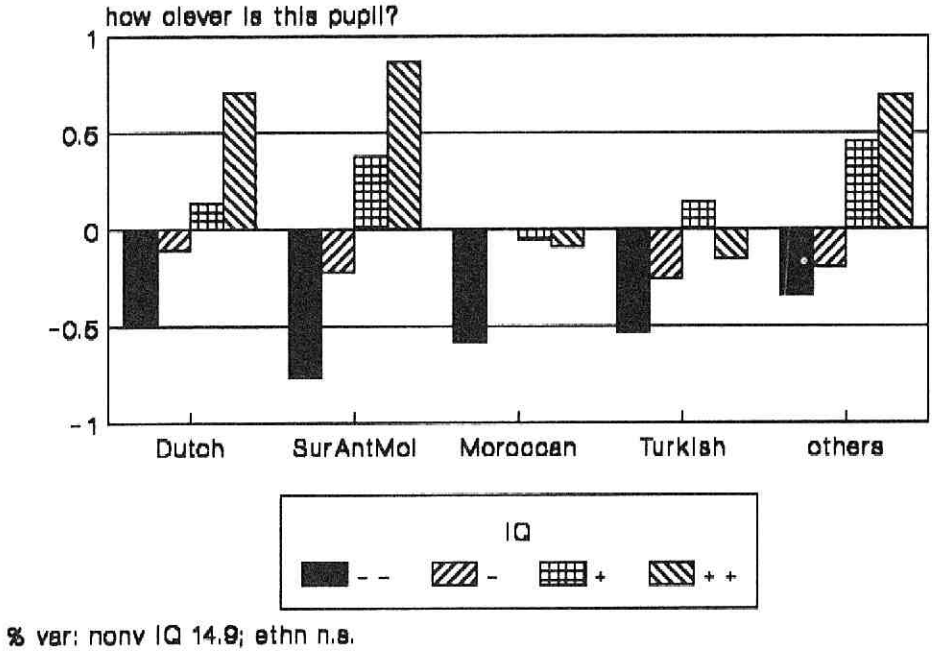
There are significant and relevant differences in achievement between pupils from different ethnic minority groups. About half of these could just as easily be interpreted as differences between pupils from different socioeconomic classes. Given the typical history of ethnic minorities, mainly in the Netherlands due to labour migration (cf. Li, 1979; Driessen 1991), they all have a specific class-composition (with the lower SES groups being overrepresented). This may partly account for those differences in achievement that at first seem to be 'ethnically' determined.

Teachers' stereotyping and covert differentiation of targets for immigrants

The first teacher variable presented here is the perception of pupil abilities from different ethnic minority groups. The second one is the educational level (covertly) set for pupils from different immigrant groups. Both variables are standardized with a mean of 0 and a standard deviation of 1. Both are shown first in graph (Figure 2), then analyzed by ANOVA with corrections for non-verbal reasoning and SES (Tables 2A and 2B). As is clear in both parts of Figure 2, compared to Dutch children, Turkish and Moroccan pupils in particular are confronted with relatively low expectations and goal levels, which show hardly any relation to their level of non-verbal reasoning. To check the apparent interaction between the ethnic factor and non-verbal reasoning by means of ANOVA, non-verbal reasoning is recoded into quintals ('IQ5'); in the graph it is presented as a quartile to assure that each bar represents 10 pupils at least.

The first analysis (Table 2A) shows that at least 11.0 per cent (up to 16.2 per cent) of perceived pupil abilities can be explained through differences in non-verbal reasoning. The 'ethnic' factor seems to be irrelevant and SES accounts for 3.7 per cent (with overlap: 8.9 per cent) of variance. There is a slightly significant interaction of SES*IQ. Since the ethnic factor in concurrence with SES looses its significance, we conclude that the tendency for teachers not to recognize the abilities of certain ethnic minority groups, mainly originates from SES biased perceived abilities (cf. Baron, Tom & Cooper, 1985).

In the analysis concerning the goals teachers set (Table 2B), pupils non-verbal reasoning



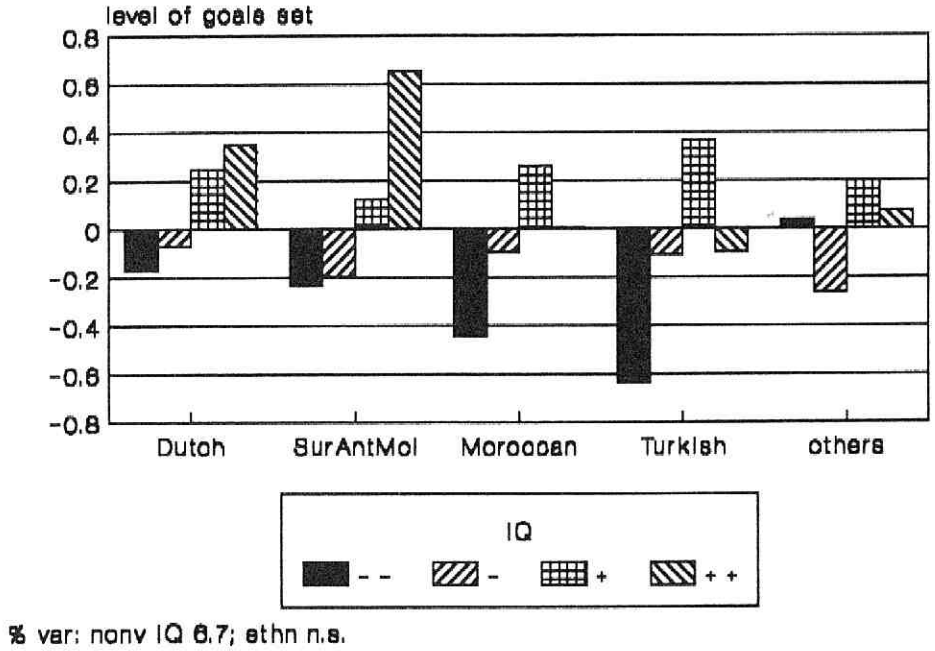


Fig. 2. Teacher's perception of a pupil's ability (left) and level of goals set for individual pupils (right) averages for different ethnic minority groups, controlling for non-verbal reasoning (using quartiles).

appears to be less important (5.0 per cent without overlap, 8.3 per cent overlap included). SES and non-verbal reasoning (both as main effects and to some degree in interaction) remove the significance of the ethnic factor. Again we conclude that the seemingly 'ethnic' bias can be considered as a social-class bias: high capacity pupils from certain ethnic minority groups are not presented with demanding target levels, mainly because of their low socioeconomic status. One does not need the ethnic factor to explain how teachers behave towards ethnic minority pupils; non-verbal reasoning and SES together can account for the apparent ethnic bias.

Table 2A. Analysis of variance: relationship of SES, ethnic background and non-verbal reasoning with perceived abilities.

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Main Effects	154.472	11	14.043	17.696	.000
SES5	29.150	3	9.717	12.245	.000
IMMIGRANT	1.843	4	.461	.581	.677
IQ5	84.573	4	21.143	26.644	.000
2-way Interactions	31.887	40	.797	1.005	.465
SES5*IMMIGRANT	3.753	12	.313	.394	.966
SES5*IQ5	15.668	12	1.306	1.645	.075
IMMIGRANT*IQ5	9.341	16	.584	.736	.758
3-way Interactions					
SES5*IMMIGRANT*IQ5	30.388	34	.894	1.126	.287
Explained	216.748	85	2.550	3.213	.000
Residual	557.071	702	.794		
Total	773.819	787	.983		

Table 2B. Analysis of variance: relations of SES, ethnic background and non-verbal reasoning with level of goals set.

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Main Effects	67.529	11	6.139	6.729	.000
SES5	10.715	3	3.572	3.915	.009
IMMIGRANT	1.158	4	.290	.317	.866
IQ5	38.989	4	9.747	10.685	.000
2-way Interactions	46.520	40	1.163	1.275	.123
SES5*IMMIGRANT	11.939	12	.995	1.091	.365
SES5*IQ5	18.987	12	1.582	1.734	.056
IMMIGRANT*IQ5	15.011	16	.938	1.028	.424
3-way Interactions					
SES5*IMMIGRANT*IQ5	23.360	34	.687	.753	.845
Explained	137.409	85	1.617	1.772	.000
Residual	640.402	702	.912		
Total	777.811	787	.988		

Pygmalion and the educational achievement of immigrants

The next research question concerns how both indicated aspects of teacher behaviour can explain pupil achievement. To test this, again two kinds of presentation follow. First, in a combination of graphs (Figure 3), achievement is linked to both perceived abilities and teachers' hidden goal differentiation. In both graphs it is clear that the achievement of both ethnic minorities and Dutch pupils are related to both kinds of teacher perspective: high expectations and high goal levels go together with relatively high achievement. As we have seen earlier in both respects, however, ethnic minority groups are seldom confronted with such conditions. In an ANOVA analysis (Table 3) the explanatory power of both new variables is tested (used as

additional covariates besides non-verbal reasoning) in concurrency with the factors SES and IMMIGRANT. The results should be compared to those in Table 1.

The total amount of variance accounted for by the three covariates is 47 per cent, which means that both teacher behaviours have a relatively high explanatory power. The indicated stereotypes have a dominant role, whereas the level of targets set has hardly any additional value (most of the variance explained appears to be overlap). After correction for these covariates, the social class of the pupils appears to be of no further relevance: the social-class related differences in achievement can be fully accounted for by the teacher variables used. The rele-

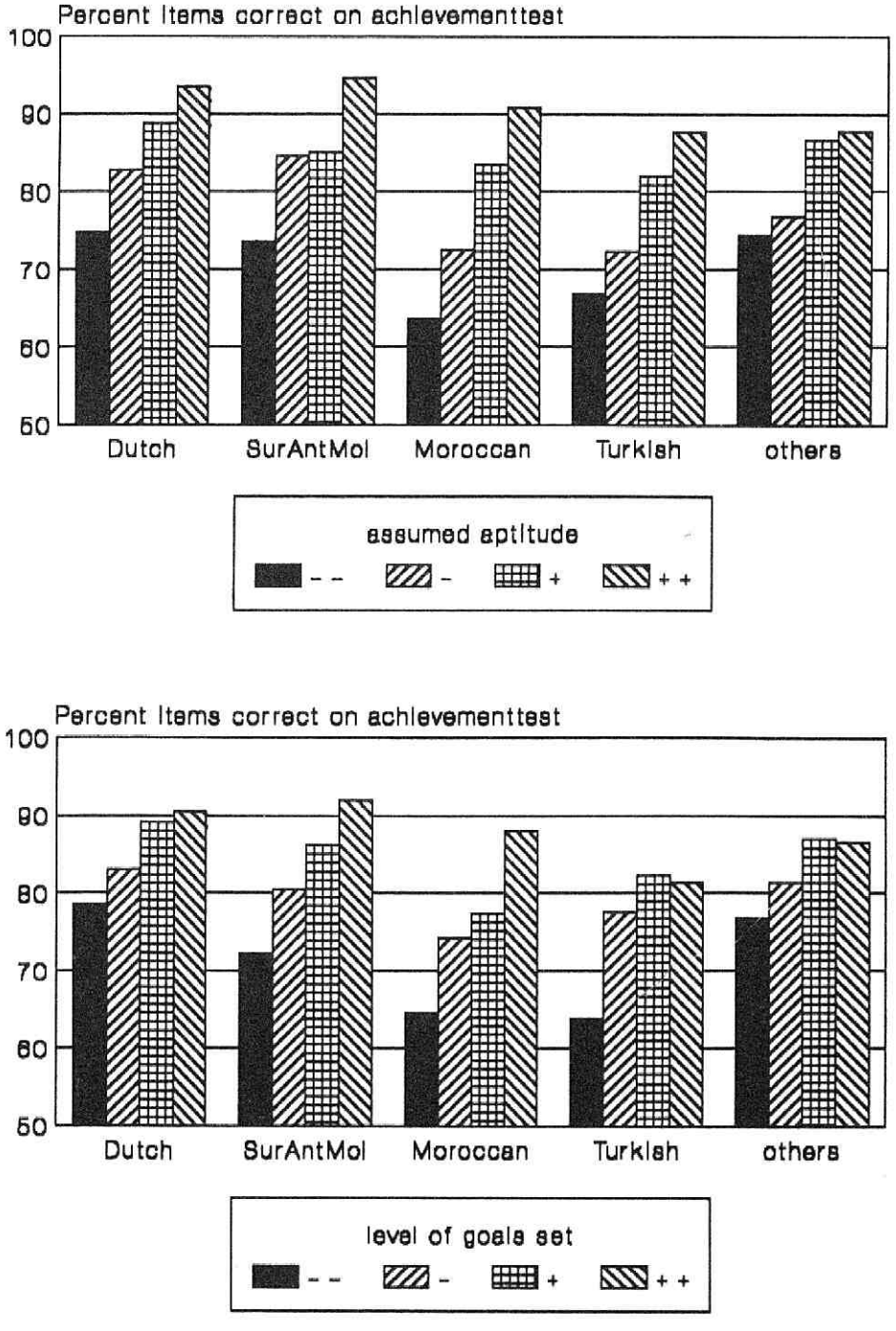


Fig. 3. Educational achievement of immigrant pupils as related to teachers' perception of pupils' abilities (left) and variational level of goals set (right).

Table 3. Analysis of variance: the relationship of SES and ethnic background with achievement, controlled for non-verbal reasoning, perceived capacities and level of goals set.

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Covariates	63695.480	3	21231.827	250.329	.000
NONVERBAL IQ	9863.235	1	9863.235	116.290	.000
COGPAS	16746.564	1	16746.564	197.447	.000
COGACT	1111.444	1	1111.444	13.104	.000
Main Effects	5656.169	7	808.024	9.527	.000
IMMIGRANT	3949.874	4	987.468	11.643	.000
SES5	270.501	3	90.167	1.063	.364
2-way Interactions					
IMMIGRANT*SES5	1235.959	12	102.997	1.214	.268
Explained	70587.608	22	3208.528	37.829	.000
Residual	64883.902	765	84.816		
Total	135471.510	787	172.137		

vance of the 'ethnic' factor appears to remain more or less equal to its original value: it is now 2.9 per cent at least, 4.2 per cent at most. Differences in achievement as related to the ethnic background of the pupils, which cannot be redefined as originating from social class, cannot be accounted for either by both teacher variables indicated.

These findings are confirmed in LISREL analyses, with and without both teacher perspectives (see Figure 4). The model without teacher variables again shows that most ethnic differences in achievement can be explained through either IQ or SES, or both (cf. Table 1). Adding both teacher characteristics does not affect the direct path between ethnic background and achievement; the direct path between SES and achievement disappears completely, whereas the relationship between IQ and achievement is strongly reduced. Apart from that it is the teacher perspective on pupils' abilities in particular that adds up essentially to explaining achievement.

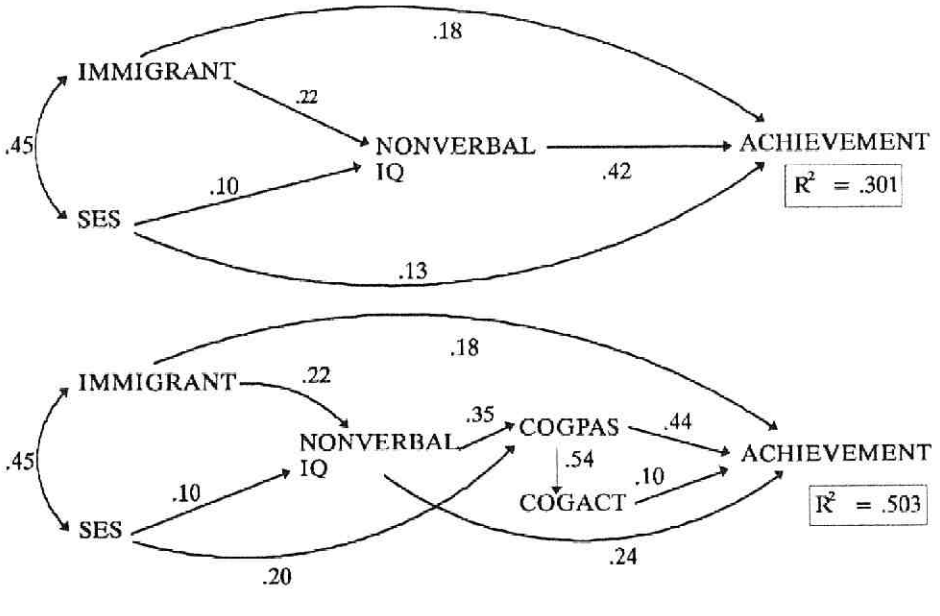


Fig. 4. LISREL analyses: achievement in relation to immigrant background, SES and IQ, before and after the introduction of teacher variables (complete models; effects >.05 only).

Pygmalion and school effectiveness

Finally, the last research question: to what extent do the two teacher variables account for the differences in effectiveness (mean achievement, corrected for IQ) between schools? In other words: does the explanatory power of both teachers' stereotyping and teachers' goal differentiation overlap the variance that appears to be linked to school level? To answer that question, the data were analyzed by multilevel-analysis.

In Figure 5 the apparent implications of school effectiveness for ethnic minority groups are indicated (with schools grouped into quartiles, according to their effectiveness). The effectiveness score of each school is computed as the mean achievement of the pupils, after correcting for non-verbal reasoning at individual level. From the graph it appears that all ethnic minorities more or less benefit from attending effective schools. Minorities at the most effective 25 per cent of schools even perform better than Dutch pupils at the least effective 25 per cent. Of course the question arises whether the teacher variables can somehow account for this variation in effectiveness between schools. In terms of VARCL, is it possible to explain the variance in achievement, located at school level by introducing the teacher variables stepwise? The results are presented in Table 4.

As was demonstrated in Table 3, both teacher variables seem to be extremely relevant in explaining educational achievement (in addition to social background and non-verbal reasoning) at individual level. Nevertheless, in Table 4 it becomes clear that this does not apply to differences at school level. To avoid technical complications, the analyses have been split up for separate groups: Turkish and Moroccon pupils on the one hand, indigenous Dutch on the other. For both groups, even after correction for non-verbal reasoning at individual level, school seems to matter. For ethnic minorities the amount of school variance is roughly what one would expect, about 11 percent (cf. Hofman, 1992); for indigenous Dutch pupils the variation at school level is relatively high, about 20 per cent. For both groups the school differences appear to depend largely on the average IQ of classmates. After that (which in itself probably overlaps the average teacher characteristics, cf. tables 2a and 2b), only for ethnic minorities can a significant extra effect of these average teacher characteristics be traced: these pupils perform significantly better at schools where teachers tend to have a positive image of this pupil group. The same

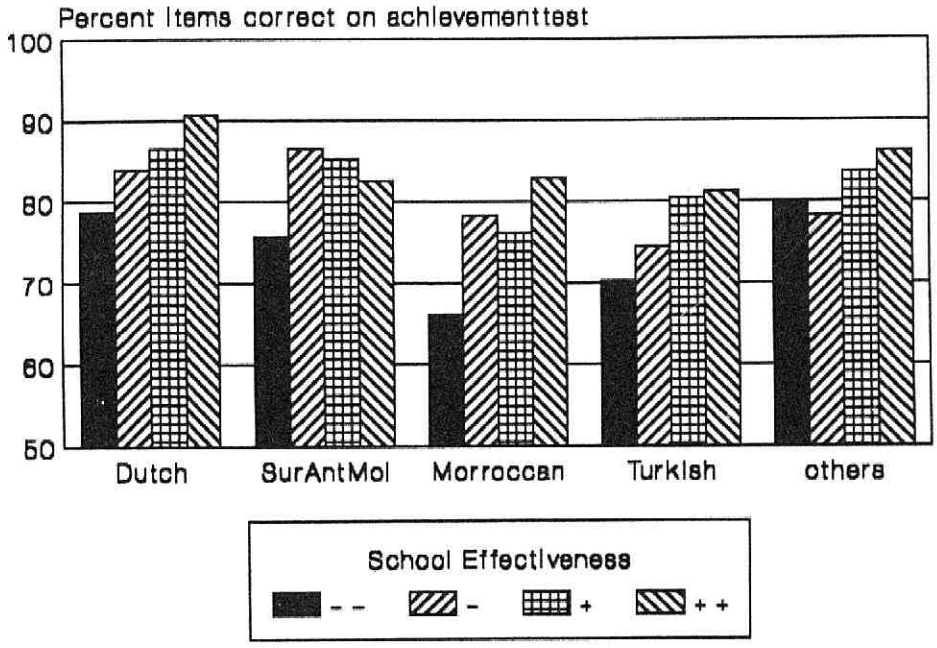


Fig. 5. Educational achievement of immigrant pupils as attained at less or more effective schools (quartiles).

Table 4. VARCL-analysis: amount of variance in achievement at school level after introduction of teacher variables.

	Variance at ind. pupil level	Remaining variance at school level	Total amount of variance	Regr. coeff.
Turkish and Moroccan pupils only (N=235)				
Zero model	200.8 (86.9%)	30.2 (13.1%)	231.0	75.4/75.2/75.4 (c)
with NONV.IQind	165.7 (71.7%)	27.5 (11.9%)		1.2/1.1/1.1
with NONV.IQmean		11.8 (5.1%)		2.3/2.1
with COGPASmean		6.3 (2.7%)		8.2
Neither COGACTmean nor any of the interaction terms introduced contributed significantly to explaining remaining variance at school level.				
Indigenous Dutch pupils only (N=416)				
Zero model	93.3 (81.5%)	21.2 (18.5%)	114.5	85.7/85.7/85.9 (c)
with NONV.IQind	79.0 (69.2%)	22.9 (20.0%)		1.1/1.1/1.1
with NONV.IQmean		9.2 (8.0%)		2.2/2.2
Neither COGPASmean nor COGACTmean nor any of the interaction terms introduced contributed significantly to explaining remaining variance at school level.				

effect cannot be found for indigenous Dutch pupils; no additional teacher effect is found at school level over and above those already described at individual level.

CONCLUSIONS AND DISCUSSION

The differences in educational achievement (controlled for non-verbal reasoning) between different ethnic groups reveal the enormous disadvantages of some second generation ethnic minorities in the Netherlands. This is even more evident when we realize that the Dutch group they were compared to in our survey was by no means representative of all Dutch pupils. Our selection of primary schools with a high number of ethnic minorities revealed an oversampling of working class Dutch children. White children attending so-called 'black' schools are mainly working class, just as their parents were when these schools were not yet black but just working class schools.

The idea behind our research questions stems from this historical view: should we look at (unintended) minority reproduction in schools as just a modern form of (equally unintended) class reproduction? To investigate that notion, we have concentrated here on just one possible explanation: the so-called Pygmalion concept. Using the same instruments in an earlier longitudinal survey in an extended form, we had previously proved that social-class related differences in achievement were to be regarded as relating to socially biased teacher perceptions and varying targets, which were secretly set for each individual. Now we tried to test whether the same applies to what is often mistaken for 'ethnical determination'. Consequently it was also tested whether the relative ineffectiveness of 'black' schools could be seen as originating from such teacher behaviour in general.

It was not possible in the present study to use a longitudinal design to measure teacher variables at an earlier stage than pupils' achievement; nevertheless teacher perceptions and the level of goals teachers set for their pupils are regarded here as determining factors of pupil achievement, not as reactions to actual achievement. Given the results from our former survey and not having the intention to find anything more than clues, this seems legitimate.

The results from the analysis reveal that the expectations are only partially confirmed. For a start, 'ethnic' differences in achievement can indeed be seen as an outcome of social class composition of certain ethnic minority groups in the Netherlands. Yet social class is not a complete determinant of these differences. We then documented other factors beside ethnic background that explained both teacher perceptions and variational targets. In as far as teachers have low expectations of the abilities of ethnic minorities or apply low goal levels, from our research it appears that these pupils suffer more from social class bias in these teacher perspectives than ethnic bias. Consequently, given our assumptions as to what constitutes a determinant and what is the result, it was found that it could once again be confirmed that both teacher perspectives go a long way in explaining educational achievement. Nevertheless, they could only account for the social-class related achievement differences in immigrant pupils. There was no additional link with the remaining 'ethnic' related variance in achievement.
Looking for an explanation for this remaining 'ethnic' variance, we would suggest that apart from the language problems already mentioned, a number of stress factors and family problems that appear to accompany immigration, and the low-paid work it entails, should be included in future research aimed at elucidating ethnic minority reproduction. In doing so, one might prevent official Dutch remedies such as 'teaching in own language' or 'intensive Dutch language courses' becoming overloaded with unrealistic pretentions.

Finally, the processes we have tried to indicate are not only relevant at individual pupil level but also at group level. First, it appears that there are significant differences between schools in terms of achievement after correction for individual pupil IQ. These can partly explained by the average IQ of pupils. For ethnic minorities the teacher perspectives in our research appear to be relevant also at school level, where the general impression teachers have of these pupils also helps explain their achievement.

As a consequence of the research design, these findings cannot be used as definite proof for the hypothesis tested. They are merely an additional indication of the validity of the concept described. Thus, they should be tested in research using longitudinal designs. Once confirmed, the Dutch Educational Priority Policy can no longer solely concentrate on the introduction or prompting of new programs or methods. There are certain covert built-in strategies in primary schools that contribute to class inequality, and should thus be combated as an essential part of that policy.

NOTE

1. The research project was subsidized by the Institute for Educational Research in The Hague (SVO projectno. 8095).

REFERENCES

Baron, Reuben M., David Y.H. Tom & Harris M. Cooper (1985). Social Class, Race and Teacher Expectations. *Teacher Expectancies*, pp. 251-269. New Jersey: Lawrence Erlbaum Ass.

Driessen, Geert (1991). *De onderwijspositie van allochtone leerlingen.* Nijmegen: ITS. Educational opportunities of migrant children.

Dusek, Jerom B. (ed.) (1985). *Teacher Expectancies*. New Jersey: Lawrence Erlbaum Ass.

Hausser, Karl (1980), *Die Einteilung von Schülern; Theorie und Praxis schulischer Differenzierung.(Categorising pupils; theory and practice of educational differentiation.)* Weinheim und Basel: Beltz.

Hoeven-van Doornum, A. van der, M. Voeten & P.Jungbluth (1990). Het effect van leerkrachtverwachtingen en streefniveaus op schoolloopbanen in het basisonderwijs.(Effects of teacher expectancies and target levels on school achievement in elementary schools.) *Tijdschrift voor Onderwijsresearch*, no. 15, 23-41.

Hofman, W.H. Adriaan (1992). *School level factors and educational performance of minority students.* Rotterdam: RISBO.

Jungbluth, Paul (1982). *Docenten over onderwijs aan meisjes. (Teachers' opinions on girls' education).* Nijmegen: ITS.

Jungbluth, Paul (1984). *Verborgen Differentiatie. Leerlingbeeld en onderwijsaanbod op de basisschool.* (Hidden Differentiation; teachers' perspectives on pupils curriculum.) Nijmegen: ITS.

Langen, Annemarie van & Paul Jungbluth (1991). *The effectiveness range of schools with a high proportion of migrant pupils.* Nijmegen: ITS.

Langen, Annemarie van & Paul Jungbluth (1992). Leerkrachtoverwegingen bij de vervolg-advisering van migranten. (Teachers' considerations behind advised level of secondary education for minority pupils.) *Tijdschrift voor Onderwijswetenschappen,* 22, 69-79.

Li, Peter S. (1979). An historical approach to ethnic stratification: the case of the Chinese in Canada. *Canadian Review of Sociology and Anthropology,* 16, 320-332.

Moelands, A. (1988). *Entreetoets. Basisvaardigheden taal, rekenen en informatieverwerking. Verantwoording. (Entrance test. Basic skills for language, math and information processing. Clarification.)* Arnhem: CITO.

Rosenthal, R. & L. Jacobson (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* New York: Holt, Rinehart & Winston.

Tesser, Paul, Lia Mulder & Greetje van der Werf (1991). *De eerste fase van de longitudinale OVB-onderzoeken. Het leerlingenonderzoek. (First phase results of the National Evaluation Survey on Dutch Educational Priority Policy: the pupils.)* Nijmegen/Groningen: ITS/RION.

Verhaak, Chr. & Jungbluth, P. (1990). Extra leerkrachten voor arbeiderskinderen en migranten; de formatieregeling basisonderwijs als beleidsinstrument voor onderwijsvoorrang. (Additional teachers for working class children and ethnic minorities, distribution of teachers over elementary schools as a policy instrument in education priority policies.) Klaassen, C. & Jungbluth, P. (eds), *OnderwijsResearchDagen 1990 Onderwijs en Samenleving* , 337-350. Nijmegen: ITS.

Zinnecker, Jürgen (1973). *Sozialgeschichte der Mädchenbildung. (Social history of girls' education.)* Weinheim und Basel: Beltz.

Manuscript ontvangen 17-9-1992
Definitieve versie ontvangen 19-4-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 2, pp. 111-119

Secondary school recommendations in relation to student and school characteristics

Lia Mulder
*Institute for Applied Social Sciences (ITS), University of Nijmegen**

ABSTRACT

In recent years various researchers have come to the conclusion that secondary school recommendations for immigrant primary school pupils are relatively high. Data from the national evaluation of the Dutch Educational Priority Policy (EPP) was the basis for research into whether recommendations are related to pupils' ethnic background or to the average level of achievement of the school which pupils attend. Multilevel analyses were carried out to separate individual effects from school effects. It was shown that factors from both levels contribute to recommendation differences. Pupils from low achieving schools are given relatively higher recommendations than those from other schools. Moreover, schools tend to make higher recommendations for immigrant pupils than for Dutch pupils with similar achievements and a comparable socioeconomic background.

INTRODUCTION

Immigrant pupils are generally advised to choose lower secondary school types, though the recommendations they receive are relatively high compared to recommendations for Dutch pupils. Their average language and arithmetic achievements, as well as their intelligence-scores, are lower than those of Dutch pupils given the same school advice (De Jong, 1987; Driessen, 1990; Mulder, 1991). No single explanation for this phenomenon has as yet been identified. Some researchers believe that it is a matter of positive discrimination; the rationale being that, because of their ethnic background, immigrant pupils are favoured above Dutch children in the sense that, for them, threshold requirements are lower, and the benefit-of-the-doubt principle is more readily applied than to Dutch pupils (Driessen 1991, Kerkhoff 1988). According to De Jong this is because teachers make allowances for language deficiency and take into account the abilities immigrant pupils may have once a higher level of language skills has been acquired. In anticipating this, they therefore recommend a higher secondary school level than their actual primary school achievements would indicate.

This explanation would imply that it is *within* a school that immigrant pupils are advised to attend higher secondary school types than Dutch children with similar achievements. It can also be supposed, however, that the explanation is to be found not within, but *between* schools. Saris and Blok (1982) have demonstrated that schools employ different recommendation strategies, depending on the school's social composition. Immigrant pupils usually attend schools with a high percentage of disadvantaged children, thus lowering the average achievement level at those schools. A pupil doing well at such a school might perform similarly to one attending a school with a high average achievement level. Nevertheless, both schools will apply a ranking order; more capable pupils being recommended for higher secondary school types than less able ones. The result is that able pupils at schools showing a low average level of achievement will be given higher recommendations than mediocre ones attending schools with higher average levels of achievement, even though language and arithmetic performance may be the same in both cases.

The aim of this article is to find out to what extent differences between pupils being recommended for different types of secondary education are related to ethnic background, and to what

* Institute for Applied Social Sciences (ITS), University of Nijmegen, Toernooiveld 5, 6500 KJ Nijmegen.

extent these differences are related to the average achievement level of schools. The research questions are:

1. Are there any differences between schools when recommendating for a specific secondary school type?
2. To what extent are these differences related to pupils' educational achievements?
3. To what extent are these differences related to schools' average achievements?
4. To what extent are differences in recommendations related to pupils' ethnic background or/ and schools' average achievement level? In case of finding only a positive effect concerning pupils' ethnic background, the 'positive discrimination' hypothesis is supported. In case of finding only a negative effect concerning schools' average achievement the hypothesis of different recommendation strategies – depending on the average achievements of the pupil population – is supported. When finding both effects the relative contribution of both variables should be analyzed to discover which hypothesis is supported the most. Besides, it might be that there is a differential effect concerning pupils' ethnic background on recommendations, indicating that some schools adjust for the ethnic origin of pupils, while others do not. It is also hypothised that this differential effect is related to average school achievement level, in the sense that schools with a low average achievement level adjust more for the ethnic background of their pupils than schools with a high average achievement level. Thus, there are two additional research questions:
5. Are there any differences between schools concerning the effect of ethnic background on recommendations?
6. Are these differences related to the average school achievement level?

METHOD

Variables and data

The data to be analyzed were derived from the national evaluation of the Dutch Educational Priority Policy (EPP). For an extensive description of the design and method employed in this evaluation, we refer to Tesser, Mulder & Van der Werf (1991).

The data set contains information about 12,382 grade 8 (last primary school year) pupils, collected during the 1988/89 school year in some 700 schools. The following variables were analyzed:

A. Pupils' ethnic origin.

Priority was given to the father's country of origin. If the father was not present, or his country of origin unknown, the mother's country of origin was taken, which was also the case if the father's country was the Netherlands, and the mother's a different country. The countries of origin were classified into four categories: Turkey/Morocco; Surinam/the Antilles/the Moluccas; the Netherlands; other countries. From the 12,382 pupils 767 were Moroccan, 730 Turkish, 628 Surinamese, 83 Antillean, 45 Moluccan and 1,135 others (including Europeans, Africans and Asians). For 347 pupils the country of origin was unknown.

B. Educational achievement.

Pupils' scores on language skills, arithmetic and intelligencetests were taken as indicators of their educational achievements. All tests were especially developed for the EPP-evaluation (Bergen, 1988; Doddema-Winsemius & Van der Werf, 1988). Scores were converted into T-scores: standardized normal scores with a mean of 50 and a standard deviation of 10.

C. School's achievement level.

The mean language skills, arithmetic and intelligence scores of a pupil's class were taken as indicators for the school's average achievement level.

D. Recommendations.
Recommendations for various levels of secondary education were rated as follows, from low to high:[1]
1 = IBO; 2 = IBO/LBO; 3 = LBO; 4 = LBO/MAVO; 5 = MAVO;
6 = MAVO/HAVO; 7 = HAVO; 8 = HAVO/VWO; 9 = VWO

Data analysis
In order to answer the research questions multilevel analysis was applied, using the multilevel VARCL programme. The first question was answered by specifying a so-called empty model to determine the relative contributions of schools and pupils to the differences in recommendations. To answer Question 2, pupils' achievements were added to the model and a school's average achievement level specified as a predictor in the model (research Question 3). Question 4 was answered by analyzing the relative contributions of ethnic background and a school's average achievement level to the differences in recommendations. Therefore ethnic background was added as a 'fixed' covariate to the model. Finally, to answer the two additional questions, pupils' ethnic background as a random covariate and the interaction between ethnic background and a school's average achievement level were specified as predictors in the model.

RESULTS

Differences between schools and within schools
In specifying the 'empty' model it appeared that recommendation differences within schools far exceed the differences between schools (see table 1). Just over one tenth of the variance of principals' advice is associated with differences between schools. The differences between the schools' average recommendation scores are small. For schools separated by two standard deviations (e.g. one standard deviation above the grand mean, and one standard deviation below the grand mean) the difference is 1.2 on the scale of the advice. This amounts to the difference between, for instance, an LBO recommendation and an LBO/MAVO recommendation.

Pupils' educational achievements and school differences
If schools differ in their average recommendations, this does not necessarily imply a difference in the way they recommend. The differences may relate to the fact that one school has better students than the other. Question 2 is answered by including individual pupil scores for language, arithmetic and intelligence tests in the multilevel model. Any remaining between-schools differences are shown in Table 1. By way of comparison, the 'empty' model (without predictors) has also been included.

When taking pupils' language achievement into account, between-school variance drops by 44 per cent ((.39-.22)/.39), implying that almost half of the between-schools average recommendation differences can be attributed to the fact that one school has better language perform-

Table 1. Between-school and within-school differences in recommendations after correcting for language, arithmetic and intelligence. Variance and percentages of the total variance.

	between schools		within schools		total
	var.	per cent	var.	per cent	var.
empty model	.39	11.2	3.1	88.8	3.49
with language	.22	6.3	2.0	57.3	2.22
and arithmetic	.29	8.3	1.5	43.0	1.79
and IQ	.29	8.3	1.5	43.0	1.79

ers than another. Between-schools 'differences' might be expected to become even smaller when pupils' arithmetic and intelligence achievements are taken into account. This, however, is not the case. In fact, between-schools differences are increased. Thus, it is in comparable arithmetic achievement rather than comparable language achievement that schools' recommendations differ.

Intelligence does not contribute at all to reducing school differences, presumably because it is too closely connected to language and arithmetic performance.

After correcting for language, arithmetic and intelligence, the total variance has been reduced by almost half ((3.49-1.79)/3.49): between-schools variance now only constitutes 43 per cent of the between-schools variance in the empty-model. Between-schools variance, initially 11 per cent of the total variance, has decreased by roughly a quarter to 8 per cent of the total.

These results indicate, as might be expected, that between-schools average recommendation differences are largely associated with differences in individual pupil achievement. However, even after controlling pupil performance, a considerable amount of between-schools variance still remains unaccounted for. This could mean that schools differ as to the recommendation strategies they use. Pupils with the same level of achievement might receive different recommendations depending on which school they attend, as suggested by Saris and Blok (1982).

Schools' average achievement level and school differences

To answer Question 3, the schools' average achievement level was added to the model in order to explain differences in recommendations between schools that remained after controlling for individual level co-variates. As indicator for the schools' average achievement level, every grade 8 pupil's class mean achievement score was computed.

The analyses do not show the effects of average language and intelligence achievement level to be significant, therefore these variables were removed from the prediction model. The only variable remaining is the mean arithmetic achievement level. What this variable contributes to between-schools average recommendation differences can be found in Table 2.

The arithmetic achievement mean accounts for 17 per cent ((.29-.24)/.29) of the difference between schools that remained after individual pupil achievements had been taken into account. Between-schools variance now falls to less than 7 per cent. The total variance remaining after controlling for schools' average achievement level has decreased by 2.8 per cent.

The effects of pupils' ethnic background and schools' average arithmetic achievements

To establish the effect of ethnic background on recommendation, the variable 'parents' country of origin' was split into four categories: Turkey/Morocco, the Antilles/Surinam/the Moluccas, the Netherlands and "other" countries. The four categories were transformed into three dummy-variables, with the Netherlands acting as reference category. The effects of the three categories,

Table 2. Between-school and within-school differences in recommendations after controlling for language, arithmetic and intelligence and mean arithmetic achievements. Variance and percentages of the total variance.

	between schools		within schools		total
	var.	per cent	var.	per cent	var.
empty-model	.39	11.2	3.1	88.8	3.49
with language	.22	6.3	2.0	57.3	2.22
and arithmetic	.29	8.3	1.5	43.0	1.79
and IQ	.29	8.3	1.5	43.0	1.79
+ mean grade level arithmetic	.24	6.9	1.5	43.0	1.74

Table 3. Unstandardized regression coefficients of school and pupil characteristics on recommendation for secondary education.

	recommendation	
	b	st. error
language	.06	.002
arithmetic	.08	.002
IQ	.01	.002
EPP group	-.72	.038
country:other	.25	.052
Antilles/Surinam/Moluccas	.30	.066
Morocco/Turkey	.23	.050
mean arithmetic grade level	-.05	.005

Turkey/Morocco, the Antilles/Surinam/the Moluccas and "other" countries, are represented in the analysis as the difference from the reference category. If the effect of the country-of-origin category is positive, after controlling for individual achievement, this means that pupils from these countries receive higher recommendations in comparison with Dutch students.

Finding a positive country-of-origin effect on recommendation, however, immediately raises the question whether this is due to a pupil's ethnic background or social background because most immigrant children have a lower socioeconomic background then Dutch pupils. Thus, an additional pupil characteristic indicating socioeconomic background was included in the analysis. The indicator used was whether pupils' belonged to one of the Educational Priority Policy target groups.[2] Dutch and immigrant pupils within such groups are distinguished by the fact that they have at least one parent with a low occupational and/or educational level. The variable "EPP target group" is also a dummy-variable. Pupils belonging to an EPP target group (scored as 1) were compared with pupils not indicated as such (scored as 0).

Firstly, the 'fixed' effects of individual pupil' achievement, socioeconomic background, ethnic background and a school's average arithmetic achievement level were computed. Table 3 presents the regression coefficients for these variables.

When comparing the effects of language, arithmetic and intelligence on recommendation, arithmetic is the most decisive. The influence of language achievement is slightly less and intelligence hardly plays any independent role. EPP pupils receive recommendations which are on average .72 lower than recommendations given to non-target-group pupils with comparable achievements.

The answer to the question to what extent differences in recommendations are related to ethnic background, taking socioeconomic background and achievements into account is to be found in the country-of-origin effects. These effects are positive for all three categories and vary from .23 to .30. This means that immigrant pupils receive a recommendation that is approximately a quarter recommendation point higher than comparable Dutch pupils.

The last regression coefficient indicates the effect on recommendation of the schools' average arithmetic level. The effect is negative, meaning that the lower a school's mean arithmetic level, the higher the recommendation; or the higher a school's mean arithmetic level, the lower the recommendation.

To form an idea of the consequences of the negative effect of socioeconomic background, the positive effect of ethnic background and the negative effect of schools' average arithmetic level on recommendations for immigrant and Dutch pupils with comparable achievements and socio-

Table 4. Mean schools' arithmetic achievement for Dutch and immigrant pupils who do/do not belong to EPP target groups.

Netherlands, non-EPP	50.0
Netherlands, EPP	47.6
Immigrants, non-EPP	49.7
Immigrants, EPP	44.6

economic background, the separate effects of both variables were included in one single equation. Firstly, pupils were categorised into four groups:

1. Dutch pupils not belonging to an EPP target group.
2. Immigrant pupils not belonging to an EPP target group.
3. Dutch pupils belonging to an EPP target group.
4. Immigrant pupils belonging to an EPP target group.

To equate individual pupil achievement, imaginary language, arithmetic and intelligence scores of 50 were assumed. To estimate schools' average arithmetic levels for each of the four groups, a mean arithmetic achievement level of the individual schools (grade 8) these pupils attended was computed. These mean arithmetic achievement scores are to be found in Table 4.

The recommendation for Dutch pupils not belonging to an EPP target group is predicted to be:

RECOMMENDATION = .06*50 + .08*50 + .01*50 – **.05*50.0** = 5.00

A Moroccan pupil not belonging to an EPP target group has a predicted recommendation level of:

RECOMMENDATION = .06*50 + .08*50 + .01*50 – **.05*49.7**... + **.23** = 5.25

A Dutch pupil not belonging to an EPP target group is predicted to have the following recommendation:

RECOMMENDATION = .06*50 + .08*50 + .01*50 – **.05*47.6 – .72** = 4.40

The recommendation for a Moroccan student not belonging to an EPP target group is predicted to be:

RECOMMENDATION = .06*50 + .08*50 + .01*50 – **.05*44.6 – .72 + .23** = 4.78

The first three elements of the equation are the same for all pupils regardless of country or their belonging to an EPP target group. There are, however, differences to be found in the last part of the equation. The fourth element reflects the effects of the schools' mean arithmetic achievement level, the fifth shows the effect of belonging to an EPP target group while the sixth shows the effect of ethnicity.

The first thing one notices is that a pupil's socioeconomic background affects recommendation much more strongly than ethnicity and a school's mean arithmetic achievement level. Recommendations made to disadvantaged students are almost three-quarters of a recommendation point (-.72) lower than for other, identically achieving, students. The effects of ethnicity and mean achievement level is much lower. A comparison of the latter two variables shows that the contribution of ethnicity amounts to about a quarter recommendation point (.23), and that the contribution of a school's mean achievement level does not exceed .27 (the highest mean achievement level minus the lowest mean achievement level = .05*50 – .05*44.6). In other words, the two factors contribute more or less equally to the recommendations of students.

Table 5. Slope variance, standard deviation and covariance between intercept and slope of the variables 'EPP group' and 'country-of-origin', after correction for language, arithmetic and intelligence.

	variance	sd	covariance
EPP group	.11	.33	-.008*
country: other	.10	.32	.025*
Antilles/Surinam/Moluccas	.11	.33	.039*
Morocco/Turkey	.002*	–	.019*

* = not significant (p > .01)

Difference between schools in the effects of students' characteristics on recommendations
To answer the last two research questions, the individual student level variables socio-economic background (whether students belong to an EPP target group) and ethnic origin were included as random co-variates in the multilevel model. The slope variances are presented in Table 5.

Schools differ significantly in the extent to which a pupils' socioeconomic/ethnic background affects recommendations; for disadvantaged pupils it makes a difference which schools they attend. Given equal achievement at one school they receive higher recommendations than at another. The difference is at most a .66 recommendation point. The same applies to pupils where a parent or parents is/are from Surinam, the Antilles, the Moluccas or another country. They, too, will receive a recommendation at a school that is up to a .64 or a .66 recommendation point higher than at another. For Turkish or Moroccan pupils, however, it does not make any difference which school they attend.

In order to investigate whether the between-schools differences in the effect of socio-economic and ethnic background on recommendation is related to a school's average arithmetic achievement level, the interaction between both these variables and a school's mean arithmetic achievement level was computed. None of the interaction effects proved to be significant.

CONCLUSIONS

The aim of this article was to find out whether the reason for comparatively high recommendations for a particular secondary school type made to immigrant pupils on leaving primary school is to be found within or between schools. Multilevel analyses were applied to the data set of the national evaluation of the Educational Priority Policy. The results show that slightly over 10 per cent of the recommendation differences can be attributed to the fact that at one school higher or lower recommendations are made than at another. These differences are partly attributable to differences in language and arithmetic achievements of the individual pupils attending these schools.

However, even after controlling for these pupil characteristics, differences in average recommendation between schools remained unexplained. In order to investigate whether these remaining differences can be accounted for by a school's average achievement level, a school's mean language, arithmetic and intelligence achievements were included in the prediction model. The mean arithmetic achievement level proved to reduce the between-schools variance significantly.

The effect of a school's average arithmetic achievement level is negative. The lower the average arithmetic level, the higher the recommendation, and the higher the average arithmetic level the lower the recommendation. This finding supports the hypothesis that schools differ in their recommendation strategies, depending on the mean achievement level of their pupils, thus immigrant students, who mostly attend schools with a low average achievement level, receive relatively higher recommendations than Dutch pupils with equal achievements. This does not,

however, preclude the possibility that within schools ethnic minorities receive higher recommendations than Dutch nationals. A comparison of the effects of the average arithmetic level and the effect of a pupil's ethnicity on recommendations, show both effects to be approximately the same. Thus the reason for the comparatively high recommendations given to ethnic minorities seems to lie in both a between-schools and in a within-schools context. Schools with a low achievement level give recommendations that are higher in comparison than those given by other schools and schools generally give higher recommendations to immigrant pupils than to Dutch ones with identical achievement levels and a comparable socioeconomic background. Additional results of multilevel analyses showed, however, that the positive effect of a pupil's ethnic background was not the same for all schools. Apparently some schools adjust more for a pupil's ethnic background than others. This differential effect of ethnic background on recommendations was not related to a school's average achievement level. What recommendation strategies play their part to explain this phenomenon might be an interesting research goal for the near future.

NOTES

1. IBO: individualized lower secondary vocational education
   LBO: lower secondary vocational education
   MAVO: intermediate general secondary education
   HAVO: higher general secondary education
   VWO: pre-university education

2. Socioeconomic background is defined in terms of parents' occupation and educational level. In the VARCL-analyses, we cannot use these variables because too many data are missing, and VARCL is unable to handle these. If only includes those respondents in the analysis which show a value for all variables. Including occupation and educational level in the analysis would result in a drastic exclusion of respondents and hence a biased representation.

REFERENCES

Bergen, J.B.A.M. van (1989). *Verantwoording constructie toetsen voor de evaluatie van het onderwijsvoorrangsbeleid. (Account for the test design for the evaluation of the educational priorities policy)* Arnhem: CITO.

Doddeman-Winsemius, H. & M.P.C. van der Werf (1988). *Het meten van intelligentie en sociale redzaamheid in de evaluatie van het onderwijsvoorrangsbeleid. (Measuring the intelligence and social competence in the evaluation of the educational priorities policy.)* Groningen: RION.

Driessen, G.W.J.M. (1990). *De onderwijspositie van allochtone leerlingen. De rol van sociaal-economische en etnisch-culturele factoren, met speciale aandacht voor Onderwijs in Eigen Taal en Cultuur. (Socioeconomic and ethnocultural factors in the educational position of non-indigenous students.)* Nijmegen: ITS.

Driessen, G.W.J.M. (1991). Discrepanties tussen toetsresultaten en doorstroomniveau. Positieve discriminatie bij de overgang basisonderwijs-voortgezet onderwijs? (Discrepancies between secondary school entrance test results and school type.) *Pedagogische Studiën, 68, 1*, 27-35.

Jong, M.-J. de (1987). *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs. (Background, capacities and opportunities and the transfer from primary to secondary school.)* Lisse: Swets & Zeitlinger.

Kerkhoff, A. (1988). *Taalvaardigheid en schoolsucces. De relatie tussen taalvaardigheid Nederlands en schoolsucces van allochtone en autochtone leerlingen aan het eind van de basisschool. (Language skills and school succes – indigenous versus non-indigenous students.)* Amsterdam/Lisse: Swets & Zeitlinger.

Mulder, L. (1991). *De overgang van basis- naar voortgezet onderwijs van de OVB-doelgroepen. Advies, schoolkeuze en rapportcijfers in het eerste leerjaar. (OVB target groups – recommendations, choice of school level and achievement in first form)* Nijmegen: ITS.

Saris, W.E. & H. Blok (1982). Het onderwijzersadvies nader bekeken. (A screening of headmasters' recommendations.) *Tijdschrift voor Onderwijs Research, 7*, 49-59.

Tesser, P., L. Mulder & M.P.C. van der Werf, (1991). *De eerste fase van de longitudinale OVB-onderzoeken. Het leerlingenonderzoek. (First stage results of the National Evaluation Survey on the Dutch Educational Priority Policy; Student variables.)* Nijmegen/Groningen: ITS/RION.

Manuscript ontvangen 17-9-1992
Definitieve versie ontvangen 5-4-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 2, pp. 120-129

Cooperation between Educational Priority Policy areas and basic adult education

Gerry J. Reezigt
*RION, Institute for Educational Research**

ABSTRACT

Basic adult education (BE) and the Dutch Educational Priority Policy (EPP) share a target group: low-educated parents of children attending primary or secondary schools. Secondary analyses of the sample of the evaluation of EPP show that there is a large group of potential BE students, especially immigrant parents. Formally educating these parents might promote their children's school achievement. Cooperation between BE and EPP is often mentioned as a method to reach this goal. However, most EPP areas are not currently involved in any form of cooperation. Case studies in six areas where there is cooperation show that most joint activities comprise courses in basic skills, attended by parents who would not easily apply for regular BE courses.

INTRODUCTION AND RESEARCH QUESTIONS

Official documents often state that cooperation between the Dutch national Educational Priority Policy and basic adult education is necessary. There are several both pragmatic as well as more content-related arguments for this.

The pragmatic argument is that the EPP areas and BE institutions share a target group. Parents of children attending EPP schools have often had little formal education themselves, and are therefore potential BE students. As both EPP areas and BE institutions offer educational activities for these parents, cooperation, or at least some coordination of activities, is indicated.

There are also two content related arguments for cooperation. A major goal of the Dutch educational priority policy is to improve the educational opportunities of disadvantaged children. Schooling of parents may be one way to do this. It is well-known that the higher the educational level of parents, the higher the achievement scores of their children (Van der Velden, 1991). Raising the educational standard of low-educated or illiterate parents might be directly relevant to the school career of their children. For this reason, schools that want to offer parental activities that are beneficial to their children's progress, might consider educating the parents.

A final argument for cooperation between EPP and BE is the prevention of illiteracy. As yet, Dutch primary schools have failed in providing all students with basic skills in reading and writing (Wesdorp et. al, 1986). Although compulsary education also involves secondary schooling, this does not automatically imply that students leaving primary school without basic skills will catch up in secondary education. On the contrary, they sometimes fall behind even further (National Steeringgroup, 1990) and later become potential BE students. To break this vicious circle, cooperation between EPP and BE could be useful: by sharing knowledge about the teaching of reading (Hammink, 1986) and by improving this in primary schools (Tubbing, 1986), the number of illiterate adults applying for BE courses could be reduced.

Traditionally, primary schools with high numbers of children from low income families offer various parental activities which often included courses for parents. Part of the current educational program BE institutions offer is directly evolved from such courses. Cooperation between EPP and BE was not an issue, as joint activities were already common. This situation changed

* RION, P.O. Box 1286, 9701 BG Groningen.

some years ago due to new EPP and BE legislation. Cooperation was still possible, but it took more effort and had to be planned more specifically by EPP and BE staff. For this reason, it might prove less attractive to cooperate, and EPP and BE might decide to deal with their own target groups independently. If this happened, coordinating activities could come to an end and the advantages of cooperation disappear. The larger the target group shared by EPP and BE, the greater the wasted energy by not cooperating would be.

Because information on the size of the target group was lacking and because it was unclear how often cooperation took place, a research project was started in 1989 to find answers to two questions:

1. What is the size of the adult target group shared by EPP and BE?
2. Is there any cooperation between EPP areas and BE institutions, and, if so, how do they cooperate?

The project was a part of the national evaluation of the Dutch Educational Priority Policy (Kloprogge, 1986; Reezigt, 1991).

HISTORY OF COOPERATION AND CURRENT SITUATION IN EDUCATIONAL PRIORITY POLICY AREAS AND BASIC ADULT EDUCATION

Between 1965 and 1985, several educational activities that are now part of standard BE courses arose from parent activities in primary schools. In contacts with parents, teachers with high numbers of pupils from low income families often noticed a lack of knowledge of the possibilities offered by the educational system. They also noticed that many parents were not sufficiently fluent in basic skills; some parents were even illiterate, although they had actually finished primary school (Tubbing, 1986). Teachers who became aware of the problems of these parents started to develop courses for them, often in cooperation with social workers. Many schools involved in earlier forms of the Educational Priority Policy offered refresher courses for parents who wanted to improve their reading, writing and mathematical skills, courses on pedagogical and educational issues and sometimes courses for illiterate parents. The need for these courses increased over the years when it became clear that many adults, even in an affluent society as the Netherlands, were in fact illiterate. Another reason was the influx of ethnic minorities with low levels of formal education, which also led to Dutch language courses especially aimed at non-Dutch parents.

Because of the large numbers of adults who showed an interest in adult education, separate organizational bodies gradually emerged. In 1987 the separation became official when new BE legislation came into force outlining the conditions and content of BE courses and separating the financing of parent activities in primary schools from the funding of standard adult education courses. Cooperation between EPP and BE, as it could still be found in certain areas, often fell apart.

In 1986, the educational policy for Dutch disadvantaged children was integrated with that of ethnic minorities and became the Educational Priority Policy. Part of this program included the targeting of priority areas where schools cooperated with welfare institutions. Such areas were provided with extra funding on the basis of drawing up a plan of activities, which they had to submit in 1986 and which would be valid until 1990. These plans had to specify how the extra money would be spent and the target groups were children attending primary or secondary school, and their parents.

Parents have nearly always played a part with regard to policies for disadvantaged children, the reason being that in order to improve educational opportunities for children, something should also be changed in the home environment. In the past, several attempts were undertaken

(Leseman, 1989), with an emphasis on parental reading behavior (Meijnen, 1979). All attempts originate from Bourdieu's theory on educational inequality, which stresses differences in cultural patterns as one of the causes for inequality (Van der Velden, 1981).

An analysis of all 70 plans (Vermeulen & Van der Werf, 1989) shows that all areas wanted to pay some attention to the parents of disadvantaged children. In fact, the relation between schools and parents was the topic most often mentioned. According to the plans, problems that should be dealt with were a cultural gap between school and family, a lack of concern for childrens' school progress, a lack of knowledge about educational opportunities (and by that, not making use of available opportunities), and the low education levels of parents. The EPP areas planned many activities involving parents. 20 per cent of all activities were aimed at parents. This is a rather high percentage; for comparison, only 9 per cent of all activities were aimed at immigrant children (one of the EPP target groups). The activities included promoting parental involvement in school (for instance by assisting teachers in reading lessons, informing parents on the education system (for instance by organizing meetings where secondary education is discussed) and improving the basic skills of parents (for instance by offering courses in reading, writing and mathematics and courses for illiterate adults).

These activities, apart from those aimed at parents' involvement in school, are also often offered by basic adult education institutions.

After official legislation was passed on basic adult education in 1987, several forms of adult education became integrated, such as courses for illiterate adults and "second chance" education for adults who had little formal education. BE offers courses in mathematics, language, social skills, English and computer technology, all designed specifically for adults with a minimum of formal schooling. More than 250 institutions have been established since 1987 (Ruijs, 1991) and some 80,000 adults are now involved in BE courses (Ruijs, 1991), while 10,000 students are waiting to enroll. The number of BE students may grow dramatically in the future according to Hammink (1987), who claims that 24,000 children each year leave primary school without sufficient reading and writing skills, many of whom may need BE courses later. About half of all BE students are immigrants, mostly immigrant men, while Dutch students are mostly women. Most finished primary school, 21 per cent did not or had no education at all. Most students are relatively young: 65 per cent under 40 (Doets & Huisman, 1988, Ruijs, 1991).

METHOD

To answer the first research question (the size of the adult target group shared by EPP and BE), secondary analyses were performed on data of the national EPP evaluation study (Tesser, Mulder & van der Werf, 1991).

To answer the second research question (checking the actual existence of cooperation between EPP areas and BE institutions and describing forms of cooperation), several case studies were conducted.

Secondary analyses

In 1988/1989 data were collected in some 700 Dutch primary schools, the sample for the national EPP evaluation. In this evaluation study, several cohorts of pupils were tested every two years. In 1988/1989 (the first time the sample was tested) a questionnaire was given to all parents of pupils in grades 4, 6 and 8 (age groups 8, 10 and 12), which asked questions on the education and occupation of both parents and their involvement in the school career of their children. Some 20 per cent of all schools refused to hand over the questionnaire to parents, mostly schools in the larger cities where most immigrants are living. This could explain the low response rate of immigrant parents (Tesser, et al., 1991): on average 36 per cent of immigrant parents (of grade 4, 6 and 8 children) responded, while for Dutch low SES parents the average response rate was 66 per cent and for Dutch middle/high SES parents the rate was 73 per cent. In

total 38,541 questionnaires were distributed and 22,521 (58 per cent) were returned. Most of the questionnaires gave information on both parents. Of these questionnaires, 2,775 (12 per cent) were returned by immigrant parents. Parents from the former Dutch colonies (Antilles, Moluccas, Surinam) responded far more often than parents from Morocco and Turkey. The selective response rates make it clear that the data from the parent questionnaire are not representative and can only be used to give a rough indication of the overlap between EPP and BE. As other datasets that might provide the information needed to estimate this overlap were not available, it was decided to perform a secondary analysis but to report all results with great caution.

Case studies

To locate EPP areas where some form of cooperation with BE institutions may be found, no use could be made of any existing dataset. The national EPP evaluation study did not include relevant data at EPP area level. Consulting researchers who had access to data collected in studies on BE institutions or BE students did not result in any useful information about cooperation with EPP areas either, simply because questions on this topic had never been asked.

At the time when the study described in this article started, another study was carried out to check the actual situation in EPP areas (Verhaak & Brandsma, 1991). For this study, coordinators of 29 EPP areas were interviewed. It was possible to add some questions on cooperation with BE institutions to the interview checklist, and the data from these 29 interviews formed the basis for the selection of cases.

Contacts with BE representatives were reported in 22 of these interviews. To find out whether these contacts implied any structual cooperation, the following questions were asked:

- are there any parental activities organised jointly by EPP and BE, or is there any form of mutual consultation on activities?
- does the coordinator think the cooperation is regular and of any importance?
- are there any links at organizational level (for instance, a BE representative is a member of an EPP council)?

There were four cases where EPP and BE cooperated. Inspection of brochures and articles in relevant EPP and BE journals resulted in another two cases. None of the six selected cases came from one of the larger Dutch cities, which is somewhat surprising as these cities comprise nearly half of all EPP areas. Therefore the search for cases in these cities continued. Contacts with EPP coordinators showed that cooperation did not exist, at least not in any structural form. This again was surprising, because former joint projects were almost always located in these cities, and because EPP schools with the highest numbers of potential BE parents (immigrants) are also most often found in these cities.

Reasons mentioned for the absence of cooperation were:

- practical problems in BE-institutions; many institutions were still involved in merging with other institutions;
- waiting lists for BE courses; most institutions cannot provide courses for all adults who want to register and are using waiting lists, which reduces the need for cooperation with other agencies;
- lack of money and personnel for extra activities;
- different ideas on adult education; BE institutions want to focus exclusively on own learning experiences of adults and sometimes fear an emphasis on the parental role in EPP activities.

In the six cases where cooperation actually took place, key persons (EPP area coordinators, BE personnel, sometimes local authorities) were interviewed. In total, 14 interviews were conducted on the basis of a checklist with open-ended questions.

Topics were derived from the analysis of EPP area plans (Vermeulen & Van der Werf, 1989), including some facts on the actual state of EPP and BE (number of schools involved,

courses offered by BE), the way cooperation was initiated, joint activities, and opinions on cooperation. Recorded interviews generally took about two hours.

RESULTS

Overlapping target group

About 10 per cent of parents who returned the questionnaire had only been educated to primary school level, while 40 per cent had had a few years of lower vocational schooling. Only those adults who left secondary school without a certificate were taken into consideration. According to this criterion (a formal one for eligibility for BE courses), 24 per cent of the mothers and 20 per cent of the fathers are potential BE students (Table 1).

There are large differences between parents of different countries. Percentages of parents who were not educated beyond primary school level are presented for several countries in Table 2. Because of missing values, the total n in Table 2 is slightly different from Table 1.

Table 2 shows that fathers generally have a somewhat higher level of education than mothers. The Dutch group has the lowest proportion of parents with primary education only. Most parents from the former Dutch colonies were also educated beyond primary school level; the

Table 1. Potential BE students.

	mothers		fathers	
	n	%	n	%
elementary school/ some years of secondary education, no certificate, potential BE students	4,405	24	3,334	20
higher level of education	13,800	76	12,999	80
	18,205	100	16,333	100

Table 2. Total number of mothers/fathers per country, and percentages with primary education only.

	mothers		fathers	
	n	% primary	n	% primary
Antilles	91	13	58	5
Moluccas	60	10	58	7
Surinam	351	24	232	25
Morocco 269	65	330	52	
Turkey	569	79	646	68
Netherlands	15,475	5	13,707	7
Other	919	17	884	15
	17,734		15,915	

Table 3. Characteristics of potential BE parents vs. other parents.

% of parents stating that they	mothers BE	mothers other	fathers BE	fathers other
- never read books	21	8	37	19
- never buy books	45	26	61	40
- never visit concerts, theaters, museums	80	58	84	63
- hardly ever discuss school events with children	6	2	13	7
- never attend parent-teacher meetings	16	5	33	20
- are unemployed	78	69	28	8

proportion that did not attend secondary school is about 9 per cent for parents from the Antilles and Moluccas and about 25 per cent for parents from Surinam. The situation for the Moroccans and Turks is very different: more than half of all parents did not attend secondary school. In interpreting these figures, it should be kept in mind that the actual educational level in some ethnic minority groups may be even more unfavorable than suggested by Table 2. For instance, many Moroccans did not finish primary school or even attend one at all. Because data on these parents are lacking, the table may show a somewhat flattering picture.

The parental level of education is a powerful indicator of their childrens' school success (Tesser et al., 1991). Concepts often used to explain the influence of the educational level of parents are cultural capital and involvement in childrens' school career (Meijnen, 1979; Leseman, 1989; Van der Velden, 1991). The parental questionnaire of the EPP evaluation study offers some information in this respect.

When potential BE parents are compared with parents who reached higher levels of education, the situation in families of potential BE parents is less favorable than in other families (Table 3).

Potential BE students indicate more often that they do not buy or read any books and that they do not visit concerts, theaters or museums. They are also less inclined to discuss school with their children or to attend parent-teacher meetings. They are also more often unemployed than parents who reached higher levels of education, which means children of potential BE students are growing up in households with characteristics that are unfavorable for their school careers. However, it is hard to say how important each of these characteristics really is. In general correlations of .25 with achievement scores are reported (Tesser et al., 1991). It is not certain that further schooling of parents would result in more favorable home conditions for children, although there are some indications that this might happen. Parents who enroll in BE courses often become more involved in their childrens' school careers; because of their own educational experiences, their self-confidence grows and they find it easier to contact teachers and to discuss their childrens' work (Hammink, 1987). They also show more interest in the homework assignments of their children.

Due to the selective response on the parent questionnaire, the data reported in this section are not representative for all parents. The numbers of potential BE parents will also vary from school to school. What became clear is that Moroccan and Turkish parents are the most prominent potential BE group. Women from these countries have generally a very low level of education and might benefit from BE courses. Men from these countries are more often involved in BE courses than women (Van der Horst, 1990; Ruijs, 1991).

Cooperation between EPP areas and BE institutions

Main results on the contact between EPP and BE in the case studies are summarized in Table 4.

When contacts exist for a long time, they are mostly informal: colleagues have known each other for many years and sometimes those who originally worked in EPP areas switch to BE and stay in touch with their former EPP colleagues. New contacts (since 1986/1987) are more formal and institutionalized. In both cases where BE representatives are members of EPP councils, the contacts are relatively new. Only once did a local authority actively stimulate cooperation, although in most cases it is responsible for both EPP and BE policies. Financing of joint activities is different in the six cases; sometimes an activity is financed by regular BE funding, while sometimes a lot of trouble is taken to finance the same activity because regular BE funding is considered inappropriate. Table 5 offers information on the content of joint activities.

Refresher courses are given to small groups (about 10-15 parents), often parents of one and the same school. Groups meet each week, and sometimes stay together for as long as two years. The aim of the courses is to offer parents knowledge of language and maths at primary level in order for instance to help their children with homework assignments. These courses are mostly taught traditionally: class teaching, seatwork, testing and homework. Most parents prefer this to grouping or cooperative learning. In most courses, educational and pedagogical topics are discussed (such as toys, television, health care, sleeping problems of children). In one case, a separate course on such issues had been developed.

Dutch courses for immigrants focus on reading and writing skills. Small groups (5-10 adults) meet each week. Sometimes parents are allowed to stay in a group until they can read and write at a basic level (which may take a long time, especially for immigrant parents who cannot read or write in their own language). Immigrant mothers often apply for these courses when they feel their children are becoming estranged from them. Because of large differences in reading and writing skills, individual instruction is sometimes needed.

Table 4. Facts on cooperation between EPP and BE, 6 cases.

		found in .... cases
- contacts between EPP and BE existed before 1986/1987		3
- contacts between EPP and BE exist since EPP area formation		3
- BE representatives in EPP council		2
- local authority involvement		1
- financing of joint activities by	EPP	1
	BE	3
	both EPP and BE	1
	local authority	1

Table 5. Content of joint activities, 6 cases.

	found in .... cases
- refresher courses for parents (language skills, maths)	5
- language acquisition/reading and writing for (illiterate) immigrants	3
- information on the educational system/pedagogical issues	1
- participation in school council	1
- BE personnel instructed by EPP team on the teaching of reading	1

School council courses were designed for primary and secondary councils. All members of these courses were supposed to attend these, and even some teachers participated. Subjects were education in general, the financing of education, educational administration, dealing with laws and regulations, and discussion techniques. In one case BE personnel were schooled by EPP team members to transfer knowledge on the teaching of reading (assessment, remedial training, developing teaching aids), available in the EPP team, to BE instructors.

The courses offered are the responsibility of BE personnel, and primary school teachers stimulate parents to attend them. The courses are not part of the regular BE curriculum but are especially designed for these groups of parents. BE staff also attend primary school parent-teacher meetings. Participants are nearly always mothers; fathers only took part in the school council courses. Immigrant parents mostly enroll in Dutch courses only.

According to those interviewed, the advantages of cooperation are:
- a group of adults (immigrant mothers) that would normally not attend any course at all can be reached ;
- BE personnel is better informed on how to teach adult learners than primary school staff; courses are professionally designed;
- courses will have more impact than irregular sessions with parents;
- courses can be offered at locations (sometimes the school) close to the homes of parents; this is especially important for immigrant mothers who do not easily leave their own neighborhood;
- there are no restrictions as in regular BE courses (conditions on group size, attendance control, possibilities for child care);
- courses may stimulate parents to enroll in regular BE courses.

The effects on parents, as reported by those interviewed in the six cases, mainly concern a greater degree of self-confidence, stronger involvement in school activities and homework assignments of children, the prevention of alienating parents from their children (in immigrant families) and a more positive attitude toward education.

In one case, cooperation between EPP an BE resulted in changes in the teaching of reading in primary school. Young adults, applying for BE courses in this particular area, having left primary school some years before, already showed a loss of reading skills. This led to a new criterion in primary schools: all children are supposed to read at least 65 words per minute when they leave school, as this is the speed necessary to understand TV subtitles (a reading activity necessary for all adults). Children who fail to reach this standard one year before they leave school receive extra tuition.

DISCUSSION

EPP areas and BE institutions share the same adult target groups. There are clear indications, that schools with high numbers of Turkish and Moroccan children also have simular numbers of potential BE parents, as most immigrant parents do not have a high level of formal education. Cooperation between these schools in EPP areas with BE is especially important for immigrant mothers, who do not easily apply for courses. BE institutions do not have a lot of extra facilities enabling them to contact immigrants who do not enlist for courses. Moreover, even if they had, it might be extremely difficult to contact immigrant women, who lead secluded lives and hardly ever take part in outgoing activities. Sometimes, walking their children to and from school are the only times they leave their homes. Formal meetings between parents and teachers are often attended by immigrant fathers; for this reason these occasions are inappropriate for contacting the mothers. The only way to contact these is via classroom teachers or via other (immigrant) parents. The latter strategy has proved to be succesful in home intervention programs such as HIPPY and the Dutch version, OPSTAP (Vedder & Eldering, 1992).

In developing courses for immigrant mothers, it is important to offer extra facilities (such as child care during lessons) and to have a rather lenient attitude towards absence during courses which due to family matters always taking precedence over courses can occur rather frequently. Regular BE courses often cannot deal with these specific problems because of strict regulations, so that setting up courses with EPP offers more flexibility.

Joint activities found in the six case studies are much the same as those undertaken by primary school teachers for parents in the years before 1986/1987. It seems reasonable for BE personnel to be responsible for adult education courses because of their greater experience of adult learning. In the past, primary school teachers often taught parents of the children attending their school. However, parents did not always like to be taught by their children's teacher and use the same textbooks and materials. Teachers were not always able to assess parents' education level and to provide appropriate instructions.

Although parental education seems to have a positive influence on childrens' school progress, no research data on the impact of BE are available to support this. Recent data on the effects of other parent-directed activities in EPP areas are also lacking. For a well-documented opinion on the relative effects of both activities more research is clearly needed. Such research might also shed more light on the cost-effectiveness of parent-directed activities in EPP-areas in general.

REFERENCES

Doets, C. & Huisman, T. (1988). *De basiseducatie in cijfers (Basic adult education in figures).* Amersfoort: SVE.

Hammink, K. (1986). *Jongeren in alfabetisering (Young illiterates).* Amersfoort: SVE.

Hammink, K. (1987). *Alfabetiseren, tien jaar vechten tegen ongelijkheid (Literacy education, ten years of fighting against inequality).* Amersfoort: SVE.

Horst, J. van der (1990). Analfabeten zitten lang niet alleen bij etnische groepen (Not only ethnic minority groups have illiterates). *Samenwijs, 10,* no. 5, January 1990, 193-194.

Kloprogge, J. (1986). *Uitvoeringsplan evaluatie onderwijsvoorrangsbeleid 1987-1990 (Plan for the implementation of the evaluation of the Educational Priority Policy 1987-1990).* The Hague: SVO.

Leseman, P.P.M. (1989). Gezinsmilieu en schoolprestaties (Families and School Achievement). J.M. Slavenburg, & T.A. Peters, (Eds.). *Het Project Onderwijs en Sociaal Milieu: een eindbalans. (The Education and Social Environment Project: final conclusions).* Rotterdam: SAD.

Meijnen, G.W. (1979). *Maatschappelijke achtergronden van intellectuele ontwikkeling (Societal backgrounds of intellectual development).* Groningen: Wolters Noordhoff.

Ministry of Education and Science (1986). *Rijksregeling basiseducatie* (National regulations for basic adult education). The Hague: Staatsdrukkerij.

Reezigt, G.J. (1991). *De relatie tussen onderwijsvoorrang en basiseducatie (The relation of Educational Priority Policy and basic adult education).* Groningen: RION.

Ruijs, G.P.C.M. (1991). Basiseducatie (Basic adult education). In J.K. Koppen, H. Stroomberg & M. van der Kamp (Ed.). *Onderwijsresearchdagen 1991 (Conference on educational research, 1991).* Amsterdam: SCO.

Tesser, P., Mulder, L. & Werf, M.P.C. van der (1991). *De eerste fase van de longitudinale OVB-onderzoeken: het leerlingenonderzoek (The first phase of the longitudinal EPP evaluation: pupil study).* Nijmegen/Groningen: ITS/RION.

Tubbing, M. (1986). De aanvoer is verzekerd; over de relatie basiseducatie en onderwijsvoorrangsbeleid (Enough students; on the relation between basic adult education and Educational Priority Policy). *School, October 1986,* no. 2, 47-49.

Tweede Kamer der Staten-Generaal (1985a). *Vaststelling van een wettelijk kader voor het voeren van een samenhangend onderwijsvoorrangsbeleid (onderwijsvoorrangswet): Memorie van toelichting (A framework for a coherent Educational Priority Policy).* Vergaderjaar 1984-1985, 18775, nos. 1-3.

Vedder, P. & Eldering, L. (1992). *The effects of OPSTAP, a home intervention programme preparing children for school.* Paper for the European Conference on Educational Research, Enschede.

Velden, R.K.W. van der (1991). *Sociale herkomst en schoolsucces (Social background and school success).* Groningen: RION.

Verhaak, C. & Brandsma, H.P. (1991). *Onderwijsvoorrangsgebieden op weg (Educational Priority Areas).* Nijmegen/Groningen: ITS/RION.

Vermeulen, H. & Werf, M.C.P. van der (1989). *De plannen van de onderwijsvoorrangsgebieden* (Plans of the Educational Priority Areas). Groningen/Nijmegen: RION/ITS.

Werf, M.P.C. van der, Weide, M.G. & Tesser, P. (1991). *Het onderwijsvoorrangsbeleid in de school en in de klas (Educational Priority Policy in schools and classrooms).* Groningen/Nijmegen, RION/ITS.

Wesdorp, H., Bergh, H. van der, Bos, D.J., Hoeksma, J.B., Oostdam, R.J., Scheerens, J. & Triesscheijn, B. (1986). *De haalbaarheid van periodiek peilingsonderzoek: een vooronderzoek op het gebied van het taalonderwijs in de lagere school (The feasability of national assessment: a preliminary study on language education in primary schools).* Lisse: Swets & Zeitlinger.

Manuscript ontvangen 17-9-1992
Definitieve versie ontvangen 30-3-1993



Schoolverzuim en schooluitval in het voortgezet onderwijs

Een literatuurstudie

G.C. de Vries

Naar schoolverzuim en schooluitval wordt veel onderzoek gedaan. Jongeren, ouders en vrienden, de school, de economische sector van de samenleving en de samenleving als geheel, hebben elk hun eigen kijk op het spijbelen en het vroegtijdig verlaten van de school. Dat deze verschijnselen als een probleem ervaren worden hangt niet zozeer samen met de hoeveelheid spijbelaars en uitvallers, maar eerder met andere factoren in de maatschappij, zoals werkloosheid en het belang dat gehecht wordt aan onderwijs en het hebben van diploma's.

De auteurs onderzochten systematisch de onderwijssociologische, onderwijspsychologische, onderwijskundige en andere relevante sociaalwetenschappelijke literatuur op dit gebied en inventariseerden de resultaten van onderzoek. De lezer krijgt op deze wijze inzicht in de aard en omvang van verzuim en uitval in het voortgezet onderwijs, de achtergronden, interpretaties en verklaringen die ervoor gegeven worden op het niveau van de samenleving, het onderwijsstelsel, de school, het gezin, het individu en de samenhang van deze factoren. De auteurs geven ook een overzicht van oplossingen en maatregelen die tot nu toe zijn voorgesteld.

Serie: Forum nr. 20

ISBN 90 265 1354 2
148 pagina's, f 34,50

Verkrijgbaar in de boekhandel

SWETS & ZEITLINGER B.V.

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

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 3, pp. 133-134

De ontwikkeling van expertise in verschillende domeinen: inleiding

Henk G. Schmidt
*Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Rijksuniversiteit Limburg**

Dit nummer bevat vier bijdragen die zich bezighouden met de ontwikkeling van expertise in complexe domeinen. Onderzoek naar hoe expertise zich ontwikkelt, is voor het onderwijs van belang om twee redenen. Studies op dit terrein houden zich veelal bezig met de resultaten van leren over een lange periode, waarin de bestudeerde personen aanzienlijke hoeveelheden kennis verwerven en hun vaardigheid om de taken van het betreffende domein met succes te volbrengen, aanzienlijk toeneemt. In de studies op het terrein van de geneeskunde die in dit nummer beschreven worden, werden in een aantal gevallen de prestaties van studenten in de geneeskunde uit verschillende studiejaren vergeleken met die van internisten. Dat betekent dat met behulp van expertiseonderzoek een beeld verkregen kan worden van de veranderingen in kennis en vaardigheden die twaalf jaar onderwijs (of meer) in personen teweegbrengen. Leren wordt in dergelijke studies vooral als een longitudinaal proces beschouwd. Dat onderscheidt dit onderzoeksdomein van de meer "klassieke" benaderingen in de psychologie, waarin het leren van personen over het tijdsbestek van enige minuten tot enkele uren object van studie is. Voor de instructietheorie, die zich in het verleden sterk richtte naar dit leerpsychologisch onderzoek, vormt de studie van expertise en zijn ontwikkeling een nuttig alternatief gezichtspunt.

Een tweede reden waarom onderzoek naar verschillen tussen experts, gevorderden en novieten voor de onderwijswetenschap van belang is, hangt met het voorgaande samen. Expertiseonderzoek richt zich zonder uitzondering op terreinen waarin verwerving en gebruik van complexe kennis een belangrijke rol spelen. In de bijdragen die in dit nummer aan de orde komen, zijn dat de natuurkunde en de geneeskunde. Dit zijn beide terreinen waarop het bereiken van enig niveau van handigheid vele jaren in beslag neemt door het complexe en uitgebreide kennisbestand dat die wetenschappen definieert. In het onderwijs is het inzicht groeiende dat bij het verwerven van zelfs relatief "simpele" vaardigheden zoals rekenen op de basisschool veel meer kennis te pas komt dan eerder wel verondersteld werd. Het expertiseonderzoek, en de daarbij ontwikkelde methode van protocolanalyse, biedt daarom een aantrekkelijk model voor andere takken van de instructiewetenschap.

In de artikelen die hierna volgen komen verschillende aspecten van het expertiseonderzoek aan de orde.

Schmidt en Boshuizen beschrijven een cognitief mechanisme dat in hun opvatting een belangrijke rol speelt bij de vaak geobserveerde verkortingen in het redeneren van experts: encapsulatie van declaratieve kennis. De auteurs verstaan daaronder een voortschrijdende inkapseling of "verpakking" van concepten in een kennisbestand in meeromvattende, hogere-orde concepten onder invloed van de toepassing van die kennis. Zij beschrijven een aantal experimenten in het medische domein waarin encapsulatie geobserveerd werd.

Ook in het werk van *De Jong en Ferguson-Hessler* speelt herstructurering van het kennisbestand onder invloed van toepassing een centrale rol. Zij bespreken een aantal studies waarin prestaties van goede studenten op het terrein van de natuurkunde vergeleken werden met die van zwakke studenten in dezelfde jaargroep. Dit is een aanpak van het expertisevraagstuk die nog betrekkelijk nieuw is. Op grond van hun onderzoek maken ze een onderscheid tussen verschillende typen kennis die ieder in bepaalde fasen van het oplossingsproces van natuurkundevraagstukken een rol spelen. Daarnaast kennen ze aan die kennis bepaalde kwaliteiten toe, zoals modaliteit, structuur en mate waarin ze geautomatiseerd is.

* Adres: Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Postbus 616, 6200 MD Maastricht.

Boshuizen, Hobus, Custers en Schmidt hebben de rol van praktijkervaring op de ontwikkeling van ziekteschema's of "ziektescripts" bestudeerd. Zij manipuleerden daarbij de aard van de informatie die beginnende en zeer ervaren artsen moesten verwerken. Zij vonden onder andere dat ervaren artsen intensief gebruik maken van contextuele informatie over patiënten en dat onervaren artsen vooral afgaan op de klacht die gepresenteerd wordt. De door hen gebruikte combinatie van quasi-experimentele vergelijking en experimentele manipulatie is een recente ontwikkeling in het expertiseonderzoek en maakt het mogelijk hypothesen omtrent de aard van de kennis van personen van verschillend expertiseniveau op meer precieze wijze te testen.

Ferguson-Hessler, tenslotte, breekt een lans voor de rol van meta-kennis bij het oplossen van complexe problemen. Zij maakt een onderscheid tussen objectieve en subjectieve componenten waaruit die meta-kennis zou bestaan. De objectieve component bestaat uit kennis van het domein en van de kennisverwervingsstrategieën die in dat domein succesvol zijn. De subjectieve component daarentegen bestaat uit de kennis die mensen hebben van hun eigen kennis en verwervingsstrategieën; datgene wat door anderen metacognitie genoemd wordt. Ze bespreekt enige implicaties van dit onderscheid voor de wijze waarop instructie gegeven zou moeten worden.

Manuscript ontvangen 8-6-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 3, pp. 135-148

Kennisencapsulatie

Henk G. Schmidt en Henny Boshuizen
*Rijksuniversiteit Limburg**

ABSTRACT

In this article, an attempt is undertaken to make the case for a new mechanism of learning, encapsulation of declarative knowledge. This mechanism would explain and unify a number of as yet not related empirical findings reported in the medical expertise literature. It is demonstrated that this mechanism can be held responsible for the decrease of references to causal, underlying, biomedical principles in think-aloud studies of subjects of different levels of expertise diagnosing cases. In addition, it is shown that encapsulation of the knowledge base is the mechanism underlying such diverse phenomena as intermediate effects in post-hoc pathophysiological explanation of clinical cases and the emergence of inferences in free recall of those cases as expertise unfolds. Implications for education are discussed.

INLEIDING

Dit artikel houdt zich bezig met de kennisstructuren van experts en gevorderde studenten. Meer in het bijzonder beschrijft het hoe die kennisstructuren veranderen onder invloed van toepassing ervan op taken die voor een bepaald domein karakteristiek zijn. Het probeert te laten zien hoe ervaring met die taken niet alleen kwantitatieve, maar vooral ook kwalitatieve effecten heeft op de wijze waarop betrokkenen hun kennis van het domein representeren.

Het domein dat ons interesseert is dat van de geneeskunde. De ontwikkeling van medische expertise is een aantrekkelijk studieobject omdat het in onze opvatting een prototype vormt van hoe expertise zich in zijn algemeenheid vormt. Zoals bij andere terreinen die potentieel van belang zijn, waaronder het oplossen van natuurkundige problemen (Chi, Feltovich, & Glaser), het reconstrueren van historische gebeurtenissen (Whineburg, 1991) of het begrijpen van problemen uit de sociale wetenschappen (Voss, Greene, Post, & Penner, 1983) zijn de problemen die experts in de geneeskunde tegenkomen, moeilijk precies te definiëren en kost het vele jaren om een acceptabel beheersingsniveau te bereiken. Boshuizen (1989) heeft geschat dat ongeveer 20.00 uur studie en praktijkervaring nodig zijn om een ervaren huisarts te worden. Daarnaast bestaat de medische praktijk – althans het gedeelte daarvan dat ons interesseert – uit diagnostisch redeneren, gebaseerd op veronderstelde onderliggende principes of mechanismen. Veel problemen waarmee mensen (expert of niet) geconfronteerd worden, vereisen een soort diagnose van de situatie op basis van conceptuele kennis teneinde die situatie te kunnen karakteriseren en op een adequate wijze te kunnen handelen. Het doel van dit artikel is, te verklaren hoe medische studenten en artsen betere diagnostici worden als gevolg van ervaring. We nemen daarbij aan dat medische diagnose een activiteit is die conceptuele kennis vereist, eerder dan een verzameling heuristieken; dat studenten en artsen, als ze met een patiënt geconfronteerd worden, relevante kennis activeren die hen in staat stelt de symptomen en klachten van die patiënt te representeren in termen van een of ander onderliggend proces. Onze interesse gaat uit naar de aard van die kennisstructuren, naar hoe zij zich ontwikkelen en hoe zij gebruikt worden.

In dit artikel zullen wij data presenteren die duiden op de werkzaamheid van een cognitief mechanisme dat wij als "encapsulatie" of inkapseling van kennis hebben omschreven (Boshui-

* Adres: Vakgroep Onderwijsontwikkeling en Onderwijsresearch, postbus 616, 6200 MD Maastricht.
Dit artikel is een bewerking van een lezing, gehouden tijdens een NATO Advanced Scientific Research Workshop "Advanced Models of Cognition for Medical Training and Practice," in Castelvecchio Pasoli, Italië, Juni 1991.

zen en Schmidt, 1992; Schmidt en Boshuizen, 1993a). Wij veronderstellen dat kennisencapsulatie een sleutelrol speelt bij de transitie van gevorderde naar expert in het medische domein, en niet alleen daar. Eerst wordt echter een kort en toegespitst overzicht gegeven van wat we weten over menselijke kennis en hoe die kennis verworven en gebruikt wordt.

KENNIS, KENNISGEBRUIK EN LEERMECHANISMEN

Betrekkelijk veel is bekend over hoe bestaande kennis het begrijpen van nieuwe informatie beïnvloedt, en welke factoren een rol spelen bij de activatie van die kennis. Onderzoek van Bransford en Johnson (1972) met betrekking tot contextuele activatie, Pichert en Andersons (1977) werk met betrekking tot effecten van het innemen van een bepaald perspectief bij het bestuderen van teksten, de experimenten van Spilich, Vesonder, Chiesi en Voss (1979) over hoe proefpersonen met veel en weinig voorkennis teksten onthouden, en de studies naar verschillen tussen experts en novieten in verschillende domeinen (e.g., Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Lesgold, 1984), tonen aan dat de specifieke kennis die personen van een bepaald domein hebben in hoge mate prestaties in dat domein bepalen. Minder is echter bekend over de precieze aard van de kennisstructuren die deze prestaties mogelijk maken. Onze kennis over hoe die structuren eruit zien is grotendeels theoretisch van aard en object van voortgaand debat. Niet alleen ons begrip van de *structuur* van menselijke kennis is nog in flux; hetzelfde kan gezegd worden van wat we weten over de mechanismen die mensen in staat stellen te *leren*. Er zijn voldoende ideeën over hoe nieuwe kennis verworven wordt, maar weer zijn die ideeën vooral theoretisch van aard, hebben weinig empirische basis en hebben daarom vooral heuristische waarde. We zullen opvattingen over de structuur van menselijke kennis en de mechanismen die verantwoordelijk geacht worden voor de ontwikkeling van die kennis hieronder bespreken.

In 1978 publiceerden Rumelhart en Norman een bijdrage aan een boek die getiteld was: "Accretion, tuning, and restructuring: Three modes of learning". Dit hoofdstuk vormde een min of meer complete weergave van wat rond 1980 gedacht werd over hoe mensen nieuwe kennis verwerven en incorporeren in hun bestaande structuren. Volgens beide auteurs heeft onze kennis van de wereld een schematische vorm (Schank & Abelson, 1977; Minsky, 1975) en is het begrijpen van de wereld in essentie schema-instantiatie. Schemata bevatten variabelen en informatiehouders ("slots"). Een situatie die begrepen moet worden, activeert een passend schema en de bijzondere kenmerken van die situatie vullen de slots van de verschillende variabelen die het schema vormen. Als echter het schema niet geheel past op de situatie die begrepen moet worden, worden elementen aan het schema toegevoegd. Rumelhart en Norman noemen dat proces *accretie;* het bestaande schema wordt uitgebreid zodat het gebruikt kan worden bij het begrijpen van een bredere klasse van situaties. Als gevolg van toenemende ervaring ontwikkelen schemata zich. Daarbij is het mogelijk dat de toegestane reeks van mogelijke waarden die variabelen kunnen aannemen kleiner wordt (specialisatie) of juist groter (generalisatie). Daarnaast neemt kennis over de condities waaronder een schema toepasbaar is, toe (conditionalisatie). Specialisatie, generalisatie en conditionalisatie tezamen vormen een leermechanisme dat Rumelhart en Norman (1978) *"tuning"* noemen, een proces waarbij een schema op een iteratieve wijze steeds beter de werkelijkheid representeert en waarbij de condities waaronder dat schema geactiveerd wordt steeds scherper gedefinieerd raken. Een derde mechanisme dat ze onderscheiden is *herstructurering*. Herstructurering van het kennisbestand treedt op als bestaande schemata tekort schieten bij het representeren van een bepaalde situatie, gebeurtenis of probleem. Het daaruit voortvloeiende elaboratieproces kan ertoe leiden dat geheel nieuwe schemata geconstrueerd worden die niet meer eenvoudig herleidbaar zijn tot de bestaande.

Een belangrijk bezwaar dat tegen schematheoretische verklaringen van leren kan worden ingebracht, is dat mensen in staat blijken dingen te leren waarvoor zij geen al bestaande schemata beschikbaar hebben[1]. Daarnaast blijkt het bijzonder lastig te definiëren wat schemata eigenlijk zijn (Alba & Hasher, 1983).

Anderson (1983, 1987) heeft een theorie geformuleerd die meer expliciet de aard van de veranderingen in de kennis onder invloed van ervaring beschrijft. Zijn ACT*-theorie veronderstelt dat kennis in eerste instantie in declaratieve vorm is opgeslagen. Dergelijke kennis is niet automatisch bruikbaar. Spreidende activatie en elaboratie zijn mechanismen die mensen in staat stellen te interacteren met hun omgeving. Die interactie verloopt moeizaam omdat ze deels plaats vindt onder controle van het bewustzijn. Als gevolg echter, van herhaalde toepassing van declaratieve proposities wordt langzamerhand procedurele kennis gevormd in de vorm van *produkties*. Produkties zijn regels die specificeren welke acties ondernomen moeten worden als aan bepaalde condities voldaan is. Dit proces wordt door hem *proceduralisatie* genoemd. Anderson postuleert echter nog een ander leermechanisme dat voor dit artikel van belang is. Hij veronderstelt dat, onder invloed van verdere oefening, de ontstane procedurele kennis, gerepresenteerd als een set produkties, als het ware "indikt" tot een kleiner aantal, meeromvattende, produkties. Hij noemt dit proces *kenniscompilatie*[2]. Proceduralisatie en compilatie zijn ervoor verantwoordelijk dat taken die eerst moeizaam en met veel fouten uitgevoerd worden, na verloop van tijd automatisch, snel en zonder fouten worden uitgevoerd. Experts hebben, in deze visie, de beschikking over een uitgebreid procedureel kennisbestand dat verregaand gecompileerd is.

KENNISENCAPSULATIE

De notie van kennisencapsulatie is door ons in eerste instantie geïntroduceerd teneinde een aantal verschijnselen die in de medische expertiseliteratuur worden beschreven te kunnen verhelderen; ontwikkelingsfenomenen die moeilijk te begrijpen lijken met behulp van begrippen als proceduralisatie of compilatie. Die bevindingen worden hieronder beschreven, eerst zullen we het begrip kennisencapsulatie nader proberen te omschrijven.

In Schmidt en Boshuizen (1993a) wordt encapsulatie gedefinieerd als een leermechanisme dat bestaat uit de voortschrijdende onderordening, of "verpakking", van lagere-orde begrippen en hun relaties in een declaratief kennisbestand, onder een kleiner aantal concepten of proposities van hogere orde, die dezelfde verklaringskracht hebben. Een voorbeeld. In de context van onderzoek naar de constructie van verklarende modellen op het terrein van de biologie, werd middelbare scholieren gevraagd een verklaring te geven waarom een bloedlichaampje in zuiver water opzwelt en barst, terwijl het in een zoutoplossing krimpt (Schmidt, Spaaij, & De Grave, 1988; Schmidt, De Grave, De Volder, Moust, & Patel, 1989). Scholieren die net een introductie in het onderwerp osmose achter de rug hadden, gaven zeer gedetailleerde verklaringen die een groot aantal relevante concepten bevatten, zoals Brownse beweging van moleculen, concentratieverschillen, diffusie van stoffen in water, de semipermeabiliteit van celmembranen, osmotische druk en turgor. Scholieren echter die al meer ervaring hadden opgedaan met vergelijkbare problemen, gaven veel meer gecondenseerde verklaringen, waarbij gebruik werd gemaakt van meer omvattende concepten. Eén leerling gaf als verklaring: "Dit is een voorbeeld van osmose". Voor deze leerling was het begrip osmose een voldoende verklaring voor de verschijnselen, een verklaring die geen verdere elaboratie nodig heeft. Osmose was voor haar rijk aan betekenis; het sluit meer gedetailleerde verklaringen in; die zijn als het ware in dat begrip verpakt. Onze theorie is dat deze encapsulatie het resultaat is van herhaalde activatie van lagere-orde verklaringen in response op vergelijkbare taken. Na verloop van tijd, en als gevolg van confrontatie met vergelijkbare taken, raken die verklaringen steeds meer ingekapseld in een kleine set van encapsulerende begrippen met vergelijkbare verklaringskracht. We nemen daarbij aan dat die lagere-orde begrippen uiteindelijk niet meer geactiveerd worden in de confrontatie met nieuwe taken, omdat ze niet meer nodig zijn bij het begrijpen of oplossen van die nieuwe opgaven.

Dus, net als compilatie is encapsulatie het resultaat van herhaalde toepassing van kennis bij het begrijpen van situaties of het oplossen van problemen. Er zijn echter ook verschillen tussen compilatie aan de ene en encapsulatie aan de andere kant. Kenniscompilatie opereert op een procedureel kennisbestand waarbij een grotere set van produkties wordt gereduceerd tot een

kleinere set. Het encapsulatieconcept doet geen aannames die omzetting van declaratieve kennis in procedurele vereisen voordat de efficiëntie van het denkproces kan worden bevorderd. Het opereert in een declaratief kennisbestand, waarbij onder invloed van ervaring moleculaire concepten en hun interrelaties langzamerhand worden gecondenseerd tot een kleinere verzameling molaire begrippen. De functie van encapsulatie is niet het trimmen van het kennisbestand door middel van specialisatie en generalisatie, maar het versnellen van de verwerking van nieuwe informatie. Daar meer grofkorrelige, omvattende, begrippen gebruikt worden bij het denken over de wereld, kan het begrijpen daarvan sneller en met minder fouten plaatsvinden. Tenslotte is geëncapsuleerde kennis onmiddellijk weer beschikbaar in geëlaboreerde vorm als de taak daarom mocht vragen (Schmidt & Boshuizen, 1993b). Procedurele kennis in gecompileerde vorm, daarentegen, is ontoegankelijk geworden voor gebruik in geëlaboreerde vorm.

AANWIJZINGEN VOOR KENNISENCAPSULATIE

We hebben al gezegd dat kennisencapsulatie werd geïntroduceerd ter verklaring van een aantal empirische verschijnselen die met name in de medische expertise literatuur gerapporteerd worden. Het betreft hier drie op het eerste gezicht ongerelateerde bevindingen, die nog geen bevredigende verklaring gevonden hebben. Onze poging is gericht op unificatie. Wat we willen aantonen is dat die fenomenen een gemeenschappelijke oorzaak hebben. De drie empirische verschijnselen waarop gedoeld wordt, zijn:

1. Afwezigheid van referenties naar onderliggende biomedische mechanismen in hardop-denk protocollen van ervaren artsen. Dit in tegenstelling tot de protocollen van gevorderde studenten (Boshuizen en Schmidt, 1992; Patel, Evans en Groen, 1989),
2. "Intermediate" effecten in post-hoc causale verklaringstaken[3] (Schmidt en Boshuizen, 1993a)
3. De aanwezigheid van inferenties in de "recall" van gevalsbeschrijvingen door ervaren artsen in contrast met de "recall" van beginners en gevorderden (Boshuizen, 1989; Groen en Patel, 1988; Patel en Frederiksen, 1984).

We zullen deze verschijnselen in de navolgende paragrafen bespreken en laten zien dat ze elk verklaard kunnen worden uit de toenemende encapsulatie van het kennisbestand zodra expertise toeneemt voorbij het niveau van de gevorderde student. Tenslotte zullen we enkele implicaties bespreken.

Afwezigheid van biomedische redeneringen in hardop-denk protocollen van experts

Naturalistische observaties en hardop-denk studie laten zonder uitzondering zien dat ervaren experts maar beperkt gebruik maken van biomedische kennis bij het redeneren over patiëntenproblemen. Patel et al. (1989), die een groot aantal studies op dit terrein bespreken, concluderen dat ervaren artsen in belangrijke mate lijken te denken over patiënten in termen van klinische begrippen en klinische manifestaties van ziekte. Studenten daarentegen, maken uitgebreid gebruik van biomedische begrippen als zij over een ziektegeval redeneren. Boshuizen, Schmidt en Coughlin (1988), die een hardop-denk strategie gebruikten, lieten zien dat huisartsen zelden naar biomedische concepten refereren wanneer zij hardop over een geval nadenken, terwijl studenten dat nu juist wel doen. Figuur 1 toont een prototypische bevinding. De figuur laat de proportie biomedische proposities (= uitspraken) zien, geproduceerd door tweedejaars, vierdejaars en vijfdejaars studenten en huisartsen, terwijl ze hardop over een geval van pancreatitis nadachten. De figuur illustreert het fenomeen duidelijk. Wanneer expertise toeneemt, neemt ook het gebruik van biomedische kennis eerst toe (wat geheel volgens verwachting is) om vervolgens drastisch af te nemen.

Dit is een ietwat tegenintuïtieve bevinding, omdat theorieën die zich bezighouden met de ontwikkeling van expertise (e.g., Chi, Glaser, & Farr, 1988) alle uitgaan van een min of meer monotone uitbreiding van diepe, causale kennis van een domein en van de aanname dat het deze

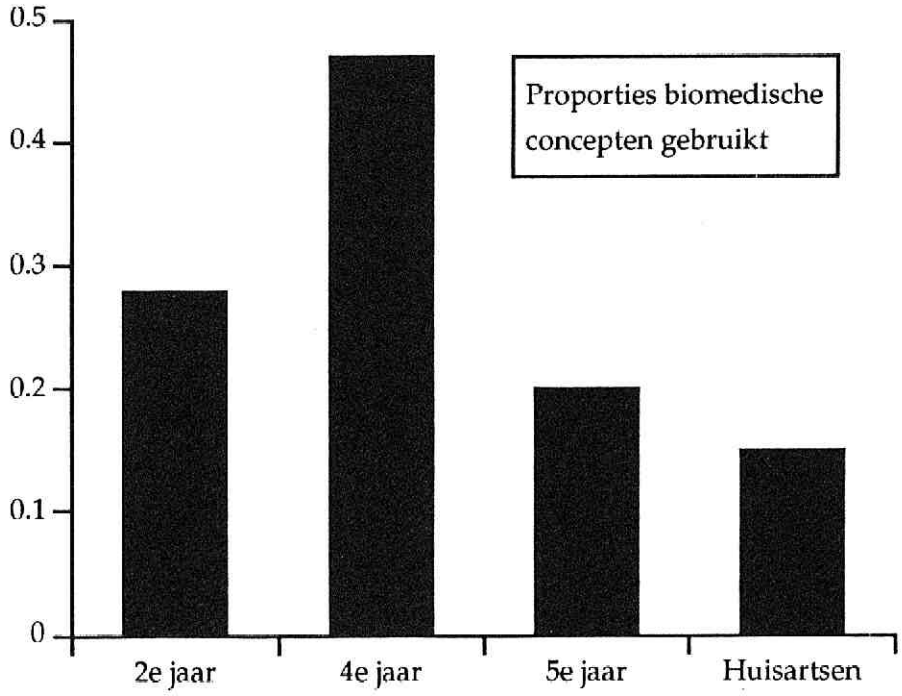


Fig. 1. Gemiddelde proporties biomedische concepten geproduceerd in hardop-denk protocollen als functie van expertise (Boshuizen, Schmidt, & Coughlin, 1988).

diepe kennis is die verantwoordelijk is voor het optreden van expertgedrag. Met andere woorden: men zou meer van dezelfde kennis verwachten bij meer ervaren personen in plaats van een andere soort kennis. Boshuizen en Schmidt (1992) veronderstellen dat drie elkaar uitsluitende cognitieve ontwikkelingsprocessen dit fenomeen kunnen verklaren. De eerste is dat voorbij een bepaald niveau van expertise biomedische kennis *rudimentair* wordt.; grote delen van die kennis (in het bijzonder gedetailleerde kennis) zou voor de expert niet meer terug te halen zijn uit het geheugen. Een tweede mogelijkheid is dat die kennis *inert* wordt; zij is er nog wel en is ook wel oproepbaar, maar wordt niet langer meer gebruikt. De kennis kan alleen geactiveerd wordt wanneer deze direct aangesproken wordt. Een derde mogelijkheid tenslotte, is dat in de loop van de jaren biomedische kennis *geëncapsuleerd* raakt en in klinische begrippen geïntegreerd wordt. Dat wil zeggen dat experts dan wel voornamelijk klinische begrippen gebruiken, maar dat die kennis gedetailleerde biomedische kennis verpakt. Deze drie hypothesen werden getest onder gebruikmaking van een gecombineerde "on-line" en "post-hoc" methodologie. Twintig proefpersonen die een verschillend niveau van expertise bezaten, kregen een verzameling van 48 kaartjes voorgelegd die de klachten en symptomen bevatten van een patiënt lijdend aan subacute pancreatitis. Hen werd gevraagd hardop te denken terwijl ze de informatie die op de kaartjes stond op een sequentiële wijze verwerkten. Dit was het on-line gedeelte van het onderzoek. Nadat deze taak was afgerond, werd hen gevraagd een diagnose te stellen. Tenslotte werd hen verzocht de klachten en symptomen die men zojuist verwerkt had, te verklaren in termen van onderliggende pathofysiologische processen. De hardop-denk en pathofysiologie protocollen van alle proefpersonen werden in proposities ontleed. Elke propositie werd vervolgens als klinisch of biomedisch geclassificeerd. Tenslotte werden beide protocollen met elkaar vergeleken. De veronderstelling was dat als biomedische kennis rudimentair geworden zou zijn bij experts, een daling gevonden zou moeten worden van deze categorie van informatie in zowel de on-line als de post-hoc verzamelde protocollen. Als echter de biomedische kennis nog wel aanwezig is, maar niet langer wordt gebruikt door experts, dan zou men een daling van deze categorie in de hardop-denk protocollen moeten verwachten (de kennis wordt immers niet meer

gebruikt), terwijl biomedische kennis wel in de post-hoc verklaringen aanwezig zou moeten zijn (er wordt immers rechtstreeks naar gevraagd). Dit zou erop duiden dat de biomedische kennis inert geworden is en niet langer meer wordt geactiveerd bij het oplossen van problemen. Een derde mogelijkheid is dat de biomedische concepten in klinische concepten geëncapsuleerd geraakt zijn. In dat geval zou men een daling van het aantal biomedische proposities in de hardop-denk protocollen verwachten, maar tevens een groei in de mate van integratie tussen online en post-hoc verklaringen. De resultaten steunden de encapsulatie-hypothese. Experts produceerden kleinere aantallen biomedische proposities – en meer klinische proposities – terwijl ze hardop nadachten over de klachten en symptomen van de patiënt, maar hun post-hoc pathofysiologische verklaringen vertoonden een sterkere "argument overlap" met de hardop-denk protocollen. "Argument overlap" is het verschijnsel waarbij proposities één of meer concepten delen. Het wordt door Van Dijk en Kintsch (1983) beschouwd als een maat voor integratie van een cognitieve representatie. In de protocollen van de experts werden dus meer biomedische en klinische proposities gevonden die concepten gemeen hadden. Bij de novieten was de overlap tussen beide protocollen slechts 14,9%. Dit percentage liep op tot 56,2 in de groep experts. Boshuizen en Schmidt (1992) concluderen daaruit dat experts biomedische kennis op impliciete ("tacit") wijze gebruiken, omdat deze kennis onderweg naar expertise in klinische kennis geëncapsuleerd raakt.

"Intermediate" effecten in post-hoc verklaringen van klinische gevalsbeschrijvingen

Aanwijzingen dat encapsulatie van het kennisbestand verantwoordelijk zou kunnen zijn voor veranderingen die in de ontwikkeling van studenten tot experts geconstateerd kunnen worden, zijn deels gebaseerd op onderzoek naar de wijze waarop personen van verschillend expertiseniveau gevalsbeschrijvingen cognitief representeren. In deze studies wordt van proefpersonen gevraagd een tekst van ongeveer een halve pagina, die de klachten en symptomen van een patiënt beschrijft, gedurende een paar minuten te bestuderen. Nadat de tekst verwijderd is, wordt hen gevraagd om (1) de meest waarschijnlijke diagnose voor deze patiënt te geven, (2) alles wat zij zich van de tekst herinneren op te schrijven, en (3) de symptomen en klachten te verklaren in termen van onderliggende causale pathofysiologisch processen, principes of mechanismen. Aangenomen wordt dat datgene wat proefpersonen zich herinneren van de tekst een afbeelding is van hun representatie van het probleem, daar de aard en de hoeveelheid herinnerde tekst afhankelijk zijn van de aard van de geactiveerde kennisstructuren die zijn ingezet bij het begrijpen van de gevalsbeschrijving (Kintsch en Greeno, 1985). Daarnaast wordt aangenomen dat het protocol dat de pathofysiologische verklaring bevat een representatie vormt van de kennis die actief was bij het verwerken van de informatie uit de gevalsbeschrijving (Patel en Groen, 1986). Gewoonlijk worden zowel de "recall" als het pathofysiologische protocol gesegmenteerd in een verzameling proposities. Het aantal proposities dat onthouden of geproduceerd is wordt geteld en hun interrelatie worden grafisch gerepresenteerd als een associatief netwerk.

In het kader van dit artikel zijn we in het bijzonder geïnteresseerd in de post-hoc verklaringen die door personen geproduceerd worden. Schmidt en Boshuizen (1993a) presenteerden een beschrijving van een door Patel en Groen (1986) ontwikkeld geval van acute bacteriële endocarditis aan vijf groepen proefpersonen, variërend van eerstejaars studenten gezondheidswetenschappen tot en met internisten. Het probleem beschrijft de lotgevallen van een werkeloze jongeman die bij de eerste hulp van een ziekenhuis komt omdat hij hoge koorts heeft. Hij klaagt erover dat hij zich uitgeput voelt en grote moeite heeft trappen te bestijgen. Hij is ook kortademig. De arts vindt puntwondjes aan de linker binnenkant van de arm van de jongeman, volgens deze veroorzaakt doordat hij door een kat gebeten is. Bij onderzoek blijkt hij 41 ° C te hebben en een zeer lage bloeddruk. In het hart is een ruis hoorbaar. In de urine worden bloedcellen gevonden. Hij klaagt verder over verlies van zicht gedurende korte perioden. Als deel van dit experiment is de proefpersonen, zoals gezegd, gevraagd een verklaring te geven voor deze symptomen in termen van hun onderliggende pathofysiologie. Schmidt en Boshuizen vonden in die verklaringen een sterk intermediate effect. Vierde- en zesdejaars geneeskundestudenten produceerden veel uitgebreidere en meer gedetailleerde verklaringen dan zowel de studenten

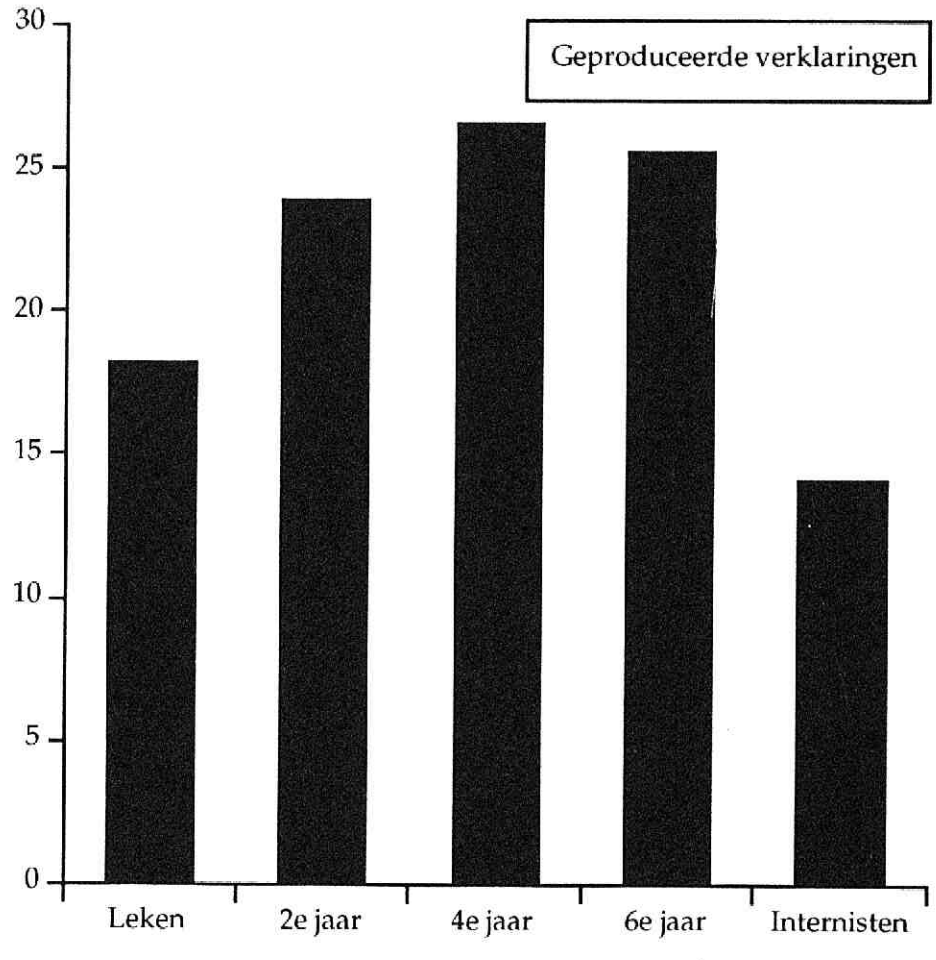


Fig. 2. Gemiddelde aantallen verklarende proposities in post-hoc protocollen (Schmidt & Boshuizen, 1993).

gezondheidswetenschappen en de artsen. In Figuur 2 zijn de aantallen verklarende proposities weergegeven die door de verschillende groepen geproduceerd werden.

Figuur 2 laat zien dat de verklaringen die door de internisten werden gegeven in hoge mate gecondenseerd waren. De vierdejaars medische studenten produceerden bijna twee keer zoveel proposities als de experts.

Figuur 3 en 4 bevatten voorbeelden van verklaringen die gegeven werden door een vierdejaars student en een internist. Het protocol van de student bestond uit 35 proposities. Elke propositie bestaat uit twee concepten of knopen ("nodes") die met elkaar verbonden zijn door middel van een kwalificatie van die relatie ("qualifier"). Deze relaties kunnen causaal (cau) van aard zijn, conditioneel (cond), temporeel (temp) of attributioneel (att). Daarnaast kunnen ze een locatie aanduiden (loc) of aangeven dat de tweede knoop een specificatie (spec) is van de eerste. Andere qualifiers zijn: negatie (neg), identiteit (iden) en lidmaatschap van een kategorie (isa). Daar knopen met meer dan één andere knoop verbonden kunnen zijn kan een protocol herschreven worden als een associatief netwerk. Bijvoorbeeld: de eerste zinnen van het protocol van gevorderde student 87-253 waren: " Door een kattebeet zijn toxinen in het bloed gekomen. Die hebben aanleiding gegeven tot antilichaamproduktie, misschien op basis van eerdere sensibilisatie tijdens drugsgebruikersverleden."

Deze proposities kunnen in het associatieve netwerk, dat op grond daarvan is geconstrueerd en dat als Figuur 3 is afgebeeld, eenvoudig teruggevonden worden. Deze Figuur laat zien dat de vierdejaars student poogt een verklaring te geven van bijna elk symptoom en elke klacht in de beschrijving van de patiënt met endocarditis (de in de gevalsbeschrijving genoemde cues zijn

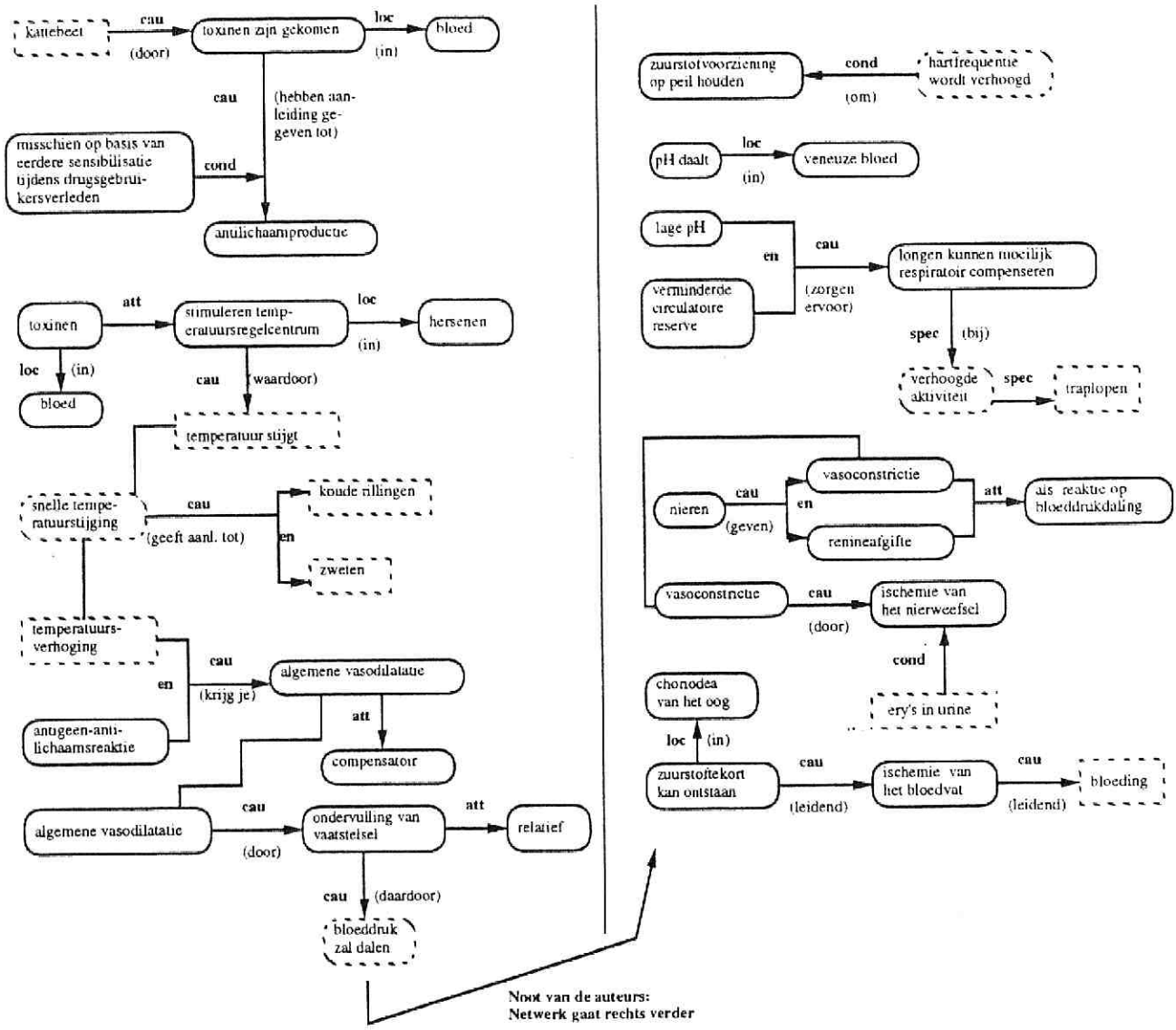


Fig. 3. Causaal netwerk van proposities ontleend aan het pathofysiologie protocol van gevorderde student 87-253.

voorzien van een box met onderbroken lijn). Dit suggereert dat hij de tekst uitgebreid cognitief verwerkt heeft. De verklaring is fijnmazig; nogal wat proposities verwijzen naar basale fysiologische mechanismen.

Figuur 4 beeldt de netwerkrepresentatie van het protocol van een aselect gekozen internist met 11 jaar ervaring af. Het contrast met het netwerk van de gevorderde student is opvallend: slechts 10 proposities worden geproduceerd. Deze proposities bevatten slechts een paar cues die direct aan de tekst ontleend zijn en de verklaring is in hoge mate geëncapsuleerd. Er wordt niet gerefereerd aan causale entiteiten zoals bacteria, noch aan een mediërend proces zoals sepsis. Zij lijken geëncapsuleerd te zijn in de paar concepten en hun onderlinge relaties die wél gegenereerd worden. Verklarende concepten op het niveau van de pathofysiologie zijn geëncapsuleerd in meeromvattende concepten. Voorbeelden zijn: "endocarditis aortaklep" en "embolieën".

We hebben de hypothese dat experts gebruik maken van kennis in geëncapsuleerde vorm getest op een meer directe wijze door alle protocollen te matchen tegen een standaardverklaring van de klachten en symptomen van de betrokken patiënt. Die standaardverklaring, bestond uit het minimale aantal proposities dat nodig is om alle klachten en symptomen te verklaren. Da

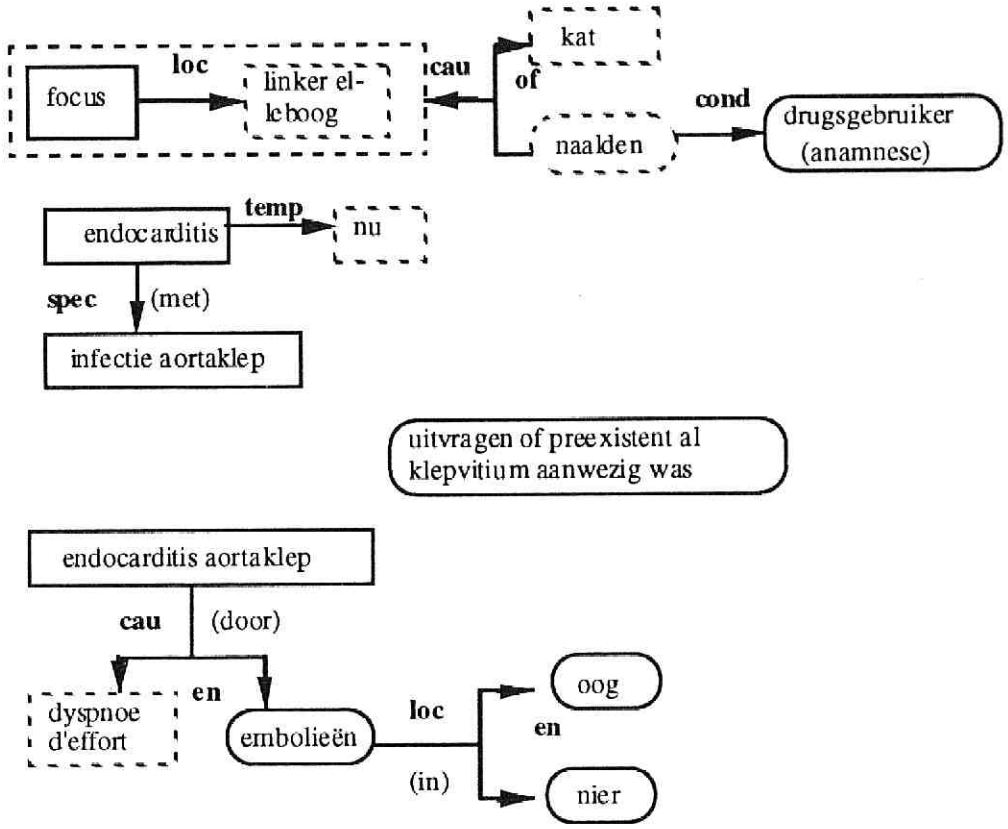


Fig. 4. Causaal netwerk van propositics ontleend aan het pathofysiologie protocol van internist 80-023.

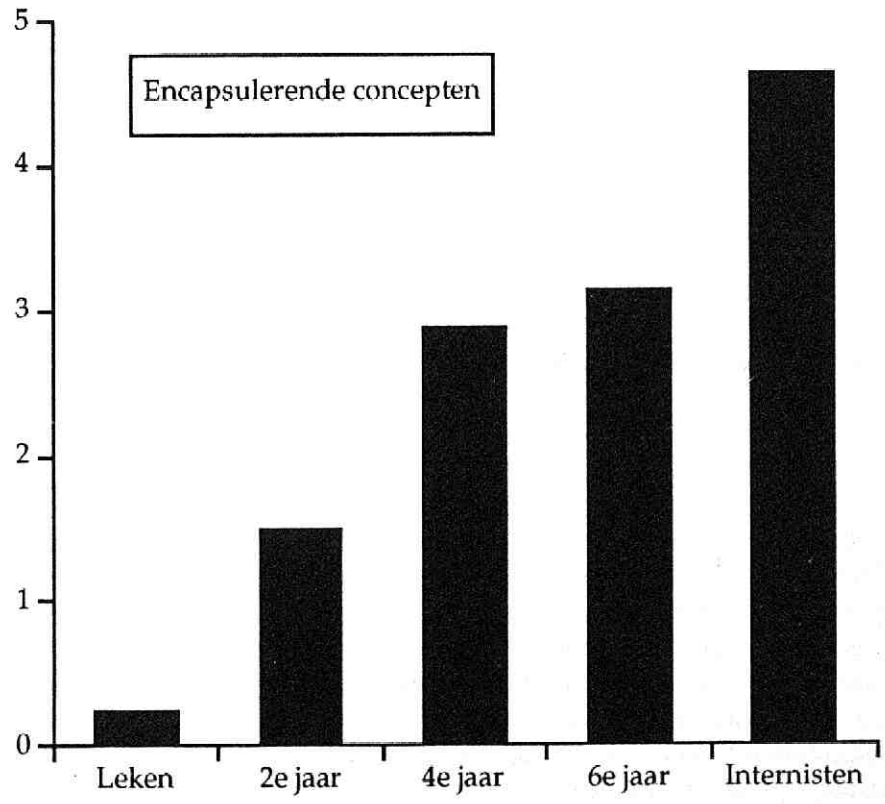


Fig. 5. Gemiddelde aantallen verklarende proposities uit de post-hoc protocollen die met de standaardverklaring matchen.

aantal was tien en elke propositie bestond uit sterk encapsulerende proposities, zoals sepsis, endocarditis en thrombo-embolieën. Dit zijn er vijf van: Het gebruik van drugs is een conditie voor het gebruik van besmette spuiten (1). De besmette spuit is de oorzaak van sepsis (4). Sepsis veroorzaakt een koortsreactie (5). Sepsis veroorzaakt endocarditis (6). Vegetatie van de aortaklep veroorzaakt thrombo-embolieën (10). De voorspelling was dat, als encapsulatie inderdaad een functie van expertise is, dat wil zeggen: als de protocollen die door de experts werden geproduceerd meer encapsulerende begrippen bevatten dan die van de studenten, een lineaire trend in de matchgegevens gevonden zou moeten worden. Figuur 5 laat de resultaten in grafische vorm zien.

Het effect is statistisch in hoge mate significant. Deze data laten zien dat de ontwikkeling van expertise gepaard gaat met een proces van voortschrijdende encapsulatie van lagere-orde concepten en proposities in een kleiner aantal meeromvattende concepten.

Inferenties in de "recall" van gevalsbeschrijvingen

Studies van Patel en haar collega's in Montreal (bijvoorbeeld, Patel & Frederiksen, 1984) documenteren dat de "recall" van een gevalsbeschrijving door artsen minder lijkt op de originele tekst dan die van meer of minder gevorderde studenten. De protocollen van experts bestaan in sterkere mate uit inferenties die op grond van de verwerking van de tekst blijkbaar gemaakt worden. Alhoewel de totale "recall" minder is , is het aantal inferenties veelal groter. Experts geven de informatie uit tekst terug in een aantal grote "chunks" die georganiseerd zijn rondom een klein aantal samenvattende constructen. Bijvoorbeeld: één van de artsen uit de Schmidt en Boshuizen (1993a) studie herinnerde zich de endocarditis casus op de volgende wijze: "De patiënt is een jongeman met een hoge koorts, die een *septisch syndroom* laat zien. Dit suggereert het *gebruik van drugs*. Hij vertoont tekenen van *thrombo-embolieën*, vanwege een *beschadigde hartklep*. De *tachycardia* klopt met het *aorta vitium*" (samenvattende concepten zijn cursief gezet). Deze "recall" van de gevalsbeschrijving bestaat bijna geheel uit inferenties. Het lijkt erop dat deze expert geen duidelijk onderscheid maakt tussen de informatie zoals gepresenteerd in de tekst en zijn eigen, onmiddellijke, interpretatie van die informatie. Dit is op zichzelf natuurlijk geen nieuwe vondst; alle informatie wordt opgeslagen, niet in originele vorm, maar als geïnterpreteerd door degene die deze informatie verwerkt heeft. Wat belangrijk is, is dat de "recall" van deze internist sterk lijkt op zijn pathofysiologisch protocol. Daarnaast is er een niet onaanzienlijke overlap tussen zijn "recall" en de standaardverklaring die in de vorige paragraaf aan de orde geweest is. Het lijkt er dus op dat de verklaring die de expert heeft voor de verschijnselen en zijn herinnering van die verschijnselen in elkaar geïntegreerd geraakt zijn. Daarnaast lijkt het erop dat de encapsulerende concepten die experts gebruiken bij het begrijpen van de casus de integrerende factor zijn. Wat wij waarnemen als de vaardigheid van experts om de kern van de gepresenteerde informatie te zien, in plaats van de oppervlaktekenmerken van die informatie, lijkt wederom een uiting van encapsulatie van het kennisbestand te zijn.

CONCLUSIE

In dit artikel hebben we een poging gedaan encapsulatie van declaratieve kennis te introduceren als een nieuw leermechanisme, dat in staat is een aantal tot op heden niet gerelateerde empirische bevindingen in de medische expertise literatuur te verklaren op een wijze die deze bevindingen identificeert als speciale gevallen van één en hetzelfde onderliggende proces. We hebben laten zien dat dit mechanisme een verklaring biedt voor de daling van het aantal biomedische concepten in hardop-denk studies in functie van het expertiseniveau (Boshuizen en Schmidt, 1992). Daarnaast hebben we laten zien dat encapsulatie van het kennisbestand het mechanisme is dat ten grondslag ligt aan zulke verschillende fenomenen als "Intermediate" effecten in post-hoc pathofysiologische verklaringen van patiëntenbeschrijvingen (Schmidt & Boshuizen, 1993a) en het toenemen van het aantal inferenties in de "recall" van dergelijke casus wanneer expertise zich ontwikkelt (Groen & Patel, 1988). Een definitieve test van onze these is overigens pas

mogelijk als we in staat zijn in longitudinale studies encapsulatie in de cognitieve structuren van individuen aan te tonen. Het expert-noviet paradigma is simpelweg niet krachtig genoeg om dergelijke ontwikkelingsgeoriënteerde hypothesen op een volledig bevredigende wijze te toetsen (zie ook Schmidt & Boshuizen 1993b).

Wat zijn de implicaties van de theorie die in dit artikel uiteengezet is en van de empirische bevindingen waarop deze steunt?

De eerste is dat encapsulatie een rol zou kunnen spelen in de relatief plotselinge verschuivingen in prestatie die bij veel expertisegerelateerde taken kan worden waargenomen. Het "intermediate" effect, het verschijnsel dat gevorderde studenten meer gedetailleerde kennis lijken te gebruiken dan experts en dat bij een groot aantal taken gedemonstreerd is, suggereert dat expertise zich in stadia ontwikkelt. We hebben elders beargumenteerd dat het nuttig kan zijn de ontwikkeling van expertise te conceptualiseren als een proces waarin op enkele tijdstippen een fundamentele herstructurering van kennis plaatsvindt van het type dat in deze bijdrage beschreven is, gevolgd door perioden van accretie (Schmidt, Norman, & Boshuizen, 1990; Schmidt & Boshuizen, 1993b). Eén zo'n periode waarin betrekkelijk massieve kennisherstructurering plaatsvindt, is wanneer medische studenten voor het eerst met "echte" patiënten in aanraking komen en zich met de gezondheidsproblemen van die patiënten gaan bezighouden. Het betreft de periode van de stages in ziekenhuizen en bij de huisarts. Alle ons bekende studies waarin ontwikkelingsgerelateerde veranderingen worden aangetoond, zoals de daling van de hoeveelheid pathofysiologische verklaringen die studenten geven in hardop-denk en verklaringstaken die in dit artikel besproken zijn, laten zien dat de transitie in die periode plaatsvindt. Dat het hier inderdaad om een verschuiving gaat die het resultaat is van de blootstelling aan nieuwe leerervaringen, wordt aannemelijk gemaakt door het feit dat waar bij Nederlandse medische studenten de verandering in het vijfde en zesde studiejaar optreedt, in de VS en Canada eenzelfde verschuiving in het derde en vierde jaar kan worden waargenomen. Dat is in die landen de periode waarin de studenten hun eerste ervaringen met patiënten opdoen.

We hebben eerder encapsulatie als een min of meer geleidelijk proces afgeschilderd, als de *voortschrijdende* onderordening van lagere-orde begrippen. Hierboven hebben we encapsulatie verantwoordelijk gesteld voor min of meer plotselinge veranderingen in het kennisbestand. Hoe zijn die twee standpunten met elkaar te rijmen? Zeker, encapsulatie is in onze opvatting een geleidelijk proces het is niet zo dat kennis van het ene op het andere moment in de tijd een volstrekt andere structuur aanneemt. Echter, encapsulatie is een response op de taak waarvoor het organisme zich gesteld ziet. Bestaat die taak er vooral uit nieuwe kennis te verwerven, dan zal er relatief weinig encapsulatie in het kennisbestand optreden. Bestaat de taak uit het herhaald toepassen van die kennis, dan zal die kennis gradueel verpakt worden in hogere-orde begrippen. Toepassing van kennis en *herhaalde* toepassing daarvan zijn dus noodzakelijke voorwaarden voor het optreden van encapsulatie. Daar in veel onderwijsprogramma's het herhaald gebruik van kennis in afgegrensde perioden plaatsvindt, zullen dat ook vooral de perioden zijn waarin encapsulatie optreedt. Vandaar dat wij encapsulatie verantwoordelijk stellen voor de relatief plotselinge verschuivingen die in die perioden optreden. Zie voor een kritisch-afwijzende discussie van dit standpunt Patel en Groen (1991).

Een tweede implicatie heeft met de eerste te maken. Daar encapsulatie door interactie met patiënten in gang lijkt te worden gezet, zou het interessant zijn de mogelijkheid te overwegen dat dit proces versneld zou kunnen worden. In andere woorden: zou het nuttig zijn studenten al veel vroeger in hun studie met echte patiënten te confronteren? Een antwoord op deze vraag is afhankelijk van de mate waarin studenten een kritische massa van biomedische kennis moeten hebben opgebouwd voordat encapsulatie kan plaatsvinden. Moet een student eerst over de hele linie een grote hoeveelheid biomedische kennis hebben opgebouwd of treden encapsulaties op zodra in een subdomein voldoende kennis verworven is? Dit is geen theoretische kwestie want op dit moment wordt aan minstens twee medische faculteiten in Nederland overwogen om studenten in een veel vroeger stadium te confronteren met patiënten. Onze positie is dat het niet zinvol is studenten intensief met echte patiënten te confronteren voordat die kritische massa van biomedische kennis verworven is. Dat wil niet zeggen dat die kritische massa over de hele linie

bereikt zou moeten zijn, wil encapsulatie kunnen optreden. Het is zeer wel mogelijk dat op deelterreinen waarop voldoende kennis bestaat, encapsulatie optreedt onder invloed van het diagnostiseren van patiëntenproblemen in dat deelgebied. Dat is een empirische kwestie die alleen op empirische wijze kan worden verhelderd (al is het de vraag welke opleiding zich voor een dergelijk experiment zal willen lenen).

Ten derde: het is onaannemelijk dat de in deze bijdrage beschreven fenomenen zich beperken tot het medische domein. Niet uitgesloten kan worden dat een bevindingen in andere domeinen op hetzelfde mechanisme teruggevoerd kunnen worden. Daarbij kan gedacht worden aan het "chunken" van informatie dat door Chase en Simon (1973) bij schaakgrootmeesters geconstateerd werd en aan het feit dat gevorderde schakers tijdens het spel meer alternatieven overwegen dan grootmeesters (De Groot, 1946). Ook het "step-skipping" fenomeen dat Koedinger en Anderson (1990) bij ervaren oplossers van geometrische problemen constateerden, lijkt kenmerken gemeen te hebben met wat wij in onze experimenten observeren.

Een laatste implicatie is wellicht dat het zinvol kan zijn onderzoek te richten op het kennisbestand van novieten, gevorderden en experts, eerder dan op hun prestaties in expertisegerelateerde taken. Het in kaart brengen van iemands kennisbestand is een lastige onderneming, temeer daar personen de neiging hebben kennis te mobiliseren afhankelijk van de specifieke taak. Niettemin is een dergelijke benadering niet geheel zonder beloning, zoals de hier gepresenteerde experimenten hopelijk hebben laten zien.

NOTEN

1. Meer recente opvattingen beschouwen kennis als gerepresenteerd in tamelijk losse associatieve netwerken (Rumelhart en McClelland, 1986). Kennis in die netwerken wordt geactiveerd en georganiseerd naar de aard van de specifieke taak die uitgevoerd moet worden. Representatie van een probleem bestaat daarbij niet uit de instantiëring van een min of meer adequaat schema maar uit de ad-hoc constructie van een taakspecifieke representatie (Kintsch, 1988). Deze ideeën hebben enige empirische steun gevonden in met name computermodelering van basale menselijke cognitie, maar lijken nog te weinig uitgewerkt om menselijke prestaties bij het uitvoeren van complexe taken te helpen begrijpen.
2. Een vergelijkbaar leermechanisme, "chunking", wordt door Newell en Rosenbloom (1981) beschreven. Volgens Rosenbloom, Laird en Newell (1988) dienen zowel compilatie als chunking ertoe om het kennisbestand af te stemmen ("tuning"), dat wil zeggen: producties te generaliseren of te specialiseren. In tegenstelling tot Anderson neemt Newell aan dat kennis onmiddellijk in procedurele vorm beschikbaar is en dat het dus niet nodig is eerst een proces van proceduralisatie van declaratieve kennis te postuleren. Zie voor een discussie over verschillen tussen encapsulatie aan de ene en compilatie en chunking aan de andere kant Schmidt en Boshuizen (1993a), noot 3.
3. Een "Intermediate" effect is het frequent optreden van een bepaald gedrag bij personen van een middelmatig kennis- of vaardigheidsniveau, terwijl datzelfde gedrag zowel bij beginners als experts weinig voorkomt. Figuur 1 en 2 tonen voorbeelden van dergelijke "Intermediate" effecten.
4. Klinische kennis wordt hier gedefinieerd als kennis van kenmerken van zieke mensen. Deze kennis heeft betrekking op de verschillende wijzen waarop ziekte zich in patiënten kan manifesteren; het soort klachten die verwacht kan worden gegeven die ziekte; de aard en de variabiliteit van de symptomen en de manieren waarop de ziekte behandeld kan worden. Biomedische kennis daarentegen bestaat uit (patho)biologische principes, mechanismen of processen die aan de manifestaties van ziekte ten grondslag liggen. Biomedische kennis wordt uitgedrukt in termen van de structuur en het functioneren van cellen, weefsels, organen en orgaansystemen en in termen van hun verstoringen door middel van entiteiten zoals bacteria of virussen, kwaadaardige groei, etc. Pathobiologie (ook wel: pathofysiologie) is dus de leer van de biologie der verstoorde functies.
5. Onder pathofysiologie wordt verstaan het causale ziekteverwekkende proces dat voor het optreden van de klachten en symptomen verantwoordelijk kan worden gesteld. Bijvoorbeeld: de bij griep horende klachten en symptomen zijn koorts, pijn in de spieren en een rode keel. Het onmderliggende pathofysiologische proces bestaat daarin dat het griepvirus via het ademhalingssysteem het lichaam binnendringt en een immuunresponse oproept.

LITERATUUR

Alba, J.W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*, 203 - 231.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192-210

Boshuizen, H. P. A. (1989). *De ontwikkeling van medische expertise; een cognitief-psychologische benadering.* Academisch Proefschrift. Haarlem: Thesis.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). The role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16,* 153-184.

Boshuizen, H. P. A., Schmidt, H. G., & Coughlin, L. D. (1988). On the application of medical basic science knowledge in clinical reasoning: Implications for structural knowledge differences between experts and novices. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 517-523). Hillsdale, NJ: Lawrence Erlbaum.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717-726.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problem by experts and novices. *Cognitive Science, 5,* 121-152.

Chi, M.T.H., Glaser, R., & Farr, M.J. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

De Groot, A. D. (1946). *Het denken van den schaker.* Den Haag: Noord Holland.

Groen, G.J., & Patel, V.L. (1988). The relationship between comprehension and reasoning in medical expertise. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.). *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109-129.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14,* 511-550.

Lesgold, A. M. (1984). Acquiring Expertise. In J. R. Anderson, & S. M. Kosslyn (Eds.), *Tutorials in learning and memory; Essays in honor of Gordon Bower* (pp. 31-60). San Francisco: Freeman & Comp.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.). *The psychology of computer vision.* New York, NY: McGraw-Hill.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.). *Cognitive skills and their acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Biomedical knowledge and clinical reasoning. In D. A. Evans, & V.L. Patel, (Eds.). *Cognitive science in medicine* (pp. 53-112). Cambridge, MA: The MIT Press.

Patel, V. L., & Frederiksen, C. L. (1984). Cognitive processes in comprehension and knowledge acquisition by medical students and physicians. In H. G. Schmidt and M. L. de Volder (Eds.), *Tutorials in problem-based learning* (pp. 143-157). Assen: Van Gorcum.

Patel, V.L., & Groen, G.J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science, 10,* 91-116.

Patel, V. L., & Groen, G. J. (1991). Developmental accounts of the transition from medical student to doctor: Some problems and suggestions. *Medical Education, 25,* 527-535.

Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology, 69,* 305-315.

Rosenbloom, P. S., Laird, J. E., and Newell, A. (1988). The chunking of skill and knowledge. In H. Bouma and B. A. G. Elsendoorn (Eds.), *Working Models of Human Perception* (pp. 391-440). London: Academic Press.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning, and restructuring: Three modes of learning. In J. W. Cotton & R. L. Klatzky (Eds.), *Semantic factors in cognition* (pp. 37-60). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1986) Parallel distributed processing. Cambridge, MA: The MIT Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Lawrence Erlbaum.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21,* 338-351.

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On acquiring expertise in medicine. *Educational Psychology Review, 5,* 1-17.

Schmidt, H. G., De Grave, W. S., De Volder, M. L., Moust, J. H.C., and Patel, V. L. (1989). Explanatory models in the processing of science text: The role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology, 81,* 610-619.

Schmidt, H.G., Norman, G. R., and Boshuizen, H. P. A. (1990). A Cognitive Perspective on Medical Expertise: Theory and Implications. *Academic Medicine, 65,* 611-621.

Schmidt, H. G., Spaaij, G., & De Grave, W. S. (1988). Opsporen van misconcepties bij middelbare scholieren. *Tijdschrift voor Onderwijsresearch, 13,* 129-140.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18,* 275-290.

Van Dijk, T. A., & Kintsch, W. (1984). *Strategies of discourse comprehension.* New York, NY: Academic Press.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 17, pp. 165-213). New York, NY: Academic Press.

Whineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83,* 73-87.

Manuscript ontvangen 17-3-1993
Definitieve versie ontvangen 8-6-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 3, pp. 149-162

Probleemoplossen, leren en onderwijzen in exacte vakken: een voorbeeld uit de natuurkunde

Ton de Jong[1*] en Monica G.M. Ferguson-Hessler[2]
[1] *Faculteit Toegepaste Onderwijskunde, Universiteit Twente**
[2] *Faculteit Wijsbegeerte & Maatschappijwetenschappen en Faculteit Natuurkunde, Technische Universiteit Eindhoven*

ABSTRACT

Through the years we have performed a number of studies in the field of problem solving, learning and instruction in physics. More specifically, we have examined these phenomena for first year University of Technology students in the domain of Electricity and Magnetism (E&M). In our studies the knowledge base, the problem solving process and learning processes of subjects were meticulously analyzed and a comparison between good and poor students was made (De Jong, 1986; Ferguson-Hessler, 1989; De Jong & Ferguson-Hessler 1983, 1984, 1985, 1986, 1987, 1988, 1991, 1993; Ferguson-Hessler & De Jong, 1986, 1987, 1989, 1990, 1993a, 1993b). A theoretical model containing three main elements was developed:

- Problem solving in semantically rich domains is considered as the *successive transformation of an internal representation* of the problem.
- In the knowledge base of the problem solver different *types of knowledge* can be distinguished, each of them playing a specific role in the problem solving process.
- The knowledge base can be described in terms of different *qualities*, such as for example the modality of knowledge, the structure and level of compiledness.

The present paper gives a summarised overview of the concepts developed and of the results of the empirical studies.

INLEIDING

Natuurkunde is een van de vakken die als moeilijk bekend staan. Er is hiervoor een drietal redenen te geven (zie Ferguson-Hessler & De Jong, 1993b):

- Natuurkunde heeft een sterke, intrinsieke structuur: begrippen hebben een ondubbelzinnige betekenis en tussen de begrippen bestaan logische verbanden; wetten hebben stringente afleidingen en geldigheidsvoorwaarden. De structuur is vaak hiërarchisch: algemene en abstracte begrippen en wetten overkoepelen een aantal specifieke definities en formules.
- Begrippen en relaties hebben zowel een formele en symbolische representatie, noodzakelijk voor wiskundige manipulatie, als een concrete fysisch-technische representatie. Men moet niet alleen beide representaties kennen, maar ook op een soepele manier van representatie kunnen wisselen, en voor iedere toepassing de meest geschikte representatie kunnen kiezen.
- Doelstellingen binnen de natuurkunde zijn vaak gericht op het uitvoeren van taken waar inzichtelijk handelen gevraagd wordt, bijvoorbeeld probleemoplossen.

Omdat natuurkundevakken aan de ene kant als complex en moeilijk worden ervaren, maar aan de andere kant een heldere en ondubbelzinnige structuur hebben, is er veel onderzoek geweest naar het oplossen van natuurkunde problemen (zie bijvoorbeeld Reif, Larkin, & Brackett, 1976;

* Adres: Vakgroep Instructietechnologie, Faculteit Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede.

Larkin, 1981; Chi, Feltovich, & Glaser, 1981; Anzai & Yokoyama, 1986), en recentelijk meer naar het verwerven van kennis in de natuurkunde (Chi,, Bassok, Lewis, Reimann, & Glaser, 1989; Ferguson-Hessler & De Jong, 1990). Een logische volgende stap is het in kaart brengen van onderwijsprocessen (Ferguson-Hessler & De Jong, 1993a).

EEN ALGEMEEN MODEL VAN PROBLEEMOPLOSSEN, LEREN EN ONDERWIJZEN

In het complexe proces van leren en onderwijzen kan een aantal hoofdaspecten onderscheiden worden:
a. Het uitvoeren van een criteriumtaak (bijvoorbeeld het oplossen van problemen);
b. Een kennisbasis, die verworven moet worden om de criteriumtaak uit te kunnen voeren;
c. Leerprocessen, die nodig zijn voor de verwerving van de kennisbasis;
d. Instructie, bedoeld om leerprocessen te faciliteren.

Bij leren en instructie is er altijd een of ander *leerdoel*. Leerdoelen kunnen vele vormen hebben, van het behendig kunnen fietsen tot het kunnen oplossen van natuurkunde problemen, maar vrijwel altijd is het leerdoel het uit kunnen voeren van een specifieke taak. Het bereiken van dit doel wordt vaak beoordeeld met behulp van een test, tentamen of examen. Een algemeen model van het uitvoeren van een taak is dat de oplosser een interne representatie van de taak creëert en deze representatie manipuleert tot de oplossing wordt bereikt (Newell & Simon, 1972).

De interne representatie van een aangeboden probleem is in eerste instantie een representatie van het probleem op zich. Deze initiële representatie komt tot stand op basis van een proces van selectieve perceptie en door het toevoegen van informatie uit het kennisbestand van de uitvoerder van de taak. *De kennisbasis* heeft zo op twee manieren invloed op de initiële representatie, door sturing van het proces van selectieve perceptie en rechtstreeks, door aanvulling van de aangeboden informatie met elementen uit de kennisbasis. Voor de verdere uitvoering van de taak wordt ook gebruik gemaakt van (andere) onderdelen van de kennisbasis (zoals bijvoorbeeld formules en methodes). De uitkomst van de taak, en ook de manier waarop de uitkomst bereikt werd, kunnen dan als nieuwe onderdelen worden toegevoegd aan de kennisbasis.

Leren is het proces van het construeren van een kennisbestand. Algemeen gesteld zijn er twee manieren van leren. De eerste manier is simpelweg het (proberen) uitvoeren van de criteriumtaak, zelfs wanneer de kennisbasis nog niet adequaat is om enig succes te garanderen. De kennisbasis verandert dan als resultaat van feedback van succes en falen bij het uitvoeren van de taak. Deze vorm van leren wordt wel aangeduid als 'learning by doing'. Deze vorm van leren vormt de kern van een invloedrijke theorie als ACT* (Anderson, 1983) en recent krijgt deze vorm van leren ook veel aandacht onder invloed van de zogenaamde 'Cognitive apprenticeship' benadering (Collins, Brown, & Newman, 1989).

De tweede manier van leren bestaat uit het verzamelen van informatie die nuttig kan zijn bij het uitvoeren van de criteriumtaak. Deze manier van leren wordt wel 'learning by instruction' genoemd. Learning by instruction kan betekenen dat een docent de lerende direct vertelt wat gekend en gekund moet worden ('expository learning'), maar het kan ook betekenen dat de lerende actief kennis construeert uit aangeboden bronnen, zoals bijvoorbeeld plaats kan vinden bij het gebruik van simulaties in het onderwijs ('exploratory learning', zie bijvoorbeeld De Jong, 1991).

Ook het proces van leren (door instructie) kan verder worden gedecomponeerd. Evenals bij het uitvoeren van de criteriumtaak wordt hier materiaal gepresenteerd en wordt door een proces van selectieve perceptie een representatie van deze informatie gecreëerd. Door specifieke leerprocessen (bv. inductie) wordt deze representatie omgezet in kennis.

Instructie is in essentie het creëren van omstandigheden waarin leren plaats kan vinden. Dit kan

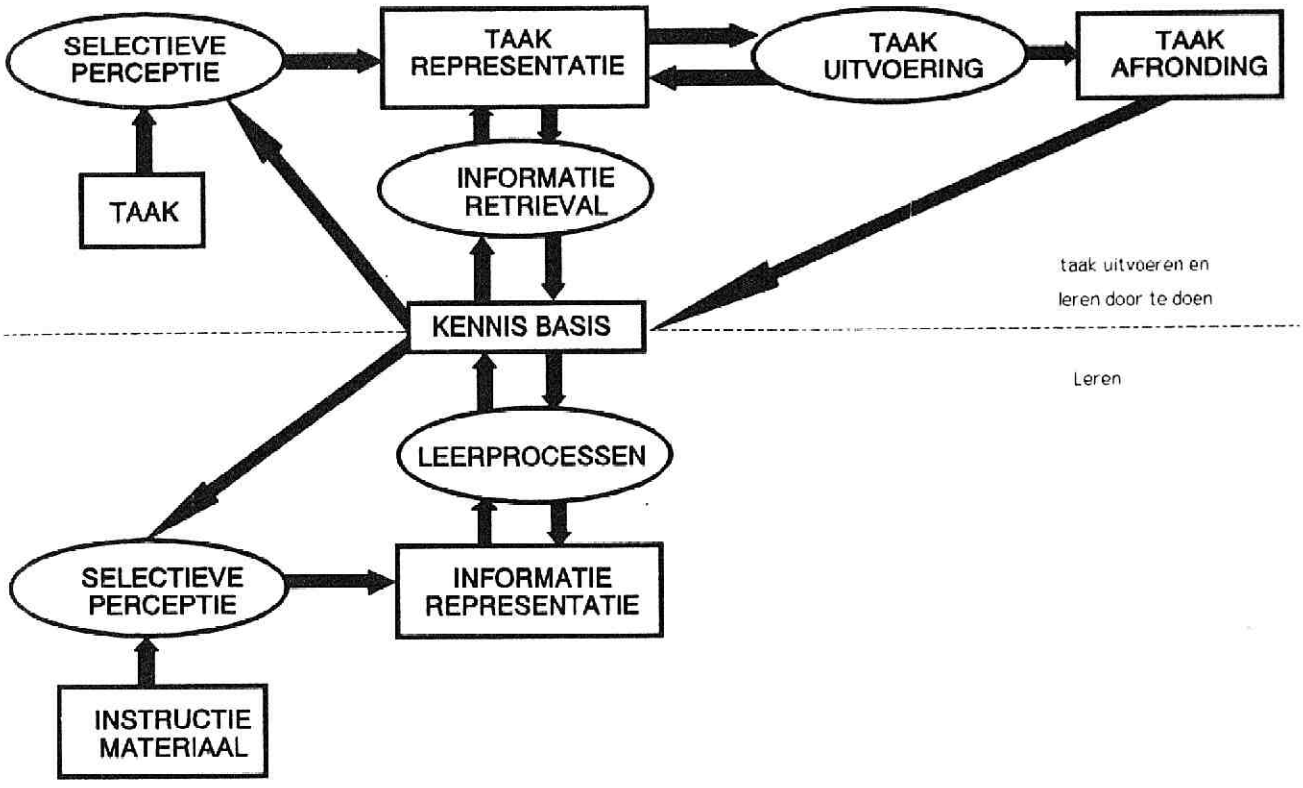


Fig. 1. Een algemeen model voor taakuitvoering, leren en instructie.

dus, gelet op het bovenstaande, bestaan uit het geven van direct onderwijs of het laten oefenen van criteriumtaken. Curricula geven vaak een combinatie van beide door bijvoorbeeld het opnemen van zowel colleges als practica.

Figuur 1 (uit De Jong & Ferguson-Hessler, 1993) geeft een samenvatting van ons model van onderwijzen, leren en taakuitvoering. In principe geldt de figuur voor een grote verscheidenheid van taken en kan deze worden toegepast op elk specifiek type taak. Figuur 1 illustreert de centrale plaats in instructie, leren en taakuitvoering van de kennisbasis. De kennisbasis resulteert uit en beïnvloedt leerprocessen en het vormt de basis van (criterium) taakuitvoering.

DE KENNISBASIS

De centrale plaats van de kennisbasis rechtvaardigt een nadere analyse. Een systematische beschrijving van de kennisbasis kan gegeven worden langs twee dimensies: de inhoud van de kennisbasis in termen van *typen kennis*, en de *kwaliteiten* van de kennisbasis.

De inhoud van de kennisbasis

In de literatuur wordt een groot aantal termen gebruikt om kennis te beschrijven. Veel van deze termen geven, impliciet, een epistemologische karakterisering van kennis en onderscheiden verschillende typen of soorten van kennis. Epistemologische typologieën hebben twee belangrijke kenmerken. Op de eerste plaats zijn de typologieën taak-afhankelijk. Dat betekent dat als we verschillende typen criteriumtaken nemen er verschillende kennistypologieën zullen ontstaan. Op de tweede plaats kan dezelfde feitelijke kennis binnen dezelfde taak verschillend worden geclassificeerd, afhankelijk van de rol die de kennis speelt in de criteriumtaak op een

bepaald moment. Een specifiek 'idee' over een patiënt in een medische diagnose taak zal in het begin de rol van 'hypothese' spelen, maar later de rol van 'gestelde diagnose' kunnen vervullen. Veelal worden typen kennis geïntroduceerd zonder dat rekening met deze twee kenmerken wordt gehouden.

In ons onderzoek hebben we als criteriumtaak probleemoplossen in de natuurkunde en in deze context onderscheiden we vier soorten kennis (zie De Jong & Ferguson-Hessler, 1986):

- *Situationele kennis* is kennis van (prototypische) situaties zoals die in een bepaald vakgebied voor kunnen komen. Kennis van probleemsituaties helpt de probleemoplosser de relevante elementen uit een aangeboden situatie te selecteren (selectieve perceptie) en, indien nodig, elementen aan de situatie toe te voegen (Braune & Foshay, 1983). Situationele kennis helpt bij het creëren van een probleemrepresentatie op basis waarvan additionele kennis (declaratief of procedureel) uit het geheugen geselecteerd kan worden. Situationele kennis wordt daarom wel aangeduid met de term 'conditionele kennis' (Alexander & Judy, 1988).
- *Declaratieve kennis*, ook wel conceptuele kennis genoemd (Greeno, 1978), is 'statische' kennis van definities, formules, wetten, verschijnselen, en verbanden die geldig zijn binnen een vakgebied. Onderdelen van de declaratieve kennis worden aan de probleemrepresentatie toegevoegd wanneer de probleemoplosser denkt deze voor de oplossing nodig te hebben.
- *Procedurele kennis* bevat acties of manipulaties die in een bepaald domein geldig zijn. Procedurele kennis helpt de probleemoplosser transities tussen probleemtoestanden te maken.
- *Strategische kennis* helpt de oplosser het probleemoplosproces te organiseren. Het kan gezien worden als een algemeen plan van actie waarin de sequentie van oplosactiviteiten is vastgelegd. In dit opzicht heeft strategische kennis een metacognitief karakter.

De eerste drie soorten kennis zijn gebonden aan een specifiek onderdeel van het domein, terwijl strategiekennis algemener is, geldig voor het gehele vakgebied, of zelfs voor een aantal aanverwante gebieden. Situationele en declaratieve kennis vormen de 'statische' aspecten van kennis (traditioneel aangeduid met kennis), terwijl procedurele en strategische kennis een 'dynamisch' karakter hebben (traditioneel eerder aangeduid met de term vaardigheden).

Een meer gedetailleerde indeling van kennistypen voor het oplossen van natuurkundeproblemen zou gemaakt kunnen worden. In ons onderzoek hebben we ons echter beperkt tot de hier genoemde vier hoofdcategorieën.

Kwaliteiten van kennis

Mogelijk succes bij het oplossen van problemen wordt niet alleen bepaald door de *aanwezigheid* van verschillende kennistypen in de kennisbasis van de probleemoplosser, maar ook door de *kwaliteiten* van die kennis. Typen en kwaliteiten worden door ons als onafhankelijke dimensies van de kennisbasis gezien, typen kennis geven de functie van kenniselementen in een criteriumtaak aan, kwaliteiten duiden op bepaalde eigenschappen. In de (recente) literatuur over leren en probleemoplossen wordt aan de kennisbasis een groot aantal kwaliteiten toegeschreven. Een selectie van bestaande termen is bijvoorbeeld: algemene en domeinspecifieke kennis, concrete en abstracte kennis, impliciete en expliciete kennis, formele en informele kennis, meta-kennis, regulatieve en metacognitieve kennis, en diepte van kennis. Alexander, Schallert, en Hare (1991) geven een uitgebreid overzicht van dit soort kwaliteiten, maar ook hun overzicht is zeker niet compleet. In De Jong en Ferguson-Hessler (1993) behandelen we een aantal van de belangrijkste kwaliteiten van de kennisbasis in de context van probleemoplossen: het niveau (oppervlakkig of diep) van kennis, domein gerelateerde of algemene kennis, geautomatiseerde (gecompileerde) kennis en niet geautomatiseerde kennis, impliciete en expliciete kennis, modaliteit van kennis, en structuur (of organisatie) van kennis.

EMPIRISCH ONDERZOEK NAAR DE KENNISBASIS

Veel van het onderzoek naar de inrichting van de kennisbasis werd en wordt verricht volgens het zogenaamde expert-novice paradigma. Dit onderzoek heeft belangrijke gegevens opgeleverd over de aard van expertise, gegevens die samengevat worden weergegeven in bijvoorbeeld Larkin, McDermott, Simon, & Simon (1980), Glaser en Chi (1988) en Elio en Scharf (1990). Het expert-novice paradigma heeft echter twee belangrijke nadelen. Het eerste nadeel is dat experts vaak als 'model' voor beginners worden gezien, terwijl echt expert gedrag voor beginners nogal eens net een stap te ver is. Het tweede nadeel is een praktisch gegeven. In veel onderzoek worden beginners geïntroduceerd die vrijwel onwetend zijn op een bepaald vakgebied. Op deze manier worden deze 'beginners' als een soort 'baseline' voor menselijk functioneren gebruikt.

Om deze redenen hebben we in ons eigen onderzoek goede en zwakke beginnende studenten vergeleken. De zwakke studenten waren studenten die zich wel voor een tentamen hadden voorbereid maar slechte resultaten haalden. Onderzoek naar verschillen tussen goede en zwakke presteerders is schaars (Silver, 1979; 1981), maar begint in de hoek van het exploratief leren opgang te maken (bv. Shute & Glaser, 1990). Uit de data van de goede studenten kunnen we kenniskenmerken afleiden die geassocieerd zijn met succesvol probleemoplossen. We hebben goede en zwakke studenten vergeleken op strategische kennis, op kennisorganisatie en situationele kennis.

Kennis van strategie

In een van onze eerste studies (De Jong & Ferguson-Hessler, 1983; 1984) hebben we gekeken naar de strategische kennis van studenten door hen te observeren terwijl zij problemen oplosten tijdens een tentamen Electriciteit en Magnetisme. De studenten (n = 16) werd geïnstrueerd hardop te denken terwijl ze de problemen van een toets oplosten. Hun uitingen werden op de band opgenomen.

In dit onderzoek introduceerden we ook een trainingsprogramma dat bedoeld was om het strategisch gedrag van de studenten te beïnvloeden. In het trainingsprogramma werd de studenten geleerd problemen op te lossen door het achtereenvolgens uitvoeren van vijf probleemoplosstappen die geïnspireerd waren op het werk van Mettes, Pilot, en Roossink (1981) en Reif, Larkin, en Brackett (1976): analyse, selecteren van kernbetrekkingen, het maken van een probleemoplosroute, uitwerking en controle.

Uit de protocollen van de studenten werd hun probleemoplosstrategie afgeleid. Dit werd gedaan voor een groep studenten die het trainingsprogramma volgde en voor een controle groep, zowel voor als na de training. Resultaten wezen uit dat voor het trainingsprogramma zowel de experimentele als de controle groep een strategie gebruikten die we als de 'kick and rush' strategie hebben gekenmerkt. Deze strategie houdt kort gezegd in dat studenten na een korte (onvolledige) analyse van het probleem vrij snel een formule (kernbetrekking) kiezen en gaan rekenen. Studenten uit de controle groep bleken deze strategie toe te passen op beide momenten van meten (voor en na het trainingsprogramma van de experimentele groep), maar ook studenten uit de experimentele groep pasten hun strategie niet aan onder invloed van de training. Bovendien, een vergelijking tussen succesvol en niet succesvol opgeloste problemen liet zien dat de kick and rush strategie in beide gevallen werd toegepast.

Wat duidelijk werd uit een analyse van de protocollen van succesvol opgeloste problemen was dat de probleemoplosstap 'het maken van een oplosroute' vaak vrij laat in het oplosproces voorkwam. Dit wijst er opdat studenten in eerste instantie vertrouwden op de 'kick and rush' strategie en alleen een meer planmatige aanpak toepasten als ze er met de kick and rush strategie niet uitkwamen. Dit betekent dat beginnende studenten kennis hebben van verschillende strategieën maar dat ze niet in staat zijn zoals experts (zie Chi, Glaser, & Rees, 1982) hun keuze van strategie aan te passen aan de complexiteit van het voorliggende probleem.

In dit onderzoek vonden we ook dat alle studenten een vrij groot aantal fouten in hun situationele kennis hadden. Studenten selecteerden vaak de verkeerde declaratieve en procedu-

rele kennis uit hun geheugen waar we de duidelijk indruk hadden dat ze wel de beschikking hadden over de juiste kennis.

Kennisorganisatie

De resultaten van ons onderzoek naar kennis en gebruik van strategie leidde er toe om meer aandacht te gaan geven aan de drie andere kennissoorten benodigd voor het oplossen van problemen en meer specifiek aan de organisatie en structuur van deze kennistypen. Een deeldomein van Electriciteit en Magnetisme werd herschreven in termen van de kennistypen 'situationele kennis', 'declaratieve kennis', en 'procedurele kennis'. Er werd ook een ideale organisatie gemaakt, gebaseerd op de schematheorie van Rumelhart (1980) en Chi, Feltovich, en Glaser (1981). Deze organisatie verdeelde het domein in zogenaamde probleemschemata. Elk schema bestond uit zowel situationele, declaratieve, als procedurele kennis en was relevant voor één type problemen. Een uitgebreide beschrijving van de verschillende kennistypen en probleemschemata is gegeven in De Jong en Ferguson-Hessler (1986), een integratie van deze schemata in een meer uitgebreid overzicht van het domein (E&M) kan gevonden worden in Ferguson-Hessler en De Jong (1987).

In een experiment ontvingen studenten ($n = 47$) elementen uit het domein op kleine kaartjes. Zij moesten deze kaartjes, die in een willekeurige volgorde lagen, sorteren in groepen. Er waren 65 kaartjes waarin in totaal 12 probleemschemata waren 'verborgen'. Figuur 2 geeft een voorbeeld van één schema samengesteld uit vijf kaartjes, zoals die aan de studenten werden gepresenteerd. Studenten ontvingen de opdracht "de kaartjes in stapeltjes te sorteren op zo'n manier dat een stapeltje kaartjes bevatte die bij elkaar horen". Theoretisch kon het aantal stapeltjes dat een student maakte dus variëren van 1 tot 65. Het kaart sorteren werd uitgevoerd kort nadat de studenten een tentamen over dit onderwerp hadden afgelegd.

---

1. Het veld van een halfoneindig lange, rechte draad, dragende een lijnlading
2. De definitie van electrische veldsterkte
3. Superpositieprincipe
4. $d\boldsymbol{F} = \dfrac{Q\mathbf{e}}{4\pi\varepsilon_0 r^2}\, dq$
5. Het vectorieel samenstellen van de bijdragen tot de electrische veldsterkte van verschillende elementen.

Fig. 2. Voorbeeld van een probleemschema (1 is situationele kennis, 2, 3 en 4 declaratieve kennis en 5 procedurele kennis).

Voor elke student berekenden we een score die een indicatie gaf van de overeenkomst van zijn of haar score met de ideale sortering in 12 probleemtypen. De correlatie tussen deze score en de score op het tentamen was .40 ($p < .01$). Een vergelijking tussen een groep studenten ($n = 13$) die hoog op het tentamen hadden gescoord en een groep die laag had gescoord ($n = 7$) werd gemaakt door een cluster analyse over de gesommeerde matrix van overeenkomsten. Deze clusteranalyses wezen op een sortering volgens de theoretische probleemschemata voor de goede studenten en een clustering volgens 'oppervlakkige' kenmerken voor de groep zwakke studenten.

Situationele kennis

Een belangrijk aspect van probleem oplossen is de selectie van de juiste probleemoploskennis op het juiste moment (zie bijvoorbeeld Perfetto, Bransford, & Franks, 1983). Ook uit de boven

beschreven studies bleek dat het selecteren van de juiste kennis voor studenten een probleem vormde en dat zwakke studenten situationele kennis niet gekoppeld hadden aan declaratieve en procedurele kennis. Elementen uit de aangeboden situatie (bv. een natuurkundeprobleem) kunnen de aanleiding vormen tot de selectie van probleemoploskennis (i.e. declaratieve en procedurele kennis) uit het geheugen. Deze 'cue' elementen zijn vaak niet in de situatie als zodanig gegeven, maar moeten er uit worden afgeleid (dit is wat Chi, Feltovich, & Glaser (1981) 'second order features' noemen). Bovendien, kunnen oplossers modellen van situaties nodig hebben in plaats van geïsoleerde kenmerken om tot een zinvolle kwalitatieve analyse te komen.

We ontwikkelden een methode voor het afleiden van in het geheugen aanwezige modellen van probleemsituaties (zie ook De Jong & Ferguson-Hessler, 1991). Proefpersonen (goede en zwakke beginners) kregen heel kort beschrijvingen van natuurkunde problemen te zien en na elke keer dat ze een probleembeschrijving gezien hadden werd hen gevraagd een reconstructie van het gepresenteerde probleem te geven. De korte 'inspectietijd' dwingt de studenten ertoe om modellen uit het geheugen te gebruiken in plaats van hun directe herinnering te gebruiken voor het geven van de reconstructie. De presentatie van situaties en de reconstructie zoals gevraagd varieerden in modaliteit (woorden, figuren of combinaties hiervan). Bij een aantal situaties werd de proefpersonen, nadat ze een reconstructie hadden gegeven, gevraagd om informatie die zij nodig dachten te hebben voor de oplossing op te schrijven.

De resultaten lieten zien dat alle studenten belangrijke informatie (componenten en relaties) beter reconstrueerden dan minder belangrijke informatie (uitwerkingen, detailinformatie), wat een aanwijzing is voor het beschikbaar hebben van modellen van situaties bij zowel goede als zwakke studenten. Een tweede aanwijzing hiervoor werd gevonden in het feit dat beide groepen een aangeboden 'onzinprobleem' veel slechter reconstrueerden dan de andere, wel zinvolle, problemen. Er waren echter ook verschillen tussen de twee groepen. Op de eerste plaats gaven de goede studenten een betere reconstructie van de vraag uit de probleemstelling dan de zwakke studenten. Op de tweede plaats was er een tendens dat goede studenten belangrijke informatie beter reconstrueerden dan zwakke wanneer er tussen aanbieding en reconstructie van modaliteit moest worden gewisseld. Deze interactie (zie Figuur 3) kan er een aanwijzing voor zijn dat

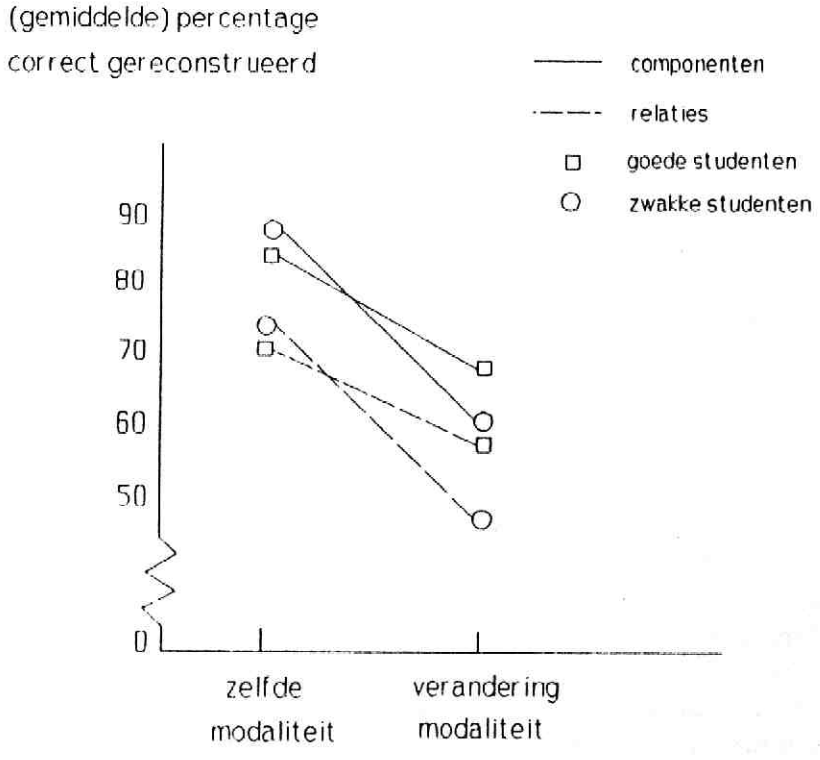


Fig. 3. Interactie-effect voor modaliteitsverandering en studentnivo voor componenten en relaties.

goede studenten 'diepere' kennis van probleemsituaties hebben dan zwakke studenten (waarmee bedoeld wordt dat ze in staat zijn aangeboden kenmerken van de situatie te vertalen naar natuurkundige concepten). Zwakke studenten lijken dit te missen. Tot slot waren goede studenten veel beter dan zwakke studenten in staat om informatie relevant voor de oplossing van het probleem te genereren. Echter, evenals de zwakke studenten genereerden de goede studenten ook veel informatie die irrelevant voor de oplossing was.

Samenvatting van onderzoek naar de kennisbasis

Het is duidelijk dat ons onderzoek verschillen in de kennisbasis tussen goede en zwakke beginnende studenten heeft kunnen aantonen. Deze verschillen zijn niet spectaculair, maar daarbij moet bedacht worden dat onze studenten een vrij coherente populatie vormen waarin niet echt grote verschillen verwacht mogen worden. Studenten lijken niet veel te verschillen in de overall probleemoplosstrategie, maar meer in de beschikbaarheid en kwaliteit van de overige drie onderscheiden kennistypen. Ondanks het feit dat met hun eigen ('kick and rush') strategie problemen succesvol opgelost blijken te kunnen worden, is deze strategie vanuit een normatief standpunt niet erg efficiënt. Verschillen tussen goede en zwakke studenten lijken te bestaan in de organisatie van kennis: Goede studenten hebben een probleem georiënteerde organisatie die zwakke studenten ontberen (zwakke studenten organiseren kennis op basis van uiterlijke kenmerken). Verder zijn er verschillen in het niveau van kennis van situaties, waarbij deze situationele kennis bij de goede studenten een dieper karakter heeft en meer is toegespitst op de oplossing van een probleem (waarbij ook bij goede studenten nog veel irrelevante kennis aan een situatie gekoppeld is). Deze laatste twee kenmerken van de kennisbasis zijn terug te vinden in het probleemoplosproces waar de zwakke studenten de juiste informatie voor het oplossen van een probleem niet kunnen vinden, terwijl de goede studenten daar, vaak na een aantal pogingen, wel in slagen.

LEREN VAN NATUURKUNDE

Als er meer inzicht bestaat in de samenstelling van de kennisbasis van goede en zwakke beginnende studenten is een logisch volgende stap te onderzoeken op welke wijze de kennisbasis wordt opgebouwd, met andere woorden onderzoek naar het *studieproces*. Als algemene trend is er momenteel ook een verschuiving in het onderzoek te zien naar het in kaart brengen van leerprocessen in plaats van probleemoplosprocessen (Chi & Bassok, 1989, De Jong & Njoo, 1992).

In een onderzoek (zie ook Ferguson-Hessler & De Jong, 1990) hebben we studenten geobserveerd die een tekst bestudeerden met als onderwerp de Aston massaspectrometer. De tekst was tien pagina's lang en bevatte vijf oefeningen. Proefpersonen waren 21 eerste jaars natuurkunde studenten die de tekst bestudeerden, terwijl ze na regelmatige intervallen verbaal rapporteerden over wat ze gedaan hadden bij de bestudering van het voorafgaande stuk tekst. De verbalisaties van de studenten werden op de band opgenomen en de protocollen van vijf goede en vijf zwakke studenten werden geanalyseerd. Dit betekende dat elke uiting geclassificeerd werd in één van 32 verschillende leerprocessen. Bovendien werd telkens bepaald op wat voor type informatie (declaratief, procedureel of situationeel) een leerproces betrekking had.

De 32 leerprocessen werden geclassificeerd in de hoofdcategorieën 'Oppervlakkige verwerking', 'Integreren', en 'Verbinden'. Een 'Rest' categorie werd toegevoegd om een complete scoring mogelijk te maken. Een leerproces dat valt onder de categorie Oppervlakkige verwerking is bijvoorbeeld 'Oppervlakkige controle', het vergelijken van symbolen in een figuur met symbolen in de tekst. 'Integreren' bevat leerprocessen die structuur geven aan de nieuw gepresenteerde informatie zoals het 'Onderscheiden van hoofd- en bijzaken'. 'Verbinden' is het associëren van nieuwe informatie met bestaande kennis zoals het 'Bedenken van voorbeelden'. De Rest categorie bevat een mengeling van processen zoals 'Het geven van een waardeoordeel' (bijvoorbeeld "dit is interessant").

De resultaten lieten zien dat goede en zwakke studenten niet verschilden in het aantal leerprocessen dat zij toepasten, wat een indicatie is voor het feit dat beide groepen de stof even actief benaderden. De twee groepen verschilden echter wel in het type leerprocessen dat zij toepasten. Leerprocessen die een 'diep' karakter hebben zoals 'Integreren op basis van eigen activiteiten', worden frequenter gebruikt door de goede studenten terwijl, aan de andere kant, de zwakke studenten meer gebruik maken van 'oppervlakkige' processen. Zwakke studenten hebben de neiging om meer aandacht te besteden aan declaratieve kennis, terwijl de goede studenten meer aandacht (dan de zwakke) besteedden aan procedurele en situationele kennis (voor situationele kennis was dit verschil niet significant). Een van de opvallende resultaten van het onderzoek was dat de opmerking "Alles is duidelijk" veel vaker werd gemaakt door de zwakke studenten dan door de goede. Dit laatste is een indicatie dat zwakke studenten een gebrek aan inzicht in hun eigen kennis hebben (meta-kennis).

Een analyse van de processen die optraden bij het oplossen van de in de tekst opgenomen oefenproblemen (zie De Jong & Ferguson-Hessler, 1989) gaf soortgelijke resultaten. Goede studenten scoorden hoger dan zwakke studenten op processen die als 'diep' kunnen worden gekenmerkt (zoals 'Geven van een verwachting over de uitkomst', of 'Relateren van de uitkomst aan de inhoud van de tekst'). Bij het oplossen van de oefenproblemen besteedden goede en zwakke studenten even veel aandacht aan declaratieve kennis (wat de meest frequente categorie was), maar goede studenten besteedden bovendien aandacht aan situationele en procedurele informatie en aan de relatie tussen de verschillende typen informatie.

Studenten in dit experiment mochten tijdens het studeren een aantekeningenblaadje prepareren dat ze later bij de toets mochten gebruiken. Deze situatie is gelijk aan die uit hun normale curriculum waar ze ook op het tentamen een zelf gemaakt aantekeningenblaadje mogen gebruiken. Een analyse van de aantekeningenblaadjes liet zien dat declaratieve informatie op deze blaadjes het meest frequent voorkomt, procedurele informatie komt nauwelijks voor (maar is dan ook moeilijker op te schrijven). Verder bleek dat goede studenten meer situationele kennis noteerden dan zwakke studenten (zie De Jong & Ferguson-Hessler, 1987).

ONDERWIJS IN DE NATUURKUNDE

Een laatste stap in het onderzoek naar probleemoplossen, leren en onderwijzen in de natuurkunde is het in kaart brengen van het onderwijsproces. In dit kader zijn we gestart met een descriptieve studie naar het onderwijsgedrag van natuurkunde docenten (Ferguson-Hessler & De Jong, 1993a). Met gebruikmaking van het in ons onderzoek naar de kennisbasis en het studiegedrag ontwikkeld begrippenkader is een twee-dimensioneel classificatieschema voor docentgedrag opgesteld. Een van de dimensies was het informatietype dat onderwerp was van de activiteit van de docent en de andere dimensie het onderwijsproces dat in de activiteit (instructiemaatregel) tot uitdrukking kwam. We onderscheidden de volgende processen:

- Creëren van voorwaarden voor leren
- Presenteren van nieuwe informatie
- Stimuleren van de integratie van informatie
- Stimuleren van het verbinden van nieuwe informatie met bestaande kennis.

De hoofdcategorieën omvatten in totaal 17 instructieprocessen. Vier colleges en vijf instructies (instructies zijn oefeningen in het oplossen van problemen, gegeven in groepen van 25-30 studenten door een instructeur) over een onderdeel uit Electriciteit en Magnetisme werden geobserveerd en opgenomen. Met behulp van het classificatieschema werd elke docentactiviteit geplaatst op de proces- en informatie/kennisdimensie.

De resultaten brachten significante verschillen tussen individuele docenten en ook tussen colleges en instructies aan het licht. De algemene trend is echter dat docenten hun aandacht verdelen over alle geïdentificeerde instructieprocessen. Sommigen, echter, hebben meer aandacht voor wat we als 'diepe' processen kunnen beschouwen (bijvoorbeeld 'Aangeven van

hoofdzaken', 'Relateren, confronteren, en concluderen', en 'Relateren aan voorkennis') dan voor andere instructieprocessen.

In de colleges vinden we meer processen die betrekking hebben op het presenteren van nieuwe informatie dan in instructies; de processen die betrekking hebben op het stimuleren van integratie en het verbinden van nieuwe informatie aan bestaande kennis komen frequenter voor in instructies.

Declaratieve informatie wordt vaker in colleges gepresenteerd en procedures en situationele informatie krijgen meer aandacht in instructies. Dit was, uiteraard, conform onze verwachtingen. Als we naar combinaties van typen informatie kijken dan vinden we een relatief hoge aandacht voor de combinatie van declaratieve en situationele kennis in beide onderwijsvormen. Andere combinaties komen veel minder voor. Drie van de geobserveerde docenten besteedden expliciet aandacht aan de combinatie van situationele, declaratieve en procedurele informatie (een combinatie zoals die van belang is voor het oplossen van problemen), maar bij andere docenten was relatief weinig sprake van instructieprocessen waarin deze combinatie betrokken was. Opvallend was de afwezigheid van aandacht voor het aanbrengen van strategische kennis in alle protocollen, met de uitzondering van één protocol.

Na deze observatiestudie hebben we een verandering in het curriculum aangebracht met het doel studenten te stimuleren de juiste leerprocessen toe te passen op het gehele scala van informatietypen zodat een voor probleemoplossen adequate kennisbasis zou ontstaan. Over dit onderzoek wordt uitgebreid verslag gedaan in Ferguson-Hessler (1989). Het onderzoek werd uitgevoerd over twee jaar met twee cohorten studenten. In het eerste jaar werd het normale, traditionele programma afgewerkt, in het tweede jaar ontvingen de studenten (eerstejaars natuurkundestudenten) experimenteel onderwijs voor het vak Electriciteit en Magnetisme.

In het experimentele onderwijs ontvingen de studenten een handleiding waarin een aantal onderdelen was opgenomen (Ferguson-Hessler, 1988). Ten eerste werd een introductie gegeven over 'Natuurkunde studeren', waarin met name verschillen tussen het 'studeren' in het middelbaar en hoger onderwijs aan de orde kwamen. Hier werd bijvoorbeeld uitgelegd dat het niet begrijpen van leerstof niet als falen moest worden gezien, maar opgevat moest worden als een eerste stap naar het verkrijgen van begrip. In een tweede deel van de handleiding werden de vier onderscheiden kennistypen geïntroduceerd en werd uitgelegd wat het belang van het juist organiseren van kennis is. Een derde hoofdstuk behandelde strategische aspecten bij het oplossen van natuurkundeproblemen. Tot slot bevatte de handleiding een gedetailleerde beschrijving van de leerstof waarin een belangrijk deel was gewijd aan het beschrijven van de stof in termen van 'wat', 'hoe' en 'wanneer'. Op deze manier werden declaratieve, procedurele en situationele elementen in de stof geïdentificeerd. Naast het verstrekken van de handleiding werden in het experimentele onderwijs de colleges en instructies aangepast. Docenten ontvingen richtlijnen voor hun instructies en colleges waarin naast de inhoud van deze colleges en instructies ook de manier van presenteren werd voorgeschreven. Deze presentatie betrof het belang van probleemschemata en het gebruik van een strategie voor het oplossen van problemen. Alle instructiemaatregelen in het experimentele onderwijs hadden de bedoeling de activiteit van studenten te stimuleren en dan met name het gebruik van 'diepe' leerprocessen.

Data werden op een aantal manieren verzameld. Er werd gebruik gemaakt van (proef)tentamens en speciaal ontworpen metingen voor het vaststellen van structuur, coherentie en compleetheid van de kennisbasis van studenten. Colleges en instructies werden geobserveerd en opgenomen en studenten moesten tijdschrijfformulieren bijhouden.

Ongelukkigerwijs werd het curriculum van de Faculteit Natuurkunde tussentijds gewijzigd, en dit beïnvloedde de mogelijkheid onze data te interpreteren. Het experimentele cohort haalde significant hogere cijfers voor het reguliere tentamen Electriciteit en Magnetisme dan het controle cohort maar dit zou deels toegeschreven kunnen worden aan het feit dat de experimentele groep (door de curriculumwijziging) meer tijd voor de voorbereiding van dit tentamen heeft gehad. De verdeling van de cijfers geeft aanleiding te veronderstellen dat met name de middelmatige studenten (en dus niet de zwakke) geprofiteerd hebben van de onderwijswijziging.

CONCLUSIE

In dit artikel hebben we getracht een kort overzicht te geven van ons onderzoek naar drie aspecten van onderwijs: taakuitvoering (probleemoplossen), leren en instructie. Er is een begrippenkader ontwikkeld dat ons in staat heeft gesteld de onderzoeken op deze drie terreinen op elkaar aan te laten sluiten. Voor de taakuitvoering is het belang van de kennisbasis voorop gesteld. Hierin hebben we een onderscheid gemaakt tussen de inhoud, uitgedrukt in verschillende *kennistypen*, en *kwaliteiten* van de kennisbasis. De kennistypen kunnen een meer statisch karakter hebben (situationele en declaratieve kennis) of een meer dynamisch karakter (procedurele en strategische kennis). Bij leren en instructie gaat het om het verwerven van de kennisbasis. Deze inhoud wordt verworven door het toepassen van *specifieke (leer- of instructie)processen*.

Proefpersonen in ons onderzoek waren altijd beginnende TU studenten en het vakgebied was telkens een onderdeel van het natuurkundevak Electriciteit en Magnetisme. In het onderzoek hebben we goede en zwakke studenten vergeleken waarbij 'goed' of 'zwak' bepaald werd (mede) aan de hand van tentamencijfers. Hierbij werden studenten die de laagste cijfers haalden, en zich dus kennelijk niet hadden voorbereid op het tentamen, uitgesloten. De reden om naar zwakke en goede beginnende studenten te kijken en niet naar beginners en experts is dat goede beginners er kennelijk in slagen een criterium taak (probleemoplossen op het tentamen) adequaat uit te voeren en qua kennisniveau dichter bij zwakke beginners staan dan experts. We gaan er daarom vanuit dat we voor het inrichten van onderwijs meer kunnen leren van goede beginners dan van experts.

Onze resultaten wijzen, zeer globaal samengevat, erop dat goede beginners (in vergelijking met zwakke beginners) een 'diepere' benadering van het vak laten zien. Dit komt zowel tot uiting in leerprocessen (meer reconstructie van aangeboden informatie) als in de resulterende kennisbasis (modelmatige kennis die niet sterk aan oppervlakkige kenmerken is gebonden). Verder is gebleken dat goede beginners in hun leerprocessen en de resulterende kennisbasis een grotere spreiding en betere verbinding over en tussen de verschillende door ons onderscheiden kennistypen hebben.

Op basis van deze resultaten zijn enkele aanwijzingen te formuleren voor onderwijs en instructie, speciaal in exacte vakken. Als actieve verwerking en diepe benadering van leerstof van essentieel belang zijn, dan dient het onderwijs dit type processen te demonstreren en stimuleren. Bij het oplossen van problemen kan dit bijvoorbeeld gebeuren door:
- het expliciet analyseren van probleemsituaties en trekken van conclusies daarover,
- het expliciet en beargumenteerd maken van keuzes ten aanzien van toe te passen declaratieve en procedurele kennis,
- het expliciet uiteenzetten van de verschillende fasen van de probleemoplosstrategie.

Het opbouwen van een goed gestructureerde kennisbasis is alleen mogelijk als de lerende enig inzicht heeft in het functioneren van de domeinkennis in typische taken, dat wil zeggen onderscheid kan maken tussen verschillende kennissoorten en relevante relaties kan leggen tussen elementen van kennis die samen functioneren. Deze meta-kennis krijgt echter meestal weinig of geen aandacht in onderwijs en leerboeken.

Ook strategieën en leerprocessen voor de verwerking van schriftelijke informatie behoeven uitleg, bijvoorbeeld door:
- het expliciet leggen van verbanden door vergelijkingen van begrippen, situaties, procedures en formules,
- het expliciet trekken van conclusies om zo nieuwe kennis te construeren,
- het opmerken van (schijnbare) tegenstellingen,
- het demonstreren van de opbouw van een probleemschema.

Een voorwaarde voor diepe verwerking van leerstof is dat de voorkennis van leerlingen/studenten voldoende is, geactiveerd wordt, en geen fundamenteel incorrecte begrippen of misconcepties bevat (Ferguson-Hessler & De Jong, 1993b; Taconis & Ferguson-Hessler, 1993).

Ons onderzoek is zeer beperkt van aard. We hebben ons beperkt tot één soort problemen (E&M problemen) en, wat de kennisbasis betreft, een beperkt aantal kwaliteiten. Het algemene raamwerk van typen en kwaliteiten kennis heeft, naar onze mening, een wijdere strekking en kan worden toegepast op problemen van uiteenlopende soort uit een verscheidenheid van domeinen. In De Jong en Ferguson-Hessler (1993) geven we een overzicht waarin de twee dimensies van de kennisbasis (kennistypen en kwaliteiten) in een matrix worden gezet. Deze matrix kan gebruikt worden om voor elke combinatie van kennistype en kwaliteit aspecten van onderzoek te identificeren. Dit kunnen onderzoektechnieken, leerprocessen of instructieprocessen zijn. Het raamwerk biedt ook de basis voor een gestructureerde beschrijving van de huidige trends in het onderzoek. Deze trends zijn het onderzoeken van steeds complexere problemen (bv. ontwerpproblemen, Goel & Pirolli, 1992), van sociale aspecten van leren en probleemoplossen (Brown & Palinscar, 1989), van de rol van interesse en motivatie (Snow, 1989; Hidi, 1990) en van exploratieve (constructivistische) leerprocessen (Njoo & De Jong, 1993).

LITERATUUR

Alexander, P.A., & Judy, J.E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*, 375-405.

Alexander, P.A., Schallert, D.L., & Hare, V.C. (1991). Coming to terms: how researchers in learning and literacy talk about knowledge. *Review of Educational Research 61*, 315-343.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.

Anzai, Y., & Yokoyama, T. (1984). Internal models in physics problem solving. *Cognition and Instruction, 1*, 397-450.

Brown, A.L., & Palinscar, A.S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L.B. Resnick (Ed.) *Knowing, Learning, and Instruction. Essays in Honor of Robert Glaser* (pp 393-451). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Braune, R., & Foshay, W.R. (1983). Towards a practical model of cognitive/information processing task analysis and schema acquisition for complex problem-solving situations. *Instructional Science, 12*, 121-145.

Chi, M.T.H., & Bassok, M. (1989). Learning from examples via self-explanations. In L.B. Resnick (Ed.) *Knowing, Learning, and Instruction. Essays in Honor of Robert Glaser* (pp. 251-282). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science 13*, 145-182.

Chi, M.T., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M.T., Glaser, R., & Rees, E. (1982). Expertise in problem-solving. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-77). Hillsdale (NJ): Lawrence Erlbaum Ass.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser.* Hillsdale, NJ: Lawrence Erlbaum.

Elio, R., & Scharf, P.B. (1990). Modelling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science, 14*, 579-639.

Ferguson-Hessler, M.G.M. (1988). *Studiehandleiding voor eerstejaars natuurkundestudenten.* Eindhoven: TUE, dictaat.

Ferguson-Hessler, M.G.M. (1989). *Over kennis en kunde in de fysica*. dissertatie, TU Eindhoven.

Ferguson-Hessler, M.G.M. (1993). Meta-knowledge and the development of expertise in complex domains. *Tijdschrift voor Onderwijsresearch* (this issue).

Ferguson-Hessler, M.G.M. & De Jong, T. (1986). On the role of knowledge structures in the field of electricity and magnetism. In: A. Art. H. Eisendrath, & F. Greandjean (Eds.) *Proceedings of the International Symposium on Physics Teaching* (pp. 183-186). Brussels, november 11-13, 1985.

Ferguson-Hessler, M.G.M., & De Jong, T. (1987). On the quality of knowledge in the field of electricity and magnetism. *American Journal of Physics, 55*, 492-497.

Ferguson-Hessler, M.G.M., & De Jong, T. (1989). *Onderwijs in de natuurkunde op basis van een abstracte beschrijving van de vakinhoud.* Paper gepresenteerd op ORD89.

Ferguson-Hessler, M.G.M., & De Jong, T. (1990). Studying physics text. Differences in study processes between good and poor performers. *Cognition and Instruction*, 7, 41-54.

Ferguson-Hessler, M.G.M., & De Jong, T. (1993a). Does physics instruction foster students' cognitive processes? A descriptive study of teacher activities. *Journal of Research in Science Teaching* (in press).

Ferguson-Hessler, M.G.M., & De Jong, T. (1993b). Het leren van exacte vakken. In: P. Span & Tomic (Eds.) *Onderwijspsychologie*. Deventer: Kluwer. (In opdracht van de Open Universiteit). (in press).

Glaser, R., & Chi, M.T.H. (1988). Overview. In: M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.) *The nature of expertise* (pp. xv-xxix). Hillsdale, NJ: Lawrence Erlbaum.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science, 16*, 395-429.

Greeno, J.G. (1978). Understanding and procedural knowledge in mathematics education. *Educational Psychologist, 12*, 262-283.

Hidi, S. (1990). Interest and its contribution as a mental resource. *Review of Educational Research 60*, 549-571.

De Jong, T. (1986). *Kennis en het oplossen van vakinhoudelijke problemen. Een voorbeeld uit een natuurkundig domein*, dissertatie, TU Eindhoven.

De Jong, T. (1991). Learning and instruction with computer simulations. *Education & Computing, 6*, 217-229.

De Jong, T., & Ferguson-Hessler, M.G.M. (1983). Het effect van een instructie van een strategie voor het oplossen van natuurkundige problemen door eerstejaarsstudenten. In J.J. Beishuizen, C. Hamaker, B. van Hout-Wolters & K.B. Koster (Eds.), *Onderwijsleerprocessen: tekstverwerking, probleemoplossen en leermoeilijkheden* (pp. 93-103). Lisse: Swets en Zeitlinger.

De Jong, T., & Ferguson-Hessler, M.G.M. (1984). Strategiegebruik bij het oplossen van problemen in een semantisch rijk domein: Electriciteit en Magnetisme. *Tijdschrift voor Onderwijsresearch, 9,* (1), 3-16.

De Jong. T., & Ferguson-Hessler, M.G.M. (1985). De cognitieve structuur van wel en niet succesvolle beginnende probleemoplossers in de natuurkunde. In P.R.J. Simons & G.C.L. Lodewijks (Eds.), *Zelfstandig leren* (pp. 183-193). Lisse: Swets en Zeitlinger.

De Jong, T., & Ferguson-Hessler, M.G.M. (1986). Cognitive structures of good and poor novice problem solvers in physics. *Journal of Educational Psychology, 78,* 279-288.

De Jong, T., & Ferguson-Hessler, M.G.M. (1987). Student note-taking behavior when preparing for an examination in science. In P.R.J. Simons & G. Beukhof (Eds.), *Regulation of learning* (pp. 107-119). The Hague: SVO-Selecta.

De Jong, T., & Ferguson-Hessler, M.G.M. (1988). The role of exercises in knowledge acquisition in a physics domain. In: Span, P., De Corte E., & van Hout-Wolters, B. (Eds.). *Onderwijsleerprocessen: Strategieën bij het verwerken van informatie.* Lisse: Swets & Zeitlinger.

De Jong, T., & Ferguson-Hessler, M.G.M. (1991). Knowledge of problem situations in physics by good and poor novice problem solvers. *Learning and Instruction*, 1, 289-302.

De Jong, T., & Ferguson-Hessler, M.G.M. (1993). Types and qualities of knowledge. *Paper presented at the AERA meeting, Atlanta.*

De Jong, T., & Njoo, M. (1992). Learning and Instruction with computer simulations: learning processes involved. In E. de Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.) *Computer-based learning environments and problem solving* (NATO ASI series F: Computer and Systems Series) (pp. 411-429). Berlin: Springer.

Larkin, J.H., MacDermott, J., Simon, D.P., & Simon H.A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.

Mettes, C.T.C.W., Pilot, A., & Roossink, H.J. (1981). Linking factual and procedural knowledge in solving science problems: a case study in a thermodynamics course. *Instructional Science, 10*, 333-36

Newell, A., & Simon, H.A. (1972). Human problem solving. Englewood Cliffs: Prentice Hall.

Njoo, M., & De Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, (in press).

Perfetto, G.A., Bransford, J.D., & Franks, J.J. (1983). Constraints on access in a problem solving context. *Memory and Cognition, 11*, 24-31

Reif, F., Larkin, J.H., & Brackett, G.C. (1976). Teaching general learning and problem-solving skills. *American Journal of Physics, 44*, 212-217.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-59). Hillsdale: Lawrence Erlbaum Ass.

Shute, V.J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments, 1*, 51-77.

Silver, E.A. (1979). Student perceptions of relatedness among mathematical verbal problems. *Journal for Research in Mathematics Education, 10*, 195-210.

Silver, E.A. (1981). Recall of mathematical problem information: Solving related problems. *Journal for Research in Mathematics Education, 12*, 54-64.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, dec '89*, 8-14.

Taconis, R., & Ferguson-Hessler, M.G.M. (1993). *A problem solving perspective on misconceptions in mechanics.* Poster presentation 1993 AERA Annual Meeting, Atlanta.

Manuscript ontvangen 17-3-1993
Definitieve versie ontvangen 8-6-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 3, pp. 163-174

Kennisstructuur en hypothesevorming; verschillen tussen beginners en experts in de geneeskunde[1]

H.P.A. Boshuizen, P.P.M. Hobus, E.J.F.M. Custers en H.G. Schmidt
*Rijksuniversiteit Limburg, Vakgroep Onderwijsontwikkeling en Onderwijsresearch**

ABSTRACT

This article discusses the structural changes in a person's medical knowledge base resulting from theoretical education and enduring practical experience. It describes the processes taking place resulting in a transition from network-like knowledge structures, requiring explicit reasoning, toward a script organisation that enables quick and flexible knowledge application in medical problem solving.
Three experiments are presented that illustrate these structural changes. Theoretical and educational implications of these findings are discussed.

INLEIDING

Het doel van medisch onderwijs is dat studenten grote hoeveelheden flexibele medische kennis verwerven met behulp waarvan bij patiënten aandoeningen kunnen worden gediagnostiseerd en zo mogelijk behandeld. In principe kan een dergelijk doel op velerlei wijzen bereikt worden. Men zou bijvoorbeeld kunnen proberen de studenten deze klinische kennis over diagnose en behandeling rechtstreeks, zonder speciale omwegen, bij te brengen. De meeste medische faculteiten in Nederland en daarbuiten verkiezen echter een andere, meer indirecte methode. Veelal wordt begonnen met wat beschouwd wordt als de 'basiswetenschappen' zoals anatomie en fysiologie. Deze zijn bedoeld om de studenten inzicht te geven in de normale structuur en funktie van (delen van) het menselijk lichaam en in de wijze waarop deze structuur en funkties gedurende een mensenleven veranderen. Vervolgens leren studenten hoe deze normale funkties verstoord kunnen worden, bij voorbeeld door toedoen van externe invloeden zoals bacteria, virussen of toxische stoffen, door degeneratieve processen, of door nieuwvormingen (de algemene pathologie). Zij leren hoe deze processen op het lichaam inwerken en hoe het lichaam zich daartegen verweert en probeert het evenwicht te herstellen. Studenten hebben een aantal jaar nodig voordat zij deze basisvakken beheersen. In een basiscurriculum van vier jaar zal globaal gesproken twee jaar aan deze onderwerpen worden besteed. De algemene kennis over de normale structuur en funktie, en algemene pathologie die in deze periode is verworven, vormt vervolgens het fundament voor de klinische vakken die daarna volgen. De klinische vakken betreffen specifieke aandoeningen, hun symptomatologie, pathofysiologie, epidemiologie, prognose en behandeling. In een dergelijk curriculum kan een student pas ervaring in de praktijk gaan opdoen wanneer deze (vermoedelijk) in voldoende mate de klinische kennis bezit die noodzakelijk is om aandoeningen bij patiënten te kunnen diagnostiseren en behandelen. Dan pas breekt de tijd aan dat de studenten hun kennis in de praktijk kunnen (leren) toepassen.

DE ONTWIKKELING VAN KENNISSTRUCTUREN

De ontwikkeling van de medische kennisstructuur weerspiegelt de standaard curriculumaanpak zoals die hierboven is beschreven. Uiteraard zal een drastisch andere curriculumaanpak tot een andere ontwikkeling van de kennis aanleiding geven. In de standaardsituatie echter ontwikkelen

* Adres: Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Postbus 616, 6200 MD Maastricht.

studenten eerst uitgebreide, geïntegreerde kennisnetwerken die na een aantal jaren studie bestaan uit vele duizenden concepten die ontleend zijn aan de basisvakken. Deze concepten zijn ondergebracht in conceptuele netwerken. Deze netwerken stellen de student instaat klinische teksten te begrijpen en over problemen te redeneren. Door op deze wijze de kennis actief toe te passen, groeit het aantal relaties tussen de concepten, worden de relaties verder versterkt, en neemt de coherentie toe. Bovendien vindt herstructurering van kennis plaats (Rumelhart & Norman, 1972). Boshuizen en Schmidt (1992) hebben in dit verband laten zien dat de herhaalde toepassing van biomedische kennis tijdens de vroege stadia van expertiseontwikkeling ertoe leidt dat deel-netwerken bestaande uit lagere-orde, detail concepten worden samengevat onder een kleiner aantal hogere-orde, vaak klinische concepten. De auteurs noemen dit proces "encapsulatie". Zij veronderstellen dat dit proces resulteert in gemakkelijk toegankelijke en flexibele kennisstructuren met korte zoekpaden (zie ook Schmidt en Boshuizen, in dit nummer). Deze hypothese is in overeenstemming met resultaten van Patel, Groen en Arocha (1990) die vonden dat experts die geconfronteerd worden met bekende problemen minder concepten gebruiken, die bovendien minder gedetailleerd zijn, dan experts die geconfronteerd worden met onbekende problemen. Kennis die in dergelijk routine problemen wordt toegepast, is reeds geëncapsuleerd, terwijl de kennis die toegepast wordt in casus waarmee men geen ervaring heeft (nog) niet is geëncapsuleerd. De resultaten van Boshuizen en Schmidt (1992) suggereren bovendien dat de gedetailleerde biomedische kennis die geëncapsuleerd is onder hogere-orde concepten beschikbaar blijft en weer teruggezocht kan worden wanneer dat nodig is. Zoiets kan bijvoorbeeld het geval zijn bij complexe problemen, of wanneer iemand iets moet uitleggen. Een voorbeeld van dit laatste zou het volgende kunnen zijn: In een onderzoek van Patel en Groen (1986) waarin ervaren artsen een casus diagnostiseerden, legde één van de experts (Expert #5) een directe verbinding tussen vermoedelijk druggebruik en 'endocarditis'. Wanneer deze expert ons zou moeten uitleggen waarom druggebruik tot endocarditis (een ontsteking van de binnenkant van het hart) kan leiden, dan zal deze de onderliggende kennis vermoedelijk meteen kunnen ontvouwen tot een keten van tenminste vijf proposities (zie figuur 1) en ons vertellen dat druggebruik vaak gepaard gaat met het gebruik van niet steriele naalden, waardoor bacteria in de bloedbaan kunnen worden gebracht, waardoor een sepsis en uiteindelijk een endocarditis kunnen ontstaan. Wanneer wij dan vervolgen vragen wat sepsis is, of waarom daarop een endocarditis kan volgen, dan zal de expert ook die delen van zijn kennisnetwerk verder ontvouwen en bij voorbeeld vertellen dat bij druggebruikers de hartkleppen vaak al enigszins beschadigd zijn, waardoor onregelmatigheden in de bloedstroom kunnen optreden, waardoor bacteria zich relatief gemakkelijk aan de aangetaste klep kunnen hechten, etc.

druggebruik
→ (cond) gebruik van niet steriele naalden
→ (cau) inbrengen van bacteria in de bloedbaan
→ (cau) sepsis
→ (cau) infectie van het endocard
→ (iden) *endocarditis*

Fig. 1. De hogere-orde propositie [*druggebruik* → *endocarditis*] uiteengevouwen tot een keten van vijf micro-proposities.

Het concept 'encapsulatie' vertoont overeenkomsten met Anderson's notie 'compilatie' van kennis (Anderson, 1982). Het verschilt echter ook op een aantal essentiële punten. Anderson veronderstelt bij voorbeeld dat declaratieve kennis wordt omgezet in procedurele kennis. Wij gaan hier niet van uit. In plaats daarvan benadrukken we de verkorting van de zoekpaden, doordat de kenmerken van de probleemsituatie (bij voorbeeld de sociale situatie of het uiterlijk van de patiënt) rechtstreeks gekoppeld worden aan de uitkomst van het probleemoplossingsproces (de diagnose of een actieplan). In ons voorbeeld wordt druggebruik gekoppeld aan endocarditis. Dit verkortingsprincipe is ook bij compilatie zeer essentieel.

Terwijl de student vaardigheid verwerft in het klinisch redeneren, vindt gelijktijdig met kennisencapsulatie een ander proces plaats. Als studenten met echte patiënten worden geconfronteerd, wordt biomedische kennis omgezet in nieuwe structuren die vervolgens gebruikt zullen gaan worden om de toestand van een patiënt te begrijpen (Schmidt, Norman, & Boshuizen, 1990). Deze structuren noemen we 'ziektescripts', een vertaling van de term 'illness scripts' die door Feltovich en Barrows (1984) werd geïntroduceerd. Ziektescripts bevatten de klinische relevante kennis over een bepaalde aandoening, de signalen en symptomen ('Consequences') en de omstandigheden waaronder die bepaalde ziekte al dan niet tot ontwikkeling komt ('Enabling Conditions'). Verder bevatten ziektescripts enige noodzakelijke kennis omtrent de pathofysiologische oorzaken van een aandoening ('Fault').

Naar men aanneemt, ontwikkelen ziektescripts zich als gevolg van praktische ervaring. Daarom zullen de inhoud en structuur ervan veranderen onder invloed van toenemende praktische ervaring. Afhankelijk van de specifieke ervaring met bepaalde (klassen van) aandoeningen worden de Consequences en Enabling Conditions in een ziektescript verrijkt, nader gedetailleerd en gespecificeerd en beter afgestemd op de daagse praktijk, terwijl de geïncorporeerde causale en biomedische kennis verder wordt geëncapsuleerd.

In deze theorie wordt het oplossen van routine-problemen beschouwd als een proces van scriptactivatie, -instantiatie en -verificatie, waarbij geen sprake is van redeneren in de strikte, bewuste betekenis van het woord. Wanneer bijvoorbeeld een ervaren arts (een huisarts of een internist) met een patiënt te maken krijgt (man, 38 jaar, sinds zes jaar alcoholist, en klagend over heftige pijn boven in de buik) dan komen er vrijwel automatisch een aantal hypothesen bij hem op (Boshuizen, 1989; zie ook Elstein, Shulman, & Sprafka, 1979). Andere informatie die vervolgens binnenkomt past wel of niet in de geactiveerde scripts, en na een paar extra gegevens zal één favoriet op de voorgrond komen, terwijl de minder waarschijnlijke naar de achtergrond verdwijnen (zie ook Pauker, Gorry, Kassirer, & Schwartz, 1976). Doorgaans zal hierbij geen sprake zijn van biomedisch redeneren. Experts in de geneeskunde hoeven niet te redeneren om uiteenlopende gegevens als alcoholisme, heftige pijn in de bovenbuik, geelzucht en een verhoogd serumamylase niveau te integreren. Zij 'weten' dat zo'n patiënt vermoedelijk aan pancreatitis zal lijden. Voor studenten daarentegen is een dergelijk probleem in het geheel geen routine. Zij moeten een redenering op basis van biomedische kennis opbouwen om er een coherente representatie van te kunnen vormen.

Doorgaans worden de eerste diagnostische hypotheses van een expert al in de eerste ogenblikken van een arts-patiënt contact gevormd (Elstein, Shulman, & Sprafka, 1979; Gale & Marsden, 1982), terwijl studenten meer informatie nodig hebben (Boshuizen, 1989; Joseph & Patel, 1987). In die eerste momenten van een consult wordt vaak alleen maar de klacht naar voren gebracht. Daarnaast heeft de arts echter nog andere informatiebronnen waarmee hij terdege rekening zal houden. De wijze waarop de patiënt gekleed is, het uiterlijk, en het dossier geven informatie over geslacht, leeftijd, ras, sociale klasse, medicatie, beroep, vroegere aandoeningen, operaties, etc. van de patiënt. In combinatie met de klacht lijkt dit soort informatie enige specifieke ziektescripts te activeren, die op hun beurt weer verwachtingen oproepen over verschijnselen en symptomen die bij anamnese, lichamelijk onderzoek en laboratoriumonderzoek gevonden zullen worden.

Deze analyse suggereert dat experts voordeel hebben van de informatie die ontleend wordt aan de context waarin de klacht wordt gepresenteerd (waar en door wie) en dat ziektescripts kunnen worden geactiveerd door een combinatie van de klacht en van context informatie die betrekking heeft op Enabling Conditions[2]. Studenten hebben meer tijd en meer informatie nodig om hun hypothesen te activeren. Zij lijken niet instaat om deze context informatie ten volle te gebruiken; hun ziektescripts (als ze deze al hebben) kunnen niet gemakkelijk via deze weg worden geactiveerd.

Hobus, Schmidt, Boshuizen en Patel (1987) toetsten deze veronderstelling door aan ervaren huisartsen en net afgestudeerde basisartsen de informatie aan te bieden die reeds in de eerste momenten van een consult beschikbaar is (dit zijn de klacht en de context) en ze te vragen wat volgens hen de meest waarschijnlijke diagnose zou zijn. Bovendien gingen ze na welke infor-

matie men nog had onthouden. In dit onderzoek bleek dat experts betere diagnostische hypothesen genereerden en dat ze meer context informatie onthielden dat onervaren artsen. Deze combinatie van resultaten suggereert dat ervaren artsen meer gebruik maken van context informatie en daardoor betere hypothesen genereren, causaliteit is er echter niet mee bewezen.

Mocht er inderdaad sprake zijn van een causale relatie, dan nog kunnen meerdere processen hieraan ten grondslag liggen. Ten eerste kan het zijn dat de ziektescripts van experts rijker zijn dan die van (gevorderde) studenten; ze bevatten meer informatie omtrent de omstandigheden die ertoe leiden dat iemand een bepaalde aandoening oploopt en omtrent de symptomen van een dergelijke aandoening. Daarom hebben ze betere hypothesen en worden deze hypothesen eerder geactiveerd, zoals in andere onderzoekingen bleek (Boshuizen, 1989; Joseph & Patel, 1987). Bovendien kunnen experts context informatie hierdoor beter onthouden, omdat hun ziektescripts meer slots hebben die overeenkomen met de aangeboden informatie. Een tweede hypothese, die de bevindingen van Hobus et al. zou kunnen verklaren, is dat de ziektescripts van experts en van studenten even rijk zijn, maar dat hun scripts verschillen in toegankelijkheid. Wellicht kunnen de ziektescripts van experts al via de Enabling Conditions worden geactiveerd, terwijl dit bij minder ervaren artsen slechts kan gebeuren via de Consequences, waarvan de klacht er één is.[3] Enabling Conditions kunnen dan nog wel een rol spelen bij het beoordelen van een patiënt, maar alleen in de zin van bijkomende evidentie die een eerder gestelde hypothese meer of minder plausibel maakt. Door nu, zoals Hobus et al. deden eerst de context en dan de klacht aan te bieden zou de taak voor beginners op twee manieren worden bemoeilijkt. Ten eerste moet de context informatie gedurende langere tijd worden onthouden, totdat de klacht wordt vernomen. Dit leidt tot een belasting van het geheugen en het informatieverwerkingsproces met ongechunkte informatie. Hierdoor wordt zowel het onthouden van die informatie en het zoeken in het geheugen naar een passend script bemoeilijkt. Zowel de diagnose als de geheugenprestatie kunnen daardoor in negatieve zin worden beïnvloed.

Om deze twee alternatieve hypothesen te onderzoeken is een aantal experimenten uitgevoerd die in dit artikel zullen worden beschreven.

EXPERIMENT 1

Het eerste experiment is bedoeld om te onderzoeken of er een causale relatie bestaat tussen het gebruik van context informatie door ervaren artsen en hun superieure diagnostische prestaties. Als ervaren artsen meer gebruik maken van context informatie bij het genereren van diagnostische hypothesen, dan zal derhalve de kwaliteit van hun vroege hypothesen meer te lijden hebben van een drastische vermindering van die informatie dan die van hun onervaren collega's.

Methode

Aan het experiment werd door 16 ervaren huisartsen en 16 net afgestudeerde basisartsen deelgenomen. Om de hypothese te toetsen werden 18 casus geselekteerd uit een bestand van casusbeschrijvingen van echte patiënten met bekende diagnose. Van deze casus werden twee varianten gemaakt: één variant bevatte uitsluitende de klacht, de andere bevatte naast de klacht ook context informatie. De klacht werd gepresenteerd zoals deze oorspronkelijk door de patiënt zelf was verwoord. Een voorbeeld van een dergelijke klacht is: "Dokter, ik heb al twee dagen last van koude koorts. Ik lig soms te schudden in mijn bed." De context informatie werd overgebracht met behulp van een portret van de patiënt en diens groene kaart. Deze bevat onder meer informatie over het geslacht en de leeftijd van de patiënt, maar ook beroep, vroegere operaties, medicijngebruik, erfelijke aandoeningen, rook- en drinkgewoonten en de aantekeningen bij eerdere consulten. De 18 casus werden op dia's gepresenteerd. Bij de korte variant was dit er steeds één, die alleen de klacht bevatte; bij de lange variant waren het er steeds drie: een portret van de patiënt, diens groene kaart en de klacht. De helft van de proefpersonen kregen alleen de korte casus gepresenteerd, de andere kregen de langere, niet gereduceerde variant. De vraag aan hen was steeds de meest waarschijnlijke diagnostische hypothese te noemen.

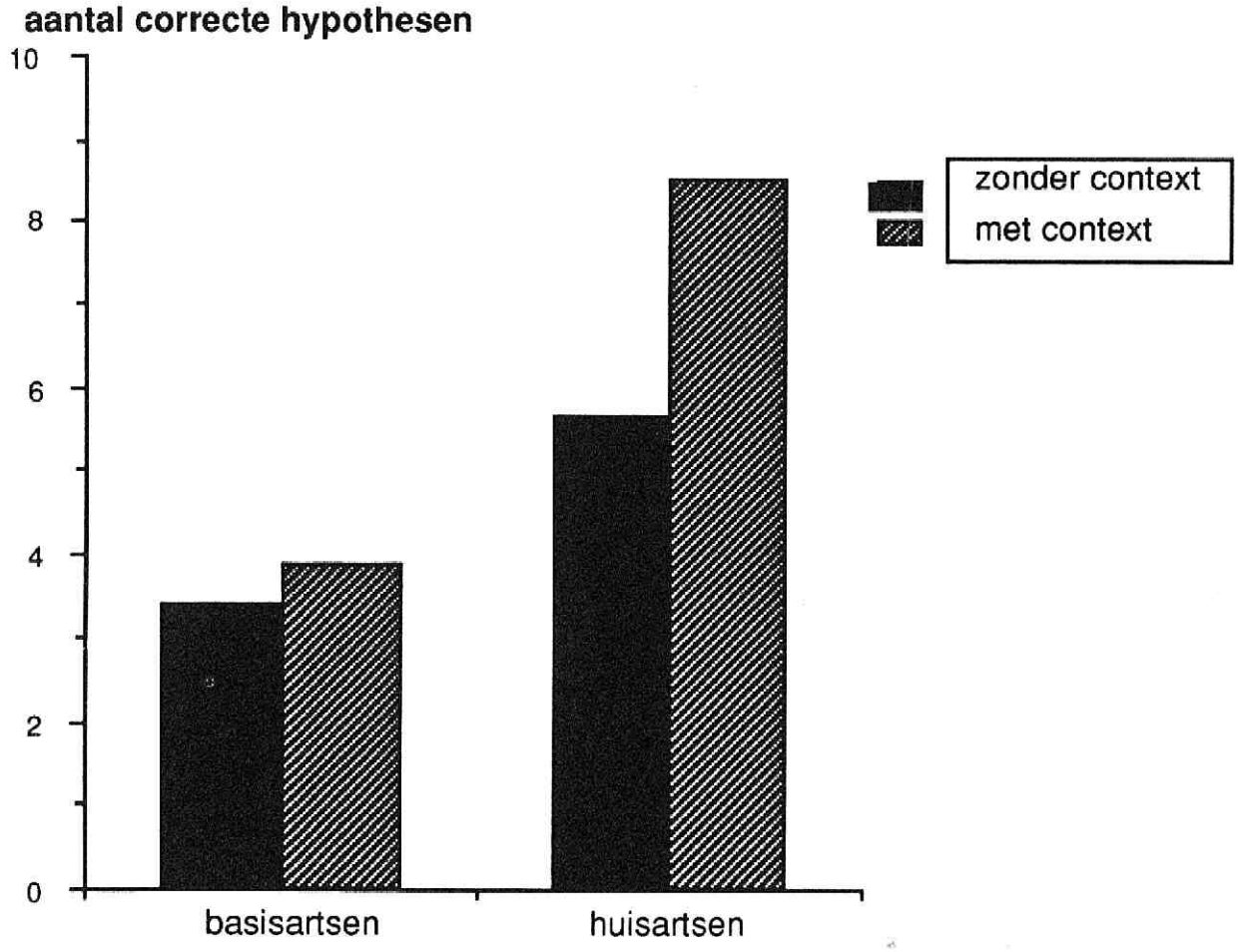


Fig. 2. Gemiddeld aantal juiste diagnostische hypothesen die gegenereerd werden door ervaren en onervaren artsen als respons op gereduceerde (zonder context) en niet gereduceerde (met context) casusinformatie.

Resultaten en discussie

Figuur 2 laat zien dat de experts meer correcte hypothesen genereerden dan de onervaren artsen ($F(1, 28) = 28.51$; $p < .001$). Dit was met name het geval wanneer zij naast de klacht ook context informatie hadden ontvangen (context: $F(1, 28) = 15.33$; $p < .001$; context * expertiseniveau: $F(1,28) = 4.56$; $p = .04$).

Deze resultaten suggereren dat onervaren artsen nauwelijks gebruik maken van context informatie in tegenstelling tot experts; presentatie van context informatie leidt ertoe dat de prestatie van experts met 50% verbetert, tegenover 12% bij de beginnende artsen. De data suggereren bovendien dat ervaren artsen betere "klachtenkennis" hebben. Zelfs wanneer zij alleen de klacht vernemen, genereren ervaren artsen betere hypothesen dan onervaren artsen. De resultaten duiden er dus op dat er bij ervaren artsen een causaal verband bestaat tussen het gebruik van context informatie en de kwaliteit van hun diagnostische hypothesen. Er kan echter nog geen uitsluitsel worden gegeven over de vraag aan welke kenmerken van de gebruikte kennisstructuren dit verschijnsel kan worden toegeschreven. Verschillen de ziektescripts van ervaren en onervaren artsen vooral in rijkdom of in toegankelijkheid? De huidige informatie kan hierin geen uitsluitsel geven.

EXPERIMENT 2

Experiment 2 is gericht op de hypothese dat met name verschillen in toegankelijkheid en activeerbaarheid van de ziektescripts de verschillen in diagnostische accuraatheid verklaren. Dit werd onderzocht door dezelfde, niet gereduceerde, casus in twee volgorden aan te bieden: eerst de klacht of eerst de context.

Als de ziektescripts van niet-experts uitsluitend door Consequences informatie kunnen worden geactiveerd, dan zal de kwaliteit van hun hypothesen en hun recall van de informatie verbeteren wanneer eerste de klacht en daarna de context wordt aangeboden. In dat geval zal de klacht namelijk één of meer ziektescripts activeren waarin de daarna volgende context informatie kan worden ingepast. Hierdoor kan én een goede selektie van het best passende script worden gedaan én de informatie blijft goed behouden. In de context-klacht volgorde zal daarentegen de context informatie gedurende enige tijd ongechunkt bewaard moeten worden, waardoor deze gemakkelijk vergeten wordt. Aan de andere kant, wanneer de ziektescripts van experts zowel door de context als door de klacht kunnen worden geactiveerd, dan zal bij deze groep de volgorde waarin context en klacht informatie wordt aangeboden geen invloed hebben op de kwaliteit van hun hypothesen, noch op de recall.

Methode

Aan Experiment 2 namen 16 ervaren en 16 onervaren artsen deel. Dezelfde 18 casus als in Experiment 1 werden gebruikt. Zij werden in twee volgorden gepresenteerd: klacht eerst of context eerst. In de context eerst conditie zagen de proefpersonen eerst het portret en de groene kaart, gevolgd door de klacht. In de klacht eerst conditie werd eerst de klacht en daarna het portret en de groene kaart getoond. Proefpersonen werd gevraagd de meest waarschijnlijke hypothese te noemen en vervolgens zich de casusinformatie te herinneren.

Resultaten en discussie

Figuur 3 laat zien dat de experts net als in Experiment 1 meer accurate hypothesen genereerden

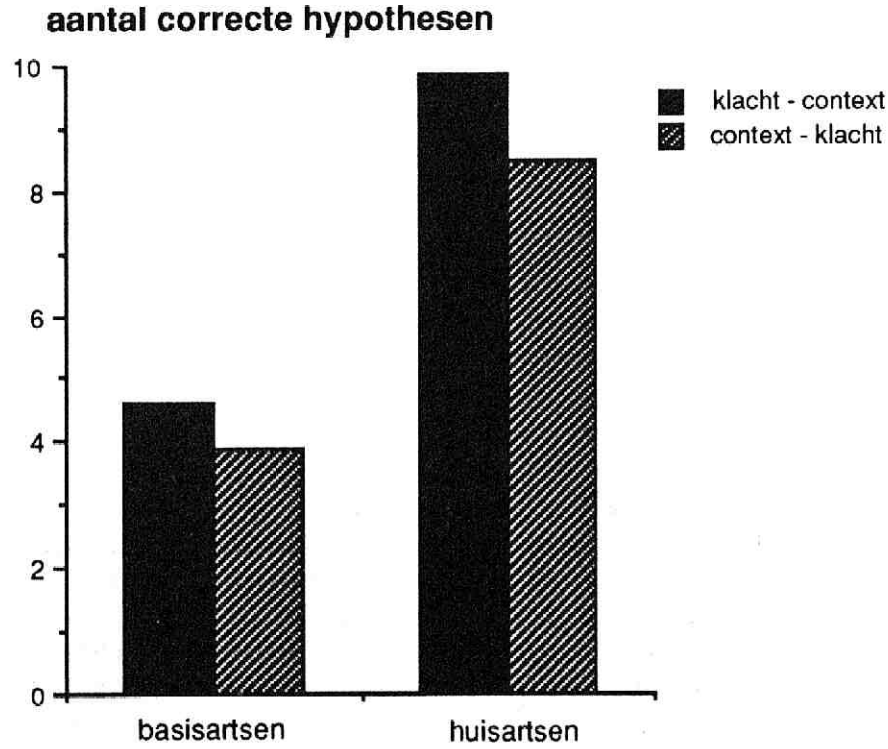


Fig. 3. Gemiddeld aantal accurate hypothesen gegenereerd door ervaren en onervaren artsen in respons op casus gepresenteerd met de klacht eerst (klacht-context), of de context eerst (context-klacht).

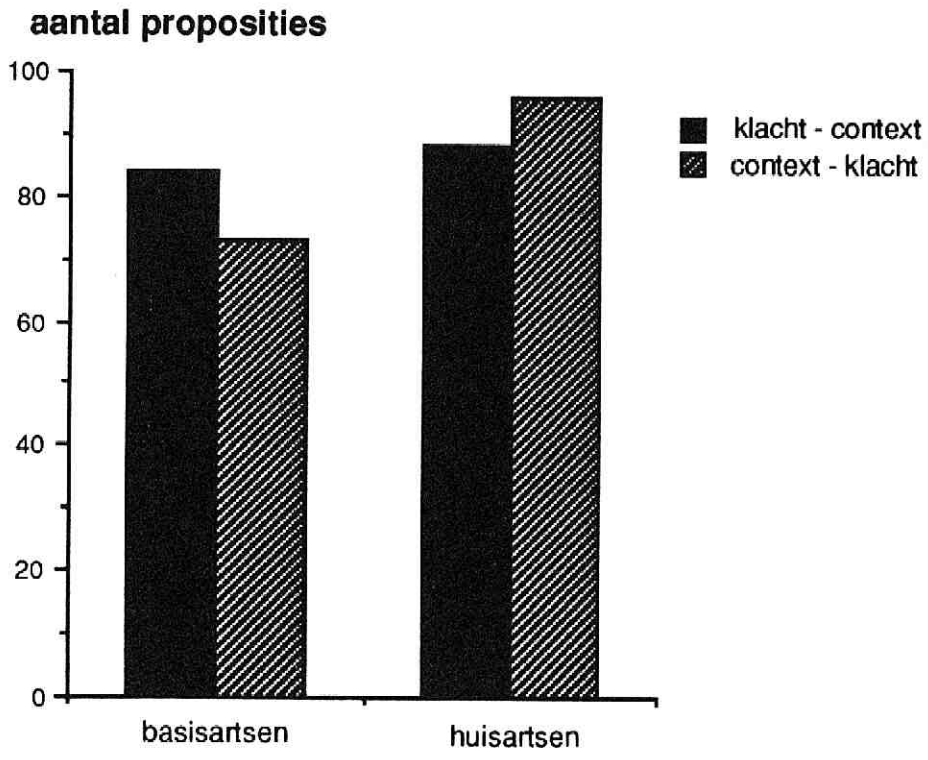


Fig. 4. Recall van casus informatie door ervaren en onervaren artsen in respons op casus gepresenteerd met de klacht eerst (klacht-context), of de context eerst (context-klacht).

dan de onervaren artsen ($F(1,28) = 31.45$; $p < .001$). Presentatievolgorde had daarentegen geen effect op de prestaties van de proefpersonen ($F(1,28) = 1.46$; $p = .24$). Er werden ook geen interacties gevonden ($F(1,28) = .126$; $p = .73$). Figuur 4 laat zien dat de experts zich ook meer herinnerden ($F(1,28) = 5.28$; $p < .05$). Ook bij de recall was er geen sprake van een hoofd- of interactie-effect van presentatievolgorde (presentatievolgorde: $F(1,28) = .07$; $p = .80$; presentatievolgorde * expertiseniveau: $F(1,28) = 2.49$; $p = .13$). Deze resultaten vormen geen ondersteuning voor de voorspelling dat beginnende artsen beïnvloed worden door variaties in de presentatievolgorde, terwijl experts dat niet worden. Zij zijn in tegenspraak met de theorie dat de ziektescripts van beginnende en ervaren artsen niet verschillen in inhoud maar wel in toegankelijkheid. Het feit dat experts in beide gevallen meer informatie onthouden dan niet-experts is daarentegen in overeenstemming met de hypothese dat de ziektescripts van ervaren en onervaren artsen verschillen in rijkdom en detail ten aanzien van de Enabling Conditions.

Alhoewel deze resultaten betreffende recall en hypothese generatie eensluidend zijn, zijn ze niet voldoende. Om meer zekerheid omtrent onze hypothesen te verkrijgen is meer, convergerende informatie nodig, bij voorkeur gebaseerd op een meer directe meting van de structuur van schema's. Daarom werd Experiment 3 uitgevoerd.

EXPERIMENT 3

Experiment 3 richt zich op de hypothese dat de ziektescripts van experts rijker zijn (waar het de Enabling Conditions betreft) dan de ziektescripts van beginnende artsen. In dit experiment wordt een methode gebruikt die ontleend is aan een onderzoek van Hobus, Boshuizen en Schmidt (1990) die ervaren en onervaren artsen vroegen de prototypische patiënten met een specifieke aandoening te beschrijven. Een dergelijke prototype beschrijving kan beschouwd worden als een directe uitlezing van het script (Abelson, 1981). De resultaten van Hobus et al. lieten zien

dat de beschrijvingen van ervaren artsen meer Enabling Conditions bevatten dan die van onervaren artsen, hetgeen suggereert dat de zicktescripts van ervaren artsen rijker zijn. Er was echter een interpretatie probleem bij deze data dat te maken heeft met de specifieke vraag die werd gesteld. De vraag om prototypische patiënten te beschrijven kan een bias veroorzaakt hebben ten gunste van de onderzochte hypothese. De nadruk op prototypische patiënten kan ertoe hebben geleid dat de proefpersonen relatief meer uitspraken met patiëntenbeschrijvingen (tegenover symptomen) deden, en aangezien de ervaren artsen nu eenmaal meer patiënten hebben gezien, kan deze specifieke vraag de twee groepen op verschillende wijzen hebben beïnvloed. Daarom is in het laatste experiment dat hier wordt beschreven aan de helft van de proefpersonen een andere, meer neutrale vraag gesteld. Bovendien zijn twee lagere expertiseniveaus in het onderzoek betrokken.

Methode

Aan dit onderzoek namen 89 proefpersonen deel: 22 vierdejaars studenten, 22 zesdejaars studenten, 23 afgestudeerde basisartsen die deelnamen aan de opleiding tot huisarts, en 22 ervaren huisartsen. De helft van iedere groep proefpersonen werd gevraagd de prototypische patiënt met een bepaalde aandoening te beschrijven, de andere helft beschreef het klinisch beeld van die aandoeningen. Daarbij werden dezelfde 18 aandoeningen gebruikt als in de Experimenten 1 en 2. De reacties werden op band opgenomen en later uitgewerkt. De teksten die op deze wijze waren verkregen, werden opgedeeld in proposities die vervolgens werden geclassificeerd als Enabling Condition, Fault, of 'andere'. De meeste van deze 'andere' proposities betroffen Consequences of behandeling van de patiënt.

Resultaten en discussie

Het totaal aantal proposities dat door proefpersonen van verschillend expertiseniveau werd gegenereerd, verschilde niet ($F(3,81) = 1.759$; $p = .18$); ook de twee typen vragen hadden geen invloed op de verbositeit van de proefpersonen ($F(1,81) = .309$; $p = .58$). Figuur 5 laat zien dat

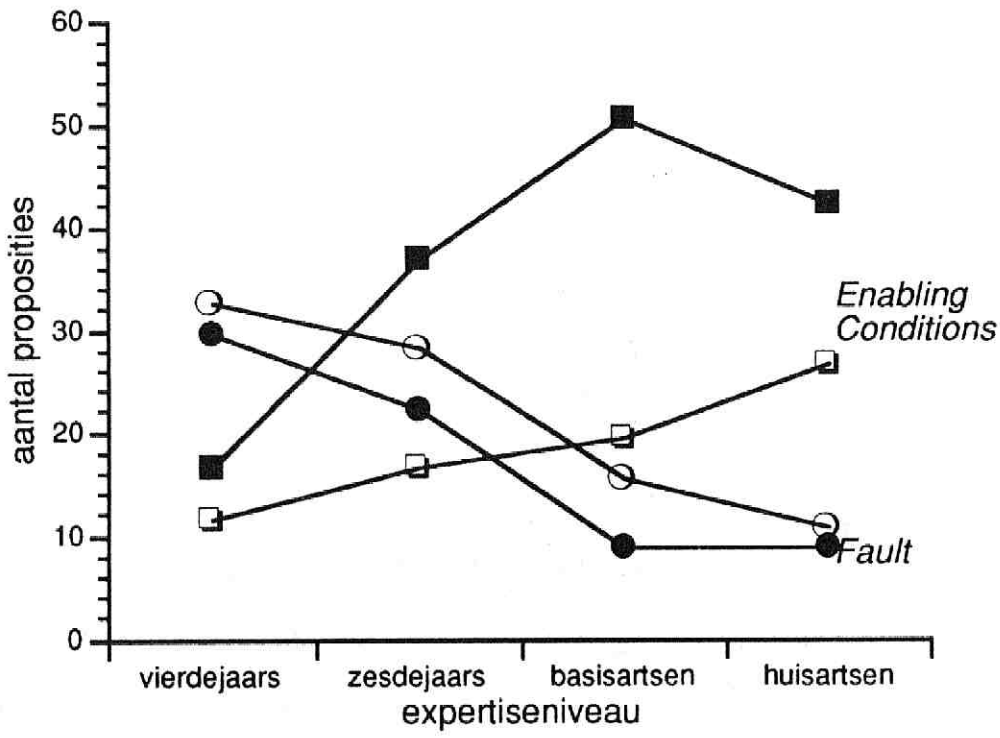


Fig. 5. Aantal Enabling Conditions en Fault proposities in de beschrijvingen van prototypische patiënten of klinische beelden (■● = prototypische patiënt, □○ = klinisch beeld).

het aantal Enabling Conditions toenam met toenemend expertiseniveau ($F(3,81) = 8.119$; $p < .001$). Bovendien had het type vraag invloed op het aantal Enabling Conditions dat werd genoemd. Proefpersonen aan wie gevraagd was prototypen te beschrijven noemden er meer dan proefpersonen die het klinisch beeld van dezelfde aandoeningen beschreven ($F(1,81) = 28.154$; $p < .001$). Verder was er sprake van een interactie-effect van expertiseniveau en type vraag ($F(3,81) = 2.617$; $p = .05$): Proefpersonen van een intermediair expertiseniveau werden veel meer beïnvloed door het type vraag dan experts en relatieve beginners.

Het aantal Fault propositities nam monotoon af met toenemend expertiseniveau ($F(3,81) = 7.873$; $p < .001$). Geen van de overige effekten was significant (vraag: $F(1,81) = 1.465$; $p = .23$; vraag * expertiseniveau: $F(3,81) = .102$; $p = .96$).

Deze resultaten suggereren dat de ziektescripts van personen met hogere niveaus van expertise rijker zijn aan Enabling Conditions. Dit is in overeenstemming met de hypothese dat toenemende ervaring leidt tot verdere verrijking van de kennis. De afname van het aantal Fault uitspraken in de beschrijvingen ondersteunt de hypothese van een toenemende encapsulatie van de biomedische kennis. Er trad echter een even opvallend als onverwacht verschijnsel op, dat noch met een hypothese omtrent verrijking van ziektescripts, noch met een hypothese omtrent veranderde toegankelijkheid van ziektescripts direct te rijmen valt. In tegenstelling tot de studenten in de pre-doctoraal fase en de ervaren huisartsen, werden de proefpersonen van een meer intermediair niveau van expertise (met name de huisartsen in opleiding) in sterke mate beïnvloed door de aard van de vraag die hen werd gesteld. In de algemene discussie zullen we hierop terugkomen.

ALGEMENE DISCUSSIE

De drie experimenten die hierboven werden beschreven hebben een aantal fenomenen laten zien: a) Experts hebben meer voordeel van context informatie bij het genereren van hypothesen dan minder ervaren personen; b) veranderingen in de volgorde waarin informatie wordt aangeboden, hebben een gelijk effect op de activatie van ziektescripts van ervaren en onervaren artsen; c) experts herinneren zich meer context informatie dan onervaren artsen; d) expertise ontwikkeling gaat gepaard met een toename van het aantal gerapporteerde Enabling Conditions en een afname van het aantal Fault propositities; en e) de door vierdejaars studenten en ervaren artsen gedane beschrijvingen van prototypen en klinische beelden, voor zo ver het de Enabling Conditions betreft, zijn meer aan elkaar gelijk dan de twee soorten beschrijvingen die door coassistenten en huisartsen in opleiding worden gegenereerd.

Deze resultaten suggereren dat praktische ervaring aanleiding geeft tot duidelijke veranderingen in de ziektescripts. Ten eerste worden de ziektescripts steeds gemakkelijker door context informatie geactiveerd. Verder bleek dat dit fenomeen niet zo zeer werd veroorzaakt door een verschuiving in het soort van informatie dat instaat is een script te activeren, maar door een verrijking van de scripts met Enabling Conditions voor een aandoening. Deze conclusie wordt echter gecompliceerd, doordat een aantal andere resultaten niet zonder meer in deze conclusies kunnen worden ingepast. Het eerste probleem is dat personen van verschillend expertiseniveau niet in gelijke mate worden beïnvloed door de bewoordingen van een vraag die tot scriptuitlezing zouden moeten leiden: Proefpersonen van intermediair expertiseniveau noemden opmerkelijk veel meer Enabling Conditions wanneer hen werd gevraagd prototypische patiënten te beschrijven dan wanneer ze klinische beelden beschreven, terwijl de verschillende vraagvormen nauwelijks enig effect hadden op de vierdejaars studenten en de ervaren artsen. Ten tweede moet verklaard worden waarom het aantal gerapporteerde Enabling Conditions van een aandoening toeneemt met de ervaring, terwijl onervaren artsen toch geen voordeel bleken te hebben van context informatie bij het genereren van vroege hypothesen omtrent een mogelijke aandoening bij een patiënt. Blijkbaar kunnen personen van intermediair expertiseniveau de Enabling Conditions van een specifieke aandoening identificeren (vooral wanneer de vraag hen daartoe direct aanzet), lang voordat dit soort van informatie al in een zeer vroeg stadium van een arts-patiënt contact relevante ziektescripts kan activeren. Aan deze discrepantie kunnen een aantal

verschillende redenen ten grondslag liggen. Een mogelijkheid is dat de Enabling Conditions die door huisartsen in opleiding worden genoemd nog niet voldoende zijn afgestemd op de alledaagse epidemiologie van de huisartspraktijk. Het is echter heel moeilijk deze hypothese te onderzoeken, aangezien Enabling Conditions meestal niet systematisch in de tekstboeken worden beschreven. Een andere reden voor deze discrepantie kan zijn dat de interne structuur van de ziektescripts van personen van intermediair expertiseniveau nog niet geheel stabiel is. Dit zou ertoe kunnen leiden dat de vraag om een prototype of een klinisch beeld te beschrijven leidt tot ziektescript constructie in plaats van rechtstreekse uitlezing van een stabiele cognitieve structuur. Experts hebben daarentegen meer stabiele en coherente kennisstructuren dan intermediairen. Dergelijke doorwrochte kennisstructuren zouden in hun geheel kunnen worden geactiveerd. Daarom maakt het ook niet uit wat voor vraag hen precies wordt gesteld: in beide gevallen wordt exact dezelfde kennis geactiveerd. Deze tweede verklaring sluit goed aan bij de bevinding dat het aantal genoemde Fault propositie afnam met een toenemend expertiseniveau. Fault proposities betreffen de pathofysiologie die ten grondslag ligt aan de symptomen van een aandoening. Volgens Feltovich en Barrows (1984) wordt de biomedische kennis gebruikt om het aantal mogelijke combinaties van Enabling Conditions en Consequences van een aandoening in te perken. Met een toenemend expertiseniveau wordt deze kennis geëncapsuleerd onder hogere-orde concepten en worden stabiele scripts gevormd. Hierdoor is het niet langer noodzakelijk om hele redeneerketens op te zetten en daarmee de verschillende onderdelen van zo'n script aan elkaar te koppelen en op elkaar af te stemmen. Deze analyse van de beschikbare data brengt ons tot de conclusie dat het toenemende gebruik van context informatie bij het activeren van ziektescripts misschien niet alleen een gevolg is van scriptverrijking, maar dat een interne herstructurering van de componenten van ziektescripts ook een rol kan spelen.

Samenvattend kunnen we concluderen dat de resultaten die in dit artikel worden gerapporteerd suggereren dat toenemende expertise gepaard gaat met kennisencapsulatie en scriptvorming, -herstructurering en -verrijking.

De kennisontwikkeling en structurele veranderingen die in dit hoofdstuk zijn beschreven, zijn een gevolg van theoretisch onderwijs en ervaringen in de praktijk. Zoals bekend bestaat het medisch onderwijs in Nederland globaal gesproken uit twee delen. Traditioneel vindt in de periode voor het doctoraal vooral de theoretische scholing plaats, terwijl in de co-schappenfase vooral de praktische ervaring moet worden opgedaan. Deze strikte scheiding wordt echter meer en meer doorbroken. Aan de Rijksuniversiteit Limburg wordt sinds de oprichting van de medische faculteit gewerkt met simulatiepatiënten, vinden korte stages in de eerste- en tweedelijnsgezondheidszorg plaats en wordt momenteel een onderwijsvorm ontwikkeld waarbij studenten vanaf het tweede jaar gekoppeld worden aan een huisartspraktijk, waar zij enige malen per maand ervaringen in de praktijk kunnen opdoen. Ook aan andere universiteiten, zoals de Universiteit van Amsterdam en de Rijksuniversiteit Groningen worden dergelijke initiatieven ontplooid. In dit laatste deel willen wij stil staan bij de drie typen kennisverandering die in het beschreven model voorkomen en nagaan welke funktie theoretische scholing en ervaring daarin vervullen.

Voor de ontwikkeling van geïntegreerde kennisnetwerken is het in de eerste plaats noodzakelijk dat de student veel theoretische kennis opdoet. Om de gewenste integratie te bewerkstelligen, is het daarnaast wenselijk dat de student daarbij zicht heeft op de praktische problemen waarvoor hij of zij die kennis opdoet en daarin ook enige oefening krijgt. Voor de encapsulatie van de biomedische kennis is het noodzakelijk dat studenten hun kennis actief toepassen in de context van diagnose en behandeling. Het is moeilijk te zeggen of daarvoor echte patiënten nodig zijn of dat simulatiepatiënten hetzelfde cognitieve effect kunnen bewerkstelligen. Het voordeel van simulatiepatiënten is dat dergelijke ervaringen al in een vroeg stadium van de opleiding geïntroduceerd kunnen worden. Echte ervaring met echte patiënten is echter de eerste en enige methode om ziektescripts te verrijken, te verfijnen en af te stemmen op de werkelijkheid. In dat opzicht is de ervaring met echte patiënten onvervangbaar (zie ook Boshuizen & Schmidt, 1993, in druk).

Een andere vraag is of dit soort praktische ervaring die vooral tijdens de co-schappen wordt opgedaan, voldoende is om tot een verdere ontwikkeling van de kennis in de gewenste richting te komen, of dat niet alleen de praktijk in het theoretische deel van de opleiding moet worden binnengebracht, maar ook de theoretische opleiding verder geïntegreerd moet worden in het praktijkdeel. Ons eigen onderzoek en onderzoek van auteurs als Collins en zijn collega's (Collins, 1990) over het verwerven van robuuste kennis suggereert dat het laatste het geval is. In ons eigen onderzoek (Boshuizen & Schmidt, 1992) bleek dat studenten gedurende de co-assistentschappen een drastische verandering in het gebruik van hun kennis vertoonden. Voor aanvang van de co-schappen gebruikten zij met name biomedische kennis in hun klinische redeneringen, nu schakelen zij over op de toepassing van kennis van de klinische vakken, terwijl deze nog lang niet voldoende is en ook nog niet geïntegreerd is met de biomedische kennis. Als de studenten in dit stadium uitsluitend ervaringskennis verwerven, lopen zij het risico dat ze een grote hoeveelheid gesitueerde kennis opbouwen die niet verbonden is met de theoretische, algemene, biomedische kennis. Het werk van Leinhardt (1987) maakt duidelijk hoe inflexibel dit soort van kennis kan zijn. Daarom menen wij dat ook in dit stadium van de ontwikkeling verder theoretisch onderwijs noodzakelijk is. Dit onderwijs dient dan met name gericht te zijn op de integratie van algemene biomedische kennis en op ervaring gebaseerde kennis, zodat robuuste medische kennis ontstaat.

NOTEN

1. Dit artikel is gebaseerd op een voordracht gehouden ter gelegenheid van de NATO advanced research workshop, Advanced Models of Cognition for Medical Training and Practice, 1991, Castelvecchio Pascoli, Italië. Voorbereiding van dit artikel is mogelijk gemaakt door een Spencer grant van de National Academy of Education aan H.P.A. Boshuizen.
2. We maken in dit artikel onderscheid tussen Enabling Conditions en Contextuele Informatie. Enabling Conditions zijn onderdeel van een kennisstruktuur (een ziektescript) en zijn dus zelf ook kennis. Context(uele) Informatie kan betrekking hebben op de Enabling Conditions van een bepaalde aandoening, maar kan er ook niets mee hebben te maken. Het gegeven dat iemand reeds twee maal een been heeft gebroken is niet relevant wanneer deze nu klaagt over oorpijn; maar is van het grootste belang wanneer de gedachten in de richting van een bot- of stofwisselingsziekte gaan.
3. De aanname dat Enabling Conditions een ziektescript kunnen activeren is niet geheel vanzelfsprekend. De meeste modellen die het genereren van diagnostische hypothese beschrijven zijn gebaseerd op de veronderstelling dat het genereren en toetsen van hypothesen volledig gestoeld is op 'harde' gegevens zoals symptomen en labwaarden. Andere gegevens zijn slechts bijkomend en kunnen niets bewijzen (cf. Pauker et al., 1976).

LITERATUUR

Abelson, R.P. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715-729.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

Boshuizen, H.P.A. (1989). *De ontwikkeling van medische expertise; een cognitief-psychologische benadering.* Academisch proefschrift Rijksuniversiteit Limburg. Haarlem: Thesis.

Boshuizen, H.P.A., & Schmidt, H.G. (1993, in druk). Teaching clinical reasoning to medical students. In: Higgs, J., & Jones, M.A. (Eds.) *Clinical reasoning skills.* Oxford: Butterworth-Heinemann.

Boshuizen, H.P.A. & Schmidt, H.G. (1992). The role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16,* 153 -184.

Collins, A. (1990). Generalizing from situated knowledge to robust understanding. Paper gepresenteerd ter gelegenheid van the Annual Conference of the American Educational Research Association. Boston, MA.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving, an analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.

Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt, & M. L. De Volder (Eds.), *Tutorials in problem-based learning; A new direction in teaching the health professions.* (pp. 128-142). Assen: Van Gorcum.

Gale, J., & Marsden, P. (1982). Clinical problem solving: the beginning of the process. *Medical Education, 16,* 22-26.

Hobus, P. P. M., Boshuizen, H. P. A., & Schmidt, H. G. (1990). Expert-Novice Differences in the Mental Representations of Patient Prototypes. Paper gepresenteerd ter gelegenheid van de 2nd International Symposium on Problem-based Learning. Yogyakarta, Indonesia: October 1990.

Hobus, P.P.M., Schmidt, H.G., Boshuizen, H.P.A., & Patel, V.L. (1987). Contextual factors in the activation of first diagnostic hypotheses: expert-novice differences. *Medical Education, 21,* 471-476.

Joseph, G., & Patel, V. L. (1987). Domain knowledge and medical problem solving: An on-line analysis. Paper gepresenteerd ter gelegenheid van the Annual Conference of the American Education Research Associations. Washington, D. C. Montreal: McGill University. (Report CME87-CS9).

Leinhardt, G. (1987). Situated knowledge: An example from teaching. Paper gepresenteerd ter gelegenheid van the Annual Conference of the American Educational Research Association. Washington, DC.

Patel, V. L., & Groen, G. J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science, 10,* 91-110.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory and Cognition, 18,* 394-406

Pauker, S. G., Gorry, G. A., Kassirer, J. P., & Schwartz, W. B. (1976). Towards the simulation of clinical cognition. *The American Journal of Medicine, 60,* 981-996.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning, and restructuring; Three modes of learning. In J. W. Cotton, & R. L. Klatzky (Eds.), *Semantic factors in cognition.* (pp. 37-53). Hillsdale, NY: Lawrence Erlbaum Associates.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). Cognitive research on clinical reasoning: theory and implications. *Academic medicine, 65(10),* 611-621.

Manuscript ontvangen 17-3-1993
Definitieve versie ontvangen 8-6-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 3, pp. 175-185

Meta-knowledge and the development of expertise in complex domains

Monica G.M. Ferguson-Hessler
Department of Philosophy and Social Sciences, Eindhoven University of Technology and Graduate School of Teaching and Learning, University of Amsterdam*

ABSTRACT

Expertise is seen as the ability to carry out certain types of tasks, requiring the application of knowledge of a complex domain. Meta-knowledge or 'knowledge about knowledge' consists of two components: a cognitive or low level component, comprising abstract and epistemological knowledge about the domain and the task performance and knowledge acquisition strategies of the domain, and a metacognitive or high level component, comprising knowledge of the person's own domain knowledge and task performance and knowledge acquisition strategies. A closer analysis of the two components and their elements reveals the role of meta-knowledge in expertise, especially for metacognitive regulation. Possible consequences for instruction in academic fields focusing on expertise are discussed.

INTRODUCTION

The development of expertise in complex domains is one of the central objectives of academic teaching. A vast amount of research has been devoted to expertise in fields like problem solving in the sciences and medical diagnosing, to the differences between expert and novice performance, and to the process of development from novice to expert. In spite of the extensive information collected on the differences between novices and experts, we have not yet wholly understood the process of development from novice to expert, and still less do we possess a general model of instruction which would foster this development. Agreement does exist, however, on the following two points:

- meaningful learning requires active cognitive processing of information, and
- efficient processing requires meta-cognitive regulation.

One aspect of knowledge, which seems to play an important role in the complicated process of meaningful learning, but is rather implicit in descriptions of this process, is meta-knowledge, taken in the literal sense of 'knowledge of knowledge'.

In the literature, the term 'meta-knowledge' is seldom found. Flavell (1976) was one of the first authors to speak explicitly of metacognition as "one's knowledge concerning one's own cognitive processes and products" (p. 232). Alexander, Schallert, and Hare, in their review article on Knowledge Terms (1991), do not include this term in the list of definitions found in the literature, but speak of 'metacognitive knowledge' as including both 'knowledge of knowledge' and knowledge of one's own cognition and the regulation of that cognition. They remark, however: "Metacognitive knowledge represents one of those critically important, frequently studied, much referred to, but seemingly ill-differentiated theoretical constructs" (p. 328).

This study attempts to make one selected aspect of this theoretically ill-defined set of concepts operational by the introduction of pragmatic and logical definitions, and to use these definitions in an analysis of task performance and learning processes. The reason for concen-

* Adres: Postbus 513, 5600 MB Eindhoven.
Author's note: This study originated from earlier research on knowledge, learning, and instruction in physics, carried out in cooperation with Ton de Jong. His contributions in the form of discussions on the concepts introduced, and comments on an earlier version of the paper are gratefully acknowledged.

trating on meta-knowledge is that in our opinion 'knowledge of knowledge' has not been given enough attention, in view of its central role in learning and task performance. We do not attempt to formulate a theory or a formal model of metacognition, but offer this analysis as a starting point for further discussion. With Alexander and her colleagues we feel the need of an "epistemological shorthand" for knowledge constructs that can be used by researchers and practitioners in the field of learning and instruction. We also agree with these authors that the introduction of terms without thorough exploration of their theoretical background and their relations to existing terminology will not contribute to better understanding in this complex field of research.

To be able to analyze and discuss the complex processes of learning and task performance we devote the first section to definitions of a number of terms used. In the next two sections we analyze the relations between these concepts and subsequently attempt to uncover the role they play in expertise and in the development thereof. A final section discusses the possibilities that are thus opened up for the application of cognitive theory in the design of instruction.

DEFINITIONS AND CLARIFICATIONS

One of the first scientific studies of expertise was the well-known research on the memory of chess players by de Groot (1946, 1965). A number of researchers have followed in his footsteps, and investigated the memory and performance of players of chess (Chase & Simon, 1973) and other games (for instance go, Reitman, 1976), but also specific technical expertise like electronic circuits (Egan & Schwartz, 1979) and expertise in more general domains, like problem solving in physics (Larkin, 1981; Larkin, McDermott, Simon & Simon, 1980; Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982). Our present study aims at a broader concept of expertise than that confined to one of these domains. Still, for a lucid discussion we need to delimit and clarify the concept of expertise, and hence introduce the following definition:
Expertise is the ability use a combination of skills in order to carry out a certain type of task in a complex domain.

Examples of relevant tasks are
- problem solving in science,
- diagnosing an X-ray photo in medicine or dental medicine,
- analyzing a law case.

These examples illustrate the level of the task: knowledge of a complex domain has to be applied in a variety of situations. One central aspect of expertise is the ability to use skills in a variety of contexts: knowledge has to be applied in situations which are new, and sometimes very different from the context in which the knowledge was acquired. Common to the tasks we concentrate on is that they play a major role in academic fields which focus on professional skills, for instance physics, chemistry, and technical sciences, medicine and law.

A *skill* can be described as the practised ability to carry out certain activities. The expert's skill is fast and accurate, without hesitation or error. In the types of task we are studying, the activities are mainly of a cognitive or metacognitive nature. As will be elaborated in the next paragraph, such an activity employs a variety of types of conceptual and operational knowledge, and hereby influences the content and structure of the knowledge base. This is the reason that we do not use the concepts of knowledge and skill as two independent and complementary entities in discussing expertise, but prefer to describe the type of skill relevant for this study as a complex interplay of various types of knowledge and activities.

In these terms, the development of expertise is seen as the acquisition of a set of cognitive and metacognitive skills, which have both domain specific and general components, and which require a scala of types of knowledge. This type of learning proceeds along two different ways: by processing external information (textbooks, lectures, demonstrations, computer programs) and by performing tasks where the final objectives or parts thereof are practised.

The two ways of learning have been elaborated in Figure 1 (from de Jong and Ferguson-Hessler, 1993), where the central role of the knowledge base is illustrated. The internal representation the learner builds of external information is influenced by the knowledge base in two ways: perception is partly controlled by prior knowledge, and elements of knowledge from memory are added to the external information to form the representation. Both task execution and learning by processing external information act on the internal representation, continually transforming it, and both lead to enrichment of the knowledge base, that is to say to learning. As a consequence, we do not speak of learning processes or learning strategies in general, but distinguish *knowledge acquisition* and *task performance* processes and strategies as two complementary aspects of learning, described as 'learning by information processing' and 'learning by doing' respectively (Ferguson-Hessler & de Jong, in press).

Cognitive activities, such as encoding external information, retrieving elements of knowledge from memory, or analyzing a given task, are distinguished from metacognitive activities, such as monitoring or evaluating, by the level of functioning: metacognition functions at a higher level, providing the regulation of cognitive activities which is necessary for successful completion. In other words, the objects of the two types of activities are different: external, content directed in the case of cognition, and internal and directed at the person's own cognition in the case of metacognition. Each of these levels covers several areas of cognitive activity, of which the three most relevant for this study are found in Figure 1: task performance, knowledge acquisition by information processing, and domain knowledge. In these activities the knowledge base plays a central role: it provides the prior knowledge which is necessary for the processes, and it is enriched as a result of the processes (Figure 1).

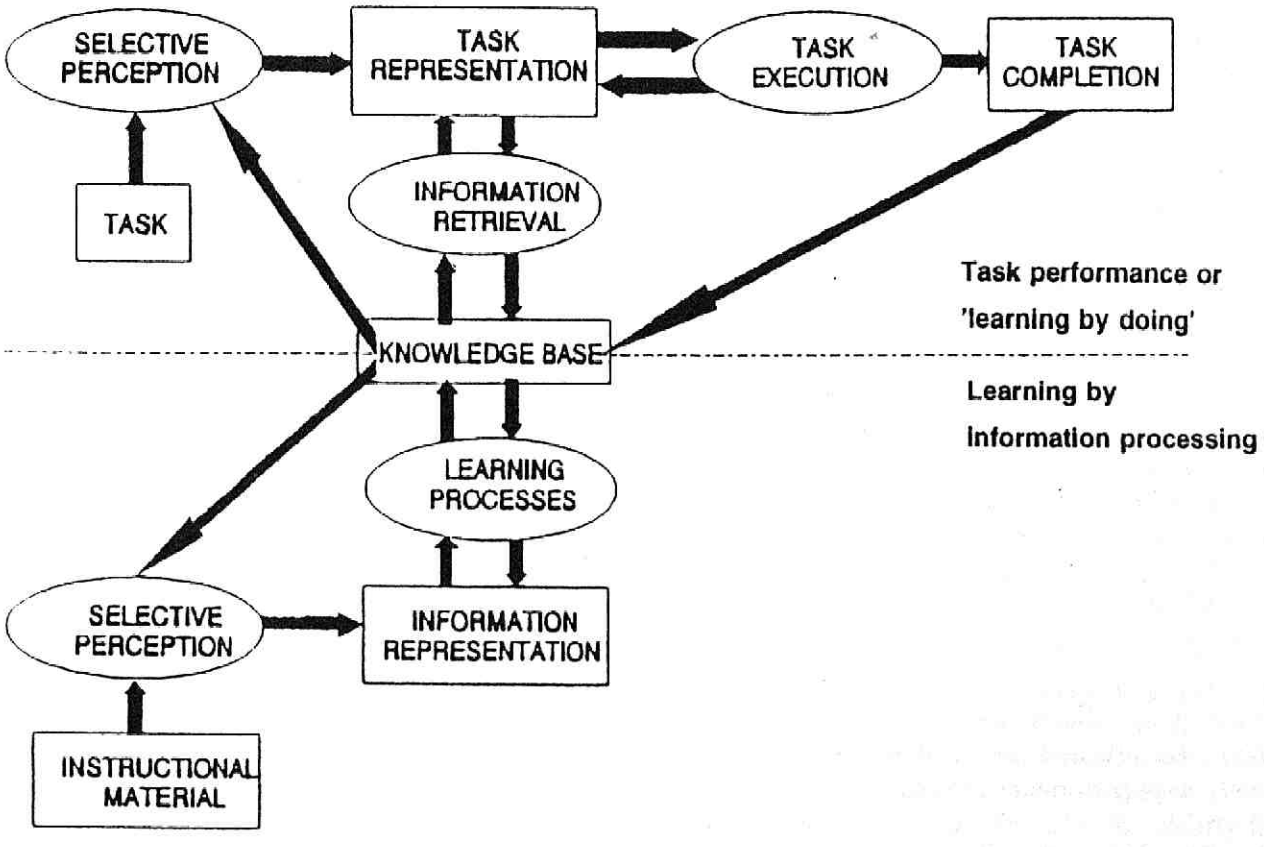


Fig. 1. Cognitive activities relevant in expertise: task performance or learning by doing, and learning by information processing.

In metacognition the aspects of *monitoring* and *regulation* are often stressed. As the literal meaning of these two words have some overlap, we will use 'monitor' to mean 'observe and analyze', and 'regulate' as a general term for all types of metacognitive activities, both executive or control activities (planning, evaluating, repairing) and monitoring. Lawson (1984), and Prawat (1989) both stress the importance of distinguishing between *knowledge of cognition* and *control of cognition*, both logically, in discussions, and empirically, in experiments on metacognition. Lawson defines metacognitive knowledge as 'knowledge of cognitive processes'. Prawat, following Gagné and Glaser (1987), introduces the concept of 'awareness' for this type of metacognitive knowledge. Alexander et al. (1991), in their conceptual framework for knowledge terms, distinguish the following forms of metacognitive knowledge: self, task, and strategic knowledge, and knowledge of goals and plans (p. 328). Their definition of metacognitive knowledge, however, includes regulation of cognition as well (p. 332).

Any form of regulation of a person's cognitive activities must be based on knowledge of the progress and quality of these activities, acquired by monitoring. Regulation, thus, is strongly dependent on the knowledge aspect of metacognition, but also depends on some aspects of cognitive knowledge. We introduce the following working definition:
Meta-knowledge is 'knowledge about knowledge' in the literal sense.

Meta-knowledge so defined has two components:
- a *cognitive component* directed at domain knowledge and task performance and knowledge acquisition strategies, and
- a *metacognitive component*, directed at the person's own domain knowledge and task performance and knowledge acquisition activities.

We label the two components low level or objective and high level or subjective meta-knowledge respectively. A schematic description of the two components is given in Table 1, where, for each component, the three areas task performance, domain knowledge, and knowledge acquisition by information processing, pictured in Figure 1, are distinguished.

1. *Low level* or *objective* meta-knowledge of *task performance and task requirements* concerns the criteria for judging the result of the task ("What sort of outcome would be an acceptable result?"), as well as knowledge of the circumstances in which the task has to be carried out, the types of strategies and activities involved, available technical aids and sources of information, whether the task is carried out individually or in a group, time limits, etcetera. The

Table 1. The two components of meta-knowledge.

	Low level meta-knowledge (Cognition)	High level meta-knowledge (Metacognition)
1	Task requirement	Criteria for judgement of success/failure of own result
2	Types of content knowledge, their function in task performance, and relations to task performance strategy	Results of monitoring: content, structure, usefulness of own domain knowledge and task performance strategy
3	Knowledge acquisition strategy Criteria for understanding	Results of monitoring: content, structure, relevance of own knowledge acquisition strategy Evaluation of own understanding

high level, or metacognitive counterpart is knowledge of the quality of one's own task performance ("How reliable is my result?", "Why do I not trust the quality of my result?", "What is the reason that I got stuck?").

2. *Low level* meta-knowledge of subject matter content is a form of *epistemological knowledge*, which concerns the types of knowledge which are typical of the domain and the task. For problem solving in a physics domain, for instance, an epistemological description would distinguish declarative, procedural and situational or conditional knowledge, all of which have to be combined with a problem solving strategy (Ferguson-Hessler & de Jong, 1987). For diagnosing an X-ray photo such a description would distinguish knowing and recognizing the types of pictures belonging to healthy bones or tissues, the normal variation in these pictures, typical deviations caused by a number of types of damage or disease, the geometrical relations between the three-dimensional body and the two-dimensional picture, the background of the actual case, and commanding a strategy for the scanning of the photo, and procedures for detailed analysis of relevant parts of the picture (Lesgold, Robinson, Feltovich, Glaser, Klopfer, & Wang, 1988). In all cases the epistemological description would also include an explicitation of the intrinsic structure of the subject matter, and a knowledge structure which is efficient for carrying out the task, and would specify the way in which this structure relates to the stages of the task performance strategy (which is part of the subject matter knowledge). At a more elementary level of subject matter, epistemological knowledge could be described as "Knowing what is relevant", or "Knowing what sort of relations are important". The *high level* counterpart is knowledge of one's own knowledge ("What exactly do I know about...?", "Is this enough for acceptable task performance?", "Is my knowledge logically structured?", "Do I master these procedures?").
3. Low level meta-knowledge of *strategies for knowledge acquisition* is different from knowledge of the strategies as such. It concerns the types of results they lead to, their appropriateness is relation to different types of task, the amount of time and effort they require, etcetera. An important aspect of these strategies are the criteria for judging understanding in the domain. The *high level* counterpart is knowledge of one's own knowledge acquisition activities ("What exactly am I doing at this moment?", "Do I see the meaning of this derivation, or am I just copying?", "Can I summarize this paragraph in my own words?")

Summing up: meta-knowledge is *both* knowledge of criteria for judging the quality of domain knowledge, task performance and knowledge acquisition processes, *and* the knowledge of these aspects of cognition in oneself.

RELATIONS BETWEEN TERMS AND CONCEPTS

It is evident from the definitions of the previous section that the various components and aspects of meta-knowledge described are strongly interrelated. Starting from knowledge of the fundamental concepts and principles of the domain, and of the activities required for the performance of typical tasks, it is possible to develop an epistemological description of a knowledge base, which is appropriate for typical tasks. Once the role of different types and elements of knowledge in task performance is thus explicitly described, it is possible to conceive strategies for the acquisition of these types of knowledge.

The three types of low level meta-knowledge are prerequisites for their high level counterparts: in order to be able to observe and analyze one's own domain knowledge, and task performance and knowledge acquisition activities, one needs the guidance of criteria for good quality knowledge and performance and knowledge acquisition processes. Finally, the ability to judge one's own success or failure in task performance grows out of a combination of these forms of meta-knowledge. This ability, not only to carry out a task, but also to evaluate the outcome and judge the reliability of the result, is a characteristic of expertise.

Low level meta-knowledge is dependent on the domain and the task, and is, in principle,

teachable. High level meta-knowledge, as defined here, corresponds to Lawson's (1984) concept of metacognitive knowledge. Like this concept it is strongly domain dependent, to some extent even task dependent, conscious and reportable. High level meta-knowledge also corresponds to the concept of awareness as introduced by Prawat (1989). According to both authors this type of metacognitive knowledge grows out of the person's reflection on his or her own cognitive activities and their outcomes, and can be developed by verbalization, for instance in the form of classroom dialogue. However, verbalization in any form requires some sort of knowledge of what counts as relevant, and words to label what counts as relevant, that is to say elements of a terminology, or low level meta-knowledge.

Returning to the concept of monitoring or 'actively observing', mentioned earlier, we conclude that this is an activity which needs a basis in low level meta-knowledge, and which leads to awareness, or higher level meta-knowledge. Hereby it facilitates the regulation of task performance and knowledge acquisition processes. These control processes have been studied extensively. Among the many examples of specific forms of regulation mentioned are: planning, analyzing, diagnosing, testing, questioning, directing, evaluating, repairing, modifying, revising, and reflecting (Lawson, 1984; Vermunt & Van Rijswijk, 1988; Simons & Beukhof, 1987; de Jong, 1987).

From the preceding discussion the place of meta-knowledge emerges as a missing link between cognition and metacognitive regulation: regulation is based on high level meta-knowledge, resulting from monitoring, which in turn is based on low level meta-knowledge. Research on learning has reached a high level of consensus on the nature of meaningful learning as an active, constructive, cumulative, and goal-oriented process (Shuell, 1986, 1990), a process requiring mental effort and motivation in the learner, but especially the ability to critically observe and regulate one's own learning processes (Vermunt & Van Rijswijk, 1988; Resnick, 1989). We see the different forms and components of meta-knowledge, which have been defined in this study, playing a central role in learning, as facilitators of various forms of metacognitive regulation processes.

WHAT IS EXPERTISE?

Glaser and Chi (1988) have summarized some characteristic features of expert performance which have been uncovered by research. They interpret the differences found between experts and novices as "reflecting the expert's possession of an organized body of conceptual and procedural knowledge that can be readily accessed and used with superior monitoring and self-regulation skills" (p. xxi).

The metacognitive component of expertise, as described by Glaser and Chi, refers to observation of the person's own cognitive activities and their outcomes, and to action taken on the result of this monitoring by a scala of control and/of executive activities. An example is the process of checking the outcome of a problem solving task: the novice is taught to carry out a check on the result as a last step in the problem solving strategy, but the expert typically keeps a more or less automatic check on his or her conclusions and calculations all the way through and parallel to other processes of solution, and immediately stops, diagnoses, and repairs if the check is not satisfactory.

Not mentioned in this description of expertise but implicitly present is the aspect of meta-knowledge. The expert's organized body of knowledge contains domain knowledge at different levels: declarative, procedural, situational, and strategical at the cognitive level, but also epistemological knowledge of this content and meta-knowledge of strategies used for task performance and knowledge acquisition in the domain. In practise, most experts probably have this type of knowledge in a tacit form (Reif, 1984), implicit, unspoken, and highly automated. Table 1 offers an explicit representation of this aspect of the expert's domain knowledge. We think that the concept of low level meta-knowledge, or what the expert knows about knowledge and performance in his or her own domain, offers an essential link between the two aspects of

expertise described by Glaser and Chi: well organized domain knowledge and superior metacognitive skills.

Transfer, or the ability to apply knowledge and skills in contexts and (sub)domains which are different from the original ones, is an important aspect of expertise. The question whether general cognitive skills exist, (and facilitate transfer), or whether cognitive skills are context bound, has been discussed at length. Perkins and Salomon (1989), in their analysis of this problem, conclude that this is a false dichotomy. There are general cognitive skills, but they always *function in a context*. Consequently they do not, as such, guarantee transfer. The authors distinguish two types of transfer: 'low road', where skills have been practised till they are automatic and at the same time somewhat flexible, and 'high road', where elements of knowledge and skills are deliberately abstracted and described in an explicit fashion, so as to become decontextualized.

Academic learning requires 'high road' transfer: selecting relevant elements from the representation of the task which has been learned, and relating these selected elements to corresponding elements of the new task (Simons & Verschaffel, 1992). The important role of metacognition in transfer has been emphasized by several authors (Perkins & Salomon, 1989; de Corte, 1991; Simons & Verschaffel, 1992). Prawat (1989) has stressed the importance of reflective awareness, (in our terms high level meta-knowledge) for promoting access to domain knowledge (see p. 178). However, the process of abstracting and selecting relevant elements requires a background of knowing the abstract properties of domain knowledge and knowing criteria for distinguishing relevant from non-relevant elements in the comparison of a new task with the task which has been learned. This type of knowledge fits exactly in the definition of low level meta-knowledge. We conclude that this type of knowledge forms a link between the cognitive and the metacognitive aspects of transfer.

INSTRUCTION AND THE GROWTH OF EXPERTISE

Academic teaching is concerned with the development of expertise in the course of the learning efforts of the student, and with the possibilities of stimulating this development by instruction. Looking back at the discussion of the relations between the various forms and components of meta-knowledge, we recognize that this description offers an image of the growth of expertise: starting from limited knowledge of concepts and task performance, learners build some low level meta-knowledge of the domain and the task requirements, observe their own knowledge and performance and processing activities, thereby enrich their domain knowledge and gain some criteria for understanding, and so forth. It is evident that we do not describe a simple sequence of processes of drawing conclusions and constructing knowledge. On the contrary, the development of expertise is a complex and lengthy process with a strongly recursive character, where elements of domain knowledge interact with elements of meta-knowledge, and cognitive activities with metacognitive activities. Resnick (1989), in a striking comparison, has likened this process to 'bootstrapping'.

Gagné and Glaser (1987) have pointed out several aspects of learning complex performances, such as the acquisition of cognitive strategies for learning and task performance, the construction of mental models that govern complex performances, and the training of self-regulation processes that influence learning and thinking. They stress the purposefulness of behaviour that is based on attentiveness for the demands of the task, the nature of the material, one's own capabilities, and state that this metacognitive awareness is typically present in experts, and to a much smaller extent in novices.

The different stages or phases through which a learner passes in meaningful learning have been discussed by Shuell (1990). In the initial phase of learning a new domain, the learner is confronted with a large amount of facts and principles, and can do little else than memorizing facts and using pre-existing schemata to interpret information. In a later, intermediate phase, the learner starts to discover an organizing structure in the domain, and becomes able to apply

elements of knowledge in new situations, and thus to learn by task performance and other more advanced knowledge acquisition processes, such as structuring and generalizing. Finally, in the terminal phase, emphasis is on performance, and learning consists of adding new facts to existing schemata, and of restructuring these schemata by discovering interrelations at increasingly higher levels.

Though Shuell discusses learning strategies and knowledge without introducing the distinction between cognition and metacognition, it is not difficult to see where the different forms of meta-knowledge play a role in his description of the development from novice to expert. Structuring of knowledge requires comprehension of the epistemology of the domain, and choice of knowledge acquisition strategies must be based on knowledge of task performance requirements and of one's own knowledge and learning.

Stimulating this complex process of development of expertise is obviously different from presenting domain knowledge and strategies for task performance. It is even different from presenting this knowledge and adding information on the epistemology of the domain.

In the initial phase of learning, according to Shuell, students would only be able to learn some fundamental facts and principles, but they would not master enough domain knowledge to be able to make sense out of a description of the organizing structure of the subject matter. They would, however, need some epistemological information on the characteristics of concepts and principles involved, in order to be able to distinguish what is relevant for task performance. Students of physics, for instance would need to know that the concept of 'force' does not designate a property of a body or a system, but an interaction between two bodies or systems. Students of law would need to learn to distinguish between elements of a case which are relevant for the application of law, and elements which speak strongly to a personal feeling of justice, but are not relevant for the application of law. In very simple words one could describe this type of elementary epistemological knowledge as 'knowing what it is all about'. Discussions and simple exercises can help the students to reflect on and verbalize their own knowledge, and so to create some elements of high level meta-knowledge.

In the intermediate phase of learning, students would be able to profit from other aspects of low level meta-knowledge: epistemological knowledge of types of domain knowledge, their intrinsic structure, and their relation to general and domain specific aspects of the task performance strategy. Also, one could expect students to able to employ more sophisticated knowledge acquisition strategies, which are tuned to the specific task performance requirements. This is in agreement with the results of experimental studies of learning in physics (Chi & Bassok, 1989; Ferguson-Hessler & de Jong, 1990). A less positive effect of meta-knowledge of task requirements on learning is reported by Entwistle & Entwistle (1992), who found that students revising for an examination tended to seek only the form of understanding of subject matter which was required by typical examination questions. In this intermediate phase, reflection on one's own knowledge and task performance strategies becomes possible. Also, knowledge acquisition strategies can be evaluated, and students would be able to reflect on the relation between task performance requirements and knowledge acquisition strategies.

The terminal phase would be characterized by 'expert learning', where the student has a certain amount of epistemological knowledge of the domain, is able to expand this knowledge by analysis and application, and thus to learn independently of instruction. As a number of cognitive processes are more or less automatic in this phase, attention is freed for reflection on task performance, knowledge and knowledge acquisition. This facilitates metacognitive regulation in the form of continuous evaluation, checking, and repair of cognitive activities and knowledge structures.

Based on the discussion of the growth of expertise, is seems reasonable to suggest the following principles for instruction fostering the development of expertise:

- When introducing a new subject, it is important to present not only domain content, but also some fundamental elements of meta-knowledge: 'what it is all about' in the new domain, and to offer students the possibility of reflecting on their own knowledge. This is the approach in the constructivist view of learning and teaching, which aims at stimulating conceptual change

in students (Smith, 1992). Duit (1991) stresses the importance of helping students to get an insight into the nature and range of scientific knowledge, that is meta-knowledge, in order to facilitate the transition from naive to scientific concepts.

- When students get to the stage of learning to apply their knowledge in typical tasks of the domain, they could profit from the other aspects of meta-knowledge: the intrinsic structure of the domain, the relations between types of knowledge and the steps of the task performance strategy, and knowledge acquisition strategies tailored to the domain. This is the moment to start stimulating systematic monitoring of task performance and knowledge acquisition activities on the basis of this low level meta-knowledge.
- Instruction in metacognitive regulation activities, such as planning or evaluating, should be preceded by explicit introduction of the relevant low level meta-knowledge of the domain, and students should be offered the experience of monitoring their own knowledge and strategies on the basis of this knowledge.
- As the development of expertise is a long and complex process, improved instructional measures need to be employed during considerable time, and across different domains studied.

Caillot and Dumas-Carré (1990) have described a teaching experiment in physics, where students were explicitly taught a few 'cognitive aids', developed to facilitate the building of a problem representation. The authors depart from an epistemological analysis of domain content, 'an expert's analysis', combined with a detailed analysis of the task performance process, in this case problem solving. The 'aids' or tools they developed have two functions, to guide students a) in collecting and organizing information, and b) in building a compact and relevant problem representation. On the basis of this representation, it becomes possible for the students to make explicit and judicious choices in planning the solution, instead of copying solutions of problems already solved. The experiment, which had other aspects than the 'cognitive aids' as well, and lasted for 4 years, was reported to be quite successful. In terms of our study of meta-knowledge, the experiment of Caillot and Dumas-Carré can be described as an effort to teach fundamental low level meta-knowledge of the domain and the specific elements of the task performance strategy to students confronted with a new type of task, problem solving in mechanics. This knowledge made it possible for the students to observe and plan their own solution activities and to make choices of their own. The authors stress that this type of expertise cannot be transmitted as such, but must be developed by the students themselves. This study of Caillot and Dumas-Carré distinguishes itself from many other experiments where students were taught a problem solving strategy in teaching explicit and detailed elements of the strategy, which bring meta-knowledge of the domain to the forefront.

In their analysis of transfer Simons and Verschaffel (1992) have also looked into conditions of instruction which could foster transfer. In addition to measures stimulating students to structure their knowledge, they refer to instructional techniques that support or induce decontextualization for instance the various forms of verbalization, mentioned by Prawat (1989), and characterize these techniques of instruction as "expanding the metacognitive consciousness of students" (p. 13). This 'metacognitive consciousness' is the component of metacognition we have labelled high level meta-knowledge. As stated earlier, this component is based on the cognitive component of meta-knowledge, and any effort to stimulate decontextualization must provide some elements of low level meta-knowledge in the form of criteria for distinguishing between context-bound and generalizable elements.

DISCUSSION

This study has attempted to uncover a link between cognitive and metacognitive activities which plays an important role in expertise, but has remained implicit in most discussions on learning and instruction: the concept of *meta-knowledge*, literally 'knowledge about knowledge. One component of meta knowledge is *cognitive* or *low level* or *objective* and comprises the

epistemology of domain knowledge, criteria for judging task performance and knowledge acquisition strategies, and criteria for evaluating understanding. A second component is *metacognitive* or *high level* or *subjective* and comprises knowledge of one's own knowledge and task performance and learning strategies.

The metacognitive component can be considered to be the result of monitoring, an aspect of metacognitive regulation which has been described by several researchers. Monitoring, however requires criteria for judging what is relevant, what is good quality domain knowledge, task performance, and knowledge acquisition strategies, in other words, cognitive meta-knowledge. It follows that efforts to teach metacognitive processes of regulation, such as planning, evaluating, or repairing, will have little result, if not accompanied by instruction in the relevant elements of objective, low level meta-knowledge, which could facilitate monitoring. Teaching which aims at stimulating the growth of expertise, be it by traditional lecturing or tutoring, or by means of the computer coaches and learning environments modern information technology offers us, must be based on a thorough analysis of the skill taught and the various types of knowledge engaged. Much work remains to be done here, and this work requires interdisciplinary cooperation by domain experts, cognitive psychologists, and computer experts.

REFERENCES

Alexander, P.A., Schallert, D.L., & Hare, V.C. (1991). Coming to terms: how researchers in learning and literacy talk about knowledge. *Review of Educational Research 61,* 315-343.

Caillot, M., & Dumas-Carré, A. (1990). Teaching decision-making to solve textbook problems. In H. Mandl, E. de Corte, S.N. Bennet, & H.F. Friedrich (Eds.), *Learning and Instruction, vol. 2:2,* (pp. 67-84). Oxford: Pergamon Press.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chi, M.T., & Bassok, M. (1989). Learning from examples via self-explanations. In L.B. Resnick (Ed.), *Knowing, learning, and instruction. Essays in honour of Robert Glaser,* (pp.251-282). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Chi, M.T., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Chi, M.T., Glaser, R., & Rees, E. (1982). Expertise in problem-solving. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence,* (pp. 7-77). Hillsdale N.J.: Lawrence Erlbaum Ass.

Corte, E. de (1991). Recent onderzoek over leren en instructie: een selectief overzicht. [Recent research on learning and instruction; a selective overview]. *Tijdschrift voor Hoger Onderwijs, 9,* 2-17.

Duit, R. (1991). Students' conceptual frameworks – Consequences for learning science. In S. Glynn, R. Yeany, & B. Britton (Eds.), *The psychology of learning science* (p. 65-85). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Egan, D.E., & Schwartz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition, 7(2),* 149-158.

Entwistle, A., & Entwistle, N. (1992). Experiences of understanding in revising for degree examinations. *Learning and Instruction 2,* 1-22.

Ferguson-Hessler, M.G.M., & Jong, T. de (1987). On the quality of knowledge in the field of electricity and magnetism. *American Journal of Physics, 55,* 492-497.

Ferguson-Hessler, M.G.M., & Jong, T. de (1990). Studying physics text. Differences in study processes between good and poor performers. *Cognition and Instruction, 7,* 41-54.

Ferguson-Hessler, M.G.M., & Jong, T. de. (in press). Het leren van exacte vakken [Learning science subjects]. In P. Spaan & W. Tomic (Eds.), *Onderwijspsychologie [The psychology of teaching].* Dutch Open University and Lemma publishers.

Flavell, J.H. (1976). Metacognitive aspects of problem solving. In L.B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-235). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Gagné, R.M., & Glaser, R. (1987). Foundations in Learning Research. In E.M. Gagné (Ed.), *Instructional Technology: Foundations* (pp. 49-83). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Glaser, R., & Chi, M.T.H. (1988). Overview. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise,* (pp. xv-xxviii). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Groot, A.D. de (1946). *Het denken van den schaker.* [Thinking of the chess player.] Unpublished doctoral dissertation, Amsterdam: Noord-Hollandsche Uitgevers maatschappij.

Groot, A.D. de (1965). *Thought and choice in chess.* Den Haag: Mouton.
Jong, F.P.C.M. de. (1987). Differences in the self-regulation processes between successful and less successful students and the prediction of learning performances in case of comprehension and learning of text. In P.R.J. Simons & G. Beukhof (Eds.), *Regulation of Learning* (pp. 33-45). 's-Gravenhage: SVO.
Jong, T. de, & Ferguson-Hessler, M.G.M. (1993). Types and qualities of knowledge. Paper presented at the 1993 AERA Annual Meeting, Atlanta.
Larkin, J.H. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 311-335). Hillsdale, N.J.: Lawrence Erlbaum Ass.
Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.
Lawson, J. (1984). Being executive about metacognition. In J.R. Kirby (Ed.), *Cognitive strategies and educational performance* (pp. 89-109). New York: Academic Press.
Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.) *The nature of expertise* (pp. 311-342). Hillsdale, N.J.: Lawrence Erlbaum Ass.
Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher,* 16-25.
Prawat, R.S. (1989). Promoting access to knowledge, strategy, and disposition in students: a research analysis. *Review of Educational Research, 59,* 1-41.
Reif, F. (1984). Understanding and teaching problem solving in physics. In *Research on Physics Education: Proceedings of the first international workshop Lalonde les Maures* (pp. 15-53). Paris: Centre National de la Recherche Scientifique.
Reitman, J. (1976). Skilled perception in go: Deducing memory structures from interresponse times. *Cognitive Psychology, 8,* 336-356.
Resnick, L.B. (1989). Introduction. In L.B. Resnick (Ed.), *Knowing, learning, and instruction. Essays in honour of Robert Glaser.* Hillsdale, N.J.: Lawrence Erlbaum Ass.
Shuell, T.J. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56,* 411-436.
Shuell, T.J. (1990). Phases of meaningful learning. *Review of Educational Research, 60,* 531-547.
Simons, P.R.J., & Beukhof, G. (1987).The regulation of learning. In P.R.J. Simons & G. Beukhof (Eds.), *Regulation of Learning* (pp. 3-11). 's-Gravenhage: SVO.
Simons, P.R.J., & Verschaffel, L. (1992). Transfer: onderzoek en onderwijs. [Transfer: research and teaching.] *Tijdschrift voor Onderwijsresearch, 17,* 3-16.
Smith, M.U. (1992). Expertise and the organization of knowledge: unexpected differences among genetic counselors, faculty, and students on problem categorization tasks. *Journal of Research in Science Teaching, 29,* 179-205.
Vermunt, J.D.H.M., & Rijswijk, F.A.W.M. van (1988). Analysis and development of students' skill in zelfregulated learning. *Higher Education 17,* 647-682.

Manuscript ontvangen 17-3-1993
Definitieve versie ontvangen 15-6-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 3, pp. 186-194

Boekbesprekingen

W.J. Nijhof, H.A.M. Franssen, W.Th.J.G. Hoeben en R.G.M. Wolbert (red.). *Handboek Curriculum. Modellen, Theorien, Technologieën.* Amsterdam/Lisse: Swets & Zeitlinger B.V., 1993. ISBN 90-265-1273-2, 428 pagina's. f 79.50

Dat sommige dingen langzaam gaan, bewijst het 'Handboek Curriculum'. Blijkens het woord vooraf is binnen de, inmiddels alweer opgeheven, Onderzoeksthemagroep Curriculum het idee voor een handboek in 1985 ontstaan, hadden in 1988 de potentiële auteurs hun medewerking toegezegd en waren in 1992, het jaar waarin het handboek verscheen, de teksten gereed. Dat het zo lang heeft geduurd, lag onder andere aan verschillen in optiek van waaruit de auteurs wilden meewerken aan het tot stand komen van het handboek. Sommigen wilden een handboek gericht op de praktijk, praktisch en met vele voorbeelden, anderen wensten de nadruk te leggen op de academische inslag met aandacht voor de praktijk (p. 8). Uiteindelijk heeft men zich kunnen vinden in het schrijven van een '... bruikbaar, overzichtelijk, wetenschappelijk en praktijkrelevant handboek' (p. 8). Het woordenboek geeft als definitie van handboek 'boek waarin een beknopte samenvatting van een tak van wetenschap wordt gegeven'. Op het einde van hun voorwoord stellen de redacteuren dat ze een 'goed boek' wilden en de reacties afwachten of dat gelukt is. De intenties van de redacteuren volgend gaat het er dus om te beoordelen of men erin is geslaagd een boek te schrijven dat het wetenschapsgebied beknopt samenvat en dat bruikbaar, overzichtelijk en wetenschappelijk- en praktijkrelevant is.

Deze boekbespreking is langer dan gebruikelijk. Dat is te rechtvaardigen op grond van een tweetal overwegingen. Ten eerste beschikken we in het Nederlands taalgebied nog niet over een standaardwerk dat het curriculumdomein in samenhang beschrijft; het handboek zou in die leemte kunnen voorzien. Ten tweede kent het handboek inhoudelijk sterk verschillende bijdragen. Om de lezer een goed overzicht van de inhoud te geven, is het noodzakelijk aan al deze bijdragen aandacht te besteden. Daarnaast dient ingegaan te worden op de vraag of de intenties van de redacteuren adequaat gerealiseerd zijn.

Het 'Handboek Curriculum' valt in vier delen uiteen. Het eerste deel is een algemeen deel waarin de rationale van het boek uiteen wordt gezet en enkele algemene vraagstukken van het curriculumdomein worden behandeld. Dit deel kent drie hoofdstukken. In de delen twee, drie en vier, met respectievelijk vijf, drie en vijf hoofdstukken, komen achtereenvolgens curriculumvraagstukken op het macro-, meso- en microniveau van het onderwijs aan de orde. Deze heeft men gekruist met taken (ontwerp/ontwikkeling, implementatie, disseminatie en evaluatie) die er ten aanzien van onderwijsontwikkeling in het algemeen en curriculumontwikkeling in het bijzonder zijn uit te voeren. Het handboek heeft één deel minder dan oorspronkelijk de bedoeling was. Een slotdeel bestaande uit '... beschouwingen en analyses van specifieke theoretische onderwerpen uit de curriculumliteratuur' (p. 7) ontbreekt.
Het eerste deel omvat drie bijdragen. Naast de inleiding en rationale van de hand van Nijhof is er een bijdrage van Franssen over de curriculumtheorie in Nederland en van Hoeben over de verhouding curriculumtheorie en -technologie.

In de 'Inleiding en Rationale' wordt de indeling in niveaus waarop curriculumvraagstukken worden onderscheiden, verdedigd. De drie onderscheiden niveaus representeren de hiërarchie van bevoegdheden die in Nederland inzake het onderwijs is te onderkennen en waarbinnen verschillende curriculumprodukten zijn te onderscheiden. Het microniveau betreft de klas en de zich daarin voltrekkende onderwijsleerprocessen. Hier is het de leerkracht die beslist over de vormgeving van het onderwijs. Deze maakt daarbij veelal gebruik van onderwijsleerpakketten/methodes/leergangen. Het mesoniveau betreft de individuele school. Op dit niveau is het bevoegd gezag van de school verantwoordelijk voor de vormgeving van het onderwijs. Bevoegde gezagen nemen deze verantwoordelijkheid waar door het vaststellen van schoolwerkplannen of instituutswerkplannen. Het macroniveau betreft de overheid, die door middel van regelgeving

de vormgeving van het onderwijs benvloedt. Er is sprake van een hiërarchie van bevoegdheden in die zin dat beslissingen op een lager niveau worden gelegitimeerd door beslissingen op een hoger niveau. In concreto betekent dit dat beslissingen op een lager niveau niet in strijd mogen zijn met de beslissingen die op een hoger niveau zijn genomen. Ten aanzien van de op de verschillende niveaus onderscheiden curriculumprodukten (leergang, schoolwerkplan, regelgeving) wordt de vraag gesteld hoe en met welk effect ze kunnen worden ontworpen, ontwikkeld, gebruikt en verspreid. Dit wordt in eerste instantie opgevat als een technologische vraag (p. 19). De curriculumtechnologie is volgens Nijhof de wetenschap die zich bezig houdt met het ontwikkelen en beproeven van methoden en technieken ter optimalisering van het curriculumprodukt (p. 19) terwijl, eveneens volgens Nijhof, de curriculumtheorie zich vooral bezighoudt met de waarden en normen die aan curricula ten grondslag liggen, of met het curriculumdesign, dat wil zeggen de selectie en organisatie van inhouden, of met het reconstrueren van causale verbanden op grond van curriculumontwikkelings- en -implementatieprocessen (p. 26).

Franssen beschrijft eerst de internationale ontwikkelingen met betrekking tot het curriculumdomein. Daarna beschrijft hij de ontwikkelingen in Nederland. Hij onderscheidt tussen didactiek en curriculumtheorie. Franssen maakt echter niet duidelijk op grond waarvan hij dit onderscheid maakt. De door hem genoemde vertegenwoordigers van de didactiek dan wel de curriculumtheorie lijken eerder op grond van vooraf gegeven 'labeling', dan op grond van een verschil in benadering en/of vraagstelling tot een bepaald gebied gerekend te worden. De beschrijving die Franssen van de ontwikkelingen op het curriculumdomein in Nederland geeft, bestaat hoofdzakelijk uit een opsomming van (ontwikkelings)projecten, uitgevoerde onderzoeken en theorieaanzetten. Er worden geen pogingen gedaan te komen tot integratie en beoordeling. Zo wordt er gewag gemaakt van de Ledo- en Curvoprojekten zonder dat wordt aangegeven waaruit de waarde van deze projecten blijkt. Bijvoorbeeld met betrekking theorievorming of voor de onderwijspraktijk. Tegen het eind van zijn bijdrage geeft Franssen een definitie van het begrip curriculumtheorie. Het gaat hem daarbij om de '... meest karakteristieke kenmerken van de curriculumtheorie' (p. 56). Curriculumproblemen hebben iets te maken met de samenleving; het gaat om produkten die het onderwijs moeten sturen; curriculumtheorie is dan '*...een formele en op procedures gerichte benadering van school en scholing binnen een brede maatschappelijke context, met het accent op het ontwikkelen, verspreiden, implementeren en evalueren van curriculumprodukten om op deze wijze innovaties sturend te beïnvloeden.*' (p. 57, cursivering door Franssen).

Hoeben gaat uitgebreid in op de verhouding tussen curriculumtheorie en curriculumtechnologie. Hij schetst eerst een logische relatie. Door middel van reconstructie van curriculumontwikkelings- en -implementatieprocessen waarin doel-middel schema's (technologie) worden gehanteerd, kan men volgens Hoeben komen tot curriculumtheorie waarin causale relaties worden beschreven (p. 69). Na empirische verificatie kunnen deze theorieën worden gebruikt om onderwijs vorm te geven (p. 74). Daarna gaat Hoeben na in hoeverre er sprake is van curriculumtheorieën in de door hem bedoelde zin. Dit blijkt tegen te vallen. De belangrijkste oorzaak zoekt Hoeben in het feit dat werkers op het curriculumdomein te veel probleemgericht en te weinig verklaringsgericht zijn (p. 75). Hoeben beperkt de curriculumtheorieën tot het domein van de kennis dat een theoretische verklaring geeft van de verschijnselen op het gebied van het curriculum (p. 74). Hij zet theorieën die iets anders pretenderen, bijvoorbeeld theorieën die een bepaalde keuze als uitgangspunt nemen of die normatief van aard zijn, buiten haakjes. Reden voor die uitsluiting is dat anders de relatie tussen theorie en technologie nodeloos zou worden gecompliceerd (p. 72). Dat is een merkwaardige reden. Hoeben had moeten aangeven dat de normatieve theorieën niet relevant zijn voor zijn probleem, of ze in het kader van zijn probleem moeten behandelen.

In het tweede deel van het boek worden curriculumproblemen op het macroniveau behandeld. Hier vinden we bijdragen van Akkermans, Van Bruggen, Hoeben, Nieuwenhuis en Plomp & Keursten. Akkermans gaat in op de juridische vragen met betrekking tot het leerplan. Hij maakt duidelijk dat het begrip leerplan in vele gedaanten in de regelgeving (schoolwerkplan, instel-

lingswerkplan) opduikt. Daarbij gaat het om wettelijke voorschriften enerzijds en niet-gereglementeerde beleidsruimte anderzijds. Met het laatste wordt geduid op de mogelijkheid om in het leerplan eisen te formuleren die boven de wettelijke minimumeisen uitgaan. Scholen die van de niet-gereglementeerde beleidsruimte gebruik maken, moeten bedenken dat zij vervolgens gedwongen kunnen worden zich aan hun woord te houden. In die zin is het leerplan een overeenkomst tussen leerling en school (p. 125).

De bijdrage van Van Bruggen draagt de titel 'Instrumenten voor inhoudelijk onderwijsbeleid'. Onder deze titel formuleert Van Bruggen vijf probleemgebieden met betrekking tot de voorschriften voor de inhoud van het onderwijs. (Hoe zien deze voorschriften eruit; is de term onderwijsleerplan geschikt om ze aan te duiden; waar zijn ze te vinden; wie ontwikkelt ze en stelt ze vast; hoe worden ze ingevoerd en verspreid). De probleemgebieden behandelt hij vanuit het ontwikkelingsperspectief en dan nog voornamelijk vanuit de werkzaamheden van de Stichting voor de Leerplanontwikkeling en vanuit de Nederlandse situatie en meer in termen van beschrijving dan analyse (p. 135). Hierdoor is een artikel tot stand gekomen dat maar gedeeltelijk de titel dekt. Het geeft eerder inzicht in de moeilijke verhouding tussen de SLO en het ministerie van onderwijs, dan dat het de mogelijkheden van de overheid tot inhoudelijke vormgeving van het onderwijs beschrijft. Een moeilijke verhouding die wordt veroorzaakt door de vrijheid van onderwijs en de mate van invloed, 'sturen' dan wel 'besturen op afstand', die de overheid op samenleving en onderwijs wenst.

Hoeben onderscheidt vier vormen van evaluatie die worden gebruikt bij het evalueren van het onderwijsbeleid. Het zijn: programma-evaluatie; peiling van het onderwijsniveau; kwaliteitsevaluatie (procesevaluatie) en analytische evaluatie. Bij elk van deze vormen beschrijft hij het onderzoek dat in Nederland met betrekking tot de evaluatie van het onderwijsbeleid is uitgevoerd. Hoeben constateert dat bij evaluatie van het onderwijsbeleid veelal wordt nagelaten toetsbare evaluatiecriteria te formuleren, zodat de overheid zich gemakkelijk kan onttrekken aan de resultaten van evaluatieonderzoek. Volgens Hofstee (1993, p. 117), kan de overheid echter wel aan geloofwaardigheid winnen door de resultaten van evaluatie-onderzoek serieus te nemen. Voorwaarde is dan wel dat de evaluator zich stelt achter het belang van de staatsburger (onderscheiden van het belang van de overheid); uit naam van de staatsburger nagaat wat redelijkerwijs de effecten van beleid hadden moeten zijn en het ideaal van onafhankelijk evaluatieonderzoek blijft nastreven.

Nieuwenhuis gaat in op curriculumontwikkelingsproblemen in het beroepsonderwijs. Het probleem om te komen van beroeps- of functiedomein tot een curriculum wordt door hem beschreven als een vertaalprobleem in vier stappen, waarbij elke stap keuzes en beslissingen veronderstelt. Er is geen sprake van deductie van onderwijsdoelen uit taakomschrijvingen of leerinhouden uit doelen. Het deductieprobleem binnen de curriculumontwikkeling is dus nog niet opgelost (vergelijk Meyer, 1971, 1972).

Plomp en Keursten houden zich bezig met informatietechnologie en het curriculum. Na een opsomming over wat er allemaal wel niet mogelijk is met een computer, gaan ze in op het feitelijke gebruik van de computer in het onderwijs. Ze concluderen dat de inzet van de computer als hulpmiddel in de klas nog zeer beperkt is. Ze verklaren dit uit implementatieproblemen.

In deel drie van het handboek staat het mesoniveau centraal. Het deel bestaat uit drie artikelen van de hand van Franssen & Van Vilsteren, Van der Werf en Jansen. Deze behandelen respectievelijk de planning van het schoolwerkplan, het functioneel gebruik van het schoolwerkplan en de evaluatie van het schoolwerkplan. De eerste twee artikelen hebben (vooral) betrekking op het basisonderwijs, terwijl het laatste artikel in principe in elke vorm van onderwijs gebruikt kan worden.

Franssen & Van Vilsteren beschrijven de vele functies die aan het schoolwerkplan worden toegedacht en het moeizame proces van schoolwerkplanontwikkeling. Ze concluderen dat er voor het schoolwerkplan vier kernfuncties overblijven: legitimering van de doelen, ondersteuning van onderwijsleerprocessen, coördinatie en zelfevaluatie. Afleggen van verantwoording zou moeten worden gerealiseerd in een door de school op te stellen jaarverslag.

Van der Werf stelt in haar bijdrage dat leerkrachten een standaardpraktijk moeten ontwikkelen waarin de kenmerken van de effectieve school en de effectieve instructie zijn verwerkt. Deze praktijk zou moeten worden voorgeschreven in het schoolwerkplan. Door nu de voorschriften van het schoolwerkplan op te volgen, zou een effectieve school kunnen ontstaan. Hierbij kan het schoolwerkplan als evaluatie-instrument dienst doen om het didactisch handelen van leerkrachten individueel en op schoolniveau te evalueren en ter evaluatie van leerprestaties op schoolniveau. Voorwaarde daarvoor is de kwaliteit van het schoolwerkplan zelf, het feit dat leerkrachten zich kunnen vinden in de afspraken zoals neergelegd in het schoolwerkplan en een schoolorganisatie waarbij het schoolteam bereid is tot overleg, intercollegiale samenwerking en peerreview van het didactisch handelen. Of de door Van der Werf voorgestane (harmonieuze) praktijk kan worden gerealiseerd is onduidelijk. Van der Werf geeft aan dat voor haar visie in het nog slechts spaarzame onderzoeksmateriaal, geen eenduidige empirische onderbouwing is te vinden. Een klein beetje vond ze echter wel. De teams in het basisonderwijs kunnen dus aan de slag.

Janssens ziet in het schoolwerkplan een voorwaarde voor het evalueren van het onderwijs. Schoolwerkplanevaluatie ziet hij als een vorm van zelfevaluatie. Deze moet door het team van onderwijsgevenden zonder professionele hulp kunnen worden uitgevoerd. Voorwaarde daarvoor is dat de dataverzameling simpel maar doeltreffend kan worden uitgevoerd. Janssens beveelt het evaluatiemodel van Stake aan.

In het laatste deel van het handboek worden vijf onderwerpen op het micro-niveau behandeld. Nijhof gaat in op het ontwerpen van onderwijsleerpakketten, Knoers op de relatie curriculum, instructie en leren, Van den Akker op de invoering van onderwijsleerpakketten, Van den Berg, Terwel & Wiersma op de evaluatie van leergangen en Suhre op de verspreiding van onderwijsleerpakketten.

Nijhof constateert dat er sinds het begin van deze eeuw steeds meer curriculumontwikkelingsstrategieën zijn ontstaan zonder dat het tot een empirisch gebaseerd, algemeen en generiek te hanteren model is gekomen. Nijhof wijt deze proliferatie van modellen gedeeltelijk aan de uniciteit van de situatie waarin onderwijsleerpakketten moeten worden ontwikkeld. Een uniciteit waarop 'de techniek' van het ontwerpen van onderwijsleerpakketten nog onvoldoende antwoord heeft. Als wordt gekeken naar de taken die de ontwikkelaars van onderwijsleerpakketten moeten uitvoeren, constateert Nijhof dat de verschillende strategieën een groot aantal elementen (o.a. identificatie van curriculumbehoeften; formuleren van eindtermen; opstellen van instructiesequenties; uitvoeren van evaluaties) gemeenschappelijk hebben en dat er sprake is van een optimale volgorde om deze taken uit te voeren (p. 315).

Knoers gaat in op de vraag naar de verhouding tussen curriculum en instructie. Vanuit een informatietheoretische benadering formuleert hij een aantal kenmerken waaraan het curriculum en de instructie moeten voldoen ten einde het leren te bevorderen.

Van den Akker brengt ordening aan in de variabelen die van invloed zijn op de implementatie van onderwijsleerpakketten. Toegespitst op de Nederlandse situatie onderscheidt hij variabelen die een directe invloed uitoefenen (kenmerken van de onderwijsleerpakketten) en variabelen die de voorwaarden bepalen waaronder de pakketten gebruikt worden (schoolkenmerken, externe ondersteuning en de landelijke beleidscontext). De ontwikkelingen in de jaren 70 en 80 overziend, komt Van den Akker tot de conclusie dat onderwijsleerpakketten redelijk overeenkomstig de bedoelingen van de ontwerpers kunnen worden geïmplementeerd, mits de kwaliteit van de produkten voldoende is. Dit laatste houdt in dat de curricula relatief gesloten zijn, duidelijke aanwijzingen voor gebruik bevatten en zijn ontworpen vanuit het perspectief van de leraar-gebruiker.

Van den Berg, Terwel & Wierstra gaan in op de evaluatie van leergangen. Hierbij gaan zij vooral in op methodologische problemen als het formuleren van criteria en standaards, theorie versus praktijkgerichtheid, het causale karakter van evaluatie-onderzoek, evaluatiedesigns en – ten behoeve van evaluatie essentieel – de noodzaak van adequate meting van leergangimplementatie. Met Van den Akker zijn ze van mening dat leergangimplementatie overeenkomstig de

bedoeling van de auteurs het beste lukt bij een gebruikersvriendelijk curriculum. Tegen het eind van hun bijdrage ontwikkelen de auteurs een conceptueel schema voor leergangevaluatie. In dit schema zijn zowel de formele leergangkenmerken die effectief onderwijs bevorderen, alsook de kenmerken van effectieve instructie ondergebracht.

Het slot van het Handboek bestaat uit een bijdrage van Suhre over de verspreiding van onderwijsleerpakketten. Suhre maakt een onderscheid tussen verspreiding op innovatiemotieven en op economische basis. Vanuit de constatering dat verspreiding van onderwijsleerpakketten in Nederland vooral gebeurt door educatieve uitgeverijen, bepleit hij het ontwikkelen van beoordelingscriteria voor onderwijsleerpakketten. Goede onderwijsleerpakketten zijn die die effectief onderwijsgedrag (doelgerichtheid, het geven van correctieve feedback, monitoring en het vergroten van de daadwerkelijk bestede leertijd) stimuleren.

Tot zover de typering van de bijdragen. Nu de vraag of men erin is geslaagd een boek te schrijven dat het wetenschapsgebied curriculum beknopt samenvat en dat bruikbaar, overzichtelijk, wetenschappelijk en praktijkrelevant is.

Beknopte samenvatting van het wetenschapsgebied
Het eerste dat om een oordeel vraagt is of het handboek een beknopte samenvatting van het wetenschapsgebied geeft. Dit valt in drie elementen uiteen, namelijk de vraag of de samenvatting beknopt is, de vraag of het gaat om een samenhangende beschrijving en de vraag of het gehele domein beschreven is (volledigheid).

Beknopt
Gezien de hoeveelheid literatuur die er in de loop van eeuwen vanuit de wetenschap met betrekking tot het curriculum is geschreven, mag 428 pagina's als een beknopte beschrijving gelden.

Samenhang
Door het kruisen van de bevoegdhedenniveaus (macro, meso en micro) met curriculumontwikkelingstaken (ontwerp/ontwikkeling, implementatie, disseminatie en evaluatie) is een matrix ontstaan aan de hand waarvan de verschillende aspecten met betrekking tot de curriculumontwikkeling goed kunnen worden besproken. Het wordt duidelijk waarom een curriculumprodukt op het ene niveau iets anders is, en noodzakelijkerwijs iets anders moet zijn, dan op het andere niveau en daaruit voortvloeiend, waarom op het ene niveau andere ontwikkelingsactiviteiten noodzakelijk zijn dan op het andere. In die zin is er samenhang. Men kan zich echter afvragen of geen appels met peren worden vergeleken als we ontwikkelingsprocessen op het microniveau met die op het macroniveau vergelijken. Hetzelfde probleem heeft gespeeld binnen de didactiek. Flechsig en Haller (1975) onderscheidden vijf niveaus van didactisch handelen. Het hoogste niveau kwam grofweg overeen met wat in het Handboek Curriculum als het macroniveau wordt aangeduid. De consequentie daarvan was bijvoorbeeld dat de Minister van Onderwijs als didacticus moest worden aangemerkt als deze wetten opstelde of uitvoerde. Nu is er veel mogelijk, maar om in het onderwijsbeleid didactisch handelen te zien, leidt tot grove vormen van cognitieve dissonantie. Hetzelfde geldt voor curriculumontwikkeling en -implementatie. Het wordt tijd de term curriculumdomein te reserveren voor de gewenste inhoud, vormgeving en uitvoering van het onderwijs zoals dat (met behulp van onderwijsleerpakketten/leergangen/methoden) gestalte krijgt op klasse/groepsniveau en deze te onderscheiden van beschrijvings- en planningsmodellen op macro- en mesoniveau. Dit voorstel sluit aan bij meer theoretische beschouwingen omtrent het curriculum (Barrow & Woods, 1988) en de opvatting dat instructie-effectiviteit een functie is van zowel de inhoud alsook het leerkrachtgedrag (Creemers, 1991). Een matrix die de essentie van het curriculumdomein beschrijft, lijkt die die in het Handboek op pagina 28 wordt beschreven. Hier worden de curriculumontwikkelingstaken gerelateerd aan de variabelen input, context, proces en resultaat (output). In deze matrix kunnen opvattingen over het curriculum tot de designtaken en daarmee tot de inputvariabelen worden gerekend.

Volledigheid

Eén manier om tegen volledigheid aan te kijken is of het curriculumdomein uit één of meerdere visies is beschreven. In het Handboek wordt het curriculumdomein benaderd vanuit een technologische kennisinteresse. Curriculumontwikkeling wordt opgevat als een techniek en de wetenschap daarvan als een technologie. Het begrip technologie wordt echter niet eenduidig gedefinieerd. Nijhof stelt de (curriculum)technologie tegenover zowel normatieve theorie alsook causale theorie. Franssen definieert curriculumtheorie als een formele op procedures gerichte benadering en Hoeben beperkt curriculumtheorie tot de reconstructie van causale verbanden op grond van doel-middelschema's (technologie). Een bonte verzameling im- en expliciete definities over wat theorie of technologie is, dan wel zou moeten zijn. De pre-occupatie met technologie heeft ertoe geleid dat normatieve theorieën in het handboek nauwelijks zijn besproken. Door Nijhof niet. Fransen komt er niet echt aan toe, omdat hij uiteindelijk op een formele definitie van curriculumtheorie (wil?) uitkom(t)en en Hoeben sluit de normatieve theorieën uit omdat ze de verhouding theorie/technologie onnodig compliceren. In deze zin is het Handboek onvolledig. Een belangrijke vraag die in deze context thuis hoort, maar niet expliciet aan de orde wordt gesteld, is of curriculumtechnologie is op te vatten als een algoritme of als een heuristiek. Uit het handboek spreekt naar mijn mening het ideaal van het algoritme, terwijl het erop lijkt dat curriculumtechnologie een heuristiek is en ook wel nooit zal uitgroeien tot een algoritme. Denk bijvoorbeeld aan het nog steeds niet opgeloste deductieprobleem. Dat curriculumtechologie in wezen een heuristiek is, is ook een verklaring voor het door Nijhof geconstateerde ontbreken van een generiek model voor curriculumontwikkeling gebaseerd op empirisch onderzoek. Deze verklaring lijkt mij plausibeler dan het wijzen op de uniciteit van de situatie waarin onderwijsleerpakketten moeten worden ontwikkeld. Bij het volgen van de heuristiek speelt onderzoek met behulp van sociaal-wetenschappelijke methoden een rol. In de ontwikkelingscontext kan onderzoek een rol spelen ten behoeve van need-assessment of formatieve evaluatie. Bij summatieve evaluatie en/of programma-evaluatie is onderzoek noodzakelijk.

Een tweede manier om tegen volledigheid aan te kijken, is om te zien of alle cellen van de matrix gevormd door de dimensies niveau (macro, meso, micro) en ontwikkelingstaken (ontwerp/ontwikkeling, implementatie, disseminatie, evaluatie) gevuld zijn. Als men deze exercitie uitvoert, blijken tien van twaalf cellen gevuld te worden. De cellen gevormd door combinaties van de waarden macro/disseminatie en meso/disseminatie worden niet gevuld. Dit zijn cellen die van nature leeg zijn. Het handboek is in deze zin dus volledig.

Bruikbaar

De vraag of het Handboek bruikbaar is hangt af van het doel waarmee men het ter hand neemt. Als men het wil gebruiken om op de hoogte te raken van de stand van zaken op het curriculumdomein (zoals gedefinieerd door de auteurs) dan biedt het boek daartoe volop de gelegenheid. Er zijn artikelen die een goed overzicht van de theorieontwikkeling geven en er zijn artikelen die de ontwikkelingen van de laatste twintig jaar in en om het onderwijs beschrijven. Beide soorten zijn goed leesbaar en up-to-date. Het is jammer dat het laatste hoofdstuk niet is geschreven. Met name de overzichtsartikelen bieden mijn inziens interessante stof voor een meta-analyse met betrekking tot de vraag waar twintig jaar aandacht voor het curriculum toe heeft geleid. Daar hadden vragen aan de orde kunnen komen zoals gesteld onder 'samenhang' en 'volledigheid'.

Als het doel is het handboek te gebruiken ten behoeve van ontwerpen en/of uitvoeren van onderzoek of het ontwikkelen van curricula, valt de vraag naar de bruikbaarheid samen met de vraag naar de wetenschappelijke en praktische relevantie van het boek. Deze vraag wordt verderop behandeld.

Overzichtelijk

Het ordenen van de artikelen met behulp van de dimensies niveau en ontwikkelingstaak, maakt het handboek overzichtelijk. Het trefwoordenregister maakt het mogelijk dat iemand die geïnteresseerd is in een bepaald onderwerp, zich snel kan oriënteren.

Wetenschappelijk en praktijkrelevant

Rest de vraag of het Handboek ook van wetenschappelijk en praktisch belang is. In de inleiding suggereren de redacteuren een tegenstelling tussen wetenschappelijke en praktische relevantie. Zo wilden sommige auteurs een handboek gericht op de praktijk, praktisch en met vele voorbeelden, terwijl anderen de nadruk wensten te leggen op de academische inslag (p. 8). Het is onduidelijk om welke praktijk het hier gaat. Gezien de definitie van het curriculumdomein als het plannen en uitvoeren van onderwijsleerprocessen door leerkrachten en studenten waarbij ook nog verschillende niveaus worden onderscheiden (p. 15), kan het om curriculumontwikkeling en daaraan te onderscheiden taken gaan, alsook om het didactisch handelen van leerkrachten (letterlijk genomen zelfs het leren van leerlingen). Gezien de inhoud van het Handboek nemen we aan dat met de praktijk de praktijk van de curriculumontwikkeling wordt bedoeld. Als we ons daartoe beperken, is de kloof tussen praktijk en theorie minder groot dan wordt gesuggereerd, althans behoort deze minder groot te zijn. Als het gaat om het ontwikkelen (met inbegrip van formatieve evaluatie) en beproeven van onderwijsleerpakketten (met inbegrip van summatieve evaluatie) dan spreken we in onze ogen over het werk van wetenschappelijk geschoolde professionals (vergelijk Westerhof, 1983). Deze dienen net als hun wetenschappelijk georiënteerde collega's op de hoogte te zijn van de wetenschappelijke stand van zaken op hun vakgebied. Dit betekent dat de oorspronkelijke vraag moet worden geherformuleerd. De vraag moet luiden of het Handboek Curriculum relevant is voor curriculumprofessionals, die als ontwikkelaar of wetenschapper werkzaam zijn. Deze vraag kan om een aantal redenen positief beantwoord worden. Ten eerste geeft het handboek een gedegen overzicht van wat er de laatste twintig jaar op het gebied van curriculum aan de orde is geweest. In het handboek worden de ups en niet het minst belangrijk ook de downs beschreven. Er valt dus uit te leren hoe het moet, maar vooral ook hoe het niet moet. Daarbij is een trend te bespeuren van een idealistische start met veel schoolnabije curriculumontwikkeling naar meer aandacht voor het produkt van de curriculumontwikkeling. Ten tweede worden belangrijke theoretische vraagstukken, ondanks een aantal kritische opmerkingen die misschien het tegendeel suggereren, op een heldere manier besproken. Deze constatering geldt zowel voor het eerste deel, alsook voor de bijdragen met betrekking tot de verschillende niveaus. Ten derde blijkt dat er een consensus groeit over de vraag wat de kenmerken van curricula (onderwijsleerpakketten) moeten zijn, willen ze worden geïmplementeerd overeenkomstig de bedoelingen van de constructeurs. Dit is een belangrijk gegeven als we bedenken dat de belangstelling voor het curriculum nauw verbonden is met de vraag hoe het onderwijs verbeterd kan worden.

LITERATUUR

Barrow, R. & Woods, R. (1988). *An introduction to philosophy of education* (3d ed.). London & New York: Routledge.

Creemers, B.P.M. (1991). *Effectieve instructie. Een empirische bijdrage aan de verbetering van het onderwijs in de klas.* 's Gravenhage: Instituut voor Onderzoek van het Onderwijs S.V.O.

Flechsig, K-H. & Haller, H-D. (1975). *Einführung in didaktisches Handeln.* Stuttgart: Ernst Klett Verlag.

Hofstee, W.K.B. (1993). Visies op de rol van de onderwijsonderzoeker in het doelstellingendebat. In B.P.M. Creemers, W.Th.J.G. Hoeben, J.L. Peschar & J. Snippe (red.). *Wat is onderwijsonderzoek waard?* De Lier: Academisch Boeken Centrum, p. 115-120.

Meyer, H.L. (1971). Das ungelöste Deduktionsproblem in der Curriculumforschung. In F. Achtenhagen & H.L. Meyer (Hrsg.). *Curriculumrevision. Möglichkeiten und Grenzen.* München: Kösel.

Meyer, H.L. (1972). *Einführung in die Curruculum-Methodologie.* München: Kösel.

Westerhof, K.J. (1983). Algemene theorie van didactiek en curriculumontwikkeling. In B.P.M. Creemers e.a. (red.). *Losbladig Onderwijskundig Lexicon.* Alphen aan de Rijn: Samson, CO 1000 p. 1-15.

K.J. Westerhof
Rijksuniversiteit Groningen
Vakgroep PA&O
Grote Rozenstraat 38
9712 TJ Groningen

Wolfhagen, H.A.P. (1993). *Kwaliteit van klinisch onderwijs*.
Proefschrift RUL, Maastricht: Universitaire Pers.
ISBN 90-5278-052-8

Op 18 februari 1993 promoveerde in Maastricht Ineke Wolfhagen. Na in Nijmegen te zijn afgestudeerd in de richting Interdisciplinaire Onderwijskunde is zij, na een vijftal jaren bij het CITO werkzaam te zijn geweest, vanaf 1988 verbonden aan de vakgroep Onderwijsontwikkeling en -research van de Rijksuniversiteit Limburg te Maastricht. Haar aandachtsgebied daar betreft vooral de programma-evaluatie van de medische opleiding, in het bijzonder van de co-assistentschappen: het onderwerp van haar promotie-onderzoek.

Onder de enigszins weids aandoende titel 'Kwaliteit van klinisch onderwijs' beschrijft het onderzoek de ontwikkeling en implementatie van een door co-assistenten in te vullen vragenlijst, te gebruiken als gestandaardiseerd instrument voor de evaluatie van de co-assistentschappen. Voor niet-ingewijden in de artsopleiding: co-assistentschappen zijn praktijkstages van wisselende omvang, die studenten geneeskunde na hun doctoraal-examen volgen in een tiental verschillende afdelingen van het academisch ziekenhuis, van regionale ziekenhuizen en van extramurale gezondheidszorginstellingen. Na het succesvol doorlopen hebben daarvan, vindt bevordering tot arts plaats. Doordat de co-assistentschappen zo gevarieerd zijn in aard, omvang en locatie is een adequate kwaliteitsbewaking ook van grote praktische betekenis.

Het proefschrift, dat in totaal 207 bladzijden telt, begint met een uitvoerige geschiedenis van het medisch onderwijs in Nederland en – vooral met betrekking tot de jongste geschiedenis – ook in internationaal perspectief. Een afzonderlijk hoofdstuk is gewijd aan de Maastrichtse opleiding, met name de plaats van de co-assistentschappen daarin; een onderwerp, dat in de vele publicaties over het Maastrichtse onderwijssysteem doorgaans wat onderbelicht blijft. Dat laatste is niet toevallig en het is daarom een gelukkige gedachte geweest, dat – helaas wat korter dan het voorafgaande – in een afzonderlijke paragraaf aandacht wordt besteed aan de verschillende soorten onderwijsorganisaties, die een interessant licht werpen op de verschillen tussen, wat voor kort nog heette, de 'eerste' en de 'tweede' fase van het medisch opleidingscontinuüm; meer inhoudelijk gezegd: tussen theorie-leren en praktijk-leren. Dit laatste dient plaats te vinden in de tweede fase gedurende de co-assistentschappen en juist dáár stelt de onderzoekster de (onderwijs)kwaliteitsvraag.

Uit de mogelijkheden van een proces-, dan wel een produkt-gerichte benadering van deze vraag, kiest de onderzoekster voor de eerstgenoemde; deels uit theoretische overwegingen, deels vanuit de meer praktische constatering, dat er nog geen concrete en duidelijk geformuleerde eindtermen bestaan. Een tweede, voor het onderzoek belangrijke, beslissing betreft de keuze van co-assistenten zèlf als beoordelaars van de kwaliteit van het genoten onderwijs. Deze beslissing komt in twee opzichten tot uitdrukking: bij de selectie van de te ondervragen domeinen en door co-assistenten de vragen te laten beantwoorden.

Na de beschrijving van de constructie van de vragenlijst en de keuze voor een (gemengd) normeringssyteem, wordt uitvoerig aandacht besteed aan de kenmerken van de organisatie waarbinnen de evaluatie dient plaats te vinden, met name de afspraken en procedures voor bespreking en follow-up, die een onmisbare (en vaak gemiste!) voorwaarde vormen voor een op verandering en innovatie gericht gebruik van evaluatiegegevens. Enige concrete casuïstiek van meer en minder geslaagde voorbeelden daarvan worden beschreven.

De studie wordt besloten met de beschrijving van het onderzoek naar de (construct)validiteit en de betrouwbaarheid (uitgaande van de generaliseerbaarheidstheorie) van het evaluatie-instrument. De conclusies van beide zijn positief.

'Kwaliteit van klinisch onderwijs' is een beknopt en helder weergegeven verslag van een elegant onderzoek, met grote praktische bruikbaarheid en ook voor niet-onderwijskundig geschoolde docenten relevant en begrijpelijk. Een schoolvoorbeeld van toegepast onderwijskun-

dig onderzoek met belangrijke lessen voor onderwijsinstellingen die van evaluatie meer dan een rituele dans willen maken. Een juweeltje dus.

J.C.M.Metz
K.U. Nijmegen
Faculteit der Medische Wetenschappen

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 4, pp. 197-207

Beboeten van spijbelaars: versterking van discipline of verzwakking van motivatie?[1]

Frank M. Weerman
*Rijksuniversiteit Groningen, Vakgroep Criminologie**

ABSTRACT

In this article, two contrasting views of the relationship between truancy and functioning at school are tested. One view, the 'discipline'-model, holds that a lack of discipline results in malfunctioning at school. The other view, the 'demotivation'-model, holds that malfunctioning at school results in less motivation and that demotivation results in truancy and rebellion. The two views are presented as causal models for the variables 'failing', 'school-appreciation', 'motivation', 'truancy' and 'rebellion'. LISREL-analysis shows that the 'demotivation'-model fits the data better than the 'discipline'-model. This result implies that harsher punishing alone is not the right solution for serious truancy.

1. INLEIDING

In november 1992 stuurde staatssecretaris Wallage een ontwerp-Leerplichtwet naar de Tweede Kamer (Volkskrant, 7-11-'92), waarin de mogelijkheid wordt gecreëerd om spijbelende scholieren een boete te geven. In het voorstel zijn niet meer alleen de ouders verantwoordelijk voor schoolverzuim, maar kunnen ernstige spijbelaars ook zelf voor de rechter worden gebracht. Deze strengere aanpak wordt noodzakelijk geacht omdat, volgens Wallage, chronisch spijbelen vaak voorafgaat aan voortijdig schoolverlaten. Verder dienen scholen in hun schoolwerkplan op te nemen hoe ze spijbelen aanpakken en worden burgemeesters en wethouders verplicht om eens per jaar in de gemeenteraad verantwoording af te leggen over hun 'anti-spijbelbeleid'.

Het voorstel sluit aan bij de discussie over normen en waarden, die door de ministers van justitie en onderwijs is aangezwengeld. Ook Wallage wil (met hardere hand) normen en waarden herstellen, in casu schoolverzuim tegengaan. Spijbelen wordt niet langer beschouwd als een probleem dat de scholen zelf moeten oplossen, maar nu ook als een wetsovertreding waar justitie op moet reageren.

Het probleem van het schoolverzuim kan in principe op twee manieren benaderd worden. In het wetsvoorstel wordt spijbelen vooral gezien als een discipline-probleem. Een andere redenering is echter ook mogelijk, namelijk dat het onderwijs cq. de school onvoldoende gezag heeft om de leerlingen aan zich te binden. Het schoolverzuim zou dan veel eerder gedefinieerd moeten worden als een gezagsprobleem.

Het is de vraag welke visie de juiste is. Daarnaast kan men zich afvragen of het beboeten van spijbelaars wel zo'n verstandig idee is. In dit artikel zal ik daarom ingaan op de volgende vragen:

- Is het juist om schoolverzuim en in het algemeen regelovertredend gedrag op school zo'n centrale plaats in de causale keten te geven? Moet schoolverzuim gezien worden als een afhankelijke variabele ten opzichte van slecht schoolfunctioneren of is spijbelen juist een uitvloeisel van slecht functioneren op school?
- Gaat het bij spijbelen om een discipline-probleem of om een gezagsprobleem?
- Valt er een remmend effect op het schoolverzuim te verwachten van een justitiële aanpak van spijbelaars?

* Adres: Vakgroep Criminologie, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen.

Bij het beantwoorden van deze vragen is gebruik gemaakt van gegevens die afkomstig zijn uit het longitudinale onderzoek 'Onderwijs en Delinquentie' van de vakgroep Criminologie van de Rijksuniversiteit Groningen. Over het gehele onderzoek wordt elders gerapporteerd. De empirische gegevens die voor dit artikel zijn gebruikt, hebben betrekking op de volgende variabelen: 'schoolprestaties', 'motivatie', 'schoolbeleving', 'wangedrag op school' en 'spijbelen'.

2. THEORIE

De aannames achter het wetsontwerp van Wallage

Achter het wetsontwerp gaat, zoals bij veel politiek beleid, een aantal onuitgesproken aannames over de werkelijkheid schuil. Eén daarvan heeft te maken met de oorsprong van chronisch schoolverzuim. De staatssecretaris lijkt te veronderstellen dat het probleem primair ligt bij de leerlingen die zich aan de regels onttrekken.

Verder wil de staatssecretaris met een hardere aanpak van het schoolverzuim tevens andere ongewenste ontwikkelingen bestrijden, zoals voortijdig schoolverlaten. Het spijbelgedrag van de leerlingen wordt centraal gesteld en voortijdig schoolverlaten wordt blijkbaar gezien als een afgeleide van schoolverzuim. Of in meer algemene bewoordingen gesteld, het overtreden van schoolregels (wangedrag in de klas, spijbelen) heeft slecht schoolfunctioneren tot gevolg (slechte prestaties, voortijdig schoolverlaten).

Tenslotte is een andere aanname die logisch aan het voorstel ten grondslag ligt het idee dat een justitiële aanpak een remmend effect zal hebben op spijbelen en voortijdig schoolverlaten. De Leerplichtwet wordt hiermee beschouwd als een nuttig instrument bij het verbeteren van het onderwijs.

Schoolverzuim als gevolg van demotivatie

Voor het onderzoek 'Onderwijs en Delinquentie' was het uitgangspunt de redenering dat 'slecht' presterende leerlingen door anderen en/of door henzelf als falers worden beschouwd. Dit kan leiden tot demotivatie en een negatieve houding ten opzichte van de school. Op zijn beurt heeft dit weer rebellie tegen de school (wangedrag in de klas en spijbelen) tot gevolg (Nijboer en Dijksterhuis, 1983).

Een iets andere nadruk legt de criminologische theorie van de maatschappelijke kwetsbaarheid, die werd ontwikkeld door Walgrave et al. (Vettenburg, Walgrave en van Kerckvoorde, 1984; Walgrave, 1993). Deze theorie heeft sterk interactionistische elementen. Er wordt vanuit gegaan dat leerlingen het onderwijs binnenkomen met verschillende startposities, doordat ze in meer of mindere mate beschikken over 'cultureel en intellectueel kapitaal'. De ene groep leerlingen krijgt vooral te maken met de belonende aspecten van het onderwijs, terwijl de andere groep (maatschappelijk kwetsbare) leerlingen vooral te maken krijgt met de bestraffende aspecten van het onderwijs. Dit leidt ertoe dat beide groepen uit elkaar groeien: de maatschappelijk kwetsbaren worden gedemotiveerd door het ontbreken van beloningen en gaan daardoor weer slechter presteren.

Een vergelijkbaar proces, waarbij door slechte prestaties de motivatie vermindert, wordt veroorzaakt door de zgn. vergelijkende beoordeling. Doordat zwakke leerlingen voortdurend minder dan gemiddeld of onvoldoende scoren, worden zij op een systematische manier ontmoedigd (De Groot, 1966; Mooij,1979).

Twee modellen

De twee hiervoor genoemde benaderingen veronderstellen verschillende causale relaties tussen de onderzochte variabelen. Ze zijn weer te geven in twee concurrerende modellen die hierna zullen worden beschreven. Met behulp van het computerprogramma LISREL kan vervolgens worden bepaald welk model het beste past bij de empirisch waargenomen verbanden.

De factor 'tijd' wordt op twee manieren in de modellen verwerkt. Ten eerste moet worden opgemerkt dat de variabelen prestaties, wangedrag en spijbelen retrospectief zijn, ze hebben

betrekking op het gedrag in de afgelopen periode, aan het interview voorafgaand. De andere variabelen, motivatie en schoolbeleving, zijn meer momentopnamen. Ten tweede wordt een tijdsverschil gecreëerd door van het ene jaar de onafhankelijke en intermediërende variabelen in de modellen op te nemen en van het erop volgende jaar de afhankelijke variabelen. In de figuren 1 en 2 is het tijdstip van meting aangegeven door de aanduidingen 'T1' en 'T2'.

Model 1 in Figuur 1 volgt de benadering van de staatssecretaris.

Wangedrag en spijbelen zijn in het 'discipline-model' beide een uitingsvorm van regelovertredend gedrag, ze hangen dus ongericht samen. Of, in termen van LISREL, ze hebben een gemeenschappelijke onverklaarde variantie. Spijbelen en wangedrag hebben beide slecht schoolfunctioneren (mindere prestaties gedurende het jaar erop) tot gevolg. Het gebrek aan discipline heeft ook een effect op de inzet op school: wangedrag en spijbelen hebben een effect op motivatie. De schoolbeleving tenslotte heeft alleen een effect op de motivatie.

Model 2 in Figuur 2 volgt de gedachtengang die de nadruk legt op demotivatie.

'Schoolprestaties' is hier de onafhankelijke variabele. Minder goed presteren heeft een negatieve invloed op schoolbeleving en op motivatie. Een negatieve schoolbeleving vermindert de motivatie. Weinig motivatie en een negatieve schoolbeleving vermeerderen het wangedrag en het schoolverzuim gedurende het jaar erna. Wangedrag en spijbelen tenslotte hangen weer ongericht samen. Ze zijn beide te zien als uitingsvorm van rebellie tegen de school.

MODEL 1

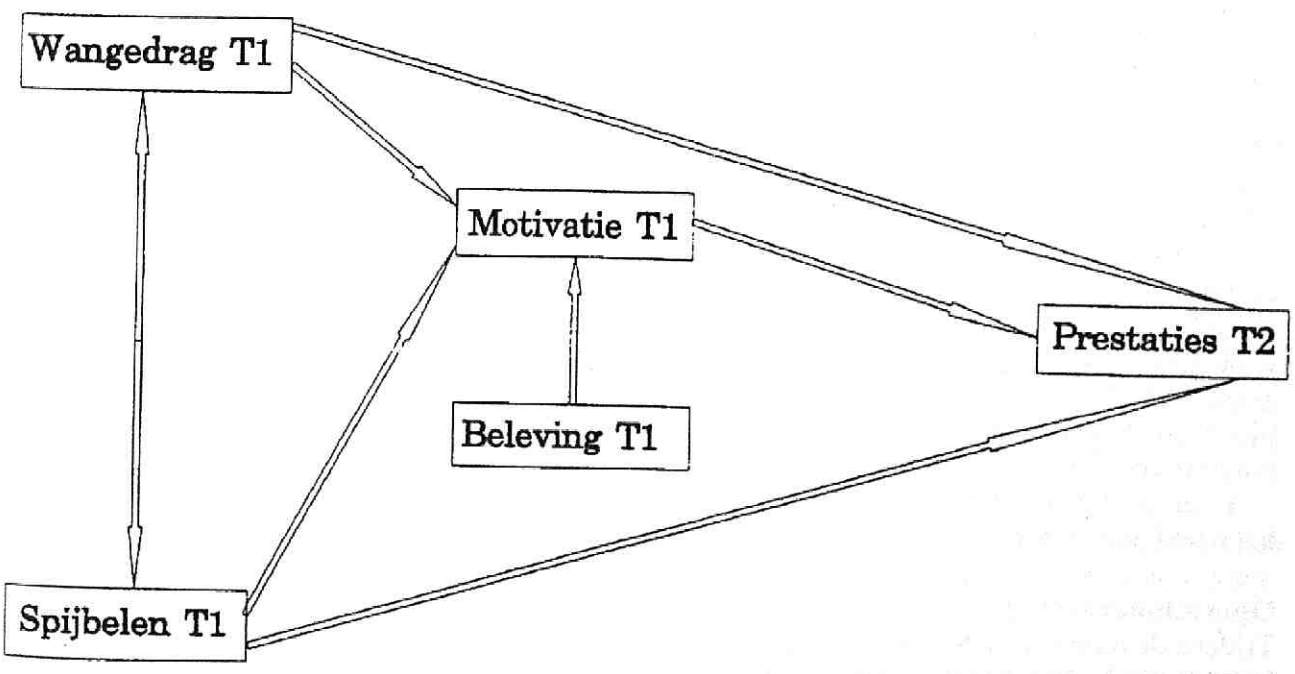


Fig. 1. Pijldiagram van model 1: 'discipline'-model.

MODEL 2

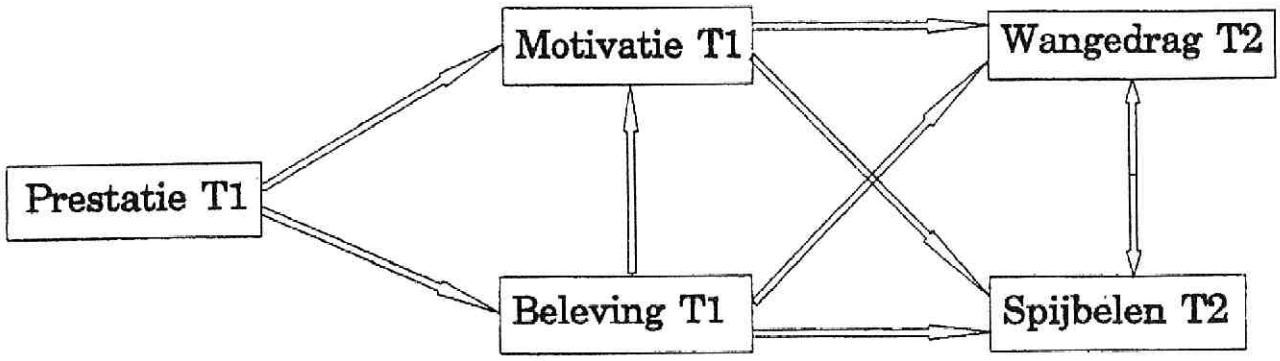


Fig. 2. Pijldiagram van model 2: 'demotivatie'-model.

3. OPZET VAN HET ONDERZOEK

Het panel

Zoals gezegd richtte het onderzoek waaruit de gegevens voor dit artikel afkomstig zijn zich in eerste instantie op delinquentie. Dit is de reden geweest voor een wat afwijkende steekproeftrekking. Slechts een klein deel van de scholieren geeft namelijk veelvuldige en ernstige delicten op. Deze groep was echter wel het meest interessant voor een onderzoek naar delinquent gedrag. Om van die categorie voldoende aantallen te krijgen en de kosten toch binnen de perken te houden werd een zogenaamde 'purposive sample' getrokken. Deze 'steekproef' vormt het panel voor de longitudinale studie.

Het panel werd in twee stappen geselecteerd. Allereerst werden in drie middelgrote gemeenten in het noorden, het oosten en het midden van het land alle (gewone) openbare basisscholen aangeschreven. Van de 83 benaderde scholen, verleenden 71 hun medewerking. Op die 71 scholen werden in principe alle leerlingen uit de hoogste groep geënquêteerd. De respons bij die enquête was 90%, het aantal ingevulde vragenlijsten bedroeg 1330. De enquête bestond uit vragen over de gezinsachtergrond en het schoolfunctioneren. Omdat een oververtegenwoordiging van delinquente leerlingen werd nagestreefd, maar feitelijke gegevens over delinquent gedrag nog niet beschikbaar waren, zijn de leerlingen ingedeeld in een aantal 'risico-categorieën'. Deze indeling was gebaseerd op de gegevens uit de enquête en werd dus gemaakt aan het eind van de basisschool. Uit de hogere risico-categorieën werden alle leerlingen geselecteerd, uit de middencategorieën de helft en uit de lagere categorieën een kwart. Aangezien het niet de bedoeling was om populatie-parameters te schatten, waren de voordelen van deze methode groter dan de nadelen.

Het panel bestond uiteindelijk uit 396 jongens en 104 meisjes. Zij werden jaarlijks – van 1986 t/m 1990 – uitgebreid en individueel geïnterviewd, waarbij onder andere gebruik werd gemaakt van vragen uit de SVL-lijst die ontwikkeld is door Smits en Vorst (1982). De tussentijdse uitval was zeer acceptabel: na vijf interviewronden was in totaal slechts 15% van de jongens en 21% van de meisjes afgehaakt.

In dit artikel worden alleen de gegevens van de jongens geanalyseerd. Het aantal meisjes in het panel is te gering om in het kader van deze vraagstelling zinvolle berekeningen uit te voeren.

Operationalisering

Tijdens de interviews zijn een groot aantal vragen gesteld, die in eerste instantie vooral criminologisch van belang waren. Daarbij is gebruik gemaakt van eerder criminologisch onderzoek. De operationalisatie wijkt daardoor in zekere mate af van hetgeen in de onderwijsresearch gebruikelijk is. Bovendien moet worden bedacht dat de opzet van het onderzoek in 1985 plaatsvond,

waardoor alleen van de op dat moment bekende schalen gebruik gemaakt kon worden. Het was verder nodig om de grote hoeveelheid gegevens uit het onderzoek samen te vatten in een beperkt aantal schalen. Deze schalen zijn in het algemeen geconstrueerd op basis van inhoudelijke overwegingen. Hieronder wordt besproken op welke wijze de onderzochte begrippen zijn geoperationaliseerd.

Prestaties: deze variabele is samengesteld uit de antwoorden op de volgende vragen: op welk schoolniveau bevindt men zich (vwo en havo zijn hoog, de anderen laag); rekent men zich tot de goede, de gemiddelde of de slechte leerlingen; is men overgegaan; gaat het al met al goed op school (nee, gaat wel, ja). Vanwege de gevoeligheid van het onderzoek (strafbare feiten) en de anonimiteit van de respondenten konden de rapportcijfers niet bij de scholen worden opgevraagd. Bovendien ging het vooral om de subjectieve waardering van de eigen prestatie. Die kon echter niet als losstaand worden beschouwd. De vraag of men zich tot de goede of minder goede leerlingen rekent, is dan niet voldoende. Een zittenblijver kan in het nieuwe jaar tot de betere leerlingen behoren, en iemand die het slecht doet op het VWO zal door de bank genomen minder het gevoel hebben te falen dan iemand die het slecht doet op het LBO. Met de geconstrueerde schaal wordt derhalve rekening gehouden met meerdere aspecten van iemands subjectieve waardering van zijn "schoolprestaties". Het betekent ook dat de gehanteerde indicatoren niet te beschouwen zijn als items op een één-dimensionele schaal.

Schoolbeleving: deze schaal bestaat uit een totaalscore van drie subschalen, waarbij leerlingen een waardering (in de vorm van een cijfer van 1 tot en met 10) moesten geven over het schoolklimaat, de leraren en de gegeven vakken. De beoordeling van het schoolklimaat bestaat uit cijfers voor 'het schoolgebouw', 'de klaslokalen' en 'hoe leuk de school is'. De beoordeling van de leraren bestaat uit cijfers voor 'duidelijkheid van uitleg', 'orde in de klas', 'hoe aardig de leraren zijn' en 'of leraren goed helpen'. De beoordeling van de gegeven vakken bestaat uit cijfers voor de interessantheid van 'praktijkvakken' (handvaardigheid,tekenen etc.), 'taalvakken' en 'exacte vakken' (wis-, natuur- en scheikunde).

De meeste schoolbelevings-vragen uit de SVL-lijst waren niet in het onderzoek opgenomen. Uit eerder onderzoek (Dijksterhuis en Nijboer, 1988) was gebleken dat van de schoolbelevingsvariabelen uit de SVL 'relatie met de leraren' de sterkste samenhang vertoonde met wangedrag op school. Om het interview niet te lang te maken is alleen die schoolbelevings-variabele uit de SVL-lijst in het interview opgenomen, en dan als indicator voor de relatie met de leerkrachten. Voor het meten van 'schoolbeleving' werd later een nieuwe schaal gemaakt.

Motivatie: Deze schaal bestaat uit de totaalscore van de SVL-schalen 'huiswerk-attitude', 'leertaakgerichtheid' en 'concentratie in de klas'. Deze drie subschalen vormen in de SVL-lijst de schaal 'motivatie'. 'Huiswerkattitude' bestaat uit acht vragen over de mate waarin een leerling zich inzet voor het huiswerk. Een uitspraak is bijvoorbeeld 'Meestal begin ik uit mezelf aan mijn huiswerk' (antwoordcategorieën: dat is zo, dat weet ik niet, dat is niet zo).

'Leertaakgerichtheid' bestaat uit acht vragen over de mate waarin de leerling zich inzet voor het schoolwerk en goed en hard werken voor school belangrijk vindt. Een uitspraak: 'ik werk hard voor alle vakken om goede cijfers te krijgen. 'Concentratie in de klas' bestaat uit acht vragen over de mate waarin de leerling zich kan concentreren op schooltaken tijdens de lesuren. Een uitspraak is: 'bij de meeste lessen let ik goed op'. Een uitvoeriger beschrijving van de drie schalen is te vinden in Smits en Vorst, 1982.

Wangedrag: deze schaal bestaat uit de vraag naar de frequentie waarmee men in het afgelopen jaar uit de klas is gestuurd (vaak, soms, nooit), hoe vaak men strafwerk heeft gekregen (vaak, soms, nooit) en of men het afgelopen jaar wel eens van school is gestuurd.

Spijbelen: de schaal voor de ernst van het spijbelen bestaat uit een vraag naar de frequentie van het spijbelen (vaak, soms, nooit) en naar de gemiddelde duur van het spijbelen (hele dagen, een dagdeel of een enkel schooluur).

Op deze manier kon niet alleen worden nagegaan of een leerling spijbelde, maar ook of hij een 'chronische' dan wel een 'incidentele' spijbelaar was.

4. RESULTATEN

Verband van spijbelen met andere variabelen

Om een eerste indruk te krijgen, worden de correlaties gepresenteerd die de variabele spijbelen in de verschillende onderzoeksrondes vertoont met de andere variabelen. De in Tabel 1 vermelde correlaties zijn significant op 1%-niveau.

Tabel 1. Relaties met spijbelen.

	1986	1987	1988	1989	1990
prestaties	n.s	n.s	n.s	-.18	-.21
schoolbeleving	n.s	-.18	-.17	-.33	-.14
motivatie	n.s	-.24	-.33	-.37	-.25
wangedrag	n.s	.28	.34	.31	.40

In het eerste jaar zijn geen significante correlaties gevonden. Dit is echter te verklaren uit het feit dat in 1986, het eerste schooljaar op het voortgezet onderwijs voor het panel, slechts door een enkeling werd gespijbeld. De schoolprestaties blijken pas in de laatste twee jaar significant samen te hangen met spijbelen. Spijbelen vertoont vooral een verband met wangedrag en motivatie en in iets mindere mate met schoolbeleving.

In Tabel 1 is de variabele 'voortijdig schoolverlaten' nog niet opgenomen. Schoolverlaten komt echter weinig voor zodat deze variabele per jaar zeer scheef verdeeld is. Het berekenen van een correlatie is dan niet zinvol. Toch is het verband tussen schoolverlaten en spijbelen interessant, omdat het als één van de redenen voor het beboeten van spijbelaars wordt genoemd in de toelichting bij het wetsontwerp voor de Leerplichtwet. Daarin wordt gesteld dat chronisch spijbelen vaak vooraf gaat aan voortijdig schoolverlaten.

In het panel kwamen in totaal 23 voortijdige schoolverlaters voor. Daarvan kan worden gesteld dat 8 in het jaar voorafgaande aan het schoolverlaten chronische spijbelaars waren. Zij spijbelden vaak hele dagen, soms hele dagen of vaak hele dagdelen. De andere 15 spijbelden in mindere mate of nooit. Zo bezien gaat de stelling dus op, relatief vaak is er sprake van chronisch spijbelen. Men kan de zaak ook van de andere kant bekijken. Van de 16 chronische spijbelaars in 1987, verlaten er 2 het jaar erop voortijdig de school, van de 16 in 1988 zijn dat er 3 en van de 23 in 1990 zijn dat er ook 3. Er is dus geen sprake van een automatisme: chronisch schoolverzuim wordt zelfs voor het grootste deel niet gevolgd door schoolverlaten.

LISREL-analyse van twee concurrerende modellen

Na het vaststellen van de sterkte van de verbanden gaat het nu om de vraag op welke manier de variabelen met elkaar samenhangen. Met het LISREL-programma kan, zoals gezegd, worden vastgesteld welk van de twee modellen het best bij de empirische data (de covariantie-structuur van de variabelen) past.

Allereerst zal worden weergegeven wat in beide modellen de grootte van de effecten zijn. De parameter-schattingen zijn gestandaardiseerd en berekend volgens de zogenaamde Maximum-Likelihood methode. In Tabel 2 staan ze vermeld. Beide modellen zijn in de onderzoeksperiode van vijf jaar elk viermaal getest, met de onafhankelijke variabelen en de afhankelijke variabelen in respectievelijk 1986 en 1987, in 1987 en 1988, in 1988 en 1989 en in 1989 en 1990. De jaaraanduiding in Tabel 2 heeft betrekking op het jaar van de afhankelijke variabele.

Tabel 2. Bèta en psi-waarden van de statistische effecten in model 1 en model 2.

MODEL 1: 'discipline-model'	1987	1988	1989	1990
wangedrag^{t1} ↔ spijbelen^{t1}	.05	.28	.34	.31
wangedrag^{t1} → motivatie^{t1}	-.17	-.26	-.26	-.31
spijbelen^{t1} → motivatie^{t1}	-.01	-.11	-.19	-.18
wangedrag^{t1} → prestaties^{t2}	.03	-.02	.06	.07
spijbelen^{t1} → prestaties^{t2}	-.04	.01	-.06	.00
beleving^{t1} → motivatie^{t1}	.40	.36	.35	.30
motivatie^{t1} → prestaties^{t2}	-.14	-.12	-.14	-.06
MODEL 2: 'demotivatie-model'	**1987**	**1988**	**1989**	**1990**
prestaties^{t1} → motivatie^{t1}	-.12	-.14	-.11	-.14
prestaties^{t1} → beleving^{t1}	-.17	-.18	-.21	-.23
beleving^{t1} → motivatie^{t1}	.40	.43	.43	.40
motivatie^{t1} → wangedrag^{t2}	-.23	-.27	-.34	-.27
motivatie^{t1} → spijbelen^{t2}	-.14	-.18	-.30	-.19
beleving^{t1} → wangedrag^{t2}	-.07	-.19	-.04	.02
beleving^{t1} → spijbelen^{t2}	-.02	-.09	-.05	-.01
wangedrag^{t2} ↔ spijbelen^{t2}	.24	.24	.20	.35

Tabel 3. Fit-waarden van de twee LISREL-modellen.

MODEL 1 (3 vrijheidsgraden)	1987	1988	1989	1990
Chi-kwadraat	7.10	35.73	34.08	46.89
(p.)	(.07)	(.00)	(.00)	(.00)
Adjusted goodness of fit	.961	.812	.819	.752
Root mean square residual	.040	.099	.099	.118
MODEL 2 (2 vrijheidsgraden)	**1987**	**1988**	**1989**	**1990**
Chi-kwadraat	.20	2.55	1.50	11.04
(p.)	(.90)	(.28)	(.47)	(.00)
Adjusted goodness of fit	.998	.978	.987	.904
Root mean square residual	.005	.021	.015	.039

Vervolgens kan nu worden getoetst in hoeverre beide modellen bij de waargenomen covariantie-structuur passen. De 'fit'-waarden van beide modellen zijn in Tabel 3 weergegeven.

Duidelijk wordt dat model 2, het 'demotivatie'-model, veel beter bij de data past dan model 1, het 'discipline'-model. Uitgaande van de χ^2-waarden en de 'adjusted goodness of fit' is dit model goed te noemen (volgens de vuistregels van Verschuren, 1991), met uitzondering van 1990. Ook de derde maat, de 'root mean square residual' is niet te hoog. Model 1 daarentegen is duidelijk slecht te noemen.

Beide modellen zijn met behulp van aanwijzingen uit het LISREL-programma te verbeteren.

In model 1 blijkt de verwijdering van de directe verbanden van wangedrag en spijbelen op de schoolprestaties een jaar later en in plaats daarvan een effect van beide eerstgenoemde variabelen op beleving een verbetering op te leveren. Verder blijkt dat een effect van schoolbeleving op

schoolprestaties moet worden toegevoegd. Het model dat nu is ontstaan, kent dus geen direct verband van regelovertredend gedrag op de schoolprestaties maar een indirect verband via de schoolbeleving en de motivatie. Dit aangepaste model past nu echter niet meer bij de aanname achter het wetsontwerp. Hoe de school door de leerlingen wordt beleefd doet daar niet terzake.

In model 2 is het effect van beleving op spijbelen in geen van de jaren significant, verwijdering hiervan levert een verbetering van het model op. Bovendien blijkt dat toevoeging van een direct effect van de schoolprestaties op wangedrag een jaar later een behoorlijke verbetering van de 'fit'-waarden geeft.

Beide modellen passen nu goed tot zeer goed bij de empirische data. In feite is een keus tussen een indirect effect van prestaties op regelovertredend gedrag en een indirect effect van regelovertredend gedrag op prestaties niet goed te maken. Waarschijnlijk doen beide mechanismen zich voor en is er sprake van een zichzelf versterkend effect.

5. DISCUSSIE

Interpretatie en empirische kritiek op het wetsvoorstel

De in de analyses gevonden correlaties zijn over het algemeen bekend. Voor een overzicht wordt verwezen naar De Vries en de Jong (1987). Het verband tussen schoolverzuim en schoolverlaten ligt wat ingewikkelder. Weliswaar melden een aantal auteurs dat voortijdig schoolverlaten vaak wordt voorafgegaan door spijbelen. Schoolverzuim zien zij echter niet als oorzaak, maar als onderdeel van een problematische schoolloopbaan (zie bv. Mooij, 1979). Uit de in dit artikel gepresenteerde gegevens wordt in ieder geval duidelijk dat een deel van de schoolverlaters ernstig spijbelde, maar dat de omgekeerde redenering, veel chronische spijbelaars verlaten de school, nauwelijks opgaat.

Met de LISREL-analyse is gepoogd duidelijkheid te brengen in de causale structuur die achter slecht schoolfunctioneren schuilgaat. Er kan worden opgemerkt dat de operationalisering van de begrippen niet optimaal is. De schaal voor prestaties bestaat bijvoorbeeld in feite uit meerdere dimensies van het begrip prestaties. Aan de schaal voor schoolbeleving kleeft het bezwaar dat wordt gewerkt met waardering in de vorm van cijfers, waardoor het mogelijk is dat respondenten verschillende betekenissen aan de cijfers toekennen. Overigens is bij het interviewen geprobeerd dit bezwaar te ondervangen door een duidelijke instructie. Bij het vergelijken van de twee modellen speelt de operationalisering echter een minder belangrijke rol.

De gevonden resultaten zijn ook niet bedoeld als harde statistische toets van de verschillende modellen. Aan de voorwaarde dat alle variabelen in het model normaal verdeeld zijn, is bijvoorbeeld niet voldaan. De meerderheid van de leerlingen houdt zich aan de regels, waardoor de variabelen 'spijbelen' en 'wangedrag' een scheve verdeling hebben. De vermelde 'fit'-waarden kunnen echter wel worden beschouwd als indicatie, en het vergelijken van modellen is goed mogelijk (Jöreskog en Sörbom,1985).

Uitgaande van de gevonden resultaten ligt empirische kritiek op de aannames achter het wetsontwerp voor de hand. In de vergelijking van de twee concurrerende modellen blijkt het model dat uitgaat van spijbelen en wangedrag als oorzaak van andere problemen, in dit geval schoolprestaties (het 'discipline-model') aanmerkelijk slechter te passen dan het 'demotivatie'-model. Wanneer het 'discipline-model' wordt aangepast blijkt geen sprake van een direct effect van spijbelen en wangedrag op de prestaties, maar alleen van een indirect effect, via motivatie alsmede via de schoolbeleving. Dit is in strijd met de gedachte dat het vooral om een disciplineprobleem gaat.

De Vries, Monsma en Mellink (1990) vinden in een regressie-analyse eveneens een statistisch effect van schoolbeleving en schoolattitude (vergelijkbaar met motivatie) van leerlingen op spijbelen. Daarnaast komt bij hen wel een effect van schoolverzuim op de schoolcijfers naar voren. Zij geven echter zelf aan dat dit verband houdt met hun aanname dat 'slechte schoolprestaties' in de tijd gezien volgt op schoolverzuim. Andere interpretaties achten zij ook mogelijk.

Wanneer men zich een combinatie van de verbeterde modellen voorstelt, lijkt het erop dat er

sprake is van een zichzelf versterkend proces. De terugkoppelingseffecten van wangedrag en spijbelen lopen dan voornamelijk via de motivatie. Dat doet vermoeden dat door de reactie van de school op het regelovertredend gedrag de binding van de leerling met de school (nog verder) vermindert, wat zich uit in toenemende demotivatie.

Het effect van een juridische reactie op schoolverzuim

De reactie van de school op regelovertredend gedrag is mede onderwerp van studie in de onderzoeken die zijn gedaan naar zgn. 'effective schools'. Gunstig voor zowel de schoolprestaties en het verzuim lijken de scholen te zijn die een evenwicht weten te vinden tussen een zekere handelingsruimte van de leerlingen en de algemene regels en normen. In aansluiting hierop pleit De Vries (1987, p.9) voor een schoolcultuur die zorgt voor een optimale binding van de leerling met de school. Zo'n goede schoolcultuur weet prestatie-eisen te verenigen met een warme ambiance en combineert orde met autonomie en verantwoordelijkheden voor de leerlingen.

Men zou kunnen betogen dat het beboeten van spijbelaars ingrijpt in een vicieuze cirkel van falen en spijbelen, en daardoor een positief effect kan uitoefenen. De theorie van de maatschappelijke kwetsbaarheid voorspelt echter een tegenovergesteld effect. Het beboeten van spijbelaars zal het proces van demotivatie eerder versterken. De groep maatschappelijk kwetsbaren zal nog meer te maken krijgen met bestraffing, en nu ook nog van het justitieel apparaat in plaats van de school. Er is kortom ook theoretische kritiek mogelijk op het wetsvoorstel.

Dat een te strenge benadering juist een tegengesteld effect kan hebben wordt eveneens duidelijk uit ander onderzoek. Reynolds (1976) concludeert dat handhaving van strenge regels tot een spiraal van conflicten leidt, en vervolgens tot hoger schoolverzuim. Ook bij Reid (1985) en Moos & Moos (1978) lijken strenge regels en strenge straffen spijbelen in de hand te werken. Aan de andere kant is het zo, dat ook het volledig ontbreken van controle op schoolverzuim ongunstig is. Wanneer de school in het geheel niet reageert op hun afwezigheid, vatten de leerlingen dat op als een gebrek aan belangstelling. Geavanceerde technieken om schoolverzuim te registreren blijken dan ook succesvol. Het lijkt echter het belangrijkste te zijn dàt er op afwezigheid wordt gereageerd, en niet dat er streng op wordt gereageerd (zie De Vries, 1987).

(Straf-)rechtspolitieke kritiek

Wanneer het wetsvoorstel om spijbelaars te kunnen beboeten wordt aangenomen, is er sprake van een verschuiving van de oorspronkelijke doelstelling van de Leerplichtwet. Terwijl deze oorspronkelijk is bedoeld als waarborg dat leerlingen niet thuis worden gehouden, wordt het nu gebruikt om onwillige leerlingen te disciplineren. De overheid neemt daarmee gedeeltelijk een taak over van de scholen. Dat betekent gelijk ook dat algemene rechtsregels gaan gelden voor spijbelaars die in het kader van de Leerplichtwet voor de rechter moeten verschijnen. Hun rechtspositie dient te worden gewaarborgd, wat inhoudt dat zij bijvoorbeeld rechtshulp zouden moeten kunnen krijgen of in hoger beroep zouden moeten kunnen gaan.

Spijbelen wordt niet langer gezien als regelovertreding binnen de school. In plaats daarvan wordt het gedefinieerd als schade toebrengend aan de maatschappij. Waar dat toe kan leiden, blijkt bijvoorbeeld uit het commentaar van het CDA op het wetsvoorstel (Volkskrant, 25-1-'93). Het CDA is van mening dat spijbelaars – die immers de leerplichtwet ontduiken – niet direct recht op een uitkering mogen hebben. De stelling dat er sprake is van een zekere criminalisering van schoolverzuim lijkt me gezien het voorgaande niet gewaagd. In andere landen is dit al zover gevorderd dat schoolverzuim als een vorm van jeugddelinquentie wordt gezien. In Amerikaanse criminologische publikaties omschrijft men spijbelen vaak als 'status offencence' of 'light delinquency'.

Het gevaar van deze criminalisering is dat alle aandacht uitgaat naar de regelovertreding en uit het oog wordt verloren dat spijbelen ook een uitingsvorm kan zijn van een verstoorde relatie van de leerling met de school. In de laatste benadering wordt spijbelen gezien als een signaal. Het onderdrukken van dit signaal betekent natuurlijk nog niet dat de achterliggende problemen zijn verdwenen.

Een ander bezwaar van criminalisering is dat een 'strafrechtelijke' aanpak wordt gekozen,

waar andere oplossingen ook mogelijk zouden zijn. Vaak wordt het adagium 'strafrecht moet ultimum remedium zijn' aangehangen, wat betekent dat pas in laatste instantie moet worden gegrepen naar het zware wapen van bestraffing door de overheid. Steeds vaker echter lijkt de overheid bestraffing te willen zien als 'primo remedium', de eerste oplossing. Naar mijn mening moet zo'n ingrijpend middel pas worden overwogen als andere maatregelen uitgeput zijn. Met betrekking tot het schoolverzuim is dit nog lang niet het geval. Uit een inventarisatie van de maatregelen die scholen daartegen nemen, blijkt dat er nog genoeg andere mogelijkheden zijn (Bos et al., 1990). Zo kan de begeleiding van chronische spijbelaars worden verbeterd. Meer dan de helft van de onderzochte scholen voert bijvoorbeeld geen gesprekken met chronische spijbelaars aan het begin van het schooljaar (Bos et al., 1990, p.58). Het zou toch wel erg gemakzuchtig zijn om die leerlingen meteen te gaan beboeten, terwijl maatregelen die de leerling minder afkeren van het onderwijs onbenut zijn gelaten.

6. CONCLUSIE

De vraagstelling uit de inleiding kan nu worden beantwoord. Spijbelen en meer het algemeen regelovertredend gedrag op school zijn niet de centrale problemen waaruit slecht schoolfunctioneren voortvloeit. Op basis van de analyses blijkt een model waarin spijbelen en wangedrag de onafhankelijke variabelen zijn minder goed bij de empirische data te passen dan een model waarin spijbelen en wangedrag de afhankelijke variabelen zijn. Verbeteringen van beide modellen wijst erop dat een zichzelfversterkend proces waarschijnlijker is. De visie dat het vooral om een discipline-probleem gaat, is minder aannemelijk. Een strengere aanpak van spijbelen zal waarschijnlijk een averechts effect hebben, doordat de leerlingen die toch al een slechte binding met de school hebben, nog meer te maken krijgen met de bestraffende elementen uit het onderwijs.

De kritiek op het plan komt er dus op neer dat de aannames erachter niet houdbaar zijn, het effect van het beboeten van spijbelaars averechts zal zijn en dat met het plan het schoolverzuim wordt gecriminaliseerd, waardoor andere oplossingen worden bemoeilijkt. Het politieke beleid zou zich daarentegen beter kunnen concentreren op het verbeteren van de binding van de leerlingen met de school. Een ander element uit het wetsvoorstel spoort wel met deze richting: het vervangen van de volledige leerplicht voor sommige leerlingen door een combinatie van werken en leren. Aangezien dit soort onderwijs soms beter past bij bepaalde leerlingen, kan worden verwacht dat de spiraal van demotivatie en falen wordt gestopt en dat het schoolverzuim daalt. Voor anderen lijken mij vooral díe maatregelen geschikt die het patroon doorbreken dat leerlingen met een slechte startpositie voortdurend te maken krijgen met bestraffing en slechte beoordeling. De Vries (1987, p.11 en p.14) suggereert een schoolcultuur waarin zulke leerlingen worden beloond voor het leren zelf, of worden gestimuleerd doordat ze andere belangrijke taken krijgen toegewezen (bv. in de vorm van een mentorschap). Vooralsnog lijkt berechting van spijbelaars de al zwakke binding met school eerder te verbreken dan te verbeteren.

NOOT

1. Met dank aan dr. J.A. Nijboer voor het lezen en becommentariëren van verschillende versies van het manuscript.

LITERATUUR

Bos, K.Tj., Kesteren, B.J. van, Stoel, W.G.R. & Vermeulen, C.J.A.J. (1990). *Schoolverzuim. Maatregelen die de scholen kunnen nemen.* Amsterdam/Lisse: Swets & Zeitlinger.

Dijksterhuis, F.P.H. en Nijboer, J.A. (1988). *Wetsopvattingen en wangedrag van scholieren.* Groningen: Criminologisch Instituut Groningen.

Groot, A.D. de (1966). *Vijfen en zessen. Cijfers en beslissingen: het selectieproces in ons onderwijs.* Groningen: Wolters-Noordhoff.

Jöreskog, Karl G. en Sörbom, Dag (1985). *Lisrel VI. Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods.* Uppsula: University of Uppsala.

Mooij, T. (1979). *Probleemanalyse van het voortijdig schoolverlaten in het LBO, MAVO, HAVO en VWO.* Nijmegen: Instituut voor Toegepaste Sociologie.

Moos, R.H. & Moos, B.S. (1978). Classroom social climate and student absences and grades. In *Journal of Educational Psychology*, 70, 2, p.263-269.

Nijboer, J.A. & Dijksterhuis, F.P.H. (1983). *Onderwijs en delinquentie: de relatie tussen het functioneren op school en delinquentie.* Groningen: Criminologisch Instituut Groningen.

Reid, K. (1985). *Truancy and school absenteeism.* London: Hodder and Stoughton.

Reynolds e.a. (1976). When pupils and teachers refuse a truce: the secondary school and the creation of delinquency. In G. Mungham & G. Pearson (eds). *Working class youth culture.* London: Routledge and Kegan Paul.

Smits, J.A.E. en Vorst, H.C.M. (1982). *Schoolvragenlijst. Handleiding voor gebruikers.* Nijmegen:

Verschuren, P.J.M. (1991). *Structurele modellen tussen theorie en praktijk.* Utrecht: Het Spectrum.

Vettenburg, N., Walgrave, L. en Kerckvoorde, J. van (1984). *Jeugdwerkloosheid, delinquentie en maatschappelijke kwetsbaarheid: een theoretisch en empirisch onderzoek naar de veronderstelde band tussen werkloosheid en delinquentie bij 17-19 jarigen.* Antwerpen/Arnhem: Kluwer,Gouda Quint.

Volkskrant (7-11-1992). *Spijbelaar dient zich straks wellicht bij de rechter te vervoegen.*

Volkskrant (25-1-1993). *CDA bepleit variabele lengte leerplichtperiode.*

Vries, G.C. de (1987). *Schoolcultuur als uitdaging en als opgave: over verzuim, delinquentie en de invloed van de school.* Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam.

Vries, G.C. & Jong, M.W. de (1987). *Schoolverzuim en schooluitval. Een literatuurstudie.* Den Haag: SVO.

Vries, G.C. de, Monsma, G., Mellink, E., m.m.v. Ruijven, N. van (1990). *Als een slak, ongaarne, schoolwaarts. Een onderzoek naar schoolverzuim van leerlingen en invloeden van de school.* Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam.

Walgrave, L. (1993). Systematische jeugddelinquentie en maatschappelijke kwetsbaarheid. In Nijboer, J.A. et al.(red.). *Criminaliteit als politiek probleem (p. 35-44).* Arnhem: Gouda Quint.

Manuscript ontvangen 9-3-1993
Definitieve versie ontvangen 7-7-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 4, pp. 208-226

De opbrengst van volwasseneneducatie voor vrouwen: economische zelfstandigheid?

Sylvia J. van Leeuwen*

ABSTRACT

This study investigates the contribution of continuing education on the social economic independence of women. Does continuing education increase the labour market participation of women? Do women who participated in continuing education get better jobs (occupational level, income) than non-participants? To answer these questions a path model is developed and a multivariate regression analysis undertaken. Use is made of longitudinal data collected from a group of 953 Dutch women born in 1940.

The findings indicate a positive effect of continuing education on the labour market participation of higher educated women, but not of women in general. The effect of the level of continuing education on the labour market position is significant and positive: the higher the level of continuing education, the higher the occupational level and income. This influence is stronger for women with a lower level of youth education than for higher educated women. Thus, the economic pay-off to the individual participating in continuing education depends on the level of youth education.

1. INLEIDING

In Nederland nemen jaarlijks ruim 2 miljoen mensen deel aan enige vorm van volwasseneneducatie. De uitgaven vormen met f 6,4 miljard 1,4 % van het bruto nationaal produkt (Cramer & Houtkoop, 1990). De baten van volwasseneneducatie kunnen zowel aan individuen als aan de werkgever of aan de samenleving als geheel toevallen. In dit artikel wordt verslag gedaan van een onderzoek naar één van deze mogelijke opbrengsten van volwasseneneducatie, namelijk de economische zelfstandigheid van vrouwen: hun arbeidsmarktdeelname en hun arbeidsmarktpositie.

Er is veel onderzoek gedaan naar de invloed van het genoten onderwijs op de individuele arbeidsmarktpositie. Het onderwijs blijkt een centrale rol te vervullen in de verdeling van sociaal-economische posities: hoe hoger de opleiding, hoe hoger het beroepsniveau en het inkomen (bijv. Blau en Duncan, 1967, Jencks et al, 1979, Peschar en Wesseling, 1985, De Graaf, 1987, Dronkers, 1986 en 1992, Meesters, Dronkers en Schijf, 1983, en Dessens et al, 1990).

In bovengenoemd statusverwervingsonderzoek wordt geen onderscheid gemaakt tussen initieel (hierna jeugdonderwijs te noemen) en post-initieel onderwijs (hierna volwasseneneducatie).[2] Over de rol die volwasseneneducatie in het statusverwervingsproces speelt, is dus nog weinig bekend.

Niet alleen vanuit een wetenschappelijke invalshoek is meer inzicht hierin interessant, ook vanuit een beleidsoptiek. In het overheidsbeleid ten aanzien van volwasseneneducatie staan twee hoofddoelen centraal (Schütze & Istance, 1987): een ongelijkheid-reducerende functie (tweede kans) en een sociaal-economische functie. In de praktijk lijkt de sociaal-economische functie te overheersen. Toerusting voor de arbeidsmarkt en arbeidsinpassingsbeleid voor mensen met een zwakke arbeidsmarktpositie krijgen een steeds vooraanstaander plaats in het beleid ten aanzien van volwasseneneducatie.

De economische zelfstandigheid en de arbeidsmarktdeelname van vrouwen heeft het laatste decennium aan actualiteit gewonnen. Onder invloed van de vrouwenemancipatie en de gevol-

* Adres: Venuslaan 2, 3721 VG Bilthoven.

gen van ontgroening en vergrijzing voor het beschikbare arbeidspotentieel, is het bevorderen van de arbeidsmarktdeelname tot doel van het overheidsbeleid geworden (zie Beleidsplan Vrouwenemancipatie "Met het oog op 1995", Ministerie van SZW, 1992 en "Naar een werkend perspectief", WRR, 1990).

De laatste jaren zijn enkele onderzoeken gepubliceerd naar de vraag in hoeverre de veronderstelde functies van volwasseneneducatie ook daadwerkelijk worden waargemaakt. De meeste onderzoeken naar de effecten van volwasseneneducatie zijn gedaan onder mannen of onder de beroepsbevolking. Het onderzoek onder vrouwen is nog beperkt. De redenen hiervoor zijn gelegen in de betere kwaliteit van de beschikbare data, maar ook omdat het voor mannen gehanteerde model van de levensloopbaan voor vrouwen slecht toepasbaar is. Immers, veel vrouwen onderbreken hun beroepsloopbaan in verband met gezinstaken of zien van verdere beroepsdeelname af. Hierdoor is hun statuspositie minder eenduidig op hun betaalde beroep terug te voeren (zie ook Acker, 1980).

De seksespecifieke rolverdeling weerspiegelt zich ook in de gedeeltelijk verschillende motieven, mogelijkheden en belemmeringen tussen mannen en vrouwen in de volwasseneneducatie. Onderzoek onder mannen kan dus niet zonder meer gegeneraliseerd worden naar vrouwen. Een aanvullend onderzoek naar de effecten van volwasseneneducatie, waarin de verschillen in levensloopbaan tussen mannen en vrouwen een plaats krijgen is dringend nodig. Dit artikel kan gezien worden als een aanzet daartoe.

2. EERDER ONDERZOEK

2.1. Arbeidsmarkttheorieën en volwasseneneducatie

Onderzoek naar arbeidsmarktdeelname en arbeidsmarktpositie vindt vooral vanuit twee theorieën plaats: de human capital theorie en de conflictsociologische theorie. Deze zijn robuust gebleken in de verklaring van individuele arbeidsmarktposities. Omdat verschillen tussen mannen en vrouwen met deze theorieën niet goed kunnen worden begrepen is in aanvulling daarop gebruik gemaakt van de theorie van het dubbele levensperspectief.

De *human capital theorie* veronderstelt dat de beloning die men voor arbeid ontvangt in overeenstemming is met iemands produktieve vermogens. Opleiding en werkervaring zijn belangrijke elementen van menselijk kapitaal, en ook investeren in volwasseneneducatie is een manier om produktieve vermogens te vergroten. Men zal eerder in volwasseneneducatie investeren wanneer men verwacht op de arbeidsmarkt te zullen blijven. Vervolgens zal men het rendement van deze investeringen ook willen incasseren. Het effect van volwasseneneducatie op de arbeidsmarktpositie en -deelname zal groter zijn, naarmate het menselijk kapitaal sterker toeneemt (bijvoorbeeld meer cursussen of cursussen op een hoger niveau). Van volwasseneneducatie kan ook een preventieve werking uitgaan ten aanzien van werkloosheid en de kans op herintrede kan er door vergroot worden.

In de *conflicttheorie* wordt de samenleving beschouwd als statusgroepen die met elkaar concurreren om schaarse en begeerde goederen, zoals geld en status. Culturele waarden en normen, behorend bij verschillende posities in de sociale structuur, worden onder andere via het onderwijs overgedragen. Volwasseneneducatie speelt een aanvullende rol ten opzichte van de jeugdopleiding in het overdragen van cultuur, ideologie en sociale posities. Naast vakinhoudelijke kennis en vaardigheden hebben werkneemsters via extra scholing meer kans zich de "ware" bedrijfsideologie eigen te maken. De werkneemster krijgt daarmee een machtiger positie op de arbeidsmarkt. Zij wordt daardoor onafhankelijker van de werkgever en kan een hoger salaris afdwingen. Volwasseneneducatie leidt volgens deze visie niet tot een hoger inkomen omdat de produktieve capaciteit van de werkneemster toeneemt, maar omdat deze meer macht krijgt (Kalleberg, 1992).

Beide theorieën kunnen in elkaars verlengde gezien worden: grotere produktieve vermogens maken een personeelslid waardevoller voor een werkgever, wat haar meer macht en een sterkere onderhandelingspositie kan bezorgen.

De *theorie van het dubbele levensperspectief* (Sanders, 1991) is onderscheidend voor onderzoek onder mannen en vrouwen. Deze theorie veronderstelt dat mensen rationeel handelen in hun streven naar inkomen, sociale goedkeuring en vrij besteedbare tijd. Vrouwen beschikken daarbij, in tegenstelling tot mannen, over twee levensperspectieven: een betaalde loopbaan en een loopbaan als huisvrouw en moeder. Naarmate vrouwen sterker op de ene loopbaan zijn gericht, zal dit in het algemeen meer ten koste gaan van de andere. Volwasseneneducatie kan bijdragen aan een grotere arbeidsmarktdeelname van vrouwen, wanneer deze de betaalde loopbaan aantrekkelijker en eenvoudiger realiseerbaar maakt. Deze veronderstelling lijkt plausibel.

Op grond van bovenstaande theorieën is dus te verwachten dat volwasseneneducatie een positief effect zal hebben op de individuele arbeidsmarktpositie en op de arbeidsmarktdeelname van vrouwen.

2.2. Empirisch onderzoek

Bij het weergegeven van de stand van zaken van het empirisch onderzoek naar de invloed van volwasseneneducatie op de sociaal-economische positie van deelnemers/sters, beperk ik mij grotendeels tot longitudinaal onderzoek. De reden is dat dit type onderzoek het meest adequaat is voor de beantwoording van dit soort vragen. Longitudinaal onderzoek is in de volwasseneneducatie nog schaars, zo blijkt uit overzichten van Houtkoop (1985), Schütze & Istance (1987) en van der Kamp & Stroomberg (1991).

De in deze paragraaf te bespreken onderzoeken zijn weergegeven in onderstaand overzicht. Het betreft economisch en sociologisch onderzoek, waarbij verschillende aspecten van de individuele sociaal-economische positie als de afhankelijke variabelen gebruikt zijn. Ook de precieze onderzoeksopzet en de onderzoekspopulaties zijn steeds verschillend. Het is uiteraard goed mogelijk dat het rendement van volwasseneneducatie afhangt van de tijd en de plaats waarin de deelnemer/ster leeft.

Auteur	Jaar	Land	Populatie	Afhankelijke variabele(n)
Felmlee	1988	VS	Jonge blanke vrouwen	beroepsstatus, inkomen
			Jonge zwarte vrouwen	beroepsstatus, inkomen
Tuijnman	1989	Zweden	Mannen	beroepsstatus, inkomen
van Leeuwen	1990	Nederland	Mannen	beroepsniveau, inkomen, werkloosheid
Gooderham	1991	Noorwegen	Mannen	beroepsstatus
			Vrouwen	beroepsstatus
Diederen	1991	Nederland	Mannen en vrouwen	beroepsstatus, arbeidsdeelname
Kalleberg	1992	Noorwegen	Mannen	inkomen
			Vrouwen	inkomen
Psacharopoulos	1992	Colombia	Mannen en vrouwen	inkomen

Vrij consequent blijkt dat volwasseneneducatie een positief (direct) effect heeft op de beroepsstatus resp. het beroepsniveau van mannen en vrouwen (Tuijnman, 1989, Van Leeuwen, 1990, Gooderham, 1991, Diederen, 1991). Via de hogere beroepsstatus heeft volwasseneneducatie ook een positieve invloed op het inkomen van mannen (indirect effect). Na controle voor beroepsstatus resteert er geen direct effect van volwasseneneducatie op het inkomen van mannen (Tuijnman, 1989, Van Leeuwen, 1990). In de onderzoeken van Kalleberg (1992) en Psacharopoulos & Velez (1992) werd de beroepsstatus/het beroepsniveau niet constant gehouden. Zij vonden wel een direct effect van volwasseneneducatie op het inkomen.

De effecten van volwasseneneducatie op beroepsstatus en inkomen zijn minder sterk dan de effecten van het jeugdonderwijs, doch zeker substantieel. Uit het onderzoek van Tuijnman (1989) blijkt voorts dat de invloed van volwasseneneducatie op beroepsniveau en inkomen in de loop van het arbeidsleven toeneemt, terwijl de invloed van het jeugdonderwijs rond het veertigste levensjaar een hoogtepunt heeft bereikt. Met een stijgende leeftijd neemt het relatieve belang van volwasseneneducatie dus toe.

Het enige onderzoek waar niet consequent een positief effect gevonden werd is dat van Felmlee (1988) onder jonge Amerikaanse vrouwen. Felmlee vergeleek de loopbanen van 'school returners' (vrouwen die een full-time baan inruilen voor een full-time studie) met die van vrouwen die bleven werken. Uit verschillende analyses concludeert ze dat terugkeer naar school alleen lonend is wanneer dat voor een vrij lange periode is, zodat het opleidingsniveau daadwerkelijk verbetert. De andere weg voor jonge vrouwen om in korte tijd meer te gaan verdienen is op de arbeidsmarkt blijven en een beter betaalde baan vinden zonder naar school terug te keren. Hierbij moet worden aangetekend dat zij effecten op de korte termijn onderzocht, terwijl het aannemelijk is dat de volle omvang van effecten pas na verloop van tijd zichtbaar wordt (zie Tuijnman, 1989).

Zijn de opbrengsten van volwasseneneducatie voor bepaalde groepen deelnemers groter dan voor andere? Hiervoor zijn inderdaad aanwijzingen.

Diederen (1991) en Van Leeuwen (1990) vonden een sterker effect van volwasseneneducatie op de beroepsstatus bij een intensievere deelname aan volwasseneneducatie (bijvoorbeeld meer cursussen, hoger niveau, grotere niveaustijging ten opzichte van het jeugdonderwijs).

De onderzoeken van Gooderham (1991) en Kalleberg (1992) leveren aanwijzingen dat mannen meer profijt hebben van deelname aan volwasseneneducatie dan vrouwen. Zij vonden voor mannen een sterker effect van volwasseneneducatie op de beroepsstatus (Gooderham) resp. het inkomen (Kalleberg) dan voor vrouwen. Echter, het verschil in stijging tussen mannen en vrouwen is in beide onderzoeken statistisch niet significant.

Bij Nederlandse mannen blijkt het financiële voordeel van volwasseneneducatie (inkomenseffect) groter te zijn naarmate de jeugdopleiding lager is (Van Leeuwen, 1990). Psacharopoulos & Velez (1992) vonden onder werknemers in Bogota, Columbia, het tegendeel: de baten van training zijn groter naarmate het opleidingsniveau hoger is. Werknemers met minder dan acht jaar onderwijs ondervinden nauwelijks financieel voordeel van aanvullende cursussen.

Alleen in de Nederlandse onderzoeken wordt deelname aan volwasseneneducatie in verband gebracht met arbeidsmarktdeelname. Het meest uitgewerkt is het van Jaar tot Jaar onderzoek van Diederen (1991). De deelname aan de arbeidsmarkt op 35-jarige leeftijd blijkt sterk samen te hangen met geslacht (mannen hebben veel vaker een baan dan vrouwen), maar daarnaast ook met kenmerken van de deelname aan volwasseneneducatie. De deelname aan het betaalde arbeidsleven is groter, wanneer men deelgenomen heeft, wanneer men een groter aantal cursussen heeft gevolgd en wanneer de belangrijkste cursus een beroepsgericht karakter had.

De mannelijke helft van het Matthijssen-Sonnemans cohort blijkt op 43-jarige leeftijd een kleinere kans op werkloosheid te hebben wanneer men heeft deelgenomen aan volwasseneneducatie en wanneer men via volwasseneneducatie een grotere niveaustijging heeft bereikt (Van Leeuwen, 1990). Vrouwen werden niet in de analyses opgenomen.

Concluderend kan worden gesteld dat het empirisch onderzoek vrij consistent laat zien dat volwasseneneducatie bijdraagt aan een hogere beroepsstatus en een hoger inkomen. Het aantal onderzoeken is nog te beperkt om conclusies te trekken over de vraag voor welke groepen het rendement van volwasseneneducatie in termen van beroepspositie en inkomen groter of kleiner is. Dit geldt nog sterker voor het effect van volwasseneneducatie op arbeidsmarktdeelname en werkloosheid, hoewel er indicaties zijn dat de invloed van volwasseneneducatie ook hier positief is.

Onderzoek onder vrouwen is het meest beperkt. De geringe aandacht voor het effect van volwasseneneducatie op de arbeidsmarktdeelname hangt hier wellicht mee samen.

3. PROBLEEMSTELLING

Een groot deel van de vrouwen verblijft, als huisvrouw, een kortere of langere tijd buiten de arbeidsmarkt en ontleent haar sociaal-economische positie aan haar partner. Deze afhankelijke

positie van huisvrouw is voor veel vrouwen in de gegeven maatschappelijke omstandigheden een aantrekkelijker alternatief dan zelf op de arbeidsmarkt te blijven. In geval van echtscheiding, overlijden, of werkloosheid van de partner blijkt echter hoe kwetsbaar deze vrouwen in sociaal-economisch opzicht zijn. Bij het onderzoeken van de sociaal-economische positie van vrouwen blijft een wezenlijk aspect buiten beschouwing wanneer geen aandacht besteed wordt aan de arbeidsmarktdeelname.

De eerste onderzoeksvraag luidt daarom: *Wat is de invloed van volwasseneneducatie op de mate van arbeidsmarktdeelname van vrouwen? Is het zo dat deelneemsters aan volwasseneneducatie vaker een betaalde baan hebben en/of een groter aantal uren per week betaald werken dan niet-deelneemsters?*

De tweede vraag is voor de hand liggend en in theoretisch opzicht eenvoudiger: *Is het volgen van opleidingen in de volwasseneneducatie van invloed op de arbeidsmarktpositie van (betaald) werkende vrouwen in termen van beroepsniveau en inkomen?*

Om deze vragen te beantwoorden is de school- en beroepsloopbaan gevolgd van een groep vrouwen, geboren in/rond 1940. Daarbij zal de sociaal-economische positie van deelneemsters aan volwasseneneducatie worden vergeleken met die van de niet-deelneemsters.

Bij de beantwoording van de vragen zal tevens aandacht worden besteed aan de interactie-effecten: *Zijn de effecten van volwasseneneducatie op de arbeidsmarktdeelname en de arbeidsmarktpositie even groot voor: a. vrouwen uit een hoog en een laag milieu? en b. vrouwen met een hoge en een lage jeugdopleiding?*

In paragraaf 5 zullen de data en de onderzoeksopzet nader worden beschreven. Van belang voor de onderzoeksopzet is de tijdsvolgorde die in de formulering van de vraagstelling besloten ligt: volwasseneneducatie gaat hier in de tijd vooraf aan de arbeidsmarktpositie.[3]

4. THEORETISCH MODEL

In dit onderzoek zijn wij geïnteresseerd in de invloed van volwasseneneducatie op de arbeidsmarktdeelname en de arbeidsmarktpositie van vrouwen. Om het effect van volwasseneneducatie te bepalen is het van belang rekening te houden met andere factoren die daar eveneens op van invloed zijn. Gebeurt dat niet, dan kan het effect van volwasseneneducatie over- of onderschat worden.

Bij de *arbeidsmarktpositie* is vooral het menselijk kapitaal van belang: opleidingsniveau (jeugdopleiding en/of volwasseneneducatie), capaciteiten, werkervaring (bij vrouwen vaak gekoppeld aan het krijgen van kinderen dan wel het uitstellen daarvan), alsmede het sociaal en cultureel kapitaal (ouderlijk milieu, persoonlijk netwerk).

Ook verschillende institutionele factoren kunnen een rol spelen, zoals de ruimte op de arbeidsmarkt (algemene werkloosheid), het arbeidsmarktsegment (overschot of schaarste in de eigen beroepsgroep), de omvang van het bedrijf (interne loopbanen). In de profit sector zijn de primaire arbeidsvoorwaarden (lonen) gunstiger dan in de non-profit sector.

Relevant bij onderzoek onder vrouwen is tenslotte het aantal uren dat men beschikbaar is voor de arbeidsmarkt, het aspiratieniveau (motivatie) en de mate waarin het dubbele levensperspectief van toepassing is (bijv. aanwezigheid van een partner, zorg voor kinderen).

Modellen ter verklaring van de verschillen in *arbeidsmarktdeelname* zijn veel minder goed ontwikkeld. Voortbouwend op de theorie van het dubbele levensperspectief kan gesteld worden dat de arbeidsmarktdeelname groter is wanneer een betaalde loopbaan aantrekkelijker of gemakkelijker te realiseren is en/of wanneer de financiële noodzaak daartoe groter is. Dit laatste kan het geval zijn wanneer de vrouw geen partner heeft of wanneer het inkomen van haar partner lager is. Het complexe stelsel van sociale zekerheid maakt het voor deze vrouwen echter niet altijd financieel de moeite waard om een betaalde baan te zoeken. Twee voorbeelden. Het inkomen van de vrouw kan een lagere uitkering van haar werkloze man tot gevolg hebben, doordat kostwinnersfaciliteiten vervallen. Een bijstandsmoeder die een part-time baan (met een beperkt inkomen) weet te vinden, is daarmee nog steeds niet onafhankelijk van de sociale dienst maar maakt wellicht wel hogere kosten voor kinderopvang.

Deelname aan de arbeidsmarkt is voor vrouwen aantrekkelijker en gemakkelijker te realiseren wanneer men een betere arbeidsmarktpositie zal kunnen verwerven (zie boven) en wanneer men meer belang hecht aan economische zelfstandigheid (emancipatie, motivatie). Naast de objectieve werkgelegenheidskansen is ook de subjectieve beleving van invloed. Voor vrouwen die veel geïnvesteerd hebben in onderwijs is het verlaten van de arbeidsmarkt ook onaantrekkelijker omdat het een grotere kostenpost betekent in termen van gederfd inkomen.

Het huwelijk en het krijgen van kinderen is voor veel vrouwen de belangrijkste reden om te stoppen met betaald werk. In hoeverre de zorg voor kinderen een belemmering vormt voor arbeidsmarktdeelname van vrouwen hangt uiteraard af van de bijdrage die hun partner levert aan huishouden en kinderverzorging, van de beschikbare kinderopvang-voorzieningen en van de sociale goedkeuring en steun die zij uit haar omgeving ontvangt. In hogere maatschappelijke lagen is de houding ten opzichte van werkende vrouwen en emancipatie in het algemeen positiever dan in de lagere milieus. Binnen de milieus zijn de onderlinge verschillen evenwel groot. In de non-profitsector is de mogelijkheid voor deeltijdarbeid groter en zijn de secundaire arbeidsvoorwaarden, die de combinatie van werk met ouderschap vergemakkelijken iets beter dan in de profit sector.

Er kunnen tenminste vier redeneringen worden gehouden waarom deelname aan volwasseneneducatie bijdraagt aan een grotere arbeidsmarktdeelname van vrouwen: 1. deelneemsters aan volwasseneneducatie kunnen een aantrekkelijker baan krijgen; 2. deelneemsters willen gedane investeringen niet verloren laten gaan; 3. deelneemsters hebben een grotere kans op het verkrijgen (denk aan herintreedsters) dan wel behouden van een baan dan niet-deelneemsters; en 4. een combinatie van educatie en werk biedt vrouwen ontplooiingskansen: het ontwikkelen van capaciteiten en groei in de betaalde loopbaan.

5. DE ONDERZOEKSOPZET

5.1. Databestand

Voor de analyse wordt het Matthijssen-Sonnemans cohort (Hartog en Pfann, 1985) gebruikt. Dit cohort bestaat uit vrouwen en mannen die in 1952 in de zesde klas van het lager onderwijs in Brabant zaten. De laatste enquête onder de leden van dit cohort vond plaats in 1983 (leeftijd 43 jaar). De non-response onder de vrouwen is in dit cohort groter dan die onder mannen, omdat alleen mannen die niet op de schriftelijke enquête reageerden, vervolgens mondeling benaderd zijn. Desondanks is het momenteel waarschijnlijk het meest geschikte Nederlandse longitudinale databestand voor dit onderzoek, vanwege de lange looptijd van het cohort. Voor de analyses is gebruik gemaakt van de gegevens van 953 vrouwen.

Van deze vrouwen heeft ongeveer de helft deelgenomen aan een of meerdere cursussen in de volwasseneneducatie. Het betreft hoofdzakelijk aanvullende cursussen ten behoeve van het werk, en in mindere mate bedrijfsopleidingen en reguliere scholen.[4]

In 1983 hadden 332 van de 953 vrouwen van dit cohort een betaalde baan.

5.2. Definities en variabelen

De gekozen operationaliseringen van de theoretische concepten sluiten zo goed mogelijk aan bij de eerdere longitudinale onderzoeken (zie paragraaf 3) van Tuijnman (1989) en Van Leeuwen (1990), zodat de uitkomsten onderling vergelijkbaar zijn.[5] De namen van de variabelen zijn tussen haakjes vermeld.

Van de vele vormen van *volwasseneneducatie* zijn alleen de scholen, bedrijfsopleidingen en cursussen opgenomen die gericht zijn op het doen van betaald werk. Over andersoortige cursussen bevat het gebruikte gegevensbestand geen informatie. In dit onderzoek gekozen voor een strenge toetsing van de effecten van volwasseneneducatie: opleidingen waarvan het onduidelijk is aan welke kant van de scheidslijn zij vallen, zijn beschouwd als jeugdopleidingen. Tot de volwasseneneducatie zijn gerekend: alle scholen, begonnen vanaf 1970 (leeftijd 30 jaar); alle scholen, begonnen tussen 1959 en 1969, mits minimaal 5 jaar na aanvang van de eerste baan of

minimaal 5 jaar na beëindiging van de vorige opleiding; alle bedrijfsopleidingen; en alle aanvullende cursussen ten behoeve van het werk. De gebruikte variabelen zijn: het niveau van de jeugdopleiding (jeugdopl), een dummy wel/geen deelname aan volwasseneneducatie (vedeelname), het niveau van volwasseneneducatie (veniveau) en de niveaustijging die via volwasseneneducatie bereikt is ten opzichte van de jeugdopleiding (ve+).

Voor de *arbeidsmarktdeelname* zijn twee variabelen beschikbaar: een dichotome variabele wel/geen betaalde baan (wel/geen baan) en het aantal uren van de baan per week (werkuren).

De *arbeidsmarktpositie in 1983* van betaald werkende vrouwen is vastgesteld aan de hand van de volgende indicatoren:

* het beroepsniveau overeenkomstig de ARBI beroepenclassificatie van het CBS (beroepsniveau);
* het opleidingsniveau dat de respondent wenselijk acht voor het uitoefenen van haar functie (opl baan);
* de natuurlijke logaritme van het bruto inkomen per uur uit arbeid (uurink).

Verder zijn de volgende variabelen in de analyse opgenomen: *ouderlijk milieu*, gemeten via het beroep van de vader in 1952 (milieu), het *aspiratieniveau* van de ouders, gemeten via de wens van de ouders ten aanzien van het vervolgonderwijs voor hun dochter aan het eind van de lagere school (wens), het gemiddelde resultaat van 3 intelligentietesten in 1952 (*IQ*), het *schooladvies* van de onderwijzer aan het einde van de lagere school (advies), de aanwezigheid, het opleidingsniveau en de natuurlijke logaritme van het jaarinkomen van de *partner* in 1983 (partner, opl partner en ink partner), de omvang van het bedrijf, de *sector* van de arbeidsmarkt waar de vrouw werkt (profit of non-profit sector), het subjectieve oordeel van de respondent over haar *werkgelegenheidskansen* (werkkans) en over de *schaarste* aan mensen voor de functie die zij beoefent (schaarste). Het is opmerkelijk dat deze twee variabelen (werkkans, schaarste) weinig met elkaar samenhangen (correlatie 0.15). Daarom zijn beide variabelen in de analyses opgenomen.

Als indicatie voor de omvang van de *zorgtaken* in de periode voor 1983 is de leeftijd van het oudste thuiswonende kind gebruikt (de duur van de zorgperiode). Voor de omvang van de zorgtaken in 1983 is de leeftijd van het jongste kind gebruikt.

De *werkervaring*, afgemeten aan het aantal arbeidsjaren tussen 1952 en 1983 (werkjaren), is (noodgedwongen) enigszins onnauwkeurig vastgesteld. Tussentijdse onderbrekingen van de betaalde loopbaan zijn namelijk onbekend. Afgemeten aan landelijke gegevens is dit een reëel probleem: ongeveer 65 % van de vrouwen die op 43-jarige leeftijd werken heeft haar beroepsloopbaan één of meerdere keren onderbroken (Dessens et al, 1990). De onbetrouwbaarheid is waarschijnlijk het grootst bij de vrouwen die in 1983 een betaalde baan en kinderen hadden. Vooral bij de analyses die beperkt zijn tot werkende vrouwen, kunnen daardoor de variabelen met betrekking tot "zorgtaken/kinderen" niet alleen de omvang van de zorgtaken weerspiegelen, maar mogelijk ook verschillen in werkervaring als gevolg van tijdelijke onderbreking van de loopbaan.

De omvang van de algemene werkloosheid (conjunctuur), het geslacht, de leeftijd en vermoedelijk ook de etniciteit zijn in dit cohort voor alle respondenten gelijk.

De ontwikkelde theoretische modellen kunnen niet volledig getoetst worden met behulp van dit databestand. Over de volgende factoren zijn geen gegevens beschikbaar: houding/ mening over emancipatie en betaald werken door vrouwen, zowel van de respondent als van de omgeving, de praktische steun die anderen daadwerkelijk in dit opzicht leveren (vooral bij huishouden en kinderverzorging). Wel kan gesteld worden dat de mening over vrouwenemancipatie en werkende moeders gemiddeld positiever is in hogere milieus en onder hoger opgeleiden.

Voorts is de financiële noodzaak voor een eigen inkomen maar ten dele bepaald door de aanwezigheid en het inkomen van de partner. Jammer is dat het gebruikte databestand geen informatie bevat over het verloop van de beroepsloopbaan tussen het verlaten van het jeugdonderwijs en 1983 (bijv. tussentijdse onderbrekingen, ontwikkeling in aantal gewerkte uren, functieniveau en inkomen). Hiermee zijn tevens de beperkingen van de analyses aangegeven.

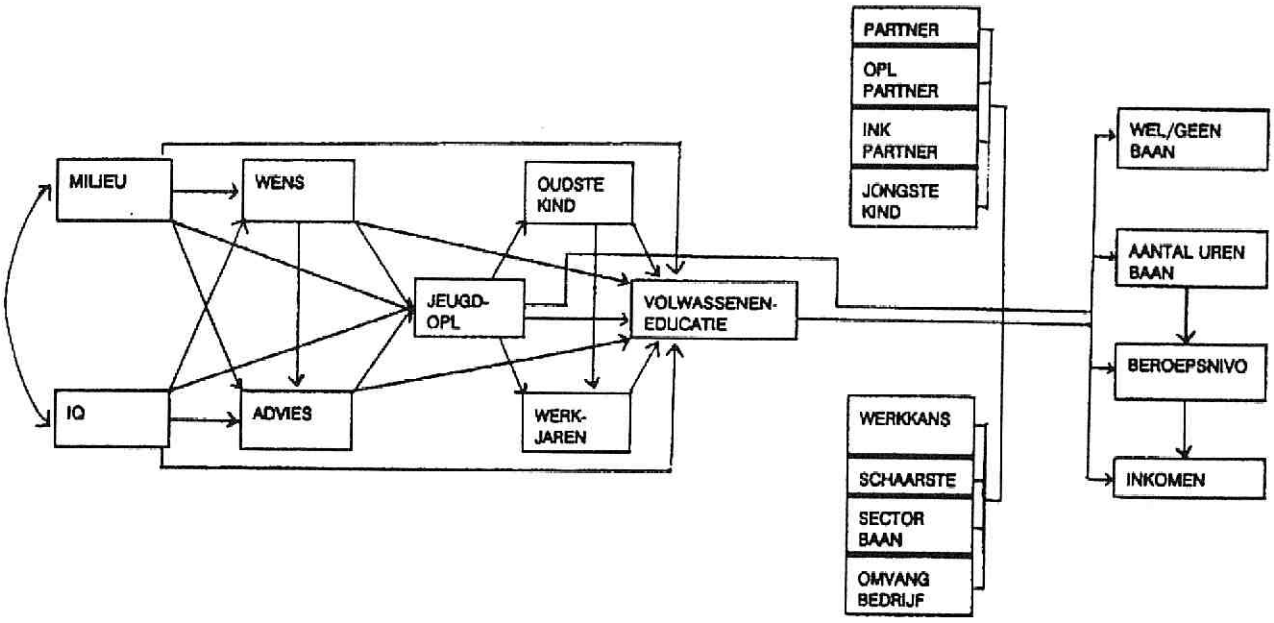


Fig. 1. Analysemodel (hypothetisch).

5.3. Analysemethode

De analyses zijn verricht via stapsgewijze multiple regressie analyse. Het analysemodel is weergegeven in Figuur 1. Bij de analyses is bij de bepaling van statistische significantie een grens van 1 % aangehouden. Telkens werd de laatste vergelijking genomen waaraan nog een significante parameter werd toegevoegd. Om te voorkomen dat teveel informatie wordt weggegooid is gekozen voor paarsgewijze verwijdering van ontbrekende gegevens.

Voor de niet-deelneemsters zijn de missende waarden voor het niveau van volwasseneneducatie (veniveau) en de niveaustijging via volwasseneneducatie (ve+) vervangen door de gemiddelde score (zie voor deze procedure Cohen & Cohen, 1975 en Meesters, 1984). De coëfficiënt van de dummy wel/geen deelname drukt het verschil uit tussen deelneemsters op een gemiddeld niveau resp. met een gemiddelde niveaustijging en niet-deelneemsters. Een zelfde werkwijze is toegepast bij het inkomen van de partner voor vrouwen zonder partner.

De analyses met betrekking tot de arbeidsmarktpositie in 1983 zijn beperkt tot werkende vrouwen (N=332). Vanwege het kleinere aantal vrouwen is hier een grens van 5 % aangehouden bij de bepaling van statistische significantie.

De onderzoeksvragen worden eerst beantwoord zonder rekening te houden met interactie-effecten. Dit geeft inzicht in de effecten van volwasseneneducatie voor de gehele groep vrouwen (1e en 2e onderzoeksvraag). In tweede instantie worden interactie-effecten aan de regressievergelijkingen toegevoegd, om na te gaan of de effecten van volwasseneneducatie gelijk zijn voor vrouwen uit verschillende ouderlijke milieus of met een verschillende jeugdopleiding (3e onderzoeksvraag).

6. RESULTATEN

6.1. Volwasseneneducatie en arbeidsmarktdeelname

Is het zo dat deelneemsters aan volwasseneneducatie vaker een betaalde baan hebben en meer uren per week (betaald) werken dan niet-deelneemsters?

Wanneer deelneemsters en niet-deelneemsters vergeleken worden zonder rekening te houden met andere factoren (bivariaat), blijkt dat deelneemsters ruim anderhalf keer zo vaak een betaal-

de baan hebben dan niet-deelneemsters (43,9 % van de deelneemsters en 27,4 % van de niet-deelneemsters heeft een baan). Deelneemsters werken gemiddeld ook een halve dag per week meer dan niet-deelneemsters (verschil 3.9 uur). De verschillen zijn statistisch significant (t-toets, p = 0.00 respectievelijk p = 0.01).

Met behulp van multiple regressie-analyse is nagegaan of het effect van volwasseneneducatie blijft bestaan na controle voor de overige factoren die de arbeidsmarktdeelname van vrouwen beïnvloeden (multivariaat). De resultaten staan vermeld in Tabel 1, vergelijkingen 1 en 3. Multivariaat wordt geen direct effect gevonden van volwasseneneducatie op de omvang van de arbeidsmarktdeelname van vrouwen. Ook de indirecte effecten van volwasseneneducatie zijn zeer gering (via sector en werkkans, indirecte effecten tussen -0.04 en -0.05). Het is niet zo duidelijk waarom het verschil tussen deelneemsters en niet-deelneemsters hier niet meer teruggevonden wordt.

Vrouwen hebben vaker een betaalde baan wanneer hun ouders een hoger aspiratieniveau (WENS) hadden, hetgeen te begrijpen is als indicatie dat er meer sociale steun en goedkeuring in de directe omgeving is. Andere indicatoren voor sociale steun, het opleidingsniveau van de partner en het ouderlijk milieu, blijken echter geen invloed te hebben op de arbeidsmarktdeelname van vrouwen. Overigens gaat het hier wel om zeer gebrekkige operationaliseringen van de werkelijke sociale steun.

Verder zijn vooral factoren van invloed die samenhangen met de realiseerbaarheid van een betaalde loopbaan: geen/oudere kinderen, betere arbeidsmarktkansen, meer schaarste in de eigen functie, en een hogere jeugdopleiding. De veronderstelling dat er in de non-profitsector gunstiger secundaire arbeidsvoorwaarden voor vrouwen zijn (bijv. gelegenheid tot part-time werk), lijkt door de resultaten te worden ondersteund. Vrouwen in de non-profitsector hebben vaker een baan maar werken gemiddeld minder uren per week.

Tabel 1. Volwasseneneducatie en arbeidsmarktdeelname in 1983. Gestandaardiseerde multiple regressiecoëfficiënten zonder en met interactietermen. (N=953 resp. 332).

	alle vrouwen (N=953)		werkende vrouwen (N=332)	
Afh. variabele:	wel/geen baan#	wel/geen baan#	werkuren#	werkuren#
Onafh. var:				
milieu	*	*	*	*
IQ	*	*	*	*
wens	-0.17	*	*	*
advies	*	*	*	*
jeugdopl	*	*	0.18	*
vedeelname	*	*	*	*
jeugdopl*vedeelname		-0.18		0.22
veniveau	*	*	*	*
jeugdopl*veniveau		*		*
partner	*	*	-0.36	-0.36
logink partner	*	*	*	*
opl partner	*	*	*	*
jongste kind	-0.20	-0.19	0.26	0.25
werkkans	-0.14	-0.14	*	*
sector	-0.17	-0.16	-0.25	-0.24
schaarste	*	*	0.20	0.18
R^2	0.14	0.14	0.34	0.36

* = Niet significant bij p < 0.01 (wel/geen baan) respectievelijk bij p < 0.05 (aantal uren).
# = Wanneer ve+ en jeugdopl*ve+ in de regressievergelijkingen worden opgenomen in plaats van veniveau en jeugdopl*veniveau is het resultaat gelijk.

De financiële noodzaak, gemeten via aanwezigheid en inkomen van de partner, blijkt geen invloed te hebben op het al dan niet hebben van een baan, maar wel op de omvang van de baan. Vrouwen zonder partner werken, als zij een baan hebben, een groter aantal uren per week dan vrouwen met een partner. Mogelijk is betaalde arbeid voor deze vrouwen alleen financieel aantrekkelijk wanneer zij daarmee daadwerkelijk economische zelfstandigheid verwerven. Een bescheiden inkomen uit een part-time baan zal vaker ten koste gaan van uitkering en/of alimentatie. Overigens heeft de helft van de vrouwen zonder partner in 1983 nog de zorg voor thuiswonende kinderen.

De verklaarde variantie van de variabele "wel/geen baan" is niet erg hoog, 14 %. Dat betekent dat er wellicht andere factoren van invloed zijn dan de factoren die hier in de analyses zijn opgenomen. Eerder is reeds opgemerkt dat de operationalisering gebrekkig was t.a.v. de financiële noodzaak tot betaalde arbeid en t.a.v. de sociale steun uit de omgeving. Er waren geen gegevens over de eigen motivatie en houding ten opzichte van emancipatie en betaald werk, ofwel: in hoeverre het 'dubbele levensperspectief' van toepassing is.

6.1.1. Verschillen in het effect van volwasseneneducatie (interacties)

"Zijn de effecten van volwasseneneducatie op de arbeidsmarktdeelname even groot voor: a. vrouwen uit verschillende ouderlijke milieus? en b. vrouwen met een verschillende jeugdopleiding?" In deze vraag gaat het om de omvang van de verschillen tussen deelneemsters en niet-deelneemsters. Om deze vraag te beantwoorden zijn aan de regressievergelijkingen interactie-effecten toegevoegd. Eerst van milieu met volwasseneneducatie, vervolgens van jeugdopleiding met volwasseneneducatie.

De interactie-effecten van milieu met volwasseneneducatie zijn niet significant. De resultaten zijn daarom niet in een tabel opgenomen. Het betekent dat het ouderlijk milieu geen (lineaire) invloed heeft op de relatie tussen volwasseneneducatie en arbeidsmarktdeelname. Het is niet zo dat de invloed van volwasseneneducatie op de arbeidsmarktdeelname sterker of minder sterk is bij meisjes uit hogere milieus dan bij meisjes uit lagere milieus.

Het ouderlijk milieu lijkt zelf ook op geen enkele manier invloed te hebben op de arbeidsmarktdeelname van vrouwen; Het hoofdeffect van milieu is niet significant.

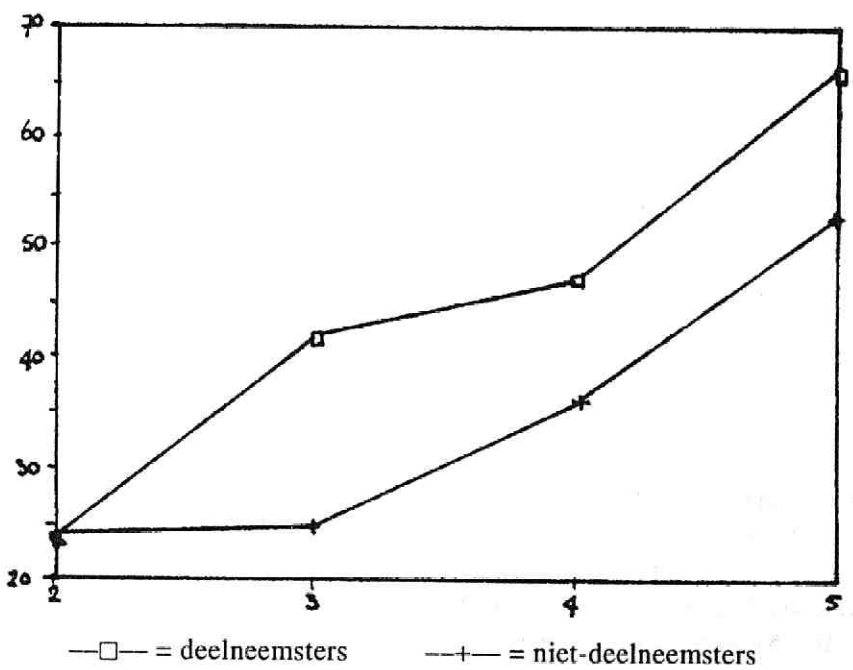


Fig. 2. Percentage werkende vrouwen onder deelneemsters en niet-deelneemsters aan volwasseneneducatie naar jeugdopleiding (N=799).

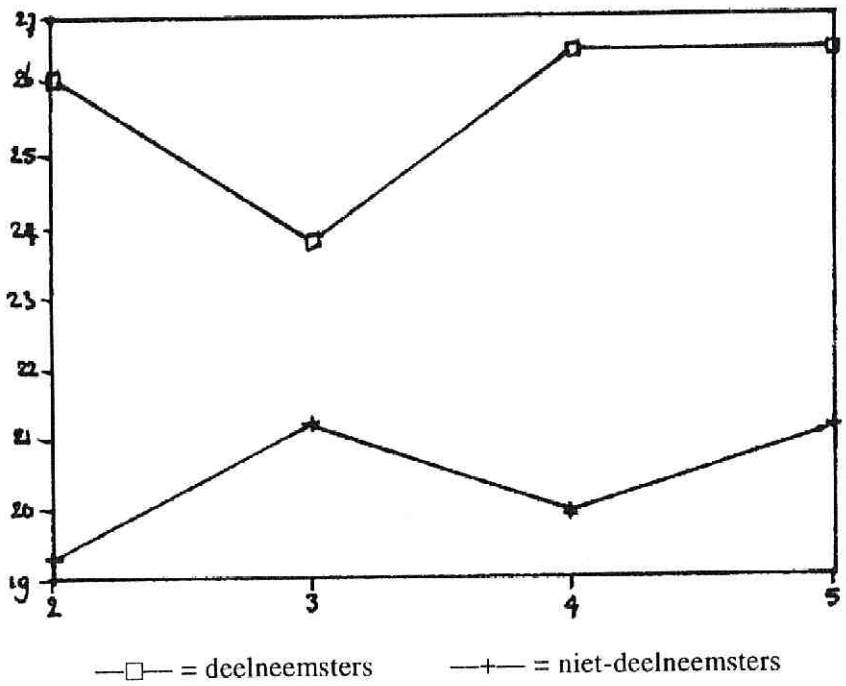


Fig. 3. Gemiddeld aantal werkuren onder deelneemsters en niet-deelneemsters aan volwasseneneducatie naar jeugdopleiding (N=250).

Er zijn wel verschillen tussen meisjes met een hoge en een lage jeugdopleiding (zie Tabel 1, vergelijkingen 2 en 4). Het niveau van de jeugdopleiding blijkt te interacteren met volwasseneneducatie. Het verschil in arbeidsmarktdeelname tussen deelneemsters en niet-deelneemsters neemt toe bij het stijgen van de jeugdopleiding. Volwasseneneducatie heeft dus vooral een positieve invloed op de arbeidsdeelname van vrouwen met een hogere jeugdopleiding.

Om het interactie-effect jeugdopleiding * deelname aan volwasseneneducatie te illustreren is per jeugdopleidingsniveau het percentage werkende vrouwen en het gemiddeld aantal arbeidsuren per week berekend, voor deelneemsters en niet-deelneemsters apart. De resultaten zijn weergegeven in Figuren 2 en 3.

Bij Figuur 2 valt allereerst op dat het percentage werkende vrouwen toeneemt naarmate het jeugdopleidingsniveau hoger is. Dit geldt zowel voor deelneemsters als voor niet-deelneemsters. Daarnaast hebben deelneemsters in alle jeugdopleidingscategorieën vaker een betaalde baan dan de niet-deelneemsters, met uitzondering van het laagste niveau. Vrouwen met alleen lagere school die aan volwasseneneducatie hebben deelgenomen hebben ongeveer even vaak een betaalde baan als niet-deelneemsters met lagere school. Het aantal uren dat vrouwen met een betaalde baan werken (Figuur 3) lijkt echter niet duidelijk samen te hangen met het niveau van de jeugdopleiding, maar wel met deelname aan volwasseneneducatie. Dit in tegenstelling tot de resultaten in Tabel 1.

6.2. Volwasseneneducatie en arbeidsmarktpositie

Is het volgen van opleidingen in de volwasseneneducatie van invloed op de arbeidsmarktpositie van (betaald werkende) vrouwen? Hebben deelneemsters aan volwasseneneducatie betere banen (hoger beroepsniveau en inkomen) dan niet-deelneemsters?

Uit bivariate analyses blijkt dat deelneemsters als groep inderdaad een betere arbeidsmarktpositie hebben dan niet-deelneemsters. Ze hebben een hoger beroepsniveau en een hoger inkomen (Bijv. gemiddeld bruto uurloon deelneemsters f 21,04, van niet-deelneemsters f 17,35). De verschillen zijn statistisch significant (t-toets, $p < 0.05$). De vraag of deze verschillen daadwer-

Tabel 2. Volwasseneneducatie en de arbeidsmarktpositie van werkende vrouwen in 1983. Gestandaardiseerde multiple regressiecoëfficiënten zonder en met interactietermen (N=332).

Afh. var:	beroepsniveau #		opl baan #		uurinkomen #		opl baan	uurink
Onafh. var:								
milieu	*	*	*	*	*	*	*	*
IQ	*	*	*	*	*	*	*	*
wens	*	*	*	*	*	*	*	*
advics	*	*	*	*	*	*	*	*
jeugdopl	0.58	0.66	0.40	0.66	0.30	0.50	0.52	0.44
vedeelname	*	*	*	*	*	*	*	*
jeugdopl * vedeelname							*	*
veniveau	*		0.33		0.21		0.56	0.42
jeugdopl * veniveau							−0.34	−0.30
ve+		0.19		0.33		0.23		
werkjaren	*	*	*	*	*	*	*	*
jongste kind	*	*	*	*	*	*	*	*
oudste kind	*	*	*	*	−0.13	−0.13	*	−0.12
werkkans	*	*	*	*	*	*	*	*
schaarste	0.19	0.19	*	*	*	*	*	*
bedrijfsomvang	*	*	*	*	*	*	*	*
sector	0.17	0.14	*	*	*	*	*	*
werkuren	0.15	0.14	*	*	*	*	*	*
partner	*	*	*	*	*	*	*	*
beroepsniveau					0.23	0.19		0.20
R^2	0.51	0.54	0.38	0.39	0.44	0.45	0.40	0.46

= Bij deze afhankelijke variabelen staan steeds twee regressievergelijkingen. In de linker vergelijking is het niveau van volwasseneneducatie (veniveau) opgenomen, in de rechter vergelijking de niveaustijging (ve+).
* = Niet significant bij $p < 0.05$.

kelijk met deelname aan volwasseneneducatie te maken hebben of dat deze veroorzaakt worden door andere factoren, kan beantwoord worden met behulp van multiple regressie-analyse. De resultaten van de analyse staan in Tabel 2, vergelijkingen 1 tot en met 6.

De verklaarde varianties (R^2) van het beroepsniveau en het inkomen zijn vrij hoog. In die situatie is de kans groter dat het effect van volwasseneneducatie zo goed mogelijk wordt geschat.

Deelname aan volwasseneneducatie heeft op zichzelf geen effect op de arbeidsmarktpositie van vrouwen. Het beroepsniveau en het inkomen zijn wel hoger wanneer vrouwen volwasseneneducatie op een hoger niveau gevolgd hebben (veniveau) en wanneer hun opleidingsniveau meer is gestegen door volwasseneneducatie (ve+). In alle gevallen is het effect van volwasseneneducatie positief. Naast het directe effect van volwasseneneducatie op het uurloon is er ook een klein indirect effect van ve+ op het uurloon via het beroepsniveau (0.19*0.19 = 0.04) De overige effecten van volwasseneneducatie (via de arbeidsmarktsector en het aantal uren van de baan) zijn eveneens gering (0.02 tot 0.03).

Ook het niveau van de jeugdopleiding heeft invloed op de arbeidsmarktpositie van vrouwen. Het effect van de jeugdopleiding is duidelijk groter dan dat van volwasseneneducatie. Dit neemt niet weg dat het effect van volwasseneneducatie substantieel is.

Welke factoren zijn verder van invloed op de arbeidsmarktpositie van deze vrouwen? Het beroepsniveau van vrouwen is hoger wanneer er meer schaarste is, wanneer zij een groter aantal

uren werken en wanneer zij in de non-profit sector werken. Op het uurloon hebben deze factoren alleen een indirecte invloed. Het uurloon is hoger wanneer vrouwen geen kinderen hebben of wanneer de komst van het eerste kind is uitgesteld (oudste kind). In paragraaf 5.2. is reeds gewezen op de mogelijkheid dat de variabele "oudste kind" wellicht ook gedeeltelijk een verschil in werkervaring door loopbaanonderbreking weerspiegelt.

Tegen de verwachting in hebben de aanwezigheid van een partner, de leeftijd van het jongste kind, de kans op werk en de omvang van het bedrijf geen invloed op de arbeidsmarktpositie van deze vrouwen. Ook de individuele capaciteiten (IQ, advies) zijn nauwelijks van invloed op hun arbeidsmarktpositie. Wanneer deze capaciteiten niet omgezet worden in de onderwijsloopbaan komen ze op de arbeidsmarkt niet tot uitdrukking of blijven ze onbenut.

6.2.1. Verschillen in het rendement van volwasseneneducatie (interacties)

Zijn de effecten van volwasseneneducatie, op de arbeidsmarktpositie van vrouwen even groot voor: a. vrouwen uit een hoog en een laag milieu? en b. vrouwen met een hoge en een lage jeugdopleiding? De methodische aanpak is dezelfde als in de vorige paragraaf: aan de regressievergelijkingen worden interactie-effecten toegevoegd van volwasseneneducatie met het ouderlijk milieu resp. met de jeugdopleiding.

De interactie-effecten van milieu met volwasseneneducatie blijken niet significant te zijn. De resultaten zijn daarom niet in een tabel opgenomen. Het betekent dat het rendement van volwasseneneducatie, in termen van beroepsniveau en inkomen, gelijk is voor vrouwen uit verschillende ouderlijke milieus. Hierbij is rekening gehouden met andere factoren zoals de jeugdopleiding. Ook het hoofdeffect van het milieu is niet significant.

Overigens bereiken meisjes uit lagere milieus een lager jeugdopleidingsniveau, en dat is wel van invloed op het rendement van volwasseneneducatie. Per saldo hebben vrouwen uit de lagere milieus daardoor wel meer profijt van volwasseneneducatie op de arbeidsmarkt.

Dit blijkt wanneer de interactie-effecten van jeugdopleiding en volwasseneneducatie aan de regressievergelijkingen worden toegevoegd (Tabel 2, vergelijkingen 7 en 8). Er zijn significante interactie-effecten van jeugdopleiding * veniveau.[6] Een significant interactie-effect betekent dat het effect van volwasseneneducatie op OPL BAAN en op UURINK verschillend is voor vrouwen met verschillende jeugdopleidingsniveaus.

De interactieterm jeugdopleiding * veniveau is negatief: het niveau van volwasseneneducatie weegt zwaarder bij een lagere jeugdopleiding dan bij een hoge jeugdopleiding. Volwasseneneducatie heeft dus het hoogste rendement voor vrouwen met een lagere jeugdopleiding. Met het rendement wordt hier de stijging in het niveau van de baan resp. in de hoogte van het uurloon bedoeld, die het volgen van volwasseneneducatie op 1 niveau hoger teweeg brengt. Bijvoorbeeld: bij vrouwen met een jeugdopleiding op lagere school niveau betekent elk extra niveau volwasseneneducatie een stijging van het uurloon met gemiddeld 33,6 % (logaritme 0,29 hoger). Voor vrouwen met een HBO-opleiding is dat 22,1 % (logaritme 0,20 hoger).

Door de toevoeging van interactie-effecten veranderen de overige parameters in het model nauwelijks. Het niveau van de jeugdopleiding en van de volwasseneneducatie en de interactie daartussen hebben de grootste invloed op het niveau van de baan en het inkomen per uur in 1983.

In Figuren 4 en 5 worden de gevonden (hoofd- en interactie-) effecten geïllustreerd. Het aantal deelneemsters is te klein om onderscheid te maken naar de verschillende niveaus van volwasseneneducatie. Daarom zijn de verschillen tussen wel en niet-deelneemsters per jeugdopleidingsniveau weergegeven.

Figuren 4 en 5 laten zien dat vrouwen een hoger beroepsniveau en inkomen hebben, naarmate ze een hogere jeugdopleiding hebben. Het interactie-effect komt tot uitdrukking in het opmerkelijke verschil tussen deelneemsters en niet-deelneemsters. De lijnen van de deelneemsters en de niet-deelneemsters kruisen elkaar namelijk. Vrouwen die na de lagere school geen verder onderwijs hebben gevolgd en vrouwen met een jeugdopleiding op LBO/MAVO-niveau hebben een betere arbeidsmarktpositie wanneer zij aan volwasseneneducatie hebben deelgenomen. Hun

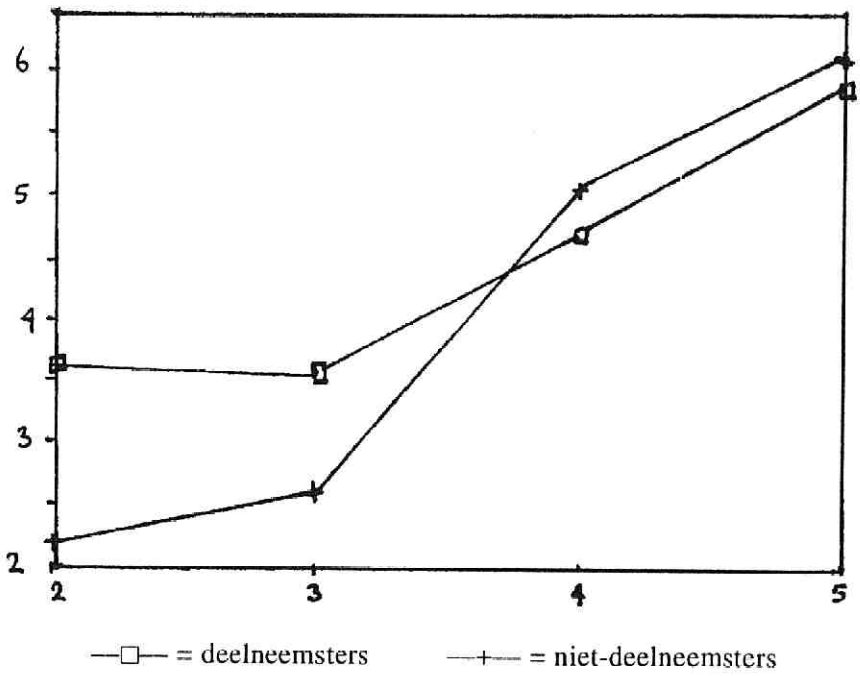


Fig. 4. Gemiddeld beroepsniveau van deelneemsters en niet-deelneemsters naar jeugdopleiding (N=305).

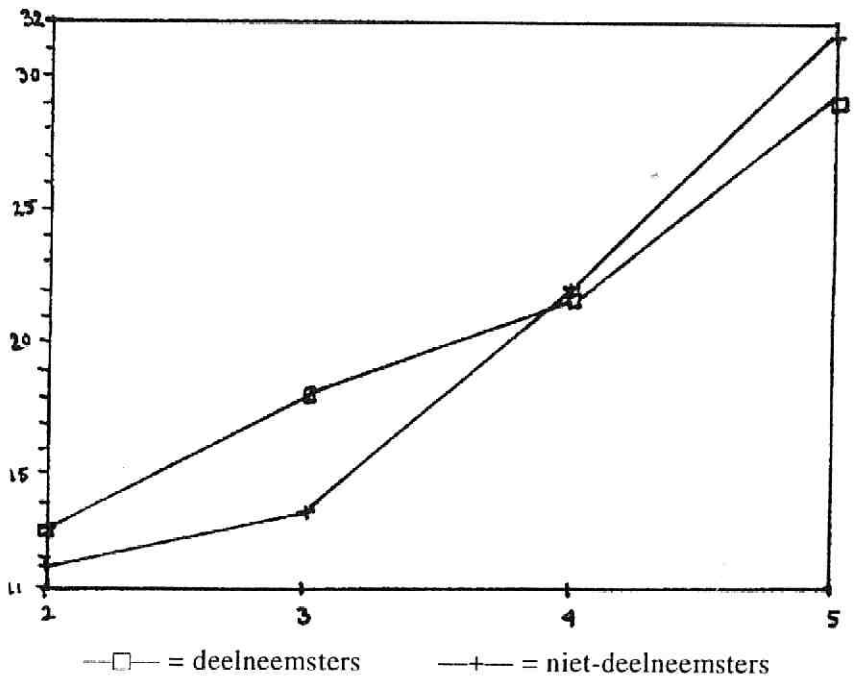


Fig. 5. Gemiddeld uurinkomen van deelneemsters en niet-deelneemsters naar jeugdopleiding (N=250).

beroepsniveau en hun uurloon zijn hoger dan dat van niet-deelneemsters met dezelfde jeugdopleiding. Voor vrouwen met een hogere jeugdopleiding (van HAVO tot WO) blijkt het rendement niet alleen kleiner te zijn dan voor vrouwen met een lagere jeugdopleiding, het rendement is zelfs negatief: deelneemsters hebben een lager beroepsniveau en een lager inkomen dan niet-deelneemsters.

7. SAMENVATTING RESULTATEN EN DISCUSSIE

De vrouwen uit het onderzochte cohort zijn geboren in/rond 1940. Het merendeel van hen is opgevoed met het idee dat hun belangrijkste taak zal liggen in het gezin. Onderwijs werd voor meisjes daarom minder belangrijk gevonden dan voor jongens, en een betaalde loopbaan lag voor de meeste vrouwen niet in het verwachtingspatroon. In het verdere leven van de meeste vrouwen uit het cohort heeft het gezin inderdaad een centrale plaats gekregen. Op 43-jarige leeftijd hebben de meeste vrouwen één of meer thuiswonende kinderen. 36 % van de vrouwen heeft een betaalde baan, meestal part-time. Tegen deze achtergrond moeten de uitkomsten van dit onderzoek gezien worden.

In dit artikel ging het om de vraag of volwasseneneducatie bijdraagt aan een grotere economische zelfstandigheid van vrouwen. Bij het interpreteren van de resultaten wil ik nog eens opmerken dat voornamelijk gegevens gebruikt zijn over deelname aan arbeidsmarktrelevante vormen van volwasseneneducatie. Juist van deze vormen van volwasseneneducatie kan invloed op de arbeidsmarktpositie verwacht worden. Wanneer ook andere, meer sociaal-culturele vormen van volwasseneneducatie in het onderzoek betrokken zouden zijn, dan was de omvang van de deelname waarschijnlijk veel groter, doch het effect ervan op de arbeidsmarktpositie vermoedelijk geringer. Overigens is het effect van volwasseneneducatie vervolgens 'streng' getoetst, doordat opleidingen bij twijfel beschouwd zijn als jeugdopleiding.

Over de invloed van volwasseneneducatie op de *arbeidsmarktdeelname* van vrouwen kunnen de volgende conclusies worden getrokken. Vrouwen die hebben deelgenomen aan volwasseneneducatie hebben vaker een baan en werken gemiddeld meer uren dan niet-deelneemsters. Wanneer rekening gehouden wordt met de overige factoren die van invloed zijn op de arbeidsmarktdeelname van vrouwen, verdwijnt het verschil tussen wel- en niet-deelneemsters echter. Volwasseneneducatie draagt dus niet bij aan de arbeidsmarktdeelname van vrouwen in het algemeen.

Het verband tussen volwasseneneducatie en arbeidsmarktdeelname blijkt af te hangen van het niveau van de jeugdopleiding (interactie-effect). Bij vrouwen met een lage jeugdopleiding is er weinig verschil in arbeidsmarktdeelname tussen deelneemsters en niet-deelneemsters aan volwasseneneducatie. Naarmate het jeugdopleidingsniveau hoger wordt neemt het verschil tussen deelneemsters en niet-deelneemsters toe. Deelneemsters met een hogere jeugdopleiding hebben vaker een baan en werken meer uren dan niet-deelneemsters met een hogere jeugdopleiding.

Grotere arbeidsmarktdeelname van vrouwen blijkt vooral samen te hangen met de afwezigheid van een partner, geen/minder zorgtaken, een hogere jeugdopleiding, een hoger aspiratieniveau van de ouders, een grotere kans om aan een baan te komen en de sector waarin de vrouw haar beroep uitoefent. Een aantal in het model opgenomen factoren blijken, nadat met de hiervoor genoemde zaken rekening gehouden is, niets meer aan de verklaring toe te voegen: het milieu, de individuele capaciteiten, het niveau van volwasseneneducatie, en het opleidingsniveau en het inkomen van de partner hebben geen invloed op de arbeidsmarktdeelname van vrouwen.

Tot slot kwam de *arbeidsmarktpositie* van werkende vrouwen aan de orde. Hier blijken deelneemsters in een gunstiger positie te verkeren dan niet-deelneemsters: ze hebben een hoger beroepsniveau en een hoger inkomen per uur.

Deelname aan volwasseneneducatie draagt niet zonder meer bij aan een betere arbeidsmarktpositie, alleen een flinke investering in volwasseneneducatie (hoger niveau of daadwerkelijke niveaustijging) lijkt te tellen. Dit sluit aan bij de bevindingen van Felmlee (1988). De neiging van vrouwen om in de volwasseneneducatie een niveau te kiezen dat even hoog of zelfs lager is dan de jeugdopleiding lijkt niet in hun voordeel te werken. Mannen hebben deze neiging veel minder dan vrouwen.

Het rendement van volwasseneneducatie is niet voor alle vrouwen gelijk. Vooral vrouwen

met een jeugdopleiding die niet hoger is dan het MAVO-niveau verwerven met behulp van volwasseneneducatie een betere arbeidsmarktpositie. Deelneemsters met een hogere jeugdopleiding hebben in dit opzicht niet veel baat bij volwasseneneducatie. Er zijn zelfs aanwijzingen dat de "opbrengst" van volwasseneneducatie voor vrouwen met een hogere jeugdopleiding negatief is.

Met enige voorzichtigheid kunnen de bevindingen vergeleken worden met die van het onderzoek onder de mannen van dit cohort. Een voorbehoud is nodig omdat in het analysemodel voor de arbeidsmarktpositie van werkende vrouwen een groter aantal variabelen opgenomen is dan bij mannen het geval was. Ten tweede hebben mannen veel intensiever gebruik gemaakt van volwasseneneducatie dan de vrouwen van hun generatie. Het percentage deelnemers is niet alleen hoger, mannen hebben ook meer cursussen gevolgd en een grotere niveaustijging bereikt. Hierdoor is het verschil tussen gemiddelde deelnemers en niet-deelnemers bij mannen groter dan bij vrouwen. Wat betreft niveau en niveaustijging in de volwasseneneducatie zou dit niet tot effecten van verschillende sterkte hoeven te leiden, omdat in beide groepen dezelfde meeteenheid gebruikt is.

Wat blijkt uit deze vergelijking? Net als bij de vrouwen heeft volwasseneneducatie bij de mannen een positieve invloed op het beroepsniveau en inkomen, vooral bij mannen met een lagere jeugdopleiding. Het effect van volwasseneneducatie op het beroepsniveau is bij mannen duidelijk sterker dan bij vrouwen.

Daarentegen hebben vrouwen juist meer financieel voordeel van volwasseneneducatie. Vooral het directe effect van volwasseneneducatie op het inkomen van vrouwen is opmerkelijk. Dit werd noch bij de Nederlandse mannen van het zelfde cohort, noch bij de Zweedse mannen van het Malmö-cohort gevonden. In deze onderzoeken werden alleen indirecte inkomenseffecten gevonden die verliepen via het beroepsniveau (Van Leeuwen, 1990 en Tuijnman, 1989).

Ook de jeugdopleiding heeft bij de vrouwen meer invloed op hun beroepsniveau en inkomen. Het ouderlijk milieu heeft bij de mannen een direct effect op beroepsniveau en inkomen op 43-jarige leeftijd, bij vrouwen niet.

Globaal kan gesteld worden dat het onderwijsniveau in sterkere mate bepalend is voor het inkomen van werkende vrouwen dan voor werkende mannen het geval is. Mede daardoor heeft het ontwikkelde model voor de vrouwen mogelijk ook wat meer verklaringskracht (hogere R^2) dan voor de mannen.

Voor het statusverwervingsonderzoek lijkt het toevoegen van volwasseneneducatie aan de verklaringsmodellen zinvol te zijn. De effecten van volwasseneneducatie zijn substantieel en de verklaringskracht kan er iets door verbeteren.

Vanuit de theorie waren de voorspellingen eenduidig, namelijk dat volwasseneneducatie een positieve invloed heeft op de arbeidsmarktdeelname en op de arbeidsmarktpositie van werkenden. De praktijk blijkt complexer dan de theorie. In twee gevallen waren de resultaten namelijk anders dan verwacht.

Ten eerste is er bij vrouwen met een lagere jeugdopleiding geen verschil in arbeidsmarktdeelname tussen deelneemsters en niet-deelneemsters. Dit zou verband kunnen houden met het grotere aanbod van algemeen vormend onderwijs voor deze vrouwen, in de vorm van het VAVO. Deelneemsters aan algemeen vormend onderwijs hebben mogelijk minder vaak arbeidsmarktgerichte motieven. Het zou interessant zijn dit verschil tussen algemene en beroepsgerichte volwasseneneducatie nader te onderzoeken. Toch is het maar de vraag of dit inderdaad de verklaring levert. Het merendeel van de door de respondenten opgegeven activiteiten in de volwasseneneducatie valt onder de noemer "aanvullende cursussen ten behoeve van het werk". Ook blijken lager opgeleide vrouwen in termen van beroep en inkomen juist meer voordeel te hebben van volwasseneneducatie.

Een tweede opmerkelijke bevinding is de uitblijvende of zelfs negatieve "opbrengst" van volwasseneneducatie voor vrouwen met een hoge jeugdopleiding, in termen van beroepsniveau en inkomen. Deelneemsters met een HBO of universitaire opleiding verdienen per uur ruim een

rijksdaalder (bruto) minder dan niet-deelneemsters met dezelfde jeugdopleiding. Ze hebben ook een lager beroepsniveau. Ook voor deelneemsters met een jeugdopleiding op HAVO/VWO/MBO-niveau heeft volwasseneneducatie een iets ongunstiger arbeidsmarktpositie tot gevolg. Voor vrouwen met een HBO of universitaire opleiding geldt dat het niveau waarop zij cursussen hebben gevolgd (gemiddeld) duidelijk onder het niveau lag van hun eerdere opleiding. Dit zou een verklaring kunnen zijn voor het uitblijven van een positieve invloed van volwasseneneducatie. Het is echter niet duidelijk waarom de "opbrengst" zelfs negatief zou zijn.

Eerder is al opgemerkt dat er in de praktijk waarschijnlijk sprake is van een wisselwerking tussen volwasseneneducatie en de arbeidsmarktpositie van vrouwen. De positie die vrouwen op de arbeidsmarkt (kunnen) hebben, de beslissing om wel of geen betaald werk te verrichten en de deelname aan volwasseneneducatie staan niet los van elkaar. Invloeden kunnen beide kanten uit verlopen. Het gemeenschappelijke mechanisme dat hier achter lijkt te schuilen is dat van het dubbele toekomstperspectief. Vrouwen die sterker gericht zijn op een onbetaalde loopbaan zullen minder geneigd zijn in (beroepsgerichte) volwasseneneducatie te investeren, ze zullen sneller hun baan opgeven en zullen wanneer ze een baan hebben waarschijnlijk ook een minder hoge arbeidsmarktpositie bereiken. Vrouwen die daarentegen sterk op een betaalde loopbaan gericht zijn, zullen meer moeite doen om een plaats op de arbeidsmarkt te verwerven, te behouden of te verbeteren en zullen daarbij ook eerder deelnemen aan volwasseneneducatie. Het zou interessant zijn dit tweerichtingsverkeer nader te onderzoeken.

NOTEN

1. Met dank aan dr. J. Dessens van de vakgroep sociologie (RUU) en Prof. dr. A. van Doorne-Huiskes van de afdeling emancipatie-onderzoek Vrouwenstudies (EUR) voor hun begeleiding en kritisch commentaar.
2. Globaal kan worden gesteld dat met het eerste een (vrijwel) aaneengesloten periode van onderwijs op jonge leeftijd en voorafgaand aan de arbeidsmarktintrede bedoeld wordt, het tweede is aanvullend onderwijs dat hij/zij daarna als volwassene verwerft. Er is veel discussie mogelijk over de vraag waar de grens tussen jeugdonderwijs en volwasseneneducatie getrokken moet worden. Leeftijd, het moment van arbeidsmarktintrede en de continuïteit van de onderwijsloopbaan kunnen criteria zijn.
3. De vraag in de omgekeerde richting, naar de invloed van de arbeidsmarktpositie op deelname aan volwasseneneducatie, zou in principe een even zinvolle vraag zijn. Het bestaan van een voortdurende wisselwerking tussen volwasseneneducatie, arbeidsmarktdeelname en arbeidsmarktpositie lijkt waarschijnlijker dan een eenzijdig causaal verband. Dat in de vraagstelling toch uitsluitend aandacht besteed wordt aan de invloed van volwasseneneducatie op de sociaal-economische positie van vrouwen, heeft te maken met de beperkingen van het gebruikte databestand.
4. Voor meer informatie over de factoren die van invloed zijn op deelname aan volwasseneneducatie zie Houtkoop & van der Kamp, 1992. Over de deelname aan volwasseneneducatie door de vrouwen van dit cohort zie Van Leeuwen, Dessens en Van Doorne-Huiskes, 1992.
5. Meer gedetailleerde informatie over de constructie en schaling van de variabelen is verkrijgbaar bij de auteur.
6. Er zijn geen significante interactie-effecten van jeugdopleiding * ve+. Deze zijn daarom niet in tabel 2 opgenomen. Het zelfde geldt voor het interactie-effect van jeugdopleiding * veniveau bij de afhankelijke variabele "beroepsniveau".

LITERATUUR

Acker, J.R. (1980). Women and Stratification: A Review of Recent Literature. *Contemporary Sociology, 9*, 25-39.

Blau, P.M. & Duncan, O.D. (1967). *The American occupational structure*. New York: John Wiley and Sons.

Cohen, J. & Cohen, P. (1975). *Applied multiple regressions/correlation analysis for the behavioral sciences*. New Jersey: Hillsdale.

Cramer, G. & Houtkoop, W. (1990). *Investing in human resources. Achtergrondrapport over de Volwasseneneducatie in Nederland.* Utrecht: RVE Adviescentrum Volwasseneneducatie.

Dessens, J., van Doorne-Huiskes, J. & Mertens, E. (1990). *Arbeidsmarkt en gezin. Een studie naar de oriëntaties en het gedrag van verschillende generaties vrouwen. OSA-werkdocument W72.* Den Haag: OSA.

Diederen, J. (1991). *Loopbaan tussen 25 en 35 jaar, 4-de fase: over de effecten van geslacht, milieu van herkomst, schoolprestatie, belangstelling en gevolgd onderwijs op de beroepsloopbaan en levensloop.* Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Dronkers, J. (1986). Onderwijs en sociale ongelijkheid. In J.A. van Kemenade, N.A.J. Lagerwey, J.G.M. Leune en J.M.M. Ritzen (eds.). *Onderwijs bestel en beleid, onderwijs en samenleving A. (p. 39-147).* Groningen: Wolters-Noordhoff.

Dronkers, J. (1992). Parents, love and money: The relations between parental class, cognitive skill, educational attainment, occupation, marriage, and family income among Dutch women. In A. Yogev (Ed.). *International Perspectives on Education and Society, Volume 2 (p. 277-293).* Greenwich, Connecticut: JAI Press Inc.

Felmlee, D. (1988). Returning to school and women's occupational attainment. *Sociology of education, 61,* 29-40.

Gooderham, P.N. (1991). Socioeconomic outcomes from adult education. *Adult Education Quarterly, 41,* 203-216.

Graaf, P.M. de (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen.* Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Hartog, J. & Pfann, G. (1985). *Vervolgonderzoek Noord Brabantse zesdeklassers. Bestand P0849, Steinmetzarchief.* Amsterdam: Universiteit van Amsterdam/ Steinmetzarchief.

Houtkoop, W.A. (1985). Volwasseneneducatie en de ongelijke verdeling van educatieve kansen. In J.L. Peschar & A.A. Wesselingh (eds.). *Onderwijssociologie, een inleiding (p. 373-389).* Groningen: Wolters- Noordhoff.

Houtkoop, W.A. & Kamp, M. van der (1992). Factors influencing participation in continuing education. *International Journal of Educational Research, 17,* 537-548.

Jencks, C., Bartlett, S., Corcoran, M., Eaglesfield, D., Jackson, G., McClelland, K., Mueser, P., Olneck, M., Schwartz, J., Ward, S. & Williams, J. (1979). *Who gets ahead? The determinants of economic succes in America.* New York: Basic Books.

Kalleberg, A.L. (1992). Job training in Norway: organisational and individual differences. *International Journal of Educational Research, 17,* 565-579.

Kamp, M. van der & Stroomberg, H. (1991). Onderzoek in de volwasseneneducatie: terugblik en vooruitzicht. In J.K. Koppen, H. Stroomberg & M. van der Kamp (eds.). *Hoger onderwijs en volwasseneneducatie, ORD bundel 1991 (p. 125-134).* Amsterdam: Stichting Centrum voor Onderwijsonderzoek.

Leeuwen, S. van (1990). De meetbare effecten van volwasseneneducatie: een onderzoek naar ongelijke onderwijskansen, volwasseneneducatie en arbeidsmarktpositie. In C.A.C. Klaassen en P. L. M. Jungbluth (eds.). *Onderwijs en samenleving, ORD bundel 1990 (p.197-210).* Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Leeuwen, S. van, Dessens, J. & Doorne-Huiskes, A. van (1992). Volwasseneneducatie en de ongelijke onderwijskansen van vrouwen. In P. Dykstra, P. Kooij en J. Rupp (eds.). *Onderwijs in de tijd; Ontwikkelingen in onderwijsdeelname en nationale curricula (p.70-89).* Boeknummer van Mens en Maatschappij. Houten/Zaventem: Bohn Stafleu Van Loghum.

Matthijssen, M.A.J.M. & Sonnemans, G.J.M. (1983). *Schoolkeuze en schoolsucces bij vhmo en ulo in Noord-Brabant.* Tilburg: Zwijssen.

Meesters, M., Dronkers, J. & Schijf, H. (1983). Veranderende onderwijskansen? Een derde voorbeeld en afrondende conclusies. *Mens en Maatschappij, 58,* 5-27.

Meesters, M. J. (1984). The early school careers of children of working and non-working parents. In B.F.M. Bakker, J. Dronkers & H.B.G. Ganzeboom (eds.). *Social stratification and mobility in the Netherlands, SISWO-publication 291 (p.201-214).* Amsterdam: SISWO.

Ministerie van Sociale Zaken en Werkgelegenheid (1992). *Beleidsplan emancipatie: Met het oog op 1995.* Den Haag: Ministerie van SZW/DCE.

Peschar, J.L. & Wesselingh, A.A. (1985). *Onderwijssociologie, een inleiding.* Groningen: Wolters- Noordhoff.

Psacharopoulos, G. & Velez, E. (1992). Does training pay independent of education? Some evidence from Colombia. *International Journal of Educational Research, 17,* 581-591.

Sanders, K. (1991). *Vrouwelijke pioniers: vrouwen en mannen met een "mannelijke" beroepsopleiding aan het begin van hun loopbaan. Proefschrift Groningen.* Amsterdam: Thesis.

Schütze, H.G. & Istance, D. (1987). *Recurrent education revisited; Modes of participation and financing.* Stockholm, Sweden: Almqvist & Wiksell International.

Tuijnman, A. (1989). *Recurrent education, earnings and well-being; A fifty year longitudinal study of a cohort of Swedish men.* Stockholm, Sweden: Almqvist & Wiksell International.

Wetenschappelijke Raad voor het Regeringsbeleid (1990). *Een werkend perspectief: arbeidsparticipatie in de jaren '90.* Den Haag: SDU.

Manuscript ontvangen 5-4-1993
Definitieve versie ontvangen 13-7-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 4, pp. 227-241

De eerste gehechtheidsrelatie(s) en de cognitieve ontwikkeling van het kind: een literatuuroverzicht

Corine de Ruiter, Marinus H. van IJzendoorn en Joost Goedbloed*

ABSTRACT

This article provides a review of the empirical literature on the relationship between the quality of attachment and cognitive development. First, a brief review of attachment theory is presented and the possible influences of the attachment bond between caregiver and child on the child's cognitive development is formulated in terms of a theoretical model. Subsequently, the empirical literature is reviewed, focusing on parental quality of instruction and emotional support, on children's exploratory and problem-solving competence and metacognitive skills, and on high-risk samples. Despite a number of caveats, the authors conclude that the findings of the research reviewed are promising. At the close of the article, the authors point to possible directions for future research.

INLEIDING

In dit artikel geven we een overzicht van de gehechtheidstheoretische literatuur die de rol van de kwaliteit van de eerste gehechtheidsrelatie tussen kind en opvoeder in de cognitieve ontwikkeling van het kind beschrijft. Wij zullen eerst in het kort een overzicht geven van de gehechtheidstheorie en van recent empirisch onderzoek dat relevant is voor het vervolg van onze analyse. Vervolgens presenteren wij een heuristisch model waarin de theoretische verbanden tussen de kwaliteit van gehechtheid en aspecten van de cognitieve ontwikkeling en latere schoolprestaties uiteengezet worden. Tenslotte zullen wij de beschikbare empirische literatuur bespreken. Studies die de cognitieve ontwikkeling in relatie tot het algemene affectieve klimaat van de opvoeder-kind relatie hebben onderzocht, worden hierbij buiten beschouwing gelaten. Het beperken van onze beschouwing tot gehechtheidstheoretisch onderzoek draagt bij tot een evaluatie van de waarde die de gehechtheidstheorie heeft voor de verklaring van individuele verschillen in cognitieve ontwikkeling. Cognitieve processen die aan bod zullen komen zijn o.a. aandacht, probleemoplossend vermogen, taalvaardigheid, en metacognitieve vaardigheden, maar ook affectieve processen waarvan verondersteld wordt dat zij een belangrijke invloed hebben op de cognitieve ontwikkeling (o.a. zelfwzelfwaardering, beheersingsmotivatie) worden besproken.

GEHECHTHEIDSTHEORIE

De gehechtheidstheorie is een theorie van de sociaal-emotionele ontwikkeling van de mens. John Bowlby (1969, 1973, 1980) beweerde dat een baby geboren wordt met een "gehechtheidsgedragssysteem", waarmee het zich van de nabijheid van de belangrijkste opvoeders (of gehechtheidsfiguren) verzekert. Huilen, reiken, lachen en kruipen zijn vormen van gehechtheidsgedrag. Volgens Bowlby is gehechtheidsgedrag evolutionair adaptief gedrag, omdat het ons

* Mw.dr. C. de Ruiter is als KNAW-fellow verbonden aan de Vakgroep Klinische Psychologie, Universiteit van Amsterdam, Roetersstraat 15, 1018 WB Amsterdam.
Prof.dr. M.H. van IJzendoorn is werkzaam bij de Vakgroep Algemene Pedagogiek van de Rijksuniversiteit Leiden.
J. Jongbloed is student bij de Vakgroep Ontwikkelingspsychologie van de Universiteit van Amsterdam.
Adres: Roetersstraat 15, 1018 WB Amsterdam.

voorheen van bescherming tegen roofdieren in onze "omgeving van evolutionaire aanpassing" verzekerde. Daarnaast noemde hij de mogelijkheid dat gehechtheidsgedrag ervoor zorgt dat kinderen diverse noodzakelijke vaardigheden voor overleving leren van hun gehechtheidsfiguren (Bowlby, 1969/1989, p.224).

Op grond van regelmatige interactie met zijn gehechtheidsfiguren, ontwikkelt de baby een mentale representatie van deze relatie(s). Bowlby (1973,1980) noemde deze mentale representaties "interne werkmodellen", waarmee hij hun dynamische ("werk") aard benadrukte (zie ook Crittenden, 1990). Als de cognitieve vaardigheden toenemen, worden de modellen in toenemende mate "verfijnd". Mary Ainsworth was de eerste die ontdekte dat er individuele verschillen in gehechtheidsgedrag en interne werkmodellen van gehechtheidsrelaties zijn bij babies van 1 jaar (Ainsworth, Blehar, Waters & Wall, 1978). Zij ontwikkelde een laboratoriumprocedure, de zogenaamde "Strange Situation", waarbij babies worden blootgesteld aan een situatie die in toenemende mate stressvol is. Het gehechtheidsgedragssysteem van het kind wordt geactiveerd door het kind in een onbekende speelruimte te plaatsen, door interactie met een onbekende volwassene en door twee korte scheidingen van het kind van zijn gehechtheidsfiguur (meestal de moeder). Uit het gedrag van de baby tijdens de twee fasen van weerzien met de gehechtheidsfiguur is op te maken wat de aard van de relatie met de gehechtheidsfiguur is. Ainsworth e.a. (1978) onderscheidden drie vormen van gehechtheid : veilig (ook wel "B" genoemd), angstig-vermijdend (A), en angstig-ambivalent (C). Vervolgonderzoek bracht een vierde type aan het licht : angstig-gedesorganiseerd (D; Main & Solomon, 1986, 1990).

Kenmerkend voor veilig gehechte kinderen is dat zij bij weerzien met de gehechtheidsfiguur nabijheid zoeken tot deze figuur. Als zij overstuur raken door de scheiding hervinden zij zichzelf relatief snel en gaan dan verder met hun exploratie van het speelgoed en de speelruimte. Ainsworth (1973) creëerde de term "secure base from which to explore" (vert. "veilige basis van waaruit geëxploreerd wordt") om daarmee de rol van de gehechtheidsfiguur te beschrijven voor een veilig gehechte baby. Babies die angstig-vermijdend aan hun opvoeder gehecht zijn manifesteren vermijdend gedrag bij het weerzien. Dit vermijdende gedrag wordt gedemonstreerd door bijvoorbeeld het afwenden van het gezicht of door het richten van de aandacht op de speeltjes in plaats van op de gehechtheidsfiguur. Bij deze kinderen is de gehechtheid-exploratie balans doorgeslagen in de richting van exploratie. Desalniettemin is de kwaliteit van hun exploratie vaak laag in vergelijking met het exploratie gedrag van veilig gehechte kinderen (zie hierna). Angstig-ambivalent gehechte kinderen reageren met een combinatie van nabijheid zoeken en afwerend-boos gedrag op het weerzien met hun gehechtheidsfiguur. Zij zijn soms moeilijk te kalmeren en doen er over het algemeen lang over voordat zij weer verder gaan met het exploreren van hun omgeving. In dit geval is de gehechtheid-exploratie balans vooral doorgeslagen naar de kant van de gehechtheid. Angstig-gedesorganiseerde kinderen missen een consistente strategie waarmee zij de door de "Strange Situation" procedure opgewekte angst het hoofd kunnen bieden. Zij kunnen bijvoorbeeld reageren met een combinatie van vermijdend en ambivalent gedrag of met gedesorganiseerd gedrag (bijvoorbeeld verstijven of stereotypisch gedrag). Diverse onderzoeken hebben de stabiliteit van deze verschillende interne werkmodellen van gehechtheid over tijd aangetoond. Dit geldt in ieder geval voor steekproeven uit de midden-klasse (Main, Kaplan & Cassidy, 1985; Waters, 1978).

In haar baanbrekende Baltimore onderzoek relateerde Ainsworth de drie vormen van gehechtheid aan verschillen in opvoedings- en verzorgingsgedrag van de moeders van de babies (Ainsworth e.a., 1978). Gedurende intensieve thuisobservaties in het eerste levensjaar, ontdekte ze dat de moeders van veilig gehechte babies over het algemeen sensitiever en responsiever ten aanzien van de signalen van hun kinderen waren dan de moeders van angstig gehechte kinderen. Moeders van vermijdend gehechte babies waren het minst sensitief en hielden niet van fysiek contact met hun kinderen. De moeders van de ambivalent gehechte kinderen waren inconsistent responsief en nogal onhandig in hun verzorgende taken. Onafhankelijk vervolgonderzoek bevestigde de bevinding dat moeders van veilig gehechte babies sensitiever en responsiever zijn dan moeders van angstig gehechte babies (Grossmann, Grossmann, Spangler, Suess & Unzner, 1985; Smith & Pederson, 1988). Slechts weinig onderzoeken hebben zich gericht op de ver-

schillen tussen de opvoeders van de vermijdend en ambivalent gehechte categorieën. De onderzoeken waarin dat wel is gebeurd, wezen uit dat moeders van angstig-vermijdende kinderen gekarakteriseerd worden door een doorkruisende en overheersende opvoedingsstijl (Smith & Pederson, 1988; Isabella, Belsky & Von Eye, 1989; Lewis & Feiring, 1989; Isabella, 1990). Moeders van ambivalente kinderen neigen juist tot onderstimulatie van hun kinderen (Belsky, Rovine & Taylor, 1984). Omdat de definiëring van de angstig-gedesorganiseerde gehechtheidscategorie later tot stand is gekomen, is onderzoek naar de antecedenten hiervan nog schaars. Main & Hesse (1990) veronderstellen dat deze vorm van gehechtheid een gevolg kan zijn van beangstigend of angstig gedrag van de gehechtheidsfiguur. Dit gedrag zou het gevolg kunnen zijn van onverwerkte rouw als gevolg van een verlies of een trauma.

Onderzoek naar de antecedenten van de verschillende gehechtheidstypen heeft recent een nieuwe stimulans gekregen door de ontwikkeling van het 'Adult Attachment Interview' (vert. Gehechtheids Biografisch Interview; GBI) door Mary Main en collegae (AAI; George, Kaplan & Main, 1984; Main e.a., 1985; Main & Goldwyn, ter perse). Dit interview biedt de mogelijkheid om bij adolescenten en volwassenen de interne werkmodellen van gehechtheid te classificeren in vier categorieën (*Dismissing, Secure, Preoccupied, Unresolved*). Het interview richt zich op de mentale representaties die de respondent heeft van zijn of haar vroegere relaties met de ouders (of andere belangrijke gehechtheidsfiguren), in plaats van op de feitelijke biografie. Algemene beschrijvingen van de ouders (of andere gehechtheidsfiguren) worden vergeleken met beschrijvingen van meer specifieke episodes waarin de ouders een rol spelen, waarbij inconsistenties en incoherenties over het algemeen als kenmerk van een onveilige gehechtheid beschouwd worden.

Bij *Dismissing* (Ned. Gereserveerd) gehechtheid bestaat er een gunstige indruk van de ouder(s) op algemeen semantisch niveau, welke vergezeld gaat van een onvermogen deze indruk met gunstige episodische herinneringen uit het verleden te staven. Gereserveerde volwassenen hebben de neiging hun vroegere gehechtheidsfiguren te idealiseren en beweren tegelijkertijd dat zij zich weinig uit hun jeugd kunnen herinneren. De interviews met deze volwassenen geven vaak het beeld dat hun ouders afwijzend en weinig liefdevol waren. Veilig gehechte (*Secure*) volwassenen geven in het GBI een coherent beeld van hun vroegere en huidige relatie met hun ouders. Zij lijken hun gehechtheidservaringen en relaties te waarderen zonder dat er sprake is van idealisering of gepreoccupeerde woede. Gepreoccupeerd gehechte personen worden gekenmerkt door een continue gepreoccupeerde betrokkenheid met vroegere of huidige relaties met de ouders. Deze betrokkenheid manifesteert zich in gepreoccupeerde woede en/of in passieve taal en denkprocessen. Een onverwerkte vorm van gehechtheid (*Unresolved*) wordt duidelijk uit incoherente denk- en taaluitingen tijdens discussies van vroegere verliezen en/of trauma's.

Zes onafhankelijke onderzoeken komen tot een overeenstemming van gemiddeld bijna 80% tussen de vorm van gehechtheid van de opvoeder gemeten met het GBI en zijn/haar kind gemeten met de 'Strange Situation', waarbij onderscheid gemaakt werd tussen een angstig en een veilig niveau van gehechtheid. Dit levert bewijs voor de intergenerationele overdracht van interne werkmodellen van gehechtheid (zie Van IJzendoorn, 1992; Van IJzendoorn & De Ruiter, 1991 voor een overzicht). Het feitelijke mechanisme van intergenerationele overdracht is nog onbekend. Main en Goldwyn (ter perse) hebben gesuggereerd dat de veilig gehechte volwassene gehechtheidssignalen zonder al te veel vervormingen kan waarnemen, omdat deze signalen zijn/ haar bestaande mentale representatie van gehechtheid niet bedreigen zoals dat wel het geval is bij een onveilig gehechte volwassene. Het modelleren van het opvoedingsgedrag van de (groot)ouders zou ook een rol kunnen spelen. Welk mechanisme er ook aan ten grondslag ligt, het is zeer waarschijnlijk dat een gedragsconstruct als sensitiviteit/responsiviteit van de ouders in de overdracht belangrijk is. Het kind baseert zijn mentale representaties van gehechtheid immers op het opvoedingsgedrag van de ouders. De relatie tussen gehechtheid van volwassenen en de mate waarin volwassenen adequaat reageren op de signalen van hun baby (responsiviteit) is in verscheidene onderzoeken beschreven (Grossmann, Fremmer-Bombik, Rudolph & Grossmann, 1988; Haft & Slade, 1989).

De ontwikkeling van een intern werkmodel van gehechtheid loopt parallel met de ontwikke-

ling van een intern werkmodel van het zelf. Het kind dat een sensitief-responsieve verzorging gehad heeft zal een "veilig" zelfbeeld van eigenwaarde ontwikkelen. Het angstig gehechte kind, wiens signalen voor contact en troost niet zijn beantwoord met een sensitieve respons, ontwikkelt een zelfbeeld waarin het zichzelf ziet als iemand waarvan men niet houdt. Bij het type angstig-vermijdende gehechtheid lijkt dit negatieve zelfbeeld echter gemaskeerd te worden door een defensief "goed" zelfbeeld (Kobak & Sceery, 1988; Cassidy & Kobak, 1988). De ambivalente vorm van gehechtheid gaat vergezeld van een relatief negatief zelfbeeld (Kobak & Sceery, 1988).

Samengevat, de ervaringen die het kind opdoet in de eerste relaties met zijn belangrijkste opvoeders vormen de interne werkmodellen van het zelf en relaties van het kind. Deze werkmodellen zijn zelf weer van invloed op daarop volgende ervaringen, in de zin dat zij functioneren als een mentale blauwdruk waarmee het individu toekomstige interacties tegemoet treedt. Informatieverwerking, geheugen en denken worden, voor zover deze betrekking hebben op het zelf en relaties, door het model beïnvloed. Het model zorgt voor een selectieve verwerking van de input waardoor het model zichzelf versterkt en verder stabiliseert. De potentiële invloed van deze kwalitatief verschillende gehechtheidsmodellen op de sociale en emotionele ontwikkeling van een kind lijkt vanzelfsprekend en is aangetoond in een aanzienlijke hoeveelheid onderzoek (b.v., Erickson, Sroufe & Egeland, 1985; Lamb, Thompson, Gardner & Charnov, 1985; Main e.a., 1985; Sroufe, Egeland & Kreutzer, 1990).

GEHECHTHEID EN COGNITIEVE ONTWIKKELING: EEN MODEL

Hoe beïnvloedt de kwaliteit van de gehechtheidsrelatie de cognitieve vaardigheden van het kind? Ten eerste is uit recent cross-cultureel onderzoek duidelijk geworden dat de kwaliteit van de gehechtheidsrepresentatie van de opvoeder wordt overgedragen op het kind (zie Figuur 1) via het al dan niet sensitief-responsief handelen van de opvoeder (van IJzendoorn & de Ruiter, 1991). Het begrip sensitiviteit-responsiviteit verwijst naar de mate waarin de opvoeder in staat

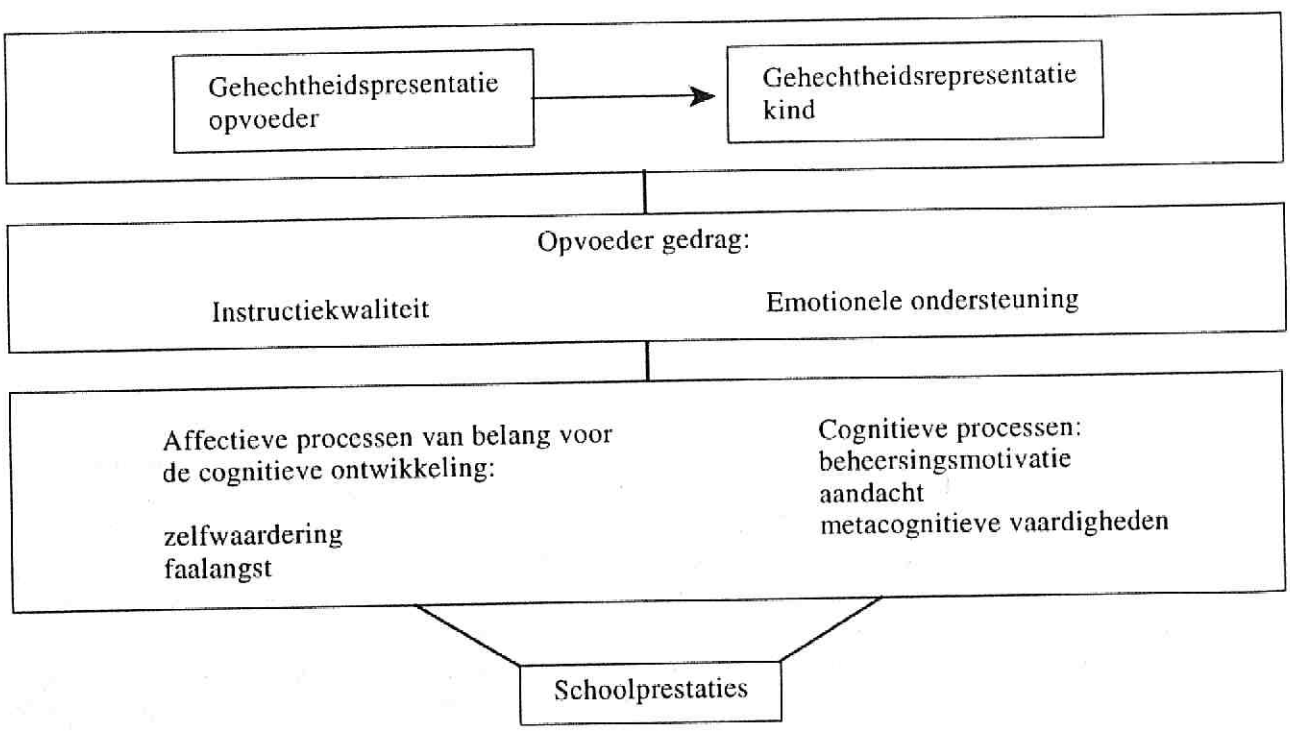


Fig. 1. Een model van de relatie tussen de kwaliteit van gehechtheid en schoolprestaties.

is de signalen die het kind geeft op te vangen, ze juist te interpreteren, en er adequaat op te reageren. Traditioneel is het begrip vooral gebruikt in verband met de emotionele signalen van het kind, maar wij veronderstellen dat sensitiviteit ook een rol speelt in situaties waarin het kind signalen geeft dat hij/zij assistentie behoeft op het cognitieve vlak. Wij verwachten dat opvoeders, afhankelijk van de kwaliteit van hun gehechtheidsrepresentatie, in meer of mindere mate in staat zijn hun kinderen *emotionele ondersteuning* en *adequate instructie* te geven bij het uitvoeren van cognitieve taken.

Hoewel het model uitgaat van intergenerationele overdracht van de gehechtheidskwaliteit van opvoeder op kind, zal duidelijk zijn dat opvoeder en kind in voortdurende transactie zijn, en dat zij elkaar wederzijds beïnvloeden. Het veilig gehechte kind kan zijn gehechtheidsfiguren gebruiken als een *veilige basis* van waaruit het de wereld kan exploreren. Zijn vertrouwen in de fysieke en psychologische beschikbaarheid van de opvoeder, vormt het fundament voor autonome exploratie en probleemoplossend gedrag (Bretherton, 1985). Daarom kunnen we verwachten dat veilig gehechte kinderen enthousiaster zijn om aan een taak te beginnen en dat zij volhardender zijn tijdens het werken aan een taak dan hun onveilig gehechte leeftijdgenootjes (*beheersingsmotivatie*, zie Figuur). Het grotere vertrouwen in hun opvoeders zorgt ervoor dat veilig gehechte kinderen de hulp van hun opvoeders ook weer gemakkelijker uitlokken en accepteren.

De nauwe relatie tussen het interne werkmodel van relaties en het werkmodel van 'het zelf' richt de aandacht op het gebied van de *zelfwaardering* (Cassidy, 1990). Anastasi (1984) geeft een samenvatting van onderzoeken naar de invloed van de algemene zelfwaardering op cognitieve prestaties. Onveilige gehechtheid kan leiden tot een lage zelfwaardering (vooral ambivalente en gedesorganiseerde kinderen) of een defensieve 'opgeblazen' hoge zelfwaardering (vermijdende kinderen). De laatste groep zou een neiging tot *faalangst* kunnen vertonen, wat op zijn beurt weer een negatieve invloed heeft op de cognitieve ontwikkeling.

Tenslotte lijken enkele specifieke cognitieve procesen beïnvloed te kunnen worden door de gehechtheidskwaliteit. Veilig gehechte kinderen kunnen hun *aandacht* volledig aan exploratie wijden omdat zij zich veilig weten. Hoe sterk de aandacht is en hoe lang de aandacht kan worden vastgehouden is van invloed op de cognitieve groei (Anastasi, 1984). Verder wordt verondersteld dat de gehechtheidskwaliteit *metacognitieve processen* beïnvloedt, dat wil zeggen kennis over cognitie en regulatie van cognitieve processen. Een veilig intern werkmodel van gehechtheid is relatief coherent, eenduidig en niet defensief, terwijl een onveilig model gekenmerkt wordt door meerdere contradictoire modellen (bijv. het idealiseren van de ouder zonder ondersteunende episodische herinneringen). Main (1991) heeft beargumenteerd dat deze multipele modellen een aanwijzing zijn dat metacognitieve kennis nog moet ontwikkelen of dat er fouten zijn opgetreden in de correctieve metacognitieve controlemechanismen.

Het model veronderstelt een aantal mediërende factoren in de relatie tussen gehechtheidskwaliteit en aspecten van de cognitieve ontwikkeling. De genoemde mediatoren zijn niet uitsluitend; andere factoren zoals gedragsproblemen kunnen ook een rol spelen. Het model kan een heuristisch doel dienen in de zin dat het een aantal mogelijke onderzoeksrichtingen aangeeft. Bovendien benadrukt het de behoefte aan verdere theorievorming omtrent psychologische (en mogelijk biologische) mechanismen welke ervoor zorgen dat affectieve factoren een invloed hebben op cognitieve processen.

Het zal duidelijk zijn dat het belang van de gehechtheidsrelatie tussen opvoeder en kind vooral relevant is voor theorieën over de cognitieve ontwikkeling waarin het effect van sociale invloeden op cognitieve processen benadrukt wordt. Vygotsky (1978) is de vaandeldrager als het gaat om de rol van sociale interactie, en dan vooral tussen volwassene en kind, in de cognitieve ontwikkeling. Volgens zijn theorie worden de hogere psychologische functies door het kind geïnternaliseerd door middel van sociale interacties met volwassenen. In de theorie van Piaget (1932, 1968) wordt sociale interactie niet van zo groot belang geacht voor de cognitieve ontwikkeling. Wel benadrukte Piaget het belang van symmetrische (leeftijdgenoten) interactie in tegenstelling tot Vygotsky's asymmetrische (kinderen en volwassenen of kinderen en oudere kinderen) interactie voor het faciliteren van de cognitieve ontwikkeling. Het is waarschijnlijk dat de kwaliteit van gehechtheid de sociaal gemedieerde cognitieve ontwikkeling zal bevorde-

ren of remmen in zowel asymmetrische als symmetrische interacties. Tot dusverre heeft het empirisch onderzoek op het gebied van gehechtheid en cognitieve ontwikkeling zich gericht op de asymmetrische (vooral moeder-kind) interactie. In het volgende literatuuroverzicht zal daarom de symmetrische interactie niet aan bod komen.

We zullen de onderzoeksliteratuur over gehechtheid en cognitie bespreken met het oog op het model. Als eerste zullen we het empirisch onderzoek naar instructiekwaliteit en emotionele ondersteuning van de opvoeder tijdens probleemoplossings taken met het kind beschrijven. Daarna zullen we het onderzoek bespreken dat betrekking heeft op de veronderstelde relatie tussen de kwaliteit van gehechtheid en aandacht en beheersingsmotivatie, waaronder wij ook het onderzoek naar exploratief en probleemoplossend vermogen zullen rekenen. Vervolgens zullen we de aandacht richten op het schaarse onderzoek op het gebied van gehechtheid en metacognitie. Tenslotte zullen we studies naar de relatie tussen gehechtheid en cognitieve ontwikkeling bij hoge-risico steekproeven afzonderlijk bespreken.

GEHECHTHEID EN COGNITIEVE ONTWIKKELING: EMPIRISCH ONDERZOEK

Hoewel het "nature-nurture" debat over cognitieve ontwikkeling bedaard is, nu beide partijen het belang van de ander onderkend hebben, moeten "nurture"-voorstanders en dus ook gehechtheidstheoretici erkennen dat genetische factoren een grote rol spelen in de cognitieve ontwikkeling. "Nurture"- voorstanders moeten bijzonder voorzichtig zijn dat zij niet aan opvoeding attribueren wat eigenlijk een gevolg is van aanleg : het probleem van indirekte genetische mediatie. Idealiter zou in alle onderzoeken naar de relatie tussen de kwaliteit van de opvoeder-kind relatie en cognitieve vaardigheden van het kind gecontroleerd moeten worden voor het IQ van de verzorger. Dit is helaas niet gebeurd in het merendeel van de onderzoeken op dit terrein. Men zal hier dus steeds alert op moeten zijn bij het evalueren van de hier gerapporteerde gegevens.

Instructiekwaliteit en emotionele ondersteuning van opvoeders

In een aantal onderzoeken naar probleemoplossingsvaardigheden van jonge kinderen, werd ook het gedrag van de ouder tijdens de probleemoplossingstaken systematisch onderzocht. Matas e.a. (1978) ontwierpen twee zevenpunts beoordelingsschalen, Ondersteunende Aanwezigheid (OA) en Kwaliteit van Assistentie (KA), welke ook in een aantal vervolgonderzoeken gebruikt werden (bijv., Crowell & Feldman, 1988; Frankel en Bates, 1990). De OA-schaal meet de mate waarin de ouder aandachtig en beschikbaar is voor het kind en zijn/haar pogingen ondersteunt. De kern van het OA-construct bestaat uit het zorgen voor een "veilige basis" door het kind op zijn gemak te stellen terwijl het aan de taak werkt en door middel van het tonen van betrokkenheid door aandacht. De KA-schaal meet de mate waarin de ouder het kind helpt de relatie te zien tussen de handelingen die nodig zijn om het probleem op te lossen en het geven van minimale assistentie die noodzakelijk is om het kind gericht op een oplossing te laten doorwerken zonder dat de ouder het oplost (bijv., ruimte geven, het 'timen' en afmeten van aanwijzingen, het verstrekken van voor het kind begrijpelijke aanwijzingen, samenwerken met het kind; Matas e.a., 1979). De KA-schaal kan worden opgevat als een maat voor sensitief ondersteunend gedrag (Bruner gebruikte voor de beschrijving van dit gedrag de metafoor "scaffolding"; Wood, Bruner & Ross, 1978).

Matas e.a. (1978) vonden dat moeders van veilig gehechte babies significant hoger scoorden op de OA- en KA-schaal dan moeders van onveilig gehechte babies. De twee onveilige groepen verschilden niet significant op de beide schalen. Zes-en-twintig van de 48 kinderen van het Matas e.a. onderzoek deden toen zij 4-5 jaar oud waren opnieuw mee aan een aantal laboratoriumtaken (Arend, Gove & Sroufe, 1979). Arend e.a. (1979) verrichtten geen onderzoek naar het gedrag van de moeders tijdens dit laboratoriumbezoek. Zij vonden echter wel dat de OA- en de KA-scores van de moeders, zoals gemeten toen hun kind 2 jaar was, de ego-veerkracht van hun kind op 5-jarige leeftijd in de laboratoriumsituatie konden voorspellen. Frankel en Bates

(1990) repliceerden de bevindingen van Matas e.a.; moeders van onveilige kinderen hadden significant lagere scores op de OA- en KA-schaal dan moeders van veilige kinderen. Interessant is dat zij ook een significante positieve correlatie vonden tussen positieve betrokkenheid in de thuissituatie, gemeten op 6, 13 en 24 maanden, en de kwaliteit van interactie tijdens probleemoplossen op 24 maanden. Crowell en Feldman (1988) bestudeerden het gedrag tijdens probleem-oplossen bij een samengestelde steekproef van klinische en niet-klinische groepen (gem. leeftijd = 37,5 mnd.). In dit onderzoek werd het interne werkmodel van de moeder, gemeten met het GBI, gerelateerd aan het gedrag van moeder en kind tijdens probleemoplossingstaken. Crowell en Feldman (1988) middelden de scores van OA en KA en creëerden zo een samengestelde variabele "ondersteuning en hulp van moeder" genaamd. Bovendien classificeerden zij de assistentiestijl van de moeder op de moeilijkste probleemoplossingstaak in één van de drie volgende groepen: (1) bevordering van autonomie en leren, (2) verwarrend of chaotisch, en (3) sturend of controlerend. Uit de resultaten blijkt dat de volgens het GBI als veilig geclassificeerde moeders significant meer ondersteunend en behulpzaam waren dan de gereserveerde en gepreoccupeerde moeders. Van de veilige moeders toonde 62% een onderwijsstijl welke het leren en zelf-exploratie bevorderde. De meeste moeders uit de gereserveerde groep (78%) waren sturend of controlerend, terwijl de gepreoccupeerde moeders zowel controlerende (35%) en verwarrende/chaotische (60%) instructiestijlen lieten zien.

Londerville en Main (1981) onderzochten vier maten van moederlijk gedrag (toon van de stem, kracht(igheid) van fysieke interventie, aantal verbale opdrachten en aantal fysieke interventies) tijdens een spelsessie van kinderen van 21 maanden oud met een onbekende vrouw. Ze vonden dat moeders van veilige kinderen een warmere toon gebruikten en minder krachtig optraden. Van IJzendoorn, Van der Veer en Van Vliet-Visser (1987) voerden een follow-up onderzoek uit bij kinderen die eerder op de leeftijd van 24 maanden met de 'Strange Situation' getest waren. Moeder-kind koppels werden tijdens vier probleem-oplossingstaken geobserveerd. Het gedrag van de moeder werd gemeten met drie schalen voor emotioneel klimaat (mate van lachen, somtotaal van positieve en negatieve opmerkingen, mate waarin fysieke afstand gehandhaafd wordt) en drie schalen voor instructiegedrag (aantal goede aanwijzingen, aantal interventies, snelheid van interventie op momenten dat het kind niet optimaal presteert). Op drie van de taken konden de vier groepen van gehechtheid (A+C, B1, B2+B3, B4) niet met behulp van de factor 'emotioneel klimaat' van elkaar onderscheiden worden. Dit lukte wel bij de moeilijkste taak, hierin werkte de A+C groep in het minst gunstige klimaat. Moeders van veilig gehechte kinderen gaven inhoudelijk geen betere aanwijzingen dan moeders van angstig gehechte kinderen.

Bus en Van IJzendoorn (1988a) onderzochten als eerste de relatie tussen kwaliteit van gehechtheid en de interactie tijdens leesopdrachten en zich ontwikkelende taalvaardigheden binnen een cross-sectioneel design. Bij kinderen van 1,5 jaar oud werd het type gehechtheid gemeten met de 'Strange Situation' en voor kinderen van 3,5 en 5,5 jaar werd gebruik gemaakt van de 1-uur herenigingsprocedure van Main e.a. (1985). Moeders van veilig gehechte kinderen gaven meer leesinstructie en disciplineerden minder tijdens leestaken. Deze moeders lijken meer van hun kinderen te eisen op het gebied van lezen, waarbij zij de nadruk leggen op leesinstructie en proto-lezen.

Aandacht en beheersingsmotivatie, exploratief en probleemoplossend vermogen

Main (1973) vond dat peuters die als baby veilig gehecht waren een langere aandachtsspanne hadden tijdens vrij spel. Tracy, Farish & Bretherton (1980) vonden in een correlationeel onderzoek met 40 babies geen bewijs voor een relatie tussen de vorm van gehechtheid en exploratieve vaardigheden. Belsky, Garduque & Hrncir (1984) vonden echter dat veilig gehechte babies competenter waren in hun vrije spel dan onveilig gehechte babies; zo vertoonden zij bijvoorbeeld minder verschil in het hoogste niveau van spontaan spelen en het hoogste niveau van het door de onderzoeker uitgelokt spelen. Hazen en Durett (1982) vonden ook dat veilig gehechte peuters aktiever waren in het exploreren van hun omgeving.

Matas, Arend & Sroufe (1978) vonden dat veilig gehechte kinderen van 2 jaar tijdens een vrij

spel sessie significant meer symbolisch spel vertoonden dan vermijdende en ambivalente kinderen. Bovendien waren de veilig gehechte kinderen in meerdere mate enthousiast, meegaand en volhardend, negeerden hun moeder minder, toonden minder frustratie gedragingen en scoorden hoger op positief affect en lager op negatief affect (jengelen/huilen) tijdens twee probleemoplossingstaken. Vaardigheden in probleemoplossen konden niet gereduceerd worden tot verschillen in Ontwikkelings Quotiënt. Een deel van de kinderen uit het Matas e.a. onderzoek werd op 5-jarige leeftijd op ego-veerkracht en ego-controle beoordeeld door hun leerkracht op de kleuterschool (Block & Block, 1979). Ego-veerkracht kan beschouwd worden als een vaardigheidsconstruct want het wordt gedefinieerd als het vermogen om flexibel, volhardend en inventief te reageren in met name probleemsituaties (Arend e.a., 1979). Kinderen die als babies veilig gehecht waren scoorden significant hoger op ego-veerkracht in zowel de leerkracht beoordeling als in de laboratorium-gebaseerde meting. Bovendien scoorden ze significant hoger op drie metingen van nieuwsgierigheid.

De Matas e.a. en de Arend e.a. onderzoeken worden veel geciteerd om de relatie tussen het type gehechtheid en interactie tijdens probleemoplossen in de peuter/kleuter leeftijd aan te tonen. Aangezien beide onderzoeken echter in hetzelfde laboratorium verricht werden, was onafhankelijke replicatie noodzakelijk. Frankel en Bates (1990) publiceerden een dergelijke replicatie en vonden dat veilig gehechte peuters meer taakgericht gedrag, minder agressief gedrag en minder verbaal negativisme vertoonden tijdens probleemoplossingstaken dan onveilig gehechte peuters. Zij vonden echter geen ondersteuning voor de Matas e.a. bevindingen van een significant verschil in meegaandheid, negeren van commando's van moeder en frustratie of jengelen/huilen.

Oppenheim, Sagi en Lamb (1988) verrichtten een onderzoek met 59 vijf jaar oude kibbutz kinderen, waarvan de gehechtheid aan moeder, vader en metapelet bepaald was met de 'Strange Situation' toen zij 11 tot 14 maanden oud waren. De kinderen werden daarnaast door hun kleuterschool leerkrachten en metaplot beoordeeld op de 'California Child Q-set' (CCQ; Block en Block, 1979) en de 'Preschool Behavior Q-set' (Baumrind, 1968). Er waren geen significante verbanden tussen moeder-kind en vader-kind gehechtheid en de beoordelingen op 5-jarige leeftijd, maar kinderen die veilig gehecht waren aan hun metaplot op één-jarige leeftijd werden als minder ego-gecontroleerd en als meer empathisch, dominant, doelbewust, prestatiegericht en onafhankelijk beoordeeld dan angstig-ambivalent gehechte babies (in de steekproef zaten geen vermijdend gehechte kinderen).

In het eerdergenoemde follow-up onderzoek van Van IJzendoorn e.a. (1987) beoordeelden ouders en kleuterschool-leerkrachten de kinderen op de Nederlandse versie van de CCQ (Van Lieshout, Riksen-Walraven, Ten Brink, Siebenheller, Mey e.a.,1983). Er werd geen significant verschil in ego-veerkracht gevonden tussen veilig en angstig gehechte kinderen, in zowel de ouder- als de leerkrachtbeoordeling. Volgens de leerkrachten hadden angstig gehechte meisjes de minst optimale ego-controle, terwijl angstig gehechte jongens een optimale ego-controle vertoonden. Het is lastig deze onderzoeksgegevens te vergelijken met de eerder beschreven onderzoeken aangezien er in de analyse gewerkt is met een verdeling van gehechtheid in vier groepen : A+C, B1, B2+B3 en B4.

In het eerdergenoemde onderzoek van Crowell en Feldman (1988) verschilde het gedrag van de kinderen met name op variabelen die het affect vaststelden en minder op taakrelevant gedrag. Kinderen van onveilig gehechte moeders waren minder aanhankelijk, meer negatief en vermijdend, meer controlerend en angstig en toonden meer onderdrukt en boos affect. Er werden echter geen verschillen in taakgedrag zoals volharding, zelfvertrouwen en enthousiasme gevonden tussen kinderen van veilig en onveilig gehechte moeders. Dat er geen verschillen in taakgedrag gevonden werden is mogelijk een gevolg van het gebruik van een gemengde steekproef. De kinderen van gepreoccupeerde moeders scoorden wel significant lager op volharding dan die van de gereserveerde moeders.

Een aantal onderzoekers hebben de relatie tussen de kwaliteit van gehechtheid en Ontwikkelings Quotiënt of Intelligentie Quotiënt onderzocht. De meerderheid is er niet in geslaagd een significant verschil in OQ te vinden tussen veilig en onveilig gehechte kinderen (Matas, Arend

& Sroufe, 1978; Joffe, 1981; Pastor, 1981; Waters, Wippman & Sroufe, 1979). In drie onderzoeken werd wel een significant verschil gevonden. Main (1983) vond dat veilige kinderen competenter waren op de Bayley test op een leeftijd van 20 maanden. Van IJzendoorn, Sagi en Lambermon (ter perse) doen verslag van een follow-up onderzoek bij Nederlandse en Israëlische kinderen, die eerder geobserveerd waren in de "Strange Situation" met hun moeder, vader en professionele verzorger. De Nederlandse kinderen werden op ongeveer vier-jarige leeftijd getest met de McCarthy Ontwikkelings Schaal (MOS; Van der Meulen & Smrkovsky, 1985) en de Israëlische kinderen werden op vijf-jarige leeftijd getest met de WPPSI test (Lieblich, 1974). In de Nederlandse steekproef vond men een lage maar significante correlatie van DQ met veiligheid in netwerkgehechtheid (dit is een samengestelde score van de gehechtheid van de drie in het netwerk voorkomende dyaden). In de Israëlische steekproef werd een iets hogere en significante correlatie met zowel het netwerk als de samengestelde gezinsscore gevonden. Tenslotte vonden Van IJzendoorn en Van Vliet-Visser (1988) dat veilig gehechte (B2+B3) 5-jarige kinderen significant hoger scoorden op een gestandaardiseerde Nederlandse IQ-test. De marginaal veilige categorieën (B1 en B4) scoorden het laagst maar verschilden niet significant van A+C kinderen.

In het eerdergenoemde onderzoek naar de zich ontwikkelende taalvaardigheden in de voorschoolse periode van Bus en Van IJzendoorn (1988a) werd gevonden dat veilig gehechte kinderen de verhalen en de illustraties meer bleken te exploreren dan angstig gehechte kinderen. Bus en Van IJzendoorn (1988b) vonden tevens een positieve relatie tussen de interesse in lezen bij kleuters en kwaliteit van gehechtheid (zoals dat drie jaar eerder was gemeten), onafhankelijk van intelligentie en mate van voorbereidende leesinstructie. Bus (1993) vond ook dat moeders met veilige mentale representaties van gehechtheid, zoals vastgesteld met behulp van het GehechtheidsBiografisch Interview, hun kinderen vaker voorlezen dan moeders met onveilige gehechtheidsrepresentaties. Veilig gehechte moeders hoeven hun kinderen ook minder te disciplineren tijdens een leestaakje, en hun kinderen hebben een grotere vaardigheid in het voorbereidend lezen.

Metacognitie

De theoretische en empirische integratie van gehechtheidstheorie met onderzoek naar de metacognitieve ontwikkeling is een zeer recent streven (Moss, 1992; Moss, Parent, Gosselin & Dumont, 1993). Er is onderzoek dat er op wijst dat het trainen van metacognitieve strategieën door de ouders invloed heeft op de metacognitieve ontwikkeling (Carr, Kurtz, Schneider, Turner & Borkowski, 1989; Moss & Strayer, 1990), maar in de literatuur komt men nog slechts één onderzoek tegen waarin de rol van de kwaliteit van gehechtheid in relatie tot de metacognitieve ontwikkeling direct is bestudeerd (Moss, Parent, Gosselin en Dumont, 1993). Deze auteurs vonden dat moeders van veilig gehechte peuters meer geneigd waren de activiteiten van hun kind tijdens een taaksituatie verbaal te begeleiden en te evalueren. De veilig gehechte kinderen maakten ook meer gebruik van metacognitieve strategieën tijdens het uitvoeren van de taak.

Hoge-risico steekproeven

Onderzoeken naar de relatie tussen kwaliteit van gehechtheid en cognitieve ontwikkeling in hoge-risico steekproeven moeten apart van lage-risico steekproeven beschouwd worden, omdat de hoge-risico omgeving een aantal risico factoren bevat die, onafhankelijk van de gehechtheidskwaliteit, van invloed zijn op de cognitieve ontwikkeling. Zulke factoren zijn bijvoorbeeld een lage levensstandaard, één-ouder gezinnen, psychiatrische stoornissen van de ouder(s), waarvan gesteld kan worden dat deze de kwaliteit van de affectieve band kunnen beïnvloeden.

In het Minneapolis onderzoek van achtergestelde gezinnen (hoge-risico steekproef) was de kwaliteit van gehechtheid systematisch gerelateerd aan de latere sociaal-emotionele en cognitieve ontwikkeling (b.v., ego-veerkracht en ego-controle). Een aantal verschillende verslagen uit dit onderzoek meldden significante voorspellingen uit de kwaliteit van de vroege gehechtheid voor de latere ontwikkeling. Daarbij moet wel worden opgemerkt dat de voorspellingen positief beïnvloed kunnen zijn door het gebruik van een selectiecriterium voor de steekproef;

stabiliteit in gehechtheid van 12 tot 18 maanden. Het is bekend dat de kwaliteit van gehechtheid minder stabiel is in hoge-risico dan in lage-risico steekproeven (zie Lamb e.a., 1985, hoofdstuk 8 voor een overzicht). Sroufe (1983) deed onderzoek bij 40 kleuters uit achtergestelde gezinnen, die deelnamen aan een speciaal kleuterschool programma gericht op voorbereiding voor de basisschool. Er werd gebruik gemaakt van de onderwijzer Q-sort voor ego-veerkracht en ego-controle uit het Arend e.a. onderzoek (1979). De bevindingen uit dit laatste onderzoek werden grotendeels gerepliceerd. Kinderen die als babies veilig gehecht waren scoorden significant hoger op ego-veerkracht dan zij die vermijdend of ambivalent gehecht waren, waarbij tussen de twee onveilige categorieën onderling geen verschillen werden gevonden. Veilig gehechte kinderen scoorden ook significant hoger op een Q-sort meting van zelfwaardering. Erickson, Sroufe & Egeland (1985) onderzochten een achtergestelde steekproef bestaande uit de 40 kinderen van het Sroufe (1983) onderzoek en 56 andere kleuterschool kinderen. Vier van de zeven observator gedragsmaten (zelfvertrouwen, afhankelijkheid, sociale vaardigheden, medewerking) afgenomen tijdens de les, leverden significante verschillen op. Geen enkele analyse maakte echter op hetzelfde moment onderscheid tussen B en zowel A als C kinderen. De onderwijzerbeoordeling van de "Preschool Behavior Questionaire" (Behar & Stringfield, 1974) levert vijf factoren op waarvan er twee significante verschillen tussen de groepen te zien gaven. Vermijdend gehechte kinderen werden als vijandiger beoordeeld dan ambivalent gehechte kinderen en zouden minder doorzettingsvermogen hebben dan veilig gehechte kinderen. Helaas werd er in het Minnesota onderzoek met de achtergestelde steekproef geen follow-up met zuiver cognitieve maten, zoals vaardigheden in probleemoplossen, uitgevoerd. Over het algemeen lijkt het erop dat de verschillen tussen veilig en onveilig gehechte kinderen nogal zwak zijn in deze steekproef in vergelijking met de gegevens uit de middenklasse steekproef (zie Matas e.a., 1978; Arend e.a., 1979).

Morisset, Barnard, Greenberg, Booth & Spieker (1990) onderzochten de invloed van een aantal omgevingsrisico factoren (sociaaleconomische status, conversatievaardigheden van moeder, en een samengestelde score van dyadische interactie en kwaliteit van gehechtheid) op de Bayley scores (24 maanden) en de "Preschool Language Scales" (Zimmerman, Steiner & Pond, 1979) bij een hoog-risico steekproef. Uit een hiërarchische regressie analyse bleek dat de voorspelling van de Bayley 24-maand score vrij zwak was. Echter, 34% van de taal-totaal-quotiënt van de "Preschool Language Scales" en 46% van het auditieve begrip werd voorspeld door de risico factoren, waarvan respektievelijk 20% en 19% volledig toe te schrijven waren aan de dyadische factor (moeder-kind interactie en gehechtheid). In een aparte analyse waarbij binnen de hoog-risico steekproef een groep kinderen met extreem risico vergeleken werd met een groep kinderen met (relatief) laag risico, werd gevonden dat veilige gehechtheid bij de extreme risicogroep als een beschermende factor werkte terwijl dit niet zo was bij de lage risicogroep.

DISCUSSIE

Indien we de hier besproken onderzoeksbevindingen samenvatten, kunnen we stellen dat ons overzicht steun verleent aan het idee dat de kwaliteit van gehechtheid van invloed is op de cognitieve ontwikkeling van het kind. Onderzoek met zowel normale als hoog-risico steekproeven laat zien dat een veilige gehechtheidsrelatie samenhangt met meer harmonieuze interacties in taaksituaties en de cognitieve vaardigheden van het kind positief beïnvloedt. Een opvoeder die zelf veilig gehecht is (en meestal ook een veilig gehecht kind heeft), neigt ertoe sensitiever ondersteunend ('scaffolding') en adequater instructie-gedrag te vertonen tijdens probleemoplossingssituaties met het kind dan een onveilig gehechte opvoeder. Het onderzoek naar de relatie tussen kwaliteit van gehechtheid en OQ/IQ was het minst eenduidig, maar dit kan een gevolg zijn van het feit dat de genetische aanleg van het kind een grotere rol speelt in de bepaling van het OQ, zoals gemeten met gestandaardiseerde tests, dan in de bepaling van exploratief gedrag en algemene probleemoplossingsvaardigheden.

Rogoff (1990) heeft beweerd dat *vrijheid van expressie* van groot belang is voor de emotio-

nele ontwikkeling en dat *vrijheid om fouten te maken* essentieel is voor de cognitieve ontwikkeling. Het hier besproken onderzoek heeft laten zien dat beide lijken samen te hangen, ieder vertegenwoordigt de acceptatie van het kind en het sensitief reguleren van de initiatieven van het kind door de ouder.

De resultaten van onderzoek naar de relatie tussen gehechtheid en cognitieve ontwikkeling zijn veelbelovend en daarom willen we afsluiten met enkele opmerkingen en aanbevelingen voor toekomstig onderzoek. Onderzoek naar de relatie tussen gehechtheid en cognitieve ontwikkeling is van relatief recente datum hetgeen ten dele een oorzaak is van het gebrek aan follow-up onderzoeken tot in de schoolleeftijd en daarna. Longitudinale onderzoeken van dergelijk lange termijn zijn noodzakelijk als we willen aantonen dat de *vroege* sociaal-affectieve band met de opvoeder van invloed is op de latere cognitieve en opleidingsontwikkeling. Het merendeel van de studies heeft betrekking op de cognitieve prestatie van peuters en kleuters, wanneer het kind nog niet is blootgesteld aan een zeer groot aantal andere mogelijk invloedrijke figuren, zoals onderwijzers en leeftijdgenootjes. Zoals reeds vermeld heeft het onderzoek zich uitsluitend gericht op asymmetrische interacties. Onderzoek naar cognitieve ontwikkeling in symmetrische relaties is derhalve wenselijk.

Lamb e.a. (1985) bekritiseerden de beweringen van gehechtheidsonderzoekers dat de vroege baby-opvoeder gehechtheidsrelatie causaal gerelateerd is aan latere ontwikkelingsuitkomsten, omdat zij in de meeste onderzoeken niet controleerden voor de huidige kwaliteit van de opvoeder-kind relatie. Deze kritiek is ook van toepassing op het leeuwedeel van de onderzoeken uit ons overzicht. Wel moet worden opgemerkt dat het controleren voor de huidige kwaliteit van de relatie alleen nodig is indien de invloed van de vroege relatie *onafhankelijk* van de latere relatie bepaald moet worden. Indien iemand geïnteresseerd is in de invloed van gehechtheid op cognitie sec, dan is het controleren voor contemporaine factoren niet van wezenlijk belang. Bovendien blijkt de kwaliteit van het interne werkmodel van gehechtheid relatief stabiel te zijn in middenklasse steekproeven (Main e.a., 1985). Het bepalen van contemporaine invloeden kan vooral van belang zijn in steekproeven waar de kwaliteit van gehechtheid grotere schommelingen ondergaat als gevolg van stressoren in de omgeving.

Een tekortkoming in vrijwel alle onderzoeken is dat verzuimd is het IQ van de opvoeder van het kind te meten. Hoewel Barocas e.a. (1991) beweerd hebben dat het IQ van de moeder waarschijnlijk geen belangrijke invloed uitoefent op de affectieve component van haar instructiestijl, is het niet ondenkbaar dat intelligentie in een aantal gevallen gerelateerd is aan de kwaliteit van het interne werkmodel van gehechtheid van de ouder. Wij speculeren dat een individu met aanzienlijke intellectuele middelen deze middelen op een dusdanige wijze kan aanwenden dat zijn/haar interne werkmodel van gehechtheid relatief open zal staan voor nieuwe ervaringen en informatie. Intelligentie zou aldus de ontwikkeling van een veilig intern werkmodel kunnen faciliteren, zelfs bij individuen waarvan verwacht zou worden dat zij zich tot onveilig gehechte volwassenen zouden ontwikkelen doordat zij afwijzende en/of inconsistente ouder(s) hadden. Met evenveel recht kan men echter het tegendeel beweren, namelijk dat een hoge intelligentie de kans op intellectuele afweermechanismen doet toenemen. Een voorbeeld hiervan is rationalisatie waardoor het interne werkmodel gestabiliseerd wordt door het niet verwerken van die informatie welke in strijd is met het bestaande model. Alleen empirisch onderzoek kan aantonen welke van deze beide speculaties de werkelijkheid het best benadert.

Gehechtheidsonderzoekers hebben zich hoofdzakelijk gericht op de verschillen tussen veilig en onveilig gehechte kinderen. Omdat er veelal kleine steekproeven gebruikt zijn, zijn gedetailleerde analyses waarbij vermijdend en ambivalent gehechte kinderen vergeleken worden of waarbij men de aandacht heeft gericht op de gedesorganiseerd gehechte kinderen, zeldzaam. Voor een theoretisch inzicht in de specifieke ontwikkelingsconsequenties van deze verschillende gehechtheidsstrategieën, zijn dit soort studies noodzakelijk. Main (1990) beweerde dat vermijdend gehechte kinderen onder stressvolle omstandigheden meer (kwalitatief lage) exploratie gaan vertonen ten koste van gehechtheidsgedrag, terwijl ambivalent gehechte kinderen juist meer gehechtheidsgedrag en minder exploratie vertonen. De gehechtheidstheorie voorspelt voor kinderen met deze tegenovergestelde strategieën verschillende uitkomsten met betrekking tot de

cognitieve ontwikkeling. Kinderen die als babies als angstig-gedesorganiseerd geclassificeerd zijn, lijken een bijzonder risico te lopen aangezien zij met een hoge frequentie in de hoge-risico steekproeven voorkomen (kinderen van depressieve moeders; Lyons-Ruth, Connell, Grunebaum & Botein, 1990; alcoholverslaafde moeders; O'Conner, Sigman & Brill, 1987; drugsverslaafde moeders; Rodning, Beckwith & Howard, 1989; zie Van IJzendoorn, Goldberg, Kroonenberg & Frenkel, 1992, voor een overzicht). Main e.a. (1985) vonden dat kinderen die als baby de classificatie gedesorganiseerd hadden gekregen in de "Strange Situation" op 6-jarige leeftijd bestraffend gedrag dan wel angstig, overdreven bezorgd gedrag naar de ouder vertoonden tijdens een hereniging volgend op een scheiding van één uur. De gedesorganiseerde kinderen presteerden in vergelijking met vermijdende, ambivalente en veilige kinderen het slechtst qua vloeiendheid van taalgebruik en emotionele openheid tijdens een interview over het gezin. Deze gedragingen (vloeiendheid, openheid en bestraffende-controlerende gedragsstijlen) zijn naar alle waarschijnlijkheid van invloed op de cognitieve ontwikkeling van het kind.

Uit het huidige overzicht blijkt dat veilig gehechte kinderen een voorsprong hebben op onveilig gehechte kinderen als zij de basisschool-leeftijd bereikt hebben. Het is niet ondenkbaar dat deze voorsprong nog verder wordt uitgebreid als de kinderen eenmaal naar school gaan, vanwege het zgn. "Matthew-effect" (Stanovich, 1986), waarbij initiële verschillen tussen leerlingen toenemen in plaats van afnemen ten gevolge van scholing. De invloed van de vroege gehechtheidsontwikkeling op leren en instructie, thuis en op school, kon derhalve wel eens groter zijn dan tot nu toe werd vermoed. Uiteraard kan alleen toekomstig empirisch onderzoek tot in de schoolleeftijd uitwijzen of onze hypothesen juist zijn.

LITERATUUR

Ainsworth, M.D.S. (1973). Systems for rating maternal care behavior. In: E.G. Boyer, A. Simon & G.R. Karafin (Eds.), *Measures of maturation: An anthology of childhood observation instruments*. Philadelphia: Research for better schools.

Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Anastasi, A. (1984). Reciprocal relations between cognitive and affective development-with implications for sex differences. In: T.B. Sonderegger (Ed.), *Psychology and gender. Nebraska symposium on motivation, 34*, 1-35.

Arend, R., Gove, F.L., & Sroufe, L.A. (1979). Continuity of individual adaptation from infancy to kindergarten: A predictive study of ego-resiliency and curiosity in preschoolers. *Child Development, 50*, 950-959.

Barocas, R., Seifer, R., Sameroff, A.J., Andrews, T.A., Croft, R.T., & Ostrow, E. (1991). Social and interpersonal determinants of developmental risk. *Developmental Psychology, 27*, 479-488.

Baumrind, D. (1968). *Manual for the Preschool Behavior Q-sort*. Department of Psychology. University of California, Berkeley.

Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601-610.

Belsky, J., Garduque, L., & Hrncir, E. (1984). Asessing performance, competence, and executive capacity in infant play: Relations to home environment and security of attachment. *Developmental Psychology, 20*, 406-417.

Belsky, J., Rovine, M., & Taylor, D. (1984). The Pennsylvania Infant and Family Development project, III. The origins of individual differences in infant-mother attachment: Maternal and infant contributions. *Child Development, 55*, 718-728.

Block, J.H., & Block, J. (1979). The role of ego control and ego-resiliency in the organization of behavior. In: W.A. Collins (Ed.), *Minnesota symposia on child psychology, Vol. 11*. Hillsdale, NJ: Erlbaum.

Bowlby, J. (1969). *Attachment and loss. Vol 1: Attachment*. New York: Basic Books.

Bowlby, J. (1973). *Attachment and loss. Vol 2: Separation: Anxiety and anger*. New York: Basic Books.

Bowlby, J. (1980). *Attachment and loss. Vol 3: Loss: Sadness and depression*. New York: Basic Books.

Bowlby, J. (1969/1989). *Attachment and loss. Vol 1: Attachment*. London: Penguin

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. In: I. Bretherton & E. Waters (Eds.).

Growing points of attachment theory and research. Monographs of the Society for Research in Child Development, Serial no. 209, Vol. 50, Nos. 1-2. pp.3-35.

Bus, A.G. (1993). Attachment and emergent literacy. *International Journal of Educational Research,* in press.

Bus, A.G., & Van IJzendoorn, M.H. (1988a). Mother-child interactions, attachment, and emergent literacy: A cross-sectional study. *Child Development, 59,* 1262-1272.

Bus, A.G., & Van IJzendoorn, M.H. (1988b). Attachment and early reading: A longitudinal study. *Journal of Genetic Psychology, 149,* 199-210.

Carr, M., Kurtz, B., Schneider, W., Turner, L., Borkowski, J. (1989). Strategy acquisition and transfer among American and German children: Environmental influences on metacognitive development. *Developmental Psychology, 25,* 765-771.

Cassidy, J. (1990). Theoretical and methodological considerations in the study of attachment and the self in young children. In: M.T. Greenberg, D. Cicchetti & E.M. Cummings (Eds.), *Attachment in the preschool years: Theory, research and intervention,* pp. 87-119. Chicago: University of Chicago Press.

Cassidy, J., & Kobak, R.R. (1988). Avoidance and its relation to other defensive processes. In: J. Belsky & T. Nezworski (Eds.), *Clinical implications of attachment,* pp. 300-323. Hillsdale, N.J.: Erlbaum.

Crittenden, P.M. (1990). Internal representational models of attachment relationships. *Infant Mental Health Journal, 11,* 259-277.

Crowell, J.A., & Feldman, S.S. (1988). Mothers' internal models of relationships and children's behavioral and developmental status: A study of mother-child interaction. *Child Development, 59,* 1273-1285.

Erickson, M.F., Sroufe, L.A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high-risk sample. In: I. Bretherton & E. Waters (Eds.), *Growing points of attachment theory and research. Monographs of the Society for Research in Child Development, Serial no.209, Vol.50, Nos.1-2,* pp. 147-166.

Estrada, P., Arsenio, W.F., Hess, R.D., & Holloway, S.D. (1987). Affective quality of the mother-child relationship: Longitudinal consequences for children's school-relevant cognitive functioning. *Developmental Psychology, 23,* 210-215.

Frankel, K.A., & Bates, J.E. (1990). Mother-toddler problem solving: Antecedents in attachment, home behavior, and temperament. *Child Development, 61,* 810-819.

George, C., Kaplan, N., & Main, M. (1984). Attachment interview for adults. Unpublished manuscript, University of California, Berkeley.

Grossmann, K., Grossmann, K.E., Spangler, G., Suess, G., & Unzner, L. (1985). Maternal sensitivity and newborns' orientation responses as related to quality of attachment in northern Germany. In: I. Bretherton & E. Waters (Eds.). *Growing points of attachment: Theory and research. Monographs of the Society for research in child development. Serial no. 209, Vol. 50, Nos. 1-2,* pp.233-256.

Grossmann, K., Fremmer-Bombik, E., Rudolph, J., & Grossmann, K. E. (1988). Maternal attachment representations as related to patterns of infant-mother attachment and maternal care during the first year. In: R.A. Hinde & J. Stevenson-Hinde (Eds.), *Relationships within families: Mutual influences,* pp. 241-260. Oxford: Clarendon Press.

Isabella, R.A. (1990, April). *Origins of attachment: Infant-mother interaction across the first year of life.* Poster presented at the Seventh biennial International Conference on Infant Studies. Montreal, Canada.

Isabella, R.A., Belsky, J., & Von Eye, A. (1989). Origins of mother-infant attachment: An examination of interactional synchrony during the infant's first year. *Developmental Psychology, 25,* 12-21.

Joffe, L. (1981). *The quality of mother-infant attachment and its relationship to compliance with maternal commands and prohibitions.* Paper presented to the Society for Research in Child Development, Boston.

Haft, W.L., & Slade, A. (1989). Affect attunement and maternal attachment: A pilot study. *Infant Mental Health Journal,* 10, 157-172.

Hazen, N.G., & Durett, M.E. (1982). Relationship of security of attachment to exploration and cognitive mapping abilities in 2-year olds. *Developmental Psychology, 18,* 751-759.

Kobak, R. R. and Sceery, A. (1988). Attachment in late adolescence: Working models, affect regulation, and representations of self and others. *Child Development, 59,* 135-146.

Lamb, M.E., Thompson, R.A., Gardner, W., Charnov, E.L. (1985). *Infant-mother attachment: The origins and developmental significance of individual differences in Strange Situation behavior.* Hillsdale, NJ: Erlbaum.

Lewis, M., & Feiring, C. (1989). Infant, mother, and mother-infant interaction behavior and subsequent attachment. *Child Development, 60*, 831-837.

Lieblich, A. (1974). *WPPSI manual.* The Psychological Corporation, The Hebrew University of Jerusalem, Israel. (in Hebrew).

Londerville, S., & Main, M. (1981). Security of attachment, compliance, and maternal training methods in the second year of life. *Developmental Psychology, 17*, 289-299.

Lyons-Ruth, K., Connell, D.B., Grunebaum, H.U., & Botein, S. (1990). Infants at social risk: Maternal depression and family support services as mediators of infant development and security of attachment. *Child Development, 61*, 85-98.

Main, M. (1973). *Exploration, play and cognitive functioning as related to child-mother attachment.* Unpublished doctoral dissertation, Johns Hopkins University.

Main, M. (1990). Cross-cultural studies of attachment organization: Recent studies, changing methodologies, and the concept of conditional strategies. *Human Development, 33*, 48-61.

Main, M. (1991). Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs. multiple (incoherent) models of attachment: Findings and directions for future research. In: P. Marris, J. Stevenson-Hinde & C. Parkes (Eds.), *Attachment across the life cycle,* pp. 127-159. New York: Routledge.

Main, M., & Goldwyn, R. (in press). Adult attachment rating and classification systems. In: M. Main (Ed.), *A typology of human attachment organization assessed in discourse, drawings and interviews (working title).* New York: Cambridge University Press.

Main, M. & Hesse, E. (1990). Parents' unresolved traumatic experiences are related to infant disorganized attachment status: Is frightened and/or frightening parental behavior the linking mechanism? In: M.T. Greenberg, D. Cicchetti, & E.M. Cummings (Eds.), *Attachment in te preschool years: Theory, research and intervention*, pp.161-182. Chicago: University of Chicago Press.

Main, M. & Solomon, J. (1986). Discovery of an insecure-disorganized/disoriented attachment pattern. In: T.B. Brazelton & M.W. Yogman. *Affective development in infancy*, pp. 95-124. Norwood, NJ: Ablex Publishing Corporation.

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth Strange Situation. In: M.T. Greenberg, D. Cicchetti, & E.M. Cummings (Eds.). *Attachment in the preschool years: Theory, research, and intervention*, pp. 121-160. Chicago: University of Chicago Press.

Main, M., Kaplan, N. & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. In: I. Bretherton & E. Waters (Eds.). *Growing points of attachment theory and research.* Monographs of the Society for Research in Child Development. Serial no. 209, Vol. 50, Nos. 1-2, pp.66-104.

Matas, L., Arend, R.A., & Sroufe, L.A. (1978). Continuity of adaptation in the second year: The relationship between quality of attachment and later competence. *Child Development, 49*, 547-556.

Morisset, C.E., Barnard, K.E., Greenberg, M.T., Booth, C.L., & Spieker, S.J. (1990). Environmental influences on early language development: The context of social risk. *Development and Psychopathology, 2*, 127-149.

Moss, E. (1992). The socio-affective context of joint cognitive activity. In: L. Winegar & J. Valsiner (Eds.), *Children's development within social contexts: Metatheoretical, theoretical and methodological issues*, pp. 117-154. Hillsdale, NJ: Erlbaum.

Moss, E., & Strayer, F. (1990). Interactive problem-solving of mothers and gifted and nongifted preschoolers. *International Journal of Behavioral Development, 13*, 177-197.

Moss, E., Parent, S., Gosselin, C. & Dumont, M. (1993). Attachment and the development of metacognitive and collaborative strategies. *International Journal of Educational Research*, in press.

O'Connor, M.J., Sigman, M., & Brill, N. (1987). Disorganization of attachment in relation to maternal alcohol consumption. *Journal of Consulting and Clinical Psychology, 55*, 831836.

Oppenheim, D., Sagi, A. & Lamb, M. E. (1988). Infant-adult attachments on the kibbutz and their relation to socioemotional development 4 years later. *Developmental Psychology, 24*, 427-433.

Pastor, D.L. (1981). The quality of mother-infant attachment and its relationship to toddlers' initial sociability with peers. *Developmental Psychology, 17*, 326-335.

Piaget, J. (1932). *The moral judgment of the child.* Glencoe, IL: Free Press.

Piaget, J. (1968). *Six psychological studies.* New York: Vintage Press.

Rodning, C., Beckwith, L., & Howard, J. (1989). Characteristics of attachment organization and play organization in prenatally drug-exposed toddlers. *Development and Psychopathology, 1*, 277-289.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford University Press.

Smith, P.B. & Pederson, D.R. (1988). Maternal sensitivity and patterns of infant-mother attachment. *Child Development, 59*, 1097-1101.

Sroufe, L.A. (1983). Infant-caregiver attachment and patterns of adaptation in preschool: The roots of maladaptation and competence. In: M. Perlmutter (Ed.), *Development and policy concerning children woth special needs. Minnesota symposia on child psychology, Vol. 16*, pp. 41-79. Hillsdale, NJ: Erlbaum.

Sroufe, L.A., Egeland, B. & Kreutzer, T. (1990). The fate of early experience following developmental change: Longitudinal approaches to individual adaptation in childhood. *Child Development, 61*, 1363-1373.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Tracy, R.L., Farish, G.D., & Bretherton, I. (1980). Exploration as related to infant-mother attachment in one-year olds. Paper presented at the International Conference on Infant Studies. New Haven, CT.

Van der Meulen, B.F., & Smrkovsky, M. (1985).*MOS 2.5-8.5. McCarthy ontwikkelingsschalen*. Lisse, the Netherlands: Swets and Zeitlinger.

Van IJzendoorn, M.H. (1992). Intergenerational transmission of parenting: A review of studies in nonclinical populations. *Developmental Review, 12*, 76-99.

Van IJzendoorn, M.H., & de Ruiter, C. (1991). Opvoedingsstijlen: Overdracht van generatie op generatie. *Gezin, 3*, 89-104.

Van IJzendoorn, M.H., Goldberg, S., Kroonenberg, P.M., & Frenkel, O.J. (1992). The relative effects of maternal and child problems on the quality of attachment: A meta-analysis of attachment in clinical samples. *Child Development, 63*, 840-858.

Van IJzendoorn, M.H., Sagi, A., & Lambermon, M.W.E. (in press). The multiple caretaker paradox: Some data from Holland and Israel. In: R.C. Pianta (Ed.), *Relationships between children and non-parental adults*. San Francisco: Jossey-Bass.

Van IJzendoorn, M.H., & Van Vliet-Visser, S. (1988). The relationship between quality of attachment in infancy and IQ in kindergarten. *Journal of Genetic Psychology, 149*, 23-28.

Van IJzendoorn, M.H., Van der Veer, R., & Van Vliet-Visser, S. (1987). Attachment three years later: Relationships between quality of mother-infant attachment and emotional/cognitive development in kindergarten. In: L.W.C. Tavecchio and M.H. Van IJzendoorn (Eds.), *Attachment in social networks*, pp. 185-224. Amsterdam: Elsevier.

Van Lieshout, C.F.M., Riksen-Walraven, J.M.A., Ten Brink, P.W.M., Siebenheller, F.A., Mey, J.T.H et al. (1983). *Zelfstandigheidsontwikkeling in het basisonderwijs*. Nijmegen (Internal report 83 ON 04, SVO-project BS560).

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Waters, E. (1978). The reliability and stability of individual differences in infant-mother attachment. *Child Development, 49*, 483-493.

Waters, E., Wippman, J., & Sroufe, L.A. (1979). Attachment, positive affect, and competence in the peer group: Two studies in construct validation. *Child Development, 50*, 821-829.

Wood, D.J., Bruner, J.S., & Ross, G. (1978). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Yarrow, L.J. , McQuiston, S., MacTurk, R.H., McCarthy, M.E., Klein, R.P., & Vietze, P.M. (1983). Assessment of mastery motivation during the first year of life: Contemporaneous and cross-age relationships. *Developmental Psycholgy, 19*, 159-171.

Yarrow, L.J., MacTurk, R.H., Vietze, P.M., McCarthy, M.E., Klein, R.P., & McQuiston, S. (1984). Developmental course of parental stimulation and its relationship to mastery motivation during infancy. *Developmental Psychology, 20*, 492-503.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (1979). *Preschool language scale*. Columbus, OH: Merrill.

Manuscript ontvangen 5-4-1993
Definitieve versie ontvangen 3-9-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 4, pp. 242-244

Notities en commentaren

PERSPECTIEF OP EEN BEROEP? EEN COMMENTAAR

A.M.P. Knoers*

De opdracht van staatssecretaris Wallage, toen nog van O.en W., aan de Commissie Toekomst Leraarschap onder voorzitterschap van mevrouw A. van Es was "te komen tot een meer fundamentele beschouwing over de rol, de positie en de waardering van het leraarschap", om een nadere beschouwing te doen plaatsvinden "op tenminste de problematiek van het loopbaanperspectief, de mobiliteit, het huidige gelaagde bevoegdhedenstelsel en de opleiding en professionalisering van leraren", ook internationaal bezien, en om die visie "empirisch te onderbouwen" en "zonodig nadere analyses te laten verrichten".

Het moet gezegd dat de commissie-Van Es zich strikt aan deze opdracht heeft gehouden. De "Feiten en cijfers" van het eerste hoofdstuk geven een met veel onderzoek, en studies, o.a. van de OECD, onderbouwde inventarisatie van problemen en knelpunten. De laatste betreffen, behalve de door de staatssecretaris genoemde, ook taakbelasting, ziekteverzuim, arbeidsongeschiktheid en salariëring. Deze problemen zijn ook al eerder in ITS-, OTO- en andere rapporten geïnventariseerd, maar worden nu meer in samenhang beschouwd. De analyse van de problematiek in het tweede hoofdstuk herleidt de knelpunten tot twee ontwikkelingen: de centrale regelgeving en het isolement van de leraar in zijn klas.

Een school die door die problemen heen wil breken "zal moeten uitgroeien tot een moderne professionele arbeidsorganisatie", zo luidt dan de conclusie. Dat kan alleen, aldus hoofdstuk drie, als die nieuwe arbeidsorganisatie gezamenlijk wordt gedragen door leraren en schoolmanagement.

Binnen raam-overeenkomsten tussen overheid, werkgevers en werknemers moet elke school een eigen personeelsbeleid voeren. Door "human resource management", dat zich bedient van instrumenten als selectie en plaatsing, beloning afgestemd op de uitgeoefende functie, beoordeling van functioneren en training en ontwikkeling, kunnen grotere mobiliteit en loopbaanontwikkeling van de leraar op gang gebracht worden. Door samenwerking, ondersteuning en overleg tussen leraren kan de taakbelasting worden verminderd. In hoofdstuk vier wordt het brede perspectief geopend in de centrale aanbeveling: nodig is (schijn)zekerheden los te laten, ruimte te creëren voor verbeelding en beweging op gang te brengen.

Deze summiere samenvatting doet geen recht aan alle ideeën die in het rapport aan de orde komen. Zij geeft de grote lijn aan in het voetspoor van de samenvatting door de commissie zelf gegeven. Belangrijke punten komen in het nu volgend *commentaar* terug.

Om te beginnen kan waardering worden uitgesproken voor de frisse kijk waarvan de commissie getuigt in de zojuist vermelde centrale aanbeveling. Die heeft iets van het reinigen van een Augiasstal: zet de deuren open en richt de hogedrukspuit op de "mess", maak schoon schip.

Veel nieuwe gegevens zijn in het rapport niet aan te treffen. Het schoongemaakte schip wordt weer geladen met de bekende gegevens, maar in een nieuwe stuwage. Alles wordt bezien onder het oogpunt van de probleemanalyse: de oorzaak van het geringe perspectief van de huidige leraar schuilt in de kloof tussen centrale regelgeving en de isolatie van de leraar in de klas. Die kloof moet overbrugd worden door de school, die daartoe een autonome school moet worden van een bepaalde signatuur, de moderne, zelfstandige, open, professionele arbeidsorganisatie. Dat moet een arbeidsorganisatie zijn waarin leraren met de schoolleiding meedenken en mede richting geven aan de ontwikkeling van het schoolbeleid. In het daartoe door de schoolleiding te ontwikkelen "human resource management" zijn de bovenvermelde instrumenten aanwezig

*Adres: Kameelstraat 19, 6531 RN Nijmegen.

voor een goed personeelsbeleid. Kennelijk is die vorm van management weer "in", hoewel Etzioni reeds in 1964 een vernietigende kritiek heeft gegeven op deze benadering uit de "human relations"-hoek. Maar ik constateerde al dat er weinig echt nieuws is onder de in dit advies verzamelde gegevens. Het gevoerde personeelsbeleid vormt ook een bijdrage tot de ontwikkeling van de loopbaan van de leraar, die gezien wordt als de carrière van verticaal op elkaar volgende functies van leraar-in-opleiding, co-leraar, leraar, senior-leraar en "buitengewoon leraar", de expert, maar ook als een door functiedifferentiatie meer flexibele en afwisselende loopbaan. De verticale functieverdeling ontstaat niet meer automatisch, op basis van anciënniteit, maar gereguleerd door de instrumenten van beoordeling en beloning. Ook geen nieuwe gedachte, maar daarom niet minder belangrijk. En de tijd is er nu kennelijk rijp voor, al zal het nog geen eenvoudige zaak zijn om een nieuwe functiewaarderings-systematiek op te zetten in tripartiet overleg van overheid, werkgevers- en werknemersvertegenwoordigers, en de functiewaardering zo waterdicht en helder te maken dat directies er ook mee kunnen werken tot tevredenheid van de leraren, want om de verbetering van hun arbeidssatisfactie gaat het daarbij toch.

Het leraars*beroep* is hier dus vervangen door de leraars*functie*. Landelijke beroepsvereisten zijn niet meer nodig, de school bepaalt op grond van het landelijke raamwerk de functie-eisen waaraan de leraren van deze school moeten voldoen. Zelfs "is een diploma van een lerarenopleiding geen vereiste voor benoeming door de school" (p. 87). Niemand zal het met de commissie, die hierbij steunt op rapportages van inspectie en visitatiecommissies, oneens zijn dat aan lerarenopleidingen een en ander te verbeteren valt. Niemand zal ontkennen dat bij uitzondering, bijvoorbeeld om gastcolleges te geven aan de hogere klassen in het v.o. en m.b.o., de inzet van vakbekwame personen met wel een h.b.o. of w.o.-opleiding, maar zonder lerarenopleiding, mogelijk moet zijn. Maar voor de professionalisering van het leraarschap, zonder welk het geen perspectief heeft, blijft de opleiding de enige, reguliere toeleidingsweg. Terecht stelt de VSNU (1992) dat er "een blijvende behoefte (is) aan universitair geschoolde leraren en aan een universitaire lerarenopleiding". En dat geldt mutatis mutandis ook voor PABO en NLO.

De gezamenlijk gedragen verantwoordelijkheid voor de in de school te verrichten taken is zeker van belang om de leraar meer het gevoel te geven gewaardeerd te worden, althans intern. Ook Hoyle (1975) meent op basis van onderzoek dat de beperkte professionaliteit de invloed en daarmee de interne positie van de leraar in de weg staat en pleit daarom voor een vakoverstijgende, meer op de meso-situatie gerichte "extended professionality" (zie ook Giesbers & Bergen, 1992). Maar zelfs een bewust op een gezamenlijk schoolbeleid gericht management en lerarenteam zullen niet verhelpen dat de leraarstaken voor een groot deel binnen de klassemuren moeten worden uitgeoefend. Thans wordt volgens het OTO-onderzoek (1988), ook door de commissie-Van Es gebruikt, 60% van de tijdsbesteding van leraren aan onderwijsgevende taken in de klas gewijd, en dat zal voor de meeste leraren niet veranderen. De school zal hier dus niet de brug kunnen zijn die de kloof waarvan eerder sprake was kan overbruggen, zelfs niet de "ideale" school à la de commissie-Van Es, waarvoor ook de oude koeien van "mutual coaching", "team teaching", intervisie door collega's, nieuwe onderwijs- en groeperingsvormen weer uit de sloot worden gehaald. (Voor deze en talloze andere ideeën uit het advies zie reeds de COLOV, 1972).

Er is terecht veel aandacht bij de commissie voor de professionalisering van het schoolmanagement dat deze moderne arbeidsorganisatie moet leiden. Er is te weinig aandacht voor de professionalisering van de leraar wiens omvangrijkste taken klassetaken blijven. En daar zit blijkens onderzoek juist de problematiek van de leraar.

Problemen in de klas, het onderwijs aan en de interactie met de leerlingen vormen circa 50% van de problematische situaties van leraren (Peters, 1985). Ordevraagstukken spelen vooral bij beginnende leraren een belangrijke rol (Veenman, 1984; Créton & Wubbels, 1984; Vonk, 1982). Dat brengt demotivatie en crisisbeleving met zich mee die ook een samenhang vertonen met een negatieve arbeidssatisfactie (De Jong, 1990).

Hoewel op verschillende plaatsen in het rapport wordt betoogd: "De hoofdtaak van de leraar is de uitvoering van het primair proces" (p. 82), ligt de nadruk in het advies zozeer op de school

als moderne *arbeidsorganisatie,* waartoe dan ook schaalvergroting moet bijdragen, dat "de ideële taakstelling van de school als pedagogisch centrum en onderwijskundige en sociaal-culturele waarden" (Verantwoordelijkheid voor onderwijs, 1989, p. 6) weinig aandacht hebben gekregen. Men kan het ook zo formuleren dat in de school de loopbaan en het welbevinden van de leerlingen op zijn minst even belangrijk zijn als de loopbaan en de arbeidsomstandigheden van de leraar. Nu gaat het in het advies wel om het leraarsperspectief, maar dat moet steeds in relatie blijven tot dat van de leerlingen. En dat is nu juist wat de in het advies te weinig uitgewerkte professionalisering van de leraar beoogt. Dat maakt een lerarenopleiding, waarin onderzoek (Knoers, 1987) en theoretisch en onderzoeksmatig ondersteunde praktijkkennis (Verloop, 1991) belangrijke elementen zijn, zo cruciaal voor een verbetering van het perspectief van de leraar in zijn dominante taken, die in interactie met de leerlingen worden uitgeoefend. Dat helpt hem ook meer, als hij de aansporing tot mobiliteit serieus neemt en van zijn vertrouwde school vertrekt naar elders.

Een gedegen professionalisering door opleiding en nascholing en wellicht ook een echte professionele organisatie en een daardoor geformuleerde beroepscode (Hoeben, 1993) zullen hem daartoe een beter perspectief bieden dan een lokale arbeidsorganisatie.

LITERATUUR

COLOV (1972). *Eerste rapport van de commissie opleiding leraren onderwijskundige voorbereiding.* 's-Gravenhage: Staatsuitgeverij.

Créton, H. en Wubbels, Th. (1984). *Ordeproblemen bij beginnende leraren.* Utrecht, Uitgeverij W.C.C.

Etzioni, A. (1964). *Modern Organizations.* Englewood Cliffs, N.J.: Prentice Hall.

Giesbers, J. en Bergen, Th. (1992). *Professionaliteit en professionalisering van leraren.* Culemborg: Educaboek.

Hoeben, W. (1993). De kwaliteit van universitaire lerarenopleidingen en de professionalisering van het leraarsberoep. *VELON-Tijdschrift voor lerarenopleiders,* 14 (3), 10-17.

Hoyle, E. (1975). Leadership and decision-making in education. In M. Hughes (Ed.), *Administering Education: international challenge.* London: The Athlone Press.

Jong, M.J.de (1990). *Docenten onder druk.* Rotterdam: RISBO

Knoers, A.M.P. (1987). *Leraarschap: amb(ach)t of professie.* Assen/ Maastricht: Van Gorcum.

Peters, V.A.M. (1985). *Docenten en hun probleemsituaties.* Diss. Nijmegen.

Taak en organisatie OTO (1988). *Leraar tot (w)elke prijs: Eindverslag van het onderzoek (OTO-rapport 10)* Tilburg: IVA.

Veenman, S. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research,* 54, 143-178.

Verantwoordelijkheid voor onderwijs (1989). Advies van de commissie voor de toetsing van wetgevingsprojecten over de ontwikkeling van de onderwijswetgeving in de jaren negentig. 's-Gravenhage: Min. van Justitie.

Verloop, N. (1991). *Praktijkkennis van docenten als deel van de onderwijskundige kennisbasis.* Leiden: R.U.L.

Vonk, J.H.C. (1982). *Opleiding en Praktijk.* Amsterdam: V.U. - uitg.

Werkgroep Toekomst Universitaire Lerarenopleiding (1992). *Rapport van de VSNU-werkgroep Toekomst Universitaire Lerarenopleiding.* Utrecht: VSNU.

Manuscript ontvangen 13-7-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 4, pp. 245-249

Boekbesprekingen

J.H.C. Moust (1993). *De rol van tutoren in probleemgestuurd onderwijs. Contrasten tussen student- en docent-tutoren.* Rijksuniversiteit Limburg: Maastricht, Academisch proefschrift.

In Probleemgestuurd onderwijs (PGO) vormen taken en problemen het startpunt voor het leren van de studenten. Op basis van de voorgelegde taken bepalen de studenten welke leerdoelen zij zich denken te moeten stellen en hoe zij deze leerdoelen zullen proberen te realiseren. Essentieel in het Probleemgestuurd onderwijs (PGO) is dan ook de autonomie van de studenten bij het bepalen van de richting en vormgeving van hun leerproces. Dit impliceert dat de docent een andere, veel minder directieve rol vervult in het leerproces dan gangbaar is in het hoger en met name in het wetenschappelijk onderwijs. De rol van de docent is veeleer faciliterend, ondersteunend daar waar de studenten niet in staat blijken te zijn de hen toegedachte verantwoordelijkheid bij het bepalen van de richting en de vormgeving van hun leerproces waar te maken. Naast de opdrachten waarmee de studenten geconfronteerd worden, vervult de docent – in het PGO tutor genaamd – daarmee een essentiële functie in het optimaal laten verlopen van het leerproces van de studenten. Veel onderzoek en publikaties over PGO hebben betrekking gehad op de werking van de verschillende soorten taken die in het PGO worden gebruikt. Veel minder aandacht is tot dusver besteed aan het functioneren van de tutor. Moust geeft met zijn proefschrift een goede aanzet om deze lacune op te vullen.

De belangrijkste invalshoek van het proefschrift is het onderzoek naar mogelijke verschillen in functioneren van staf-tutoren en student-tutoren. Omdat Probleemgestuurd onderwijs met kleine groepen werkt, is het bij grote aantallen studenten vaak niet mogelijk voldoende stafleden tegelijk te mobiliseren om alle groepen van een tutor te voorzien. Daarom worden in dat geval ook student-assistenten als tutor ingeschakeld. In veel gevallen levert dat debatten op over de vraag of daarmee de kwaliteit van het onderwijs geen geweld wordt aangedaan. Vanuit deze vraag wordt in dit proefschrift het functioneren van de tutor nader onderzocht.

Aangeknoopt wordt bij het onderzoek naar de effectiviteit van onderwijs waarbij studenten de rol van docent vervullen. De resultaten van deze onderzoeken blijken eerder de conclusie te rechtvaardigen dat er geen sprake is van verschil in effectiviteit tussen student-docenten en staf-docenten dan dat één van beide groepen het systematisch beter lijkt te doen. Hiervoor worden verschillende verklaringen geopperd, zoals:

- studenten begeleid door student-docenten werken harder om zo het gebrek aan deskundige begeleiding te compenseren;
- student-docenten maken hun kennisachterstand goed door een beter inlevingsvermogen in de problemen van 'hun' studenten.

Deze verklaringen zijn vervolgens nader onderzocht.

In twee elkaar replicerende onderzoeken is bekeken of groepen die begeleid werden door student-tutoren verschilden van groepen begeleid door docent-tutoren met betrekking tot studieprestaties en bestede studietijd. Verschillen in studieprestaties konden in geen van beide onderzoeken worden geconstateerd, terwijl in één van de twee onderzoeken een verschil kon worden vastgesteld dat groepen begeleid door student-tutoren meer tijd aan zelfstudie besteedden dan door docent-tutoren begeleide groepen. Omdat deze resultaten nauwelijks als een duidelijk verschil tussen beide groepen aangemerkt kunnen worden, is ook nagegaan of er verschillen in (ervaren) expertise en optreden van de tutoren te vinden zouden zijn.

Ten slotte is gepoogd op basis van de empirische gegevens een causaal model te ontwikkelen voor de wijze waarop het functioneren van een tutor van invloed is op het leren van de studenten, los van de vraag of het hierbij gaat om een student- of een docent-tutor.

Het belang van dit proefschrift is vooral gelegen in de bijdrage die het kan leveren aan een

totaalmodel van het functioneren van PGO, naast een meer praktische waarde in de discussie over het al dan niet inzetten van studenten als tutor. In dat verband is het jammer dat in theoretisch opzicht vrijwel uitsluitend aansluiting wordt gezocht bij de literatuur over studentgeleid onderwijs. De daar gebruikte theoretische noties zijn zeer beperkt en wortelen vaak nauwelijks in (onderwijs)leertheorieën. Cognitieve theorieën, zoals o.a. Vermunt en Simons die in de afgelopen jaren ontwikkeld hebben met betrekking tot de afstemming tussen leeractiviteiten en onderwijsmaatregelen, zouden zowel de hypothese-ontwikkeling als het onderzoek van een steviger theoretisch fundament hebben kunnen voorzien.

Desondanks is het totaalbeeld van het proefschrift positief. Het is helder geschreven en daardoor makkelijk leesbaar. Er is een duidelijke lijn in de gevolgde redenering:
- Is er verschil te vinden in de studieprestaties?
- Zo nee, is er dan verschil in de mate van studie-activiteiten?
- Zo nee, is er dan verschil te vinden in de wijze waarop de student- en docent-tutoren zich gedragen?
- In hoeverre is het gedrag van de tutor van invloed op het leren van de studenten?

Een groot voordeel daarbij is dat de analyses van de resultaten vrijwel steeds strikt beperkt worden tot hetgeen noodzakelijk is om de onderzoeksvraag te beantwoorden. Een inperking die des te meer te waarderen is, omdat nogal wat analyses niet tot significante resultaten leiden. De verleiding is dan extra groot middels nadere analyses op zoek te gaan naar significantie.

Onlogisch vond ik de redenering waarbij in het proefschrift geen verschil in kennis over de onderwerpen in het desbetreffende blok tussen student-tutoren en docent-tutoren werd gevonden op een afgenomen toets. Vervolgens werd er dan gemakshalve toch maar vanuit gegaan dat er wel degelijk verschillen in kennis bestonden om verschillen in tutoroptreden te verklaren. Deze veronachtzaming van de resultaten was m.i. ook niet nodig. Studenten blijken de kennis van student-tutoren lager in te schatten dan de kennis van docent-tutoren. Aangezien verwacht mag worden dat studenten hun eigen oordelen als correct beschouwen en daarop hun activiteiten baseren, is deze inschatting voldoende motief om eventuele verschillen in waarneming en gedrag van studenten te verklaren.

Het laatste hoofdstuk van het proefschrift betreft de toetsing van een model voor de invloed van tutorgedrag op de leeractiviteiten van studenten binnen en buiten de onderwijsgroep. De belangrijkste factoren in het tutorgedrag betreffen:
- de mate waarin de tutor zich kan inleven in de problemen die studenten met de leerstof hebben: cognitieve congruentie;
- de mate waarin de tutor gebruik maakt van vakinhoudelijke deskundigheid: expertisegebruik;
- de mate waarin de tutor aandacht besteed aan samenwerking tussen de studenten in de onderwijsgroep: sociale congruentie;
- de mate waarin de tutor een persoonlijke relatie met de leden van de onderwijsgroep tracht aan te gaan: rolcongruentie;
- de mate waarin de tutor gericht is op de toets na afloop van het blok: toetsgerichtheid;
- de mate waarin de tutor afstand schept t.o.v. de studenten: autoriteit.

Van deze factoren is nagegaan op welke wijze ze van invloed zijn op het functioneren van de onderwijsgroep, de mate van zelfstudie door de studenten, de studieprestaties en de mate waarin studenten in het onderwerp geïnteresseerd raken.

Dit model is getoetst met behulp van een padanalyse. Uit deze analyse blijkt dat de meest directe invloed op het functioneren van de onderwijsgroep uitgaat van de factor cognitieve congruentie. Het gebruik van expertise heeft hooguit een indirect effect via de cognitieve congruentie. Desondanks wordt er in de discussie van uit gegaan dat voor het adequaat functioneren van de tutor vakinhoudelijke deskundigheid primair van belang is. Op basis van overwegingen van gezond verstand lijkt mij dit een plausibele veronderstelling. In het proefschrift is daarvoor echter niet veel evidentie te vinden. Wellicht zou het gebruik van expertise differentieel bena-

derd moeten worden om wezenlijke verschillen te vinden. Als de expertise gebruikt wordt om door te vragen, zodat de studenten niet te snel tevreden zijn met hun analyse van de problematiek of met hun antwoorden daarbij, zijn wellicht andere leerresultaten te vinden dan als de expertise gebruikt wordt om de studenten van antwoorden te voorzien. Expertisegebruik om leeractiviteiten te prikkelen in plaats van om leeractiviteiten af te sluiten.

Omdat het ontwikkelde model aanleiding geeft tot het genereren van dit soort veronderstellingen is het een waardevol element in het onderzoek naar het functioneren van PGO. Belangrijk is ook dat dit model duidelijk maakt dat het functioneren van PGO geen eenvoudige zaak is. PGO biedt de mogelijkheid de studenten aan te zetten tot zeer actief leergedrag. Of en in hoeverre deze mogelijkheid gerealiseerd wordt, hangt zeker ook in sterke mate af van de wijze waarop de tutor functioneert.

J. Willems
IOWO, KU Nijmegen
Montessorilaan 3
Postbus 9104
6500 HE Nijmegen

Veugelers, W. (1993). *Pedagogische Opdracht en Arbeid. Doelstellingen van docenten in het Algemeen Voortgezet Onderwijs en het Beroepsonderwijs.* Proefschrift Universiteit van Amsterdam. Amsterdam: Thesis Publishers. ISBN 90 5170 155 1 (*f* 35,–/Bfr 700)

Een van de weinige dingen die in het slecht gevoerde, abstracte debat over de pedagogische opdracht van het onderwijs duidelijk is geworden, is dat er nog veel onduidelijkheid bestaat over de waarden-doelstellingen die docenten (zeggen te) hanteren. Het proefschrift van Veugelers zou hierin voor een deel kunnen voorzien omdat het zich richt op de waarden ten aanzien van arbeid die docenten hun leerlingen willen meegeven. Verschillende categorieën docenten zijn door Veugelers onderzocht, namelijk docenten uit vwo, havo, kmbo en het leerlingwezen. Binnen deze schooltypen heeft hij docenten economie en maatschappijleer ondervraagd, evenals decanen en docenten voor de praktijkvakken in het beroepsonderwijs. Het aardige van dit proefschrift is dat het handelt over waarden betreffende arbeid en nu eens weer niet over de waarden waarmee politici denken de talrijke ziekten van de nederlandse samenleving (criminaliteit, fraude, vandalisme, etc.) via het onderwijs te cureren. Veugelers geeft te kennen duidelijk het belang in te zien van attitudinale vorming in het onderwijs als voorbereiding op het functioneren in de arbeidssituatie.

Anders dan de titel van zijn proefschrift op het eerste gezicht doet vermoeden heeft Veugelers al ruim vóór het huidige debat oog gehad voor het ook door de onderwijskunde ruim verwaarloosde terrein van de waardenontwikkeling in het onderwijs. Veugelers startte zijn onderzoek in een tijd waarin het wisselvallige en kortzichtige onderwijsbeleid-'discours' gekenmerkt werd door het 'back-to-basics' denken van een 'no nonsense' politiek die zaken als 'waarden en normen' afdeed met de aanduiding 'franje'.

In het proefschrift maakt Veugelers duidelijk dat tot de maatschappelijke doelstellingen van het onderwijs naast kennis en vaardigheden ook het tot ontwikkeling brengen van waarden behoort.

Bij zijn proefschrift kunnen echter een tweetal belangrijke kanttekeningen geplaatst worden die betrekking hebben op enerzijds zijn theoretisch 'framework' en anderzijds de door hem (niet) uitgevoerde analyses.

Zoals blijkt uit zijn literatuurlijst heeft Veugelers zich uitvoerig gedocumenteerd op het betreffende terrein van onderzoek naar waarden, maar gaat hij enigszins opmerkelijk en weinig kritisch om met hetgeen uit empirisch voorinzicht en theoretische overwegingen bekend is over deze materie. Om een voorbeeld te geven: Veugelers bespreekt de verschillende methoden van waardenonderwijs, zoals de methode van waardenoverdracht, de waardenverheldering, de waardencommunicatie en de waardenconditionering. Opvallend daarbij is dat hij de vooruitgang die in de loop der tijd geboekt is in de ontwikkeling van deze methoden niet in zijn proefschrift vermeldt, maar daarentegen wel aan komt zetten met de methode van de 'waardenconditionering', die overigens geen methode is maar een aspect van het verborgen curriculum en niet in het bovenstaande rijtje thuis hoort. De genoemde vooruitgang bestaat erin dat nieuwe methodes proberen de tekortkomingen van vroegere methodes te voorkomen. Zo probeerde de methode van de waardenverheldering de kritiek te vermijden die op de waardenoverdracht geleverd is, namelijk dat deze de vrijheid van de leerlingen beperkt om zich een eigen kritische mening te vormen. Tussen haakjes kan worden opgemerkt dat in het huidige debat over de pedagogische opdracht menigeen nog spreekt in termen van 'overdracht'. De methode van waardencommunicatie poogt de vrijblijvendheid en het ethisch relativisme van de waardenverheldering weer te ontwijken, maar bouwt tevens de stadia- of niveautheorie van Kohlberg in de methode in. Op dit laatste aspect wordt in het proefschrift merkwaardigerwijs niet ingegaan. Wel wordt onkritisch de waardenconditionering van Romizowski in de schijnwerper gezet zonder voldoende te argumenteren waarom dat meer zoden aan de dijk zou zetten dan andere (niet) besproken benaderingen van de waardenproblematiek.

Door een zorgvuldiger benadering hadden bepaalde foutieve opmerkingen voorkomen kunnen worden. Zo staat op pagina 11 van het boek bijvoorbeeld te lezen dat het concept 'waardencommunicatie' onvoldoende tot uitdrukking brengt dat de docent de ontwikkeling van specifieke waarden stimuleert. In deze methode wordt echter bij de bespreking van morele kwesties de dialoog op basis van argumenten zodanig ingericht dat een concrete situatie wordt getoetst aan een principe,

bijvoorbeeld ontleend aan de 'rechten van de mens'. De methode van de waardencommunicatie onderscheidt zich juist van de waardenverheldering door de inbreng van principes en specifieke waarden.

Romizowski zou overigens een goed aanknopingspunt kunnen zijn voor meer aandacht voor het verborgen curriculum betreffende de waarden op het terrein van arbeid. In veel studies naar het verborgen curriculum van de schoolse socialisatie komt de werking naar voren van het onderwijs die gericht is op het stimuleren van arbeidsbereidheid, aanpassingsbereidheid, competitiebereidheid, etcetera. Veel studies illustreren hoe in de alledaagse rituelen, structuren en omgangsvormen deze impliciete schoolculturele eisen als het ware 'ingebakken' zitten. In het proefschrift wordt hieraan nauwelijks aandacht besteed. Wellicht dat de manifest uitgesproken doelstellingen van docenten op dit terrein belangrijker geacht worden.

In het onderzoek wordt aan docenten gevraagd welke doelstellingen zij bij hun leerlingen willen realiseren ten aanzien van waarden bij arbeid. Naast het vak dat de docent geeft worden in het onderzoek een aantal docentkenmerken meegenomen zoals leeftijd, werkervaring, sekse, stellingname in de relatie tussen onderwijs en arbeid, positie van de docent in de school en denominatie.

Veugelers komt in zijn dissertatie tot de conclusie dat de docenten van diverse schooltypen en schoolvakken verschillen in de waarden die zij hun leerlingen willen meegeven. Ook de leeftijd en de sekse van de docent(e) en de cultuur van de school geven verschillen in doelstellingen van de docenten te zien, zo wordt geconcludeerd. In Hoofdstuk 6 worden de verschillen tussen docenten van verschillende schooltypen beschreven en in het volgende hoofdstuk de verschillen tussen docenten van verschillende vakken binnen de diverse schooltypen. Hoofdstuk acht beschrijft de relatie tussen individuele docentkenmerken en hun doelen. In tegenstelling tot de voorgaande hoofdstukken wordt hier ineens een vorm van multi-variate analyse toegepast met als argument dat de verschillende docentkenmerken wel eens zouden kunnen samenhangen wat tot 'schijnverbanden' zou kunnen leiden. Omdat de rapportage van de multi-variate analyse uitermate summier is wordt niet duidelijk ofdat in de ANOVA analyse nu wel of niet gecontroleerd is voor 'schooltype' en 'schoolvak' naast de docentenkenmerken. Het zal duidelijk zijn dat een analyse met al deze variabelen erbij voor de hand ligt om inzicht te krijgen in mogelijke schijnverbanden. De resultaten van deze gebruikelijke 'totaalanalyse' worden in het proefschrift echter niet vermeld. Dit laatste is in de onderwijsresearch echter weer ongebruikelijk en zou door de begeleiders van deze dissertatie voorkomen hebben kunnen worden. Nu wordt in het voorliggende verslag onvoldoende duidelijk of de verschillen in doelstellingen vooral te maken hebben met het schooltype, het schoolvak of toch weer met docentkenmerken. Vanwege de mogelijkheid van 'schijnverbanden' is het jammer dat in het onderzoek te weinig op deze kwestie is ingegaan. Dit temeer daar voor het beleid van belang is te weten welke docenten nu welke opvattingen hebben over het stimuleren van welke waarden op het zo belangrijke terrein van de arbeid. Veugelers verwacht dat in de toekomst docenten minder zelf kunnen aangeven welke waarden zij willen ontwikkelen. Hun keuze van waarden zal onderdeel uitmaken van een invulling van de eigen identiteit, de eigen profilering van de school. Hij merkt echter ook op dat docenten in de interactie tussen docenten en leerlingen waarden uitdragen en reageren op de waarden van leerlingen. Daarom zal het ook in de toekomst nodig en wenselijk zijn dat onderzoeken met een vraagstelling als de onderhavige op de onderzoeksagenda blijven staan, ook al zou dan onverhoopt de pedagogische opdracht weer tot franje gereduceerd zijn door het van weer en wind bewogen onderwijsbeleid.

C. Klaassen
Vakgroep Onderwijskunde
Katholieke Universiteit Nijmegen



Een kwestie van vraag en aanbod

emancipatie en roldoorbreking in het onderwijs

A.C.A.M. Vermeulen en M. Derriks

Vernieuwing in het onderwijs op het gebied van emancipatie en het verminderen van sekse-ongelijkheid, komt niet goed op gang. Overheid, onderwijsbegeleidingsdiensten, scholen en leerkrachten staan positief ten opzichte van roldoorbrekend onderwijs, maar dit is niet voldoende om ook daadwerkelijk het onderwijs te veranderen. Dat vereist een gecoördineerde, strategische aanpak, en specifieke deskundigheid.
De begeleiding en stimulering van veranderingsprocessen in het onderwijs is een taak van de onderwijsbegeleidingsdiensten. Het in dit boek beschreven onderzoek werd gedaan om na te gaan hoe zij op het gebied van emancipatie en roldoorbreking deze taak kunnen vervullen. De auteurs onderzochten de stand van zaken op scholen en begeleidingsdiensten, de fasen van het vernieuwingsproces, welke factoren dit proces belemmeren, en wat bevorderend werkt. Op grond hiervan komt een panel van deskundigen tot aanbevelingen voor de toekomst.
Serie: Forum nr. 18

ISBN 90 265 1331 3
136 pagina's, f 42,00

Verkrijgbaar in de boekhandel

SWETS & ZEITLINGER B.V.

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

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres
Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen,
tel. 045-762293

Jrg. 17, nr. 1 1993

Inhoudsopgave

Verslag VOR-ledenvergadering 27 mei 1993 252

Lezing ORD'93 255

Verslag VOR-ledenvergadering 27 mei 1993: 12.15-12.40 uur (tijdens ORD'93 in Maastricht)

Aanwezig: leden: M. Berger, J.L. van der Linden, M. Mulder, F. Riemerstra, J. Thijssen, M. Voeten, J. Voogt.
bestuursleden: N. Lagerweij (voorzitter), F.J.J. van Bussel (penningmeester), J. Ax, W.Th.J.G. Hoeben, H. van Hout (vice-voorzitter), J.M. Pieters, R-J. Simons, P.J.J. Stijnen (secretaris).

1. Opening
De voorzitter opent de vergadering om 12.15 uur.
Afgemeld hebben zich: J. Claessen, L.F.W. de Klerk en J.H. Slavenburg.

2. Mededelingen
- In de divisie BOBO zal een nieuw bestuur aantreden. Dit betekent dat R-J. Simons zich terugtrekt uit het VOR-bestuur. Hij zal worden opgevolgd door de heer M. Mulder onder voorwaarde dat de divisievergadering (later tijdens deze ORD) het met deze voordracht eens is. De vergadering gaat hiermede accoord.
- De heer Van Wieringen gaat over van de divisie 'Beleid en Organisatie in het Onderwijs' naar 'Hoger Onderwijs en Volwasseneneducatie'.

3. Verslag van de VOR-ledenvergadering van 17 december 1992 (TOR/VOR-bulletin nr. 3 (1993)
Het verslag wordt onder dankzegging aan de secretaris ongewijzigd vastgesteld.

4. Vaststelling jaarrekening 1992
De heer Voeten doet namens de kascommissie verslag en complimenteert de penningmeester voor de wijze waarop hij de financiële situatie van de vereniging heeft verbeterd. De kascommissie stelt voor de jaarrekening goed te keuren. De vergadering neemt dit voorstel over.

Van de zijde van de kascommissie wordt nog opgemerkt dat het voor volgende jaren goed is om in de procedure van controle van de kas op te nemen dat de bescheiden vooraf kunnen worden ingezien. De vergadering neemt dit voorstel over.

5. Decharge kascommissie
Onder dankzegging door de voorzitter wordt de kascommissie gedechargeerd.

6. Instelling nieuwe kascommissie
De huidige leden van de kascommissie (M. Berger, J. Claessen en M. Voeten) worden voorgedragen als leden van de nieuwe kascommissie. De vergadering neemt dit voorstel over.

7. Verkiezing aanvulling VOR-bestuur
Door het bestuur wordt voorgesteld de heer J. Ax als lid van het VOR-bestuur te benoemen als vertegenwoordiger van de divisie 'Beleid en Organisatie in het Onderwijs'. De vergadering gaat accoord met dit voorstel.

Onder de voorwaarde dat de divisie BOBO accoord gaat met de voordracht van de heer Mulder is de bestuurssamenstelling van de VOR met ingang van 27 mei 1993 aldus:

Prof.dr. N.A.J. Lagerweij (voorzitter; UU)
Dr. J.F.M.J. van Hout (vice-voorzitter; KUN) (divisie HO+VE)
Dr. P.J.J. Stijnen (secretaris; OU)
Dr. F.J.J. van Bussel (penningmeester; KUB)
Prof.dr. J. Dronkers (lid; UvA) (divisie O&S/OOMO)
Prof.dr. W.Th.J.G. Hoeben (lid; RUG) (divisie CUR)
Prof.dr. J.M. Pieters (lid; UT) (divisie L&I)
Prof.dr. J. Scheerens (lid; UT) (divisie M&E)
Dr. M. Mulder (lid; UT) (divisie BOBO)
Prof.dr. N. Verloop (lid; RUL) (divisie LLG)
Dr. J. Ax (lid; UvA) (divisie BOO)

8. Regeling ere-leden

Het door het bestuur ingediende stuk 'Regeling ereleden en emeriti van de VOR' wordt ongewijzigd door de vergadering vastgesteld. De vaststelling van dit stuk impliceert de onderstaande wijzigingen van en aanvullingen op de teksten voor de statuten en het huishoudelijk reglement (voor emiriti met terugwerkende kracht vanaf 1-1-1993; het instituut erelid bestond al sinds 2-12-1991).

– *Teksten voor statuten:*

artikel 6:
lid a: (wijziging). De vereniging bestaat uit gewone leden, emeritileden en ereleden.

lid b: ongewijzigd (m.b.t. gewone leden).

lid c: (nieuw). Als emeriti worden toegelaten gewone leden die de leeftijd van 65 jaar hebben bereikt en onafgebroken gedurende 10 jaar gewoon lid van de VOR zijn geweest.

lid d: (wijziging oude lid c). Tot ereleden kunnen benoemde worden zij die zich op een bijzondere manier verdienstelijk hebben gemaakt voor de vereniging, dan wel zich door publikaties en/of praktische arbeid bijzonder verdienstelijk gemaakt hebben voor het onderzoek op het gebied van onderwijs, opleiding en training of de kwaliteit van de beroepsuitoefening.

lid e: (wijziging oude lid d). De benoeming van ereleden, emeriti en de wijze van toelating van gewone leden en de mogelijkheden van beroep worden geregeld in het huishoudelijk reglement.

lid d: (ongewijzigd oude lid e). Ieder verenigingslid is tevens lid van tenminste een divisie.

– *Teksten voor huishoudelijk reglement:*

art. 2: (wijziging oude lid 2). De vereniging kent gewone leden, emeriti-leden en ereleden.

art. 10: (ongewijzigd). Ereleden worden op voordracht van het VOR-bestuur door de Algemene Ledenvergadering gekozen.

art. 10a: (nieuw). Een emeritus lidmaatschap wordt aangeboden aan gewone leden die de leeftijd van 65 jaar bereikt hebben en direct daaraan voorafgaand gedurende 10 jaar onafgebroken gewoon lid van de vereniging zijn geweest. De penningmeester zal namens het Bestuur op grond van een onderzoek van de leden-administratie het betreffende lid hiervan in kennis stellen. Hiertegen is geen beroep mogelijk.

art. 10b: (nieuw). Zowel ereleden als emeriti-leden hebben alle rechten die aan gewone leden toekomen. Aan deze leden zal geen contributie in rekening gebracht worden, tenzij voor extra tijdschriften en andere extra activiteiten die niet in de normale contributie zijn verdisconteerd.

9. Wat verder ter tafel komt

De voorzitter meldt dat er in het VOR-bestuur momenteel een discussie plaatsvindt over het verenigingsorgaan. In een later stadium zal hierop tijdens een Algemene Ledenvergadering worden teruggekomen.

10. Rondvraag

Van de rondvraag wordt geen gebruik gemaakt.

11. Sluiting

Niets meer aan de orde zijnde sluit de voorzitter onder dankzegging om 12.40 uur de vergadering.

N.B.: De volgende Algemene Ledenvergadering vindt plaats op 16 december om 14.00 uur in het Stiltecentrum te Utrecht.

Lezing ORD'93

Prof.dr. N.A.J. Lagerweij
Vakgroep onderwijskunde
Universiteit Utrecht

Dames en heren,

Met enige spanning ben ik vanuit het gebied boven de rivieren naar het mooie zuidelijke Maastricht gekomen. Maastricht heeft zeker een goede naam onder de noordelingen: een gezellige bourgondische stad, met vriendelijke en gastvrije mensen. Die reputatie geldt echter niet meer onverkort en zeker niet voor mensen in het buitenland.

Het afgelopen jaar bezocht ik in mijn vakantie Frankrijk. Ik bezoek dan ook vanwege een innerlijke behoefte enkele wijnboeren. In Santenay kreeg ik behalve goede wijn ook een exposé over Europa 92. De wijnboer maakte mij zonneklaar duidelijk hoe waardeloos "Maastriek" toch in zijn ogen wel was. De hele internationalisering kon hem gestolen worden. In plaats van de gebruikelijke drie moest hij nu vijf formulieren voor dezelfde zending invullen. Dan nog afgezien van de gevreesde intocht van allerlei ongewenste buitenlanders, waartoe ook de angst voor ongebreidelde invoer van drugs behoorde. Dat alles als gevolg van "Maastriek".

Nu hoeft u niet te denken dat de Franse wijnboer mijn hooggestelde verwachtingen heeft beïnvloed. Al eerder was de ORD te gast in deze stad en daarover bestaan zeer positieve ervaringen. In 1975 startte de VOR. In 1974 de eerste ORD. In 1981 voor het eerst in Limburg en nu twaalf jaar later opnieuw in Maastricht.

De organisatiecommissie, in het bijzonder Henk van Berkel als voorzitter, verdient nu al reeds dank voor de uitstekend verlopen voorbereidingen. Het geheel zoals dat nu geprogrammeerd is, ziet er veelbelovend uit. Er is sprake van aantrekkelijke en afwisselende presentatievormen. Het boek is mooi en op tijd. De rijkgeschakeerde inhoud roept wel weer een al vaker gestelde vraag op: Is onderwijsonderzoek nog steeds een lappendeken? of nog steeds een "overmaat aan kleingoed", zoals De Groot al in 1975 constateerde.

Als voorzitter van de VOR wil ik hier vanavond graag een drietal onderwerpen kort aan de orde stellen:

onderwijsonderzoek anno 1993
de visitatie onderwijskunde
de verbreding van de VOR.

Onderwijsonderzoek anno 1993

In 1975 vroeg De Groot zich af of de uitkomsten van onderwijsonderzoek iets uithalen. Ook nu moeten wij ons die vraag stellen. Tenslotte steekt de samenleving niet voor niets miljoenen in onderwijsresearch. In 1973 voor beleidsgericht onderzoek bij SVO 4 ton gesubsidieerd. In 1978 was dat 10 miljoen. Hoe zal dat na de evaluatie van de WOV, die in 1995 expireert, eruit zien?

Nu zijn er nogal wat soorten gebruikers en soorten gebruik van de resultaten van onderwijsonderzoek. Dat varieert van beleidmakers tot schrijvers van leerboeken, van opleiders van onderwijsgevenden tot presentatoren in de media. Dat brengt mij ook op een vraag die ik tijdens mijn tocht hierheen voelde opkomen. Zouden pers en media acte de présence geven. Het zou toch voor het image van onderwijsonderzoek een goede zaak zijn als er echt wat interessant nieuws voor een groot publiek te melden zou zijn. Zo in de trant van: nieuw middel tegen zittenblijven levert aanzienlijke geldbesparingen op. Of: nieuw onderwijsmodel leidt tot een soepele overgang van basisschool naar voortgezet onderwijs. Of: opleidingspeil van Nederlanders in 30 jaar met grote sprongen vooruit (dat klopt overigens zie Rijkdom van het onvoltooide, pag. 71).

Verspreiding van de resultaten van onderwijsonderzoek verdienen in elk geval aandacht binnen de samenleving. Helaas is het karakter van het object van onderzoek niet bepaald geschikt voor het doen van spectaculaire ontdekkingen of voor het ontwikkelen van succes garanderende produkten of procedures. Er zijn steeds erg veel variabelen met elkaar in interactie en het verloop van dergelijke processen is moeilijk te voorspellen en lastig te sturen. Voor de maatschappelijke erkenning van de betekenis van onderwijsonderzoek is dat een complicerende factor. Dat verplicht ons overigens steeds goed in de gaten te houden, dat de buitenwacht wel oordeelt over ons onderzoekswerk. In dat verband is de positie van SVO voor onderwijsresearch van grote waarde. Het kabinetsstandpunt betreffende de WOV en over de toekomstige positie van SVO geeft weinig aanleiding tot tevredenheid. Het is terecht dat geld voor fundamenteel onderzoek doorgesluisd wordt naar NWO. Als dat maar gepaard gaat met de garantie van gelijkblijvende omvang en de eis voor besteding aan onderwijsvraagstukken. De procedures rond besluitvorming binnen de NWO zijn echter van dien aard dat we hier goed op onze hoede moeten zijn er niet op achter uit te gaan.

Bij de evaluatie van de WOV blijkt dat het veld- en beleidsgericht onderzoek een onzekere toekomst tegemoet gaat. Wie krijgt het geld in handen om onderzoek te gaan aanbesteden? Welke procedures zullen er ontstaan rondom de aanbestedingen? Hoe lang en in welke omvang en met welke taken blijft de SVO nog bestaan? Hoe is het gevaar van een verdergaande versnippering van het geld voor onderwijsonderzoek te voorkomen? Hoe is enige programmering te bewerkstelligen? De maatschappelijke functie van onderwijsonderzoek is de afgelopen jaren aangetoond. In noem als voorbeeld het onderzoek naar de effecten van het onderwijsvoorrangsbeleid. Het in 1988 gestarte cohorten onderzoek krijgt ook internationaal veel waardering. Voortzetting van dergelijk onderwijsonderzoek is dan ook alleen mogelijk als de subsidie daarvoor gegarandeerd wordt over meerdere jaren. De beleidsvoornemens rondom de WOV laten zien dat er voor de derde geldstromers onder ons een ongewisse toekomst te wachten staat.

Voor de eerste geldstromers staat er eveneens een belangwekkende verandering voor de deur: het fenomeen onderzoekschool. Het nationale beleid is erop gericht toponderzoekers te verenigen en daaraan een opleiding voor AIO's te koppelen. Het is een volgende fase in het streven naar programmering, planning en clustering van onderzoek. Of deze operatie wel enig nut zal opleveren is vooralsnog onzeker. De centrale opgave is de wetenschappelijke kwaliteit te optimaliseren. Inmiddels zijn er voor onderwijsonderzoek op landelijk niveau overleggremia bezig om ook te komen tot een interuniversitaire onderzoekschool voor de onderwijskunde. Dat is een verheugende en positieve ontwikkeling.

Het initiatief van ICO, oorspronkelijke een samenwerking tussen Twente en Groningen, is daarbij de basis. De toekenning van een Stimulans subsidie door NWO is in dat opzicht van grote betekenis, niet alleen voor de betreffende contractpartners maar ook voor het onderwijsonderzoek in het algemeen. Naar het zich laat aanzien zal men er in slagen voor januari 1994 een officiële aanvrage bij de KNAW in te dienen. Verschillende onderzoekers uit de vakgroepen onderwijskunde uit verschillende universiteiten willen zich verenigen op landelijk niveau tot een onderzoekschool. Daarmede is een verdergaande versnippering van het onderwijsonderzoek vooralsnog voorkomen.

Voor het erkennen van dergelijke scholen hanteert de KNAW criteria. Daar is op zich niets tegen, maar er zit naar mijn mening ook een gevaarlijk element in. Men kent namelijk uitzonderlijk veel waarde toe aan het jaarlijks internationaal publiceren. Om te mogen toetreden tot de onderzoekschool is dat criterium bepalend. Zelfs het publiceren in de eigen taal, zo bleek mij onlangs, blijkt beschouwd te worden als "een vergissing zeker". Internationaal publiceren is vooral in de bèta wetenschappen al een zeer lang bestaande gewoonte. In hoeverre is dat een maat voor toponderzoek in de sociale wetenschappen in het algemeen en voor onderwijsonderzoek in het bijzonder? Juist voor een op toepassing gerichte wetenschap is een dergelijke eis, vooral als die eenzijdig gehanteerd wordt, problematisch.

De koppeling van maatschappelijke relevantie, die voor onderwijsonderzoek geldt, aan internationaal publiceren ligt toch niet zonder meer voor de hand. Er ontstaat zo een gevaar van een van het onderwijsveld vervreemdende onderwijsonderzoeker, die uiteindelijk zelfs steeds meer

publiceert voor steeds minder mensen. Zal de belastingbetaler staan juichen bij een dergelijke ontwikkeling in het onderwijsonderzoek?

Voor de toekomst van onderwijsonderzoek is de ontwikkeling van een onderzoekschool zeker van belang. We mogen echter onze ogen niet sluiten voor de noodzaak om door onderwijsonderzoek bij te dragen aan het oplossen van nationale onderwijsproblemen en zodoende mee te werken aan onderwijsverbetering. Uiteindelijk moet de ontwikkeling van onze kennis op het gebied van onderwijs plaats vinden binnen de nationale context. Daarover kan en moet zeker vervolgens gepubliceerd worden voor een internationaal forum. Het is echter volstrekt idioot om de volgorde om te keren: eerst en vooral in een vreemde taal publiceren in internationaal erkende tijdschriften maakt de onderzoeker tot een toponderzoeker.

De positie van onderwijskunde als studierichting

Het gaat goed met de opleidingen onderwijskunde in ons land. In 1992 is door een visitatiecommissie onder voorzitterschap van collega Van den Broek een visitatie uitgevoerd voor pedagogiek en onderwijskunde tegelijkertijd. De indrukken van de commissie zijn inmiddels verschenen in de bekende VSNU rapporten. Vertegenwoordigers van de VSNU-kamer onderwijskunde hebben uitvoerig over de inhoud van de visitatie van gedachten gewisseld. Er zijn in de loop van dit jaar voortgezette bijeenkomsten gepland om in landelijk overleg verdere uitwerking te geven aan de opmerkingen van de visitatiecommissie. Twee leden van de visitatiecommissie hebben ook gesproken met het bestuur van de VOR over het rapport.

Samenvattend blijkt er een vrij goede waardering te bestaan bij de bezochte vakgroepen over de gehele visitatieprocedure. Vooral het opstellen van een zelfstudie blijkt een aanzienlijke bewustwording te bewerkstelligen over kwaliteitsaspecten van de opleiding. Daardoor kwamen de ervaren problemen en de voornemens ter verbetering al aardig in kaart. De daadwerkelijke visitatie leverde dan in de meeste gevallen geen verrassende gezichtspunten meer op. De commissie meent dat over het geheel genomen de opleiding onderwijskunde van redelijke tot goed niveau zijn (een 7 à 8 aldus de voorzitter). Zij heeft verschillende suggesties gegeven die nog tot verbetering kunnen leiden. Die zijn niet steeds in alle gevallen ook aanvaardbaar geacht, maar zij geven in elk geval wel aanleiding tot een intensieve gedachtenwisseling tussen de betrokken vakgroepen over het gehele opleidingsconcept voor onderwijskundigen.

Het is te verwachten dat de studierichting onderwijskunde steeds vaker zal gaan starten in de propaedeuse. Reeds lang is de toegepaste variant in Twente al vier jaar. Het komend cursusjaar zal dat ook in Utrecht het geval zijn. In andere instellingen is de positie van de studierichting in de propaedeuse eveneens versterkt. Dat alles kan er zeker toe bijdragen dat meer diepgang in de opleiding kan worden verkregen.

Over een kwestie zijn de vakgroepen het volledig met elkaar eens: het verdient geen aanbeveling de visitatie voor twee studierichtingen door één commissie te laten uitvoeren. Zowel het bestuur van de VOR als de Kamer onderwijskunde dringen er bij de VSNU op aan een volgende visitatie door twee gescheiden of afzonderlijke commissies te laten uitvoeren. Dit standpunt, dat overigens veel begrip krijgt bij de visitatiecommissie zelf, is onder de aandacht gebracht bij het bestuur van de VSNU. Binnen enkele jaren volgt er een visitatie voor het onderzoeksdeel van de universiteiten. Voor onderwijskunde en pedagogiek zal dat in 1995 het geval zijn. Het is ook dan zaak de twee studierichtingen afzonderlijk te visiteren en ook in een afzonderlijk rapport te evalueren.

De onderwijskunde bestaat aan de meeste instellingen nu zo ongeveer 30 jaar. De afgestudeerden blijken nog steeds vrij goed aan een baan te kunnen komen. Maatschappelijk is het beeld over de onderwijskunde wat diffuus. De positie van de onderwijskundigen in debatten over onderwijs is tegenwoordig weinig opvallend. In de beginperiode was dat wel anders. Ik denk aan bekende strijders voor beter onderwijs en onderwijsbeleid als Kohnstamm, Idenburg Van Gelder en vele anderen. Deelname van onderwijskundigen aan adviescommissies van de overheid blijft bescheiden. Van een kritische functie vanuit de wetenschap valt weinig te vernemen. De Green peace functie werkt in onderwijsland nog niet en zeker niet vanuit de wetenschapsbeoefenaren.

De verbreding van de VOR

De VOR is een bloeiende vereniging. Momenteel hebben we ruim 550 leden. De VOR is niet langer alleen een vereniging van onderwijsonderzoekers. De VOR beschouwt zich als een beroepsvereniging van personen die onderwijskundig actief zijn in het onderwijsveld. In 1990 is de verbreding ingezet. Dat houdt in dat alle professionele werkers in het onderwijsveld, met een universitaire of daaraan gelijkstaande opleiding, lid kunnen worden.

De vakgroepen onderwijskunde zijn verenigd in de Kamer onderwijskunde van de VSNU. Daar is besloten de VOR te beschouwen als de beroepsvereniging voor onderwijskundigen. Vandaar dat de VOR ook een afgevaardigde kon aanwijzen voor de visitatiecommissie onderwijskunde.

De divisiestructuur binnen de VOR is dit jaar enigszins gewijzigd, hetgeen heeft geleid tot een herformulering en een uitbreiding. Het ledenaantal steeg. Kortom het gaat goed met de VOR.

Ik kom nu tot een afronding van mijn bijdrage. Ik wens de organisatiecommissie en de bezoekers een paar inspirerende en vruchtbare conferentiedagen toe. In de folder voor deze ORD93 wordt de dubbele functie van de conferentie kort omschreven: "het is een forum voor onderwijsonderzoekers en het is een platform waar docenten, staf- en beleidsmedewerkers en andere geïnteresseerden 'wat komen halen'." Ik hoop van harte dat er veel te halen valt, zowel voor de Nederlandse als de Vlaamse bezoekers. Als het goed bevalt kan ik alvast aankondigen dat de ORD'94 zal gaan plaats vinden in Utrecht, onder auspiciën van de Vakgroep onderwijskunde aldaar. Naast de thema's gekoppeld aan de divisies zal Utrecht als eigen thema aandragen een actuele en grootschalige onderwijsvernieuwing: de basisvorming. Ongetwijfeld een thematiek waar velen in het onderwijsveld bij betrokken zijn. Ook dan hoop ik dat er weer veel te halen zal zijn. We houden die conferentie in Utrecht overigens nog gewoon in onze eigen taal.

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 5, pp. 261-278

The quality of the IEA-reading activity inventory pilot version

Martha E. Otter
*SCO-Kohnstamm Instituut, University of Amsterdam**

ABSTRACT

The quality of a questionnaire intended to measure the reading time of students in grade 2 of secondary education was piloted in the scope of the IEA Reading Literacy Study. Total reading time and the time on reading categories were measured; school, home work, books, magazines and newspapers. For the purpose of the study 276 students were administered twice to complete the questionnaire. The analysis of the data showed that the means of the main categories were about equal for both administration rounds. However, the average reading time measured by the questionnaire turned out to be very high and deviates a great deal from the results in other studies. Moreover, the 'true' correlation of the scores measured at both administrating rounds were well below 1.00 suggesting that the instrument measures other or different constructs at both administration rounds.

INTRODUCTION

One aim of the present study is an examination of the quality of a questionnaire intended to measure the leisure-time reading of students in secondary education. This questionnaire was constructed by the National Research Coordinators of the Reading Literacy Study, under auspices of the International Association for Educational Achievement (IEA) in which more than 30 countries participated (Elley 1992; Postlethwaite & Ross 1993; Lundberg & Linnakyla 1993).

Examing the quality of this questionnaire, the so-called Reading Activity Inventory (RAI), is important because, in the scope of the Reading Literacy Study, one wishes to use the questionnaire data in order to describe reading behaviour in such terms as: 'American students spend more time reading newspapers than do students in Denmark: on average American students read 35 minutes per week while their Danish peers read for only 15 minutes per week.' Furthermore, one might wish to use the data from this questionnaire to analyze the relationship between student characteristics. Examples of research questions one might wish to answer could include: 'Is there a relation between students leisure-time reading and their reading achievement in school?' or 'Is there a relation between students leisure-time reading and the educational level of their parents?' To answer such questions all instruments used has to be of good quality.

Analyzing the quality of the RAI is not only expedient for the IEA Reading Literacy Study self, but it also serves an broader purpose. Should the quality of the data provided by the RAI prove satisfactory, this could be a specific indication of an easy and cheap way of measuring *behaviour* (think of student- teacher- or consumer's behaviour) on ratio scale (compare physical measures such as length or speed, for instance). The RAI is namely aimed at measuring the reading time *in minutes per (average) week.*

On the basis of both empirical research regarding the quality of written questionnaire data and theoretical models concerning the question-and-answer process one could however entertain some doubts with respect to the quality of the RAI. From scientific literature is appears to be the case that there are still many problems attached to the questionnaire as an approach to

* Adres: Grote Bickersstraat 72, 1013 KS Amsterdam.

collect data. From research conducted by Schuman & Presser (1981), Billiet et al. (1984) and Molenaar (1986) it was shown that minor variations in formulating questions can have an effect on the answers given by respondents. Not only variations in wording have an effect but also the use or non-use of a 'don't know' filter (Schuman & Presser 1981), the question length (Bradburn & Sudman 1979) and from minor variations in the number and sequence of the answer categories (Schuman & Presser 1977; Groves & Kahn 1979; Presser & Schuman 1980; Miller 1984).

The above mentioned studies mainly concerned attitude questions for adults. However, there are indications in the literature that the problems mentioned also occur in questions about pupil behaviour (Otter, in press a; Otter submitted for publication b; Vooijs, Van der Voort en Beentjes 1987), teacher behaviour (Overmaat, 1992; Otter et al. 1992; Otter submitted for publication c) and adult behaviour about topics of such matters as visiting the family doctor (Cannell, et al. 1979), drinking habits (Bradburn & Sudman 1979) and sleeping habits (Kalton et al. 1980). It is generally assumed that questions about overt behaviours and facts are less sensitive to the problems mentioned above then are attitude questions.

Cognitive theoretical models about the question-and-answer process also raise questions about the questionnaire approach to measure behaviour (see for instance Cannel, Miller & Oksenberg 1981; Graesser & Murachver 1985 en Tourrangeau & Rasinski 1988).

Cannel, Miller & Oksenberg (1981) have distinguished three cognitive activities undertaken by respondents when searching for information in long-term memory: deciding which information must be looked for, actually searching for this information and then organizing it. The accuracy of the outcome of these cognitive activities (the response) depends on the interval between storing and recalling the required information (Cash & Moss 1972; Duncan & Hill 1985) *and* on the complexity of the task.

Task difficulty or complexity increases when the required information cannot be directly extracted from memory (Groves 1989), because for instance it refers to too many information units stored in memory. These units must be extracted from memory one by one, compared and then weighed up. The final results of this complex process is the actual answer to the question (Groves 1989). Some examples of complex questions are: 'On the average, how often do you read detectives at home for fun?'.[1] Krosnick & Alwin (1987) demonstrated that respondents tended not to put much cognitive effort in answering complex questions.

Empirical results obtained by Otter (a, b & c) and Overmaat (1992) conform the conclusion drawn by the cognitive theories that complex questionnaire items give unreliable and invalid data.

In a study conducted by Otter (a) it appeared that the mean test-hertest coefficients of non-complex (simple) questions is higher than the mean test-retest coefficient of complex questions. Expressed in an effect size (Cohens' D) the mean (test-retest coefficient) of simple questions is nearly one standarddeviation higher than the mean (test-retest coefficients) of complex questions. In an other study Otter (b) compares the distributions of leisure-time reading measured via complex questionnaire items and diaries gathered by the same group of children.[2] It appeared that the distributions differed significantly. Because the quality of the diary method seemed to be satisfactory Otter concluded that it is very plausible that complex questions are an invalid method for measuring leisure-time reading.

But not only youngers have difficulties answering complex questions. Overmaat (1992) and Otter (c) compares the questionnaire and diary method for measuring language and reading provisions in schools. Both authors express serious doubts about the quality of the questionnaire method in term of reliability and validity.

But the complexity of the questions is not the only factor that can influence the quality of the RAI. There is also the problem of the variety of literacy genres that must be recalled (see Appendix 1). Vooijs, Van der Voort & Beentjes (1987 p 78). stated that an overestimation of the reading time occurs when the number of discerned literature genres increases, 'Because, with each estimation, respondents are inclined to overestimate their reading time, all these false estimations pile up, finally resulting in a total estimation which generally turn out to be at least

more than hundred percent too high.' Bechtel, Achelpohl & Akers (1972) and Israël, Simmons & Robinson (1972) reached the same conclusion in their studies about viewing habits.

Another research approach to measure the actual (physical) duration of behaviour is the so-called diary method. This method is used in studies dealing with the way people spend their time. For a certain number of days, respondents are asked to keep a diary of daily activities. An important advantage of this method compared to that of the RAI is that the cognitive process required is minimal. Respondents can extract the required information from their memory.

Researchers like Vooijs, Van der Voort & Beentjes (1987), Knulst, Kalmijn & Van Beek (1988), Schmidbauer & Löhr (1985), Anderson et al. (1988) and Timmer, Eccles & O'Brien (1985) used diaries to measure the amount of time spent by youngsters on reading books, newspapers and magazines. Although these studies differed in many respects from one another there was a common finding across all of the studies, namely that youngsters spent between 1 1/2 and 3 hours per week on reading books, magazines and newspapers.

Overall view of the procedure

It will be recalled that the goal of the present study is to gain insight into the quality of the Pilot version of the Reading Activity Inventory. First the time spent by students on reading books, magazines and newspapers as measured by the RAI will be checked against the average reading time of 1 1/2 to 3 hours obtained in other studies. Secondly a comparison over time of the results will be made to check if the RAI fulfils its intended purpose of measuring the time spent on reading during *average-* or *standard weeks*. This will be done at the group/sample and at the individual level.

The analysis concerning the quality of the RAI will be completed by presenting some psychometric results of the questionnaire.

For the purpose of the study the RAI was administrated twice (test-retest design). Between the two administration rounds, an interval of 5 to 6 weeks was chosen. This time span is generally considered long enough to minimize memory effect and short enough not to expect any changes in the (supposedly stable) reading habits of the respondents.

PROCEDURE

Respondents

The first administration round of the RAI took place within the framework of the IEA Reading Literacy Pilot Study. For the purpose of the reading pilot study a judgement sample was drawn in two stages. In the first stage, 52 schools of secondary education were selected at random. Nearly 67% of the schools (34) were willing to cooperate in the pilot study. In the second stage, one (intact) second-grade class per school was selected at random. The procedure used has been extensively described by Ross (1991) and Schleicher (1989).

Four weeks after the first administration round of the RAI, on account of financial reasons eleven classes were selected at random to participate in the second administration round, of wich 91% was cooperative (ten classes). Cooperation meant that in the 5th or 6th week after the first administration round of the questionnaire, the respondents completed the RAI again under exactly the same (fully described) conditions as the first time. An overview of the classes and students participating in the IEA Reading Literacy Pilot Study and in the present study is presented in Table 1.

Table 1 shows that there have been more classes participating in the second administration round than actually involved in the analyses. The reason for this was that at the moment of analyses just a part of the data of the first administration round was available. By comparing information gathered in the first and second administration round concerning background variables such as age, sex and home language, it was cheched weather the pupils in the analyses differed from all pupils in the second administraion round. This was not the case.

Instruments

The RAI consists of five sections. First, a description of the sections will be given, then an examination of the scoring.

Two of the five sections refer to 'reading at school', underneath entitled as: 'School' and 'Homework'. These two sections consist each of seven similar questions. Five of them refer to the reading behaviour related to different school subjects. One question refers to learning how to read tables and graphs and another to the use of workbooks. (see Appendix 1, question 3 to 16). The other three sections refer to reading as a leisure-time activity: reading at home for one's own enjoyment. These sections are entitled as 'Books' (14 questions), 'Magazines' (7 questions) and 'Newspapers' (7 questions).

In the section 'Books', 14 different genres were given such as detective stories, science-fiction, and poetry (see Appendix 1, question 23 to 36). The questions in the sections 'Magazines' and 'Newspapers' refer to various columns in magazines and newspapers such as sport, politics or cartoons (see Appendix 1, questions 37 to 52.

Scoring

The respondents had to fill in two answers for each question. The first answer referred to the frequency of reading the specific material and the second one to the average duration (in minutes) for the chosen frequency categories. Four frequency categories were used, namely, 'almost never', 'about once a month', 'about once a week' and 'almost every day' together with five units of time: 'none', 'about 10 minutes', 'about 30 minutes', 'about 60 minutes' and '90 minutes or more'.

The questionnaire items were scored in accordance with the IEA suggested scoring procedures: the frequency categories were recoded at the *week level*: 'almost never' was recoded to 0; 'once a month', to .25; 'once a week', to 1 and 'almost every day' to 6. Subsequently, the frequency codes were multiplied by the respective time units. That means that an answer such as 'once a month for 60 minutes' obtains the score 15 (.25 x 60 = 15) 'once a week about 10 minutes' scores 10 (1 x 10) and so on. Scores at the *section* level were obtained by summing the question scores. This way of scoring implies a ratio scale (the numerical ratio corresponds to the actual-time ratio), for instance 2*15=30 minutes or 3*10=30 minutes, a scale which has not been frequently used in social science research.

One could decide to use *only* the four frequency categories, scored as 1, 2, 3, and 4. Yet, there are also some drawbacks attached to this choice. Frequency and duration categories were constructed to be dependent upon each other (see Appendix 1, p. 2). In other words, frequency categories were meant to be associated with duration categories and vice versa. Thus, supposing one might wish to use *only* the frequency categories, this might change the rank order of respondents. Compare fore instance answers as 'once a week, for 60 minutes' to answers as 'almost every day, for 10 minutes'.

PROBLEM AND ANALYSIS

Reading time measured by the RAI compared with results obtained in other studies

In order to compare the time spent on reading books, magazines and newspapers measured per week with results obtained in previous studies, univariate statistics from the RAI are presented.

Table 1. Number of participating classes and students in the first and second administration round.

	1st round	2nd round	included in analysis
classes	34	10	6
students	949	276	202

Per section, the average reading time per week (Mean), the standard deviation (Sd), the median reading time (Median) and the minimum and maximum (Min-Max) reading time in minutes per week are reported.

Fulfilment of the RAI at the group/sample level

Subsequently, a check will be made weather at, *the group or sample level*, the RAI fulfils its intended purpose of measuring the time spent on reading in an average or standard week. If the answers are consistent in the two different weeks and the *distribution* of responses in the RAI is the same at both the section and question levels, then one can make statements such as: '30% of the 14 years old pupils declare that they spend a maximum of 10 minutes per week reading books, 60% between 10 and 20 minutes and 10% of the pupils spend more than 2 hours per week on reading books'. By means of the X^2-test for marginal homogeneity (see Everitt 1977 p. 115-116) it will be checked whether these conditions have been fulfilled. In case it appears that the distribution (of the sections and/or the questions) differs substantially, it will be verified whether the *average reading time* obtained both at the section and at question level is stable enough to enable one to make statements such as: '14-years-olds spend an average of one hour per week on reading books. Most of their time was spent reading about wildlife and nature, namely 15 minutes per week'. Before presenting these results, a check will be made by means of the Kolmogorov-Smirnov test whether the scores (at section and question levels) were distributed approximately normally in the total group of respondents. If the scores are not distributed normally then instead of the t-test used for dependent samples, its non-parametric equivalent will be used: the Wilcoxens' matched-pairs signed rank test.

Fulfilment of the RAI at the individual level

In order to check whether the instrument measures the reading time at the individual level in a stable manner, models are postulated in which per administration round, the questions comprised in one section are conceived as indicators of a latent variable and the correlation between the latent variables will be estimated via LISREL (Jöreskog & Sörbom 1986). These analyses will be carried out on two data-sets: once on the ratio scores (Frequencies * Time) and the second time, only based upon ordinal scores (Frequency categories scored as 1, 2, 3 and 4).

Psychometric indicators of the RAI

The internal reliability of each section will be calculated using alpha. Also will be verified the structure of the RAI by checking whether its sections measure a single dimension or factor by means of a factor analysis in accordance with Hagglund (1982). If the questions are not essentially tau equivalent (Lord & Novick 1968) then alpha underestimates the reliability. For these reasons, the reliability will also be verified by the method of Fleishman and Benson (1987), on the basis of an one-factor model.

RESULTS

Reading time measured by the RAI

Table 2 presents univariate statistics about the sections of the RAI at both administration rounds (T1, T2).

From Table 2 it can be seen that, on average, 14-years-olds spend 289 minutes (more than 4 hours) per week on book reading at home, 277 minutes (more than 4 hours) on magazines and 166 minutes per week (almost 3 hours) on newspapers (see Table 3, T1). The median reading time om books, magazines and newspaper reading is 95, 143 and 78 minutes per week respectively.

There is a lot of variation among respondents. In Figure 1 and 2 this variation is shown. In Figure 1 the reading times obtained in 'Books', 'Magazines' and 'Newspapers' were summed. Each bar in the histogram represents the mean reading time of ten percent of the pupils (ten

Table 2. Leisure-time reading in minutes per week per section at time T1 and T2, number of students (N), mean and median per week (mean, median), standard deviation (Sd), minimum and maximum per week (min-max).

Section	N	Mean	Sd	Median	Min-Max
School					
T1	178	452	474	264	0-2650
T2	230	479	451	330	0-2220
Homework					
T1	176	384	423	249	0-2400
T2	245	332	342	225	0-2190
Books					
T1	154	289	476	95	0-3060
T2	214	253	401	115	0-3383
Magazines					
T1	163	277	338	143	0-2430
T2	253	309	470	143	0-3780
Newspapers					
T1	157	166	245	78	0-2070
T2	237	200	312	93	0-2160

Note: The sections entitled 'School', 'Homework', 'Magazines' en 'Newspapers' consist of 7 questions each; the 'Books' section contains 14 questions.

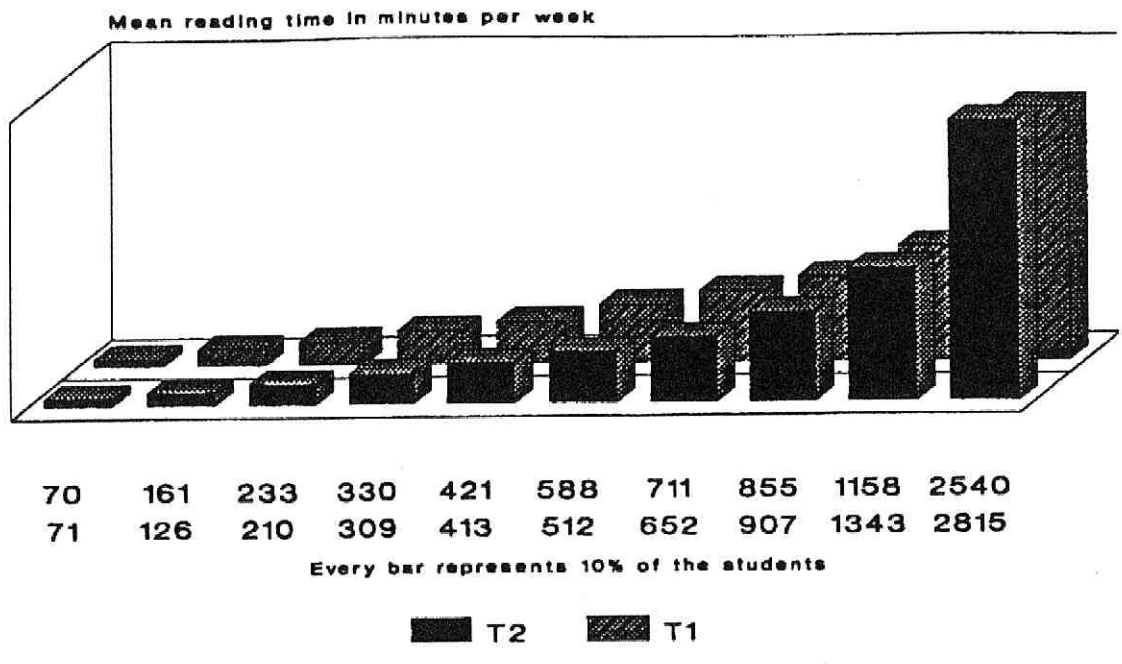


Fig. 1. Leisure-time reading.

procent is eleven pupils at T1 and ninteen pupils at T2). In Figure 2 each bar represents the median reading time of ten procent of the pupils.

From Figure 1 it can be concluded that ten procent of the respondents reporting the smallest amount of time of reading at home do this for more than an hour per week. It also appears that,

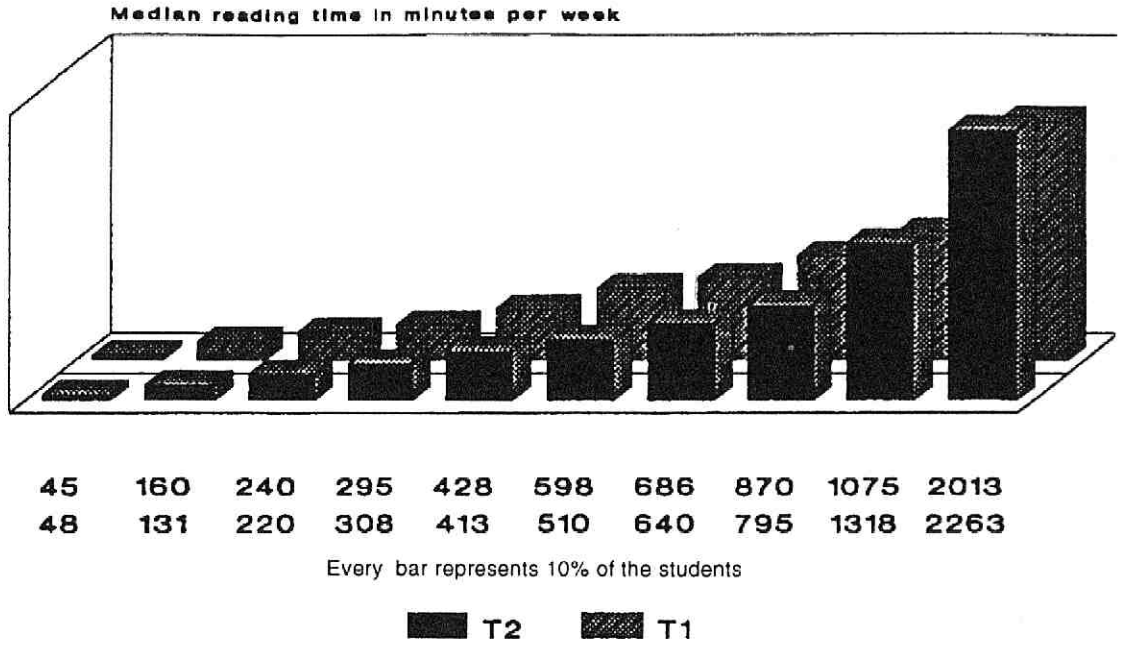


Fig. 2. Leisure-time reading.

at T2, 20% of the pupils indicate they read longer than 22 hours per week and 10% indicate they read about 46 hours per week (at least 61/2 hours a day). Figure 2 shows an identical picture.

It can be concluded that the reading time measured by the RAI appear high and deviates a great deal from the results obtained in other studies.

Fulfilment of the RAI at the group/sample level

Table 3 presents the results of the X^2 tests for marginal homogeneity for the response distribution at the *section level* at T1 and T2. For each section, it gives the X^2-value and the number of degrees of freedom. The results concerning the response distribution at the question level are summarized in Table 4. This table shows, per section, the lowest and highest X^2-value of the questions and the number of degrees of freedom.

From Table 3 it can be seen that the distributions at the section level at time 1 and time 2 do *not* significantly deviate from each other. It can be concluded that the response distribution at section level in the RAI remains stable over an interval of six weeks.

Table 3. Test for marginal homogeneity at two measurement times at section level: number of respondents (N), X^2-value and degrees of freedom (Df).

Section	N	X^2	Df
School	152	6,8	10
Homework	162	7.3	9
Books	122	6.5	10
Magazines	151	7.5	11
Newspapers	139	4.0	8

Note: The sections entitled 'School', 'Homework', 'Magazines' en 'Newspapers' consist of 7 questions each, the 'Books' section contains 14 questions.

Table 4. Test for marginal homogeneity at two measurement times at question level: number of respondents (N), X^2-value and degrees of freedom (Df).

Number of tests per section	N	X^2	Df	Number of tests where (X^2) <0.005 applies
School: 7	180-191	13.9 - 58.3	8 - 11	6
Homework: 7	187-196	9.4 - 84.1	8 - 10	5
Books: 14	168-190	1.2 - 43.4	3 - 11	6
Magazines; 7	168-189	5.3 - 37.6	7 - 10	3
Newspapers: 7	167-183	4.8 - 42.5	8 - 9	6

Table 5. Wilcoxons' tests for differences at two measurement times at question level: number of respondents (N), lowest and highest Z-values and number of test where p (Z) < .05 applies.

Number of tests per section	N	Z	Number of tests where (X^2) <0.005 applies
School: 7	180-191	-2.12 - 0.33	1
Homework: 7	187-196	-1.22 - 1.34	0
Books: 14	168-190	-0.62 - 2.54	1
Magazines; 7	168-189	-1.81 - 1.78	0
Newspapers: 7	167-183	- 2.10 - 0.10	0

Table 4 shows that the distributions in 26 out of the 42 questions at T1 and T2 *do* differ significantly. This is the case for 6 questions contained in 'School', 'Books' and 'Newspapers', 5 questions in 'Homework' and 3 questions contained in 'Magazines'. Therefore must be concluded that the response distribution at the question level is *not* stable over an interval of 6 weeks.

Using the Kolmorov-Smirnov test, it is checked whether it can be assumed that the question distributions are normal for the population. For all questions this hypothesis had to be rejected. Table 5 summarizes the Wilcoxens' matched-pairs signed rank test of the differences at T1 and T2. For each section, the lowest and the highest Z-values of the questions (Z) and the number of questions where p(Z)<.0.5 applies, are presented.

From Table 5 it can be seen that only the mean in 2 out of the 42 questions differ significantly between T1 and T2. The conclusion is that the mean reading time measured by the RAI questions remains stable over a period of six weeks.

Fulfilment of the RAI at the individual level

Table 6 presents the results of the LISREL analysis regarding the postulated models in which, for T1 and T2, the questions of one section are conceived as indicators of a latent variable. Because the variables at the question and section level were not normally distributed the unweighed-least-squares method (ULS) is used because his method does not make any assumptions about the distribution of the variables. Presented are the 'goodness-of-fit' (Gfi) and the correlation coefficients between the latent variables (Correlation).

In Table 6 the results are presented twice: once, based upon ratio scores (Frequency * Time) and the second time, based upon ordinal scores (Frequency categories scored as 1, 2, 3 and 4).

Table 6 shows that the 'true' correlation coefficients between the latent variables based upon ratio scores are .57, .74, .65, .75, .93 respectively for 'School', 'Homework', 'Books' 'Maga-

Table 6. Fit of postulated models (Gfi) and the correlation between the latent variables (correlation) based on both ratio scores and ordinal scores.

Section	Ratio scores			Ordinal Scores		
	N	Gfi	Correlation	N	Gfi	Correlation
School	152	.94	.57	174	.91	.55
Homework	159	.95	.74	178	.94	.60
Books	122	.89	.65	155	.89	.81
Magazines	151	.89	.75	178	.82	.92
Newspapers	139	.91	.93	157	.96	.78

Note: The 'School', 'Homework', 'Magazines' en 'Newspapers' sections consists each of 7 questions, the 'Books' section contains 14 questions.

zines' and 'Newspapers'. With the exception of the coefficient obtained on 'Newspapers' the 'true' correlations are all too low to presume that the instrument measures at both administration rounds. This picture does not change when only frequency scores are used in the analysis. The correlation coefficients between the latent variables are .55, .60, .81, .92 and .78 respectively for 'School', 'Homework', 'Books', 'Magazines' and 'Newspapers'.

Psychometric indicators of the RAI

Table 7 presents the fit of a one-factor model on each section of the RAI (goodness-of-fit index: Gfi), the reliability estimate (FB) according to Fleishman and Benson (1987), Alpha and a 95% confidence interval around Alpha (Int. Alpha).

From Table 7 is can be concluded that the fit of the one-factor model for 'School', 'Homework' and 'Newspapers' is rather good; the 'goodness-of-fit' (Gfi) is .97 or higher. For 'Books'

Table 7. Number of respondents (N), the fit of a one-factor model (Gfi), reliability according to Fleishman & Benson (FB), the alpha coefficient and a 95% reliability interval for alpha (Int. Alpha) for each section at times 1 and 2.

Section	N	One-factor model		Alpha	Int. Alpha
		Gfi	FB		
School					
T1	178	.984	.75	.70	.63 - .76
T2	230	.998	.79	.71	.64 - .76
Homework					
T1	176	.992	.83	.76	.71 - .81
T2	241	.998	.81	.73	.68 - .78
Books					
T1	154	.936	.73	.69	.62 - .76
T2	214	.942	.72	.68	.62 - .74
Magazines					
T1	163	.906	.60	.59	.49 - .68
T2	253	.978	.77	.78	.73 - .82
Newspapers	157	.984	.72	.70	.63 - .77
s	237	.972	.75	.74	.68 - .78
T1					
T2					

Note: The 'School', 'Homework', 'Magazines' en 'Newspapers' sections consist each of 7 questions, the 'Books' section contains 14 questions.

and 'Magazines' the fit is moderate. The sections 'School' 'Homework' and 'Newspapers' measure probably one factor or construct. For 'Books' and 'Magazines' this conclusion is not so obviously.

Table 7 also shows that the reliability of 'School' and 'Homework' is, in accordance with Fleishman and Benson, more than satisfactory, that of 'Books', 'Magazines' and 'Newspapers' reasonable. With the exception of the 'Magazines' section, the reliability is practically equal at T1 and T2.

DISCUSSION

What are the mayor results of this research? Which conclusions can be drawn about the question of whether or not the RAI is a suitable instrument for both descriptive research and research about relations between leisure-time reading and reading achievement?

At first glance, the results about the use of this instrument for descriptive purposes appear to be somewhat divergent. The fact that response *distributions* at the *section level* have proved to be stable over an interval of six weeks is in favour of the RAI. Furthermore, the *average reading time* at both the question and section levels is also stable. Although the response distribution at the question level is somewhat different at both times of administration, the fact is that statements are not (or hardly ever) based on responses to single items.

On the other hand, the average reading time measured by the RAI turned out to be very high: more than 11 hours per week. Not only does this *appear* high but it also *deviates a great deal* from the results obtained in other studies (Knulst & Van Beek 1988, Schmidbauer & Löhr 1985; Vooijs, Van der Voort & Beentjes 1987, Timmer, Eccles & O'Brien 1985). Although these studies differed from each other in many respects, there was -as already mentioned- one similar finding: on average pupils read books, magazines and newspapers for approximately 11/2 to 3 hours per week. This contrasts sharply with the average of more than 11 hours measured by the RAI. It indicates that the RAI overestimates the average reading time. A plausible explanation for this aberrative was given by Vooijs, Van der Voort & Beentjes (1987) who stated that direct time estimations (based on questions to which respondents had to indicate how many minutes they spent on reading, on average) led to an overestimation of the time spent on reading and that, the larger the number of literature genres, the stronger the overestimation. Moreover, both the results obtained by Vooijs, Van der Voort and Beentjes and those of the present research are congruent with cognitive theories about the question-and-answer process. These theories explain that the quality of answers diminishes when the complexity of questions increases. Respondents are not able to give reliable and valid answers if the requested information refers to too many information units in memory. And this is the case with *all* the questions contained in the RAI.

The fact that the RAI measures the average reading time (at both the section and question levels) in a stable manner indicates that overestimation of the reading time occurred at both administration rounds.

The above mentioned arguments militate against the use of the RAI in descriptive research.

What can be concluded with regard to the question about whether or not the RAI is a suitable instrument to make statements about relations between leisure-time reading and reading achievement? The psychometric results appear to argue in favour of the use of some sections of the instrument. The alpha reliability coefficients of some sections are more than satisfactory ('School', 'Homework'), for others reasonable ('Books', 'Newspapers'). The fitt of the one-factor model of some sections are supporting the RAI too.

However, the relatively low 'true' correlations between the latent variables argue against the use of the RAI. Moreover this does not change when only frequency scores are used. The fact that the 'true' correlations between latent variables are well below 1.00 suggests that the sections measure *other or different constructs* at both administration rounds in time.

It should also be noted that the latent variables 'School' and 'Homework' (sections showing the best psychometric qualities: one-factor structure and coefficient alpha) show the lowest 'true' correlation between the two administration rounds. This suggests that an evaluation of an instrument that relieves only on internal consistency measures is an *insufficient* approach when one wishes to be sure that one has measured what one actually intended to measure.

The general conclusion of this study is that the Reading Activity Inventory doesn't have sufficient psychometric quality to be used in either descriptive research or in research about relations between student leisure-time reading and reading achievement. A further development of the Reading Activity Inventory (RAI) is in the eyes of the author not advisable: the questions are complex and the amount of literature genres too many, thus almost *forcing* the respondents to give unreliable and/or invalid answers. Reconstructing the RAI questions into simple ones (did you read yesterday (last week) ...) doesn't seem a solution because it is plausible that respondents read, in such a short perod of time, only in one or two of the fourteen different genres of books the RAI discriminates (see Appendix 1 question 37 to 52). Moreover, such an instrument suggests that leisure-time reading is a stable personal construct, like intelligence, that doesn't change over a period of weeks. Otter (c) makes clear that this assumption should be called in question.

But are we really left empty-handed in measuring the time spent on reading? This does not seem to be the case. From a two year longitudinal study conducted by Otter (submitted for publication d) with a sample of 40 school (nearly 1000 pupils) it appears that, when diaries are used to measure reading behaviour the obtained information proves to be reliable and the results concerning the validity of the instrument can be considered to be very promising.

ACKNOWLEDGMENT

The author thanks Prof. Dr. G.J. Mellenbergh, Prof. Dr. K. de Glopper, Prof. Dr. T. N. Postlethwaite, Dr. M. Eiting, Drs. L.M. Lamers & Drs. G. de Jong for their comments on an earlier draft of this paper.

NOTES

1. Romantic stories/sport reports/stories of adventure/science fiction/poetry/biographies/travel books are other (RAI) examples.
2. In this study the diary method (for measuring leisure-time reading) was empirical evaluated in terms of reliability and validity. The reliability of the diary method appeared to be very good, the (construct)validity amply sufficient. The test-retest coefficient (after six weeks) of the complex questionnaire items (for measuring leisure-time reading) appeared to be unsatisfactory.

REFERENCES

Anderson, R.C., Wilson, P.T., Fielding, L.G. (1988). Grow in reading and how children spend their time outside of school. *Reading Research Quarterly, 1988, 13*, 285-303.

Bechtel, R.B., Achelpohl, C., & Akers, R. (1972). Correlates between observed behaviour and questionnaire responses on television viewing. In: E.A. Rubinstein, G.A. Comstock, & J.P. Murray (Eds.), *Television and social behaviour: Vol. 4. Television in day-to-day life: Patterns of use*. Washington, D.C.: Government Printing Office.

Billiet, J., Loosveldt, G., & Waterplas, L. (1984). *Het survey-interview onderzocht* [An analysis of the survey interview]. Leuven: Sociologisch Onderzoeksinstituut K.U.

Bradburn, N.M., & Sudman, S. (1979). *Improving interview method and questionnaire design*. San Francisco: Jossey-Bass.

Cannell, C.F., Miller, P.V., & Oksenberg, L. (1979). *Research on Interviewing techniques*. Ann Arbor, Institute for Social Research, The University of Michigan.

Cannell, C.F., Miller, P.V., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhart (Ed.), *Sociological Methodology* (389-437). San Francisco: Jossey-Bass.

Elley, W. B. (1992). *How in the World do Students read? IEA Study of Reading Literacy.* The Hague: IEA.

Everitt, B.S. (1977). *The analysis of contingency tables.* London: Wiley.

Fleishman, I., & Benson, J. (1987). Using LISREL to evaluate measurement models and scaling reliability. *Educational and Psychological Measurement, 47,* 925-939.

Graesser, A.C., & Murachver, T. (1985). Symbolic procedures of question answering. In: A.C. Graesser & J.B. Black (Eds.), *The Psychology of Questions.* Hillsdale: Lawrence Erlbaum.

Groves, R.M., & Kahn, R.L. (1979). *Surveys by telephone.* New York: Academic Press.

Groves, R.M. (1989). *Survey errors and survey costs.* New York: Wiley.

Hagglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika, 47,* 209-222.

Israel, H., Simmons, W.R., & Robinson, J. (1972). Demographic characteristics of viewers of television violence and news programs. In: E.A. Rubinstein, G.A. Comstock, & J.P. Murray (Eds.), *Television and social behaviour: Vol. 4. Television in day-to-day life: Patterns of use.* Washington, D.C.: Government Printing Office.

Jöreskog, K.G., & Sörbom, D. (1986). *LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood (Version VI).* Uppsala: University of Uppsala

Kalton, G., Roberts, J., & Holt, D. (1980). The effects of offering a middle response option with opinion questions. *The statistican, 29* (1), 65-78.

Knulst, W., Kalmijn, M., & Beek, P. van (1988). *Van woord naar beeld? Onderzoek naar verschuivingen in de tijdbesteding aan media in de periode 1975-1985.* [From word to image? Survey of the shifts occurring in the amount of time spent on media in the period 1975-1985]. Rijswijk: Sociaal en Cultureel Planbureau.

Krosnick, J.A., & Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51,* (2), 201-219.

Lord, M.F., Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading: Addison-Westley.

Molenaar, N.J. (1986). *Formuleringseffecten in survey-interviews* [Wording effects in survey-interviews]. Amsterdam: VU-Uitgeverij.

Lundberg, I. & Linnakyla, P. (1993). *Teaching Reading around the World.* The Hague, IEA.

Miller, P.V. (1984). Alternative question forms for attitude scale questions in telephone interviews. *Public Opinion Quarterly, 48* (4), 766-768.

Otter, M.E., Mellenbergh, G.J., Blok H., & Bergh, H. van den (1992). Measuring time spent on teaching native language. *Journal of Experimental Education, 60,* (3), 281-287.

Otter, M.E. (a). *Pupils answer Questions. Information Processing Theories forecast the Test-Retest Association Coefficients of the Questions.* (Accepted: Journal of Educational Measurement).

Otter, M.E. (b). *Leisure-time reading measured via questionnaires and diaries. Reliability and validity indices of both methods.* Manuscripts submitted for publication (Reading Research Quarterly).

Otter, M.E. (c). *Language and reading provision measured via questionnaires and diaries.* Manuscript submitted for publication (Journal of Experimental Education).

Otter, M.E. (d). *Relations between reading performances, reading provision and leisure-time reading. A longitudinal study in grade five and six in elementory education.* Manuscript submitted for publication (Reading Research Quarterly).

Overmaat A. M. (1992). *De kwaliteit van meetinstrumenten voor het aanbod van taal en lezen* [The Quality of instruments measuring language and reading provision in school]. S.C.O., Amsterdam.

Postlethwaite T. N. and Ross, K. (1992). *Effective Schools in Reading: Implications for Educational Planners. An Exploratory Study.* The Hague: IEA.

Presser, S., & Schuman, H. (1980). The measure of a middle position in attitude surveys. *Public Opinion Quarterly, 44,* 77-85.

Robinson, J.P. (1977). *How Americans use time: A social-psychological analysis of everyday behaviour.* New York, NY: Praeger.

Ross, K. (1991). *Sampling manual for the IEA International Study of Reading Literacy.* International Coordinating Center. IEA International Study of Reading Literacy, University of Hamburg.

Schmidbauer, M., & Löhr, P. (1985). *Der Markt der kommerziellen Kindermedien.* München: Saur.

Schuman, H., & Presser, S. (1977). Question wording as an independent variable in survey analysis. *Sociological Methods and Research, 6,* 151-170.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Timmer, S.G., Eccles, J., & O'Brien, K. (1985). How children use time. In: Juster, F.T. & Stafford, F.P. (Eds.), *Time, goods and well-being.* Ann-Arbor, University of Michigan.

Tourangeau, R., & Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*, 299-314.

Vooijs, M., Voort, T. van der, Beentjes, H. (1987). De geschiktheid van verschillende typen vragen om de kijktijd en leestijd van kinderen te meten: een validatie onderzoek. [The quality of different types of questions in measuring children's viewing and reading time: a validation survey]. *Massacommunicatie, I*, 65-80.

Manuscript ontvangen 30-3-1993
Definitieve versie ontvangen 17-9-1993

APPENDIX

Reading Literacy:
Reading Activity Inventory,
Population B

Pilot Test Version

Directions:

The first two questions are very easy.
In Question 1 you should check whether you are a boy or a girl.
In Question 2 please check how often you use Dutch at home.
If you need help, ask your teacher.

1. Are you a boy or a girl?

   [] boy
   [] girl

2. Do you use Dutch at home ...
   (Check one only.)

   [] always?
   [] almost always?
   [] sometimes?
   [] hardly ever?
   [] never?

Directions:

The charts below are designed to ask several questions about your reading activities. On the left, the Activities are listed. In the middle are the Frequency columns. On the right are the Amount of Time columns.
*For each activity put a cross (**X**) in the correct box to say how **frequently** you do this activity. Then place a (**X**) in the correct place in the Amount of Time section. This tells how much time you usually give to each activity.*

Examples:

a) Suppose you play a sport once a week, for about one hour each time. You would place a cross under **about once a week** in the Frequency column, and under **about 60 minutes** in the Amount of Time column.

b) Now you fill out by yourself example b) for watching TV.

Ask the teacher if you have a question.

Activities	**Frequency**				**Amount of Time**				
	almost never	about once a month	about once a week	about every day	none	about 10 min.	about 30 min.	about 60 min.	90 min. or more
How often do you do this?									
Examples:									
a) Play a sport	☐	☐	☐	☐	☐	☐	☐	☐	☐
b) Watch TV	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Frequency				Amount of Time				
	almost never	about once a month	about once a week	about every day	none	about 10 min.	about 30 min.	about 60 min.	90 min. or more
(A) Reading in school									
How often do you read books and textbooks in the following areas in school?									
3. Stories/Literature/Fiction in Dutch	☐	☐	☐	☐	☐	☐	☐	☐	☐
4. Science/Mathematics	☐	☐	☐	☐	☐	☐	☐	☐	☐
5. Foreign Language	☐	☐	☐	☐	☐	☐	☐	☐	☐
6. History/Geography/Civics/Economics	☐	☐	☐	☐	☐	☐	☐	☐	☐
7. Vocational-Technical/Home Economics	☐	☐	☐	☐	☐	☐	☐	☐	☐
How often do you read the following materials in any subject?									
8. Reference/Directory/Dictionary or Encyclopdia	☐	☐	☐	☐	☐	☐	☐	☐	☐
9. Workbooks and Exercises	☐	☐	☐	☐	☐	☐	☐	☐	☐
(B) Reading for homework									
How often do you read texts and textbooks in the following areas for homework?									
10. Stories/Literature/Fiction in Dutch	☐	☐	☐	☐	☐	☐	☐	☐	☐
11. Science/Mathematics	☐	☐	☐	☐	☐	☐	☐	☐	☐
12. Foreign Language	☐	☐	☐	☐	☐	☐	☐	☐	☐
13. History/Geography/Civics/Economics	☐	☐	☐	☐	☐	☐	☐	☐	☐
14. Vocational-Technical/Home Economics	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Frequency				Amount of Time				
	almost never	about once a month	about once a week	about every day	none	about 10 min.	about 30 min.	about 60 min.	90 min. or more
How often do you read the following materials for homework?									
15. Reference/Directory/Dictionary or Encyclopedia in any subject	☐	☐	☐	☐	☐	☐	☐	☐	☐
16. Workbooks and Exercises in any subject	☐	☐	☐	☐	☐	☐	☐	☐	☐
(C) Reading these types of materials for enjoyment									
How often do you read these materials for enjoyment?									
17. Books	☐	☐	☐	☐	☐	☐	☐	☐	☐
18. Magazines	☐	☐	☐	☐	☐	☐	☐	☐	☐
19. Newspapers	☐	☐	☐	☐	☐	☐	☐	☐	☐
20. Comic books	☐	☐	☐	☐	☐	☐	☐	☐	☐
21. 'Thrillers'/History	☐	☐	☐	☐	☐	☐	☐	☐	☐
22. Encyclopedia/Dictionary/Manual/Maps	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Frequency				Amount of time				
	almost never	about once a month	about once a week	about every day	none	about 10 min.	about 30 min.	about 60 min.	90 min. or more
(D) Reading at home or in community for enjoyment									
How often do you read these texts for enjoyment?									
23. Mystery	☐	☐	☐	☐	☐	☐	☐	☐	☐
24. Romantic	☐	☐	☐	☐	☐	☐	☐	☐	☐
25. Sports	☐	☐	☐	☐	☐	☐	☐	☐	☐
26. Adventure	☐	☐	☐	☐	☐	☐	☐	☐	☐
27. Science fiction	☐	☐	☐	☐	☐	☐	☐	☐	☐
28. Horror	☐	☐	☐	☐	☐	☐	☐	☐	☐
29. Poetry	☐	☐	☐	☐	☐	☐	☐	☐	☐
30. History	☐	☐	☐	☐	☐	☐	☐	☐	☐
31. Humor	☐	☐	☐	☐	☐	☐	☐	☐	☐
32. Science	☐	☐	☐	☐	☐	☐	☐	☐	☐
33. Travel	☐	☐	☐	☐	☐	☐	☐	☐	☐
34. Wildlife and nature	☐	☐	☐	☐	☐	☐	☐	☐	☐
35. Life stories	☐	☐	☐	☐	☐	☐	☐	☐	☐
36. Classic literature	☐	☐	☐	☐	☐	☐	☐	☐	☐
(E) Read these topics in magazines in your leisure time									
How often do you read these topics in magazines in your leisure time?									
37. Sports/Recreation	☐	☐	☐	☐	☐	☐	☐	☐	☐
38. Movies/TV	☐	☐	☐	☐	☐	☐	☐	☐	☐
39. Romance/Fashion	☐	☐	☐	☐	☐	☐	☐	☐	☐
40. News	☐	☐	☐	☐	☐	☐	☐	☐	☐
41. Politics	☐	☐	☐	☐	☐	☐	☐	☐	☐
42. Pop music	☐	☐	☐	☐	☐	☐	☐	☐	☐
43. Wildlife and nature	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other (please add!)									
44. _____	☐	☐	☐	☐	☐	☐	☐	☐	☐
45. _____	☐	☐	☐	☐	☐	☐	☐	☐	☐

(F) Read these sections of newspapers

How often do you read these sections of newspapers?

	Frequency				Amount of Time				
	almost never	about once a month	about once a week	about every day	none	about 10 min.	about 30 min.	about 60 min.	90 min. or more
46. Sports	☐	☐	☐	☐	☐	☐	☐	☐	☐
47. Comic strips	☐	☐	☐	☐	☐	☐	☐	☐	☐
48. Movies/TV	☐	☐	☐	☐	☐	☐	☐	☐	☐
49. News	☐	☐	☐	☐	☐	☐	☐	☐	☐
50. Politics	☐	☐	☐	☐	☐	☐	☐	☐	☐
51. Romance/Personality	☐	☐	☐	☐	☐	☐	☐	☐	☐
52. Classified Advertisements	☐	☐	☐	☐	☐	☐	☐	☐	☐

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 5, pp. 279-294

Coördinatie en de effectiviteit van het onderwijs op het vakgebied wiskunde

B. Witziers
*Onderzoekscentrum Toegepaste Onderwijskunde, Universiteit Twente**

ABSTRACT

Coordination of the curricular domain in secondary schools is the main theme of this article. Special attention is given to the role of departments and school management. The results presented in this article show that departments are important units within secondary schools with respect to the curricular domain. As such, they do have influence on what teachers do in their classroom. Furthermore, the results show that some aspects related to coordination of the curricular domain – i.e. educational leadership and departments with strong policies – are positively associated with student achievement.

INLEIDING

Thema van het hier te presenteren onderzoek is coördinatie van het curriculaire domein in scholen voor voortgezet onderwijs. Onder coördinatie wordt veelal de afstemming en ordening van de activiteiten van organisatieleden verstaan. Integratie is het doel, coördinatie het middel om dat te verwezenlijken (cf. Firestone & Wilson, 1985; Kieser & Kubicek, 1983; Mintzberg, 1979; Pelkmans & Vrieze, 1987). Uit veel publikaties van onderwijskundige aard kan afgeleid worden dat coördinatie binnen schoolorganisaties geen vanzelfsprekend gegeven is. Te wijzen valt bijvoorbeeld op het onderzoek naar het beroep van leraar. In deze studies komt niet alleen naar voren dat docenten een grote vrijheid van handelen hebben, maar ook dat dit samengaat met het afschermen van het eigen functioneren, een gebrekkige samenwerkings- en overlegcultuur en een solistische taakopvatting (cf. Lortie, 1975; Prick, 1983; Vonk, 1986). In de vele modellen van schoolorganisatie die in de literatuur te vinden zijn worden deze kenmerken daarbij in verband gebracht met andere aspecten van de schoolorganisatie, zoals het gemis van sturingselementen als controle en evaluatie en een gebrek aan overeenstemming over doelen en inhoud van het onderwijs. Het zijn deze kenmerken waar begrippen als 'loosely coupled system ' (Weick, 1976), 'segmentale organisatie' (Marx, 1975) en 'professionele bureaucratie' (Mintzberg, 1979) naar verwijzen.

Voor onderzoekers houden dergelijke beelden de opgave in om de juistheid ervan empirisch te toetsen. Dit is niet alleen van belang vanuit de vraag hoe schoolorganisaties getypeerd moeten worden, maar ook van belang vanuit het oogpunt van schoolverbetering. Zo komt uit de literatuur veelal de veronderstelling naar voren dat naarmate scholen meer geïntegreerde organisaties zijn, het beleidsvoerend vermogen van scholen groter is (cf. Pelkmans & Vrieze, 1987; Sleegers, Bergen & Giesbers, 1992). Daarnaast komt uit de literatuur naar voren dat coördinatie een relatie heeft met de kwaliteit van het onderwijs. Als zodanig leidt het bovenstaande tot een tweetal vragen. De eerste is die naar de aard en mate van coördinatie binnen scholen, de tweede die naar de relatie tussen coördinatie en de kwaliteit van het onderwijs. In de volgende paragraaf zal hierop nader worden ingegaan.

* Adres: Postbus 217, 7500 AE Enschede.

THEORETISCH KADER

De vraag of scholen voor voortgezet onderwijs al dan niet geïntegreerde organisaties zijn en of dit kenmerk een relatie heeft met de effectiviteit van de organisatie, wordt aangepakt door het onderzoek te richten op de wijze waarop vaksecties hun onderwijs coördineren. Het curriculaire domein binnen de schoolorganisatie staat dus centraal. Daarmee is tevens de eerste theoretische invalshoek genoemd.

Vaksecties kunnen omschreven worden als taakgroepen binnen de schoolorganisatie die (idealiter) verantwoordelijk zijn voor het beleid, de uitvoering en evaluatie van het (vak)onderwijs. De keuze voor deze organen is met name ingegeven vanuit de constatering dat het functioneren van vaksecties een verwaarloosd terrein van onderzoek vormt, op een enkele uitzondering na (cf. Hylkema, 1990). Over de reden hiervan kan alleen maar gespeculeerd worden. Zo is de dominantie in de schoolorganisatieliteratuur van beelden en modellen (zie boven), waarin scholen naar voren komen als cellulaire organisaties waarin leerkrachten individueel en in betrekkelijk isolement hun onderwijstaken uitvoeren, een mogelijke oorzaak. Dergelijke typeringen bevorderen immers niet of nauwelijks onderzoek naar verschijnselen als taakdifferentiatie en groepsformatie.

Ondanks deze verwaarlozing kan niettemin bij voorbaat gesteld worden dat vaksecties belangrijke organen kunnen zijn binnen scholen voor voortgezet onderwijs en de bovengenoemde taken meer ter hand nemen dan veelal wordt verondersteld. In de eerste plaats kan daarbij gewezen worden op het feit dat vaksecties binnen scholen voor voortgezet onderwijs een 'normaal' verschijnsel zijn. Siskin (1991) stelt dat in de Verenigde Staten sinds de dertiger jaren de opdeling van de schoolorganisatie in vaksecties de dominante wijze van organiseren is in scholen voor voortgezet onderwijs. Voor de Nederlandse situatie stellen Van Vilsteren en Witziers (1990) dat sinds de zestiger jaren, zowel gestimuleerd door de toegenomen schoolgrootte als door omvattende innovatie- en fusieprocessen, vaksecties een wezenlijk onderdeel zijn gaan uitmaken van scholen voor voortgezet onderwijs. In de tweede plaats kan het potentiële belang van secties nader onderstreept worden door te wijzen op het feit dat het bij vaksecties gaat om een bundeling van specialisten die specifieke vakkennis en een verwante opleiding delen en gemeenschappelijke referentiekaders hebben. Deze achtergrond zal de interactie tussen docenten kleuren en in dit opzicht is het niet vreemd te veronderstellen dat, voorzover er samenhang binnen schoolorganisaties bestaat, deze vooral binnen vaksecties gezocht moet worden (cf. Siskin, 1991). In zekere zin wordt deze veronderstelling bevestigd door een onderzoek van Van Vilsteren (1986) waaruit bleek dat docenten van alle lesoverstijgende activiteiten het werk in en voor de eigen sectie het meest waarderen. Een interessante implicatie hiervan is dat hiermee de relatie tussen het functioneren van vaksecties en de effectiviteit van de schoolorganisatie in het geding komt. Een zeker bewijs hiervoor kan ontleend worden aan een onderzoek van Cuttance (1987), die binnen scholen voor voortgezet onderwijs aanmerkelijke verschillen in effectiviteit tussen vakken vond. Dit gegeven roept de vraag op of, naast vakgebonden verschillen in kenmerken van docenten, functioneringskenmerken van vaksecties wellicht een rol spelen in het realiseren van leerlingprestaties. Duidelijk is dat een positieve beantwoording van deze vraag consequenties heeft voor de theorievorming over effectieve schoolorganisaties, met name omdat hierin kenmerken die bijdragen aan de effectiviteit van de school voornamelijk gesitueerd worden op het algemene niveau van de school of op het niveau van de individuele docent (cf. Scheerens, 1989).

Een andere theoretische invalshoek dringt zich op als in ogenschouw genomen wordt dat organisaties meerdere middelen – of coördinatiemechanismen – ter beschikking staan om meer samenhang te bewerkstelligen. Bij vaksecties kan hier gedacht worden aan zaken als overleg en samenwerking, consensus en de aanwezigheid van een onderwijskundig beleid dat het gedrag van docenten reguleert. Van belang hierbij is dat, naast het functioneren van vaksecties, in dit onderzoek tevens de aandacht wordt gericht op de rol van de schoolleiding en van vaksectieleiders binnen het curriculaire domein. Concreet betekent dit dat integrerende elementen als controle en evaluatie en overleg tussen schoolleiding en docenten in dit onderzoek aan de orde komen.

Vanuit de schoolorganisatieliteratuur is het nu de vraag of al deze verschillende wijzen van coördinatie even succesvol zijn bij het realiseren van leerlingprestaties. De literatuur hierover is nogal ambivalent. Aan de ene kant is er een sterk geloof dat de sleutel tot kwaliteitsverhoging van het onderwijs ligt in het verbeteren van de coördinatie. In de Verenigde Staten komt dit vooral naar voren in meer theoretische beschouwingen over effectieve schoolorganisaties. Hierin wordt veelal een onderscheid gemaakt tussen losjes gekoppelde/niet-effectieve scholen aan de ene kant en stevig gekoppelde/wel-effectieve scholen aan de andere kant (cf. Edmonds, 1979; Rosenholtz, 1985). Voor deze stelling is ook wel de nodige empirische steun te vinden. Zo blijkt uit de omvattende reviews van Scheerens (1989) en Levine en Lezotte (1990) van onderzoek op het terrein van effectieve scholen dat in verscheidene onderzoeken een samenhang wordt gerapporteerd tussen integrerende elementen als overleg en samenwerking, consensus en onderwijskundig leiderschap enerzijds en leerlingprestaties anderzijds. Aan de andere kant zijn in de literatuur over deze aanname ook de nodige tegenwerpingen te vinden. Zo menen Astuto en Clarke (1985) op basis van een meta-analyse van 41 gevalstudies dat uit de onderzochte studies in kwalitatieve en kwantitatieve zin een dermate grote verscheidenheid aan vormen van coördinatie naar voren komt, dat het maar de vraag is of het succes van bepaalde scholen hieraan toegeschreven moet worden. In dit opzicht zou het prescriptieve element, dat zo vaak in de organisatieliteratuur aanwezig is, het descriptieve overheersen, een geluid dat ook te horen valt bij Tyler (1988) in zijn waardering van het effectiviteitsonderzoek. Dat een dergelijke stelling niet geheel uit de lucht komt vallen blijkt bijvoorbeeld uit Nederlands onderzoek op dit terrein. Regelmatig hebben kenmerken als onderwijskundig leiderschap, consensus en overleg en samenwerking geen, of zelfs een negatieve, relatie met indicatoren die inzicht geven in de effectiviteit van de school (cf. Brandsma & Stoel, 1986; Van der Grift & Akkermans, 1991; Hofman & Lugthart, 1991; Stoel, 1986; Tesser, 1986). Een andere kwestie in dit verband is de vraag of er niet een duidelijker onderscheid gemaakt moet worden tussen de verschillende wijzen van coördinatie, elk met mogelijk andere gevolgen voor de kwaliteit van het onderwijs. Zo pleit Rosenholtz (1989) op grond van haar onderzoeksresultaten weliswaar voor een meer geïntegreerde school, maar waarschuwt zij tegelijkertijd voor een (te) sterke rol van de schoolleiding. In haar ogen vormen bureaucratisering en een top-down benadering een belemmering voor de betrokkenheid van docenten bij hun school en als zodanig een ongunstige conditie voor het ontstaan van een effectieve school. Daarentegen zou samenwerking – in brede zin des woords – tussen docenten wel bevorderlijk zijn voor de kwaliteit van het onderwijs. Dergelijke pleidooien kunnen ook gevonden worden bij auteurs als Little (1982), Ashton en Webb (1986) en Smith en Scott (1990). Een laatste kwestie tenslotte is of er tussen kenmerken die betrekking hebben op integratie enerzijds en effectiviteit anderzijds geen lineair, maar een curvilineair verband bestaat. In dit licht valt bijvoorbeeld de opmerking van Creemers en Scheerens (1991) te plaatsen dat (te)veel overleg tijd weg kan nemen van activiteiten die direct in verband staan met het onderwijs en daardoor schadelijk kan uitwerken op de effectiviteit van de schoolorganisatie. In dit opzicht is, zo stellen zij, de relatie tussen integratie binnen de schoolorganisatie en effectiviteit veel complexer dan veelal wordt verondersteld.

VRAAGSTELLING

Uit het bovenstaande wordt duidelijk dat het accent in dit artikel zal liggen op de vraag of er een verband bestaat tussen de mate waarin schoolorganisaties ten aanzien van hun onderwijs als geïntegreerde organisaties getypeerd kunnen worden en de effectiviteit van de schoolorganisatie. Aangezien dit onderzoek zich richt op de coördinerende rol van vaksecties en van het (midden)management en als indicator voor effectiviteit een toets is gebruikt die inzicht geeft in de leerprestaties van leerlingen op het vakgebied wiskunde, gaat het er meer in het bijzonder om of er een verband bestaat tussen de wijze waarop scholen het vakonderwijs coördineren en leerprestaties. Voorafgaand aan deze vraag zal echter eerst kort worden ingegaan op de beschrijvende gegevens. Centrale vraag hier is of scholen voor voortgezet onderwijs al dan niet geïnte-

greerde organisaties zijn. De nadruk zal daarbij vooral vallen op het functioneren van vaksecties. Het primaire doel daarbij is te kijken of vaksecties gezien de wijze waarop zij functioneren überhaupt belangrijke organen (kunnen) zijn met betrekking tot de effectiviteit van scholen.

ONDERZOEKSOPZET

Steekproef

Voor de steekproef is aansluiting gezocht bij het onderzoek 'Uitgangssituatiemeting eerste fase voortgezet onderwijs' uitgevoerd door het Cito. Dit onderzoek kan getypeerd worden als een 'assessment'-onderzoek, waarbij op grootschalige wijze empirische gegevens over de resultaten van het onderwijs verzameld worden (cf. Kremers, 1990). Voordeel van deze aansluiting was dat hierdoor beschikt kon worden over precieze en betrouwbare gegevens over de prestaties en achtergrondkenmerken van leerlingen. Concreet betekende deze aansluiting dat in principe alle scholen die aan dit onderzoek deelgenomen hadden tot de steekproefpopulatie behoorden. Aangezien het oorspronkelijke onderzoek betrekking had op alle schooltypen en meerdere vakken, was al snel duidelijk dat hierbinnen een keuze gemaakt moest worden. Daartoe werd besloten het onderzoek te beperken tot het schoolvak wiskunde en een tweetal schooltypen, te weten MAVO en HAVO. Voor de steekproeftrekking had deze keuze tot gevolg dat in totaal 93 secties die aan het 'Uitgangssituatie'-onderzoek hebben meegedaan, aangeschreven zijn. Van deze groep hebben 39 wiskunde-secties (20 'HAVO'-secties, 19 'MAVO'-secties) aan dit onderzoek deelgenomen.

Een belangrijk probleem inzake de steekproeftrekking was welke docenten bij het onderzoek betrokken moesten worden. Uiteraard zou het wenselijk zijn als van de deelnemende secties alle leden aan het onderzoek mee zouden doen. Bij voorbaat werd dit – mede op basis van ervaringen van ander onderzoek op dit terrein (Hylkema, 1990) – echter niet realistisch geacht, zeker niet bij grote secties. Dit heeft er toe geleid dat bij de verspreiding van de vragenlijsten aan de verantwoordelijke contactpersoon (meestal de vaksectieleider) werd gevraagd aan alle docenten een vragenlijst te geven in het geval zijn sectie kleiner was dan zes personen en alleen aan die docenten die (mede) werkzaam waren op de HAVO- of MAVO-afdelingen wanneer dit niet het geval zou zijn. Deze werkwijze had als nadeel dat onbekend is hoeveel vragenlijsten precies uitgezet zijn. Wel kon op basis van het aantal geretourneerde vragenlijsten (n=169) en gegevens over de omvang van de secties berekend worden dat 66% van alle bij deze secties werkzame docenten een vragenlijst ingevuld heeft. Gelet op de beperking in het onderzoek tot die docenten die betrokken zijn bij het MAVO- dan wel HAVO-onderwijs, lijkt de conclusie gerechtvaardigd dat inderdaad die docenten bereikt zijn die tot de doelgroep behoorden.

Wat de deelnemende leerlingen betreft kan vermeld worden dat de toetsscores van 975 leerlingen beschikbaar waren. Uit een analyse van de toetsscores blijkt dat het prestatieniveau van deze leerlingen min of meer gelijk is aan die van de totale groep leerlingen die aan het 'Uitgangssituatie'-onderzoek meegedaan heeft. Het gemiddelde van de toetsscores van deze leerlingen bedraagt -0.08, terwijl de standaarddeviatie 0.98 bedraagt. Op dit punt is er dus geen sprake van een afwijkende groep scholen. Dit kan nader geaccentueerd worden door te wijzen op het feit dat ook ten aanzien van de achtergrondkenmerken van leerlingen (sexe, opleiding ouders en dergelijke) de steekproef vrijwel identiek is aan de oorspronkelijke leerlingpopulatie.

Voor de volledigheid dient nog vermeld te worden dat aan het onderzoek op basis van een andere steekproef ook nog 31 vaksecties geschiedenis en 34 vaksecties Engels hebben meegedaan. Aangezien de deelname van deze secties vooral van belang is geweest voor andere vraagstellingen in het onderzoek, zal hier niet verder op worden ingegaan. Relevant is dit gegeven hier dan ook alleen in zoverre, dat bij een aantal belangrijke analyses – zoals het berekenen van de betrouwbaarheid van de variabelen – alle docenten en secties betrokken waren.

De laatste opmerkingen in deze paragraaf heeft betrekking op het feit dat de aansluiting bij een bestaand databestand tot gevolg heeft gehad, dat de onderzochte schoolkenmerken een (klein) jaar later zijn vastgesteld dan de leerlingprestaties. Aangezien de deelnemende scholen

ten tijde van dit onderzoek echter nog niet op de hoogte waren van de resultaten van het 'Uitgangssituatiemeting'-onderzoek, is het (vrijwel) uitgesloten dat kennis hiervan invloed had op de wijze waarop scholen/secties hun onderwijs coördineren. In dit opzicht is de koppeling van leerlinggegevens en organisatiekenmerken die een jaar later zijn vastgesteld nauwelijks problematisch. Iets anders ligt dat ten aanzien van de vraag of de gemeten organisatiekenmerken op zich in de tijd stabiel zijn. Hoewel het in dit onderzoek in verreweg de meeste gevallen gaat om duurzame en 'robuuste' eigenschappen die slechts moeizaam veranderen, kan niet uitgesloten worden dat bepaalde kenmerken wel onderhevig zijn geweest aan verandering. Enige voorzichtigheid op dit punt is dus wel geboden.

Concepten en variabelen

In het bovenstaande is naar voren gekomen dat een belangrijke reden om aansluiting te zoeken bij het onderzoek 'Uitgangssituatiemeting eerste fase onderwijs' het feit was dat gebruik gemaakt kon worden van bestaande toetsscores. Deze scores fungeren dan ook als afhankelijke variabele in dit onderzoek. Resumerend gaat het om scores van leerlingen op een wiskundetoets afgenomen in het derde leerjaar van het voortgezet onderwijs. Voor nadere informatie over de constructie en inhoud van deze toets wordt verwezen naar het rapport handelend over bovengenoemd onderzoek (Kremers, 1990).

Als covariaten dienen in dit onderzoek kenmerken op leerling-, op docent- en op schoolniveau. Bij de covariaten op leerlingniveau gaat het om de achtergrondkenmerken etniciteit, geslacht, doublurestatus en het opleidingsniveau van de vader en de moeder. Ook deze gegevens zijn afkomstig uit het databestand van het bovengenoemd onderzoek. De gegevens over de covariaten op het niveau van de docent en de school zijn tijdens het onderzoek verzameld. Bij de docentkenmerken gaat het om ervaring en prestatiegerichtheid, bij de schoolkenmerken om de kenmerken denominatie (openbaar/bijzonder), de structuur van de school (categoraal/scholengemeenschap) en schooltype (MAVO/HAVO).

Als verklarende variabelen fungeren in dit onderzoek die variabelen die inzicht geven in de mate waarin scholen als geïntegreerde organisaties getypeerd kunnen worden. Conceptueel is daarbij gebruik gemaakt van de literatuur over coördinatie (cf. Firestone & Wilson, 1985; Kieser & Kubicek, 1983; Mintzberg, 1979). Concreet heeft dit geleid tot een zestal concepten die vervolgens elk uiteengelegd zijn in een aantal variabelen die als indicator gezien worden van het betreffende concept. Voor deze nadere invulling is vooral gebruik gemaakt van de meer specifiek -al dan niet expliciet- op coördinatie binnen scholen gerichte literatuur. In het navolgende zullen deze concepten en de (belangrijkste) variabelen die daaronder vallen kort beschreven worden.

Het eerste concept (*beleid*) heeft betrekking op het door vaksecties gevoerde onderwijsbeleid. In feite gaat het hier om de mate waarin secties expliciet vastleggen wat de aard en bedoeling is van hun onderwijs. Dit is onder andere vastgesteld door na te gaan welke (bindende) afspraken sectieleden maken ten aanzien van hun onderwijs en hoeveel gemeenschappelijke proefwerken door de sectie per leerjaar worden gegeven. Vanuit deze laatste variabele wordt duidelijk dat dit concept verwantschap vertoont met het in de coördinatieliteratuur gebruikte begrip standaardisatie (cf. Mintzberg, 1979). Bij gemeenschappelijke proefwerken gaat het immers veelal om een standaardisatie van doelen, inhoud en toetsing van het onderwijs.

Het tweede concept (*overleg en samenwerking*) is gerelateerd aan het door Mintzberg (1979) gebezigde begrip wederzijdse afstemming; de afstemming van de activiteiten van de organisatieleden tijdens het werkproces door middel van onderling overleg en informatie-overdracht. Variabelen die hier een rol spelen zijn onder andere de frequentie van overleg, de aard van overleg en het samenwerkingsklimaat. Deze laatste variabele geeft inzicht in de openheid van de communicatie, de kwaliteit van de onderlinge samenwerking en de mate waarin er binnen secties een zekere controle bestaat op elkaars werk.

Het concept *consensus* bouwt voort op de notie dat de mate waarin organisatieleden overeenkomstig denken en handelen een belangrijk vorm van coördinatie is. Mintzberg (1979) reserveert hiervoor het begrip coördinatie door middel van 'ideology'. Dit concept wordt geïndiceerd

door een viertal variabelen die meten in hoeverre docenten binnen een sectie gemeenschappelijk denkbeelden koesteren over zaken als de taken en functie van de school, vakdoelen en -didactiek en de (gewenste) leerinhouden.

Het concept *besluitvorming* sluit aan bij de gedachte dat coördinatie nauw verbonden is met de besluitvormingsstructuur. Kieser en Kubicek (1983) stellen dat in veel organisaties leidinggevenden verantwoordelijk zijn voor de coördinatie en daartoe uitgerust zijn met een bepaalde beslissingsbevoegdheid. Deze bevoegdheid kan echter ook toegewezen worden aan specifieke organen en/of individuen die formeel gezien geen hiërarchische functie vervullen. In veel schoolorganisatiemodellen komt dit laatste naar voren in de nadruk op de professionele autonomie van docenten (cf. Mintzberg, 1979). Andere auteurs wijzen er daarentegen op dat ook andere participanten in de besluitvorming – zoals vaksecties – een belangrijke invloed kunnen hebben op de dagelijkse gang van zaken in het onderwijs (cf. Marx, 1975). In dit opzicht mag dus verwacht worden dat scholen op dit punt zullen verschillen. Het concept besluitvorming is uiteengelegd in een tweetal variabelen die inzicht geven in de mate waarin individuele docenten zelfstandig kunnen beslissen over hun onderwijs – de befaamde professionele autonomie – dan wel de mate waarin de vaksectie hier een belangrijke rol speelt.

De laatste twee concepten (*schoolleiding* en *sectieleider*) hebben respectievelijk betrekking op de onderwijskundige rol van de schoolleiding en van de vaksectieleider. Deze concepten bouwen voort op begrippen als 'direct supervision' (Mintzberg, 1979) en 'Koordination durch personliche Weisungen' (Kieser & Kubicek, 1983). Voor een nadere invulling van het eerste concept is uitgegaan van het begrip onderwijskundig leiderschap zoals dat door Firestone en Wilson (1985) is uitgewerkt. Zij onderscheiden een drietal wijzen waarop de schoolleiding het onderwijs kan coördineren, waaronder controle en regelgeving. Voor het tweede concept is aansluiting gezocht bij Smets (1986) die verschillende rollen onderscheidt die een vaksectieleider kan vervullen. Hier gaat het om de mate waarin de sectieleider zich als onderwijskundig leider van de sectie manifesteert. Bij beide concepten spelen daarbij nog variabelen een rol die inzicht geven in de invloed van deze actoren op een aantal onderwijskundige beslispunten.

Een overzicht van bovenstaande concepten en variabelen wordt weergegeven in Tabel 1. Tevens geeft deze tabel inzicht in het aantal per variabele ontwikkelde items, de betrouwbaarheid van de afzonderlijke schalen, de gemiddelde scores, de schaalbreedtes, de ladingen van de variabelen op de concepten en bij wie welke informatie verzameld is. Op deze twee laatste aspecten zal in volgende paragrafen nog teruggekomen worden. Voor nadere informatie over de inhoud van de vragen wordt verwezen naar het oorspronkelijke onderzoeksrapport (Witziers, 1992).

De laatste opmerking in deze paragraaf heeft betrekking op de correlaties tussen de indicatoren. Per concept bedraagt de gemiddelde correlatie tussen de indicatoren 0.43.

Dataverzameling en -verwerking

Bij de dataverzameling heeft niet elke respondent met dezelfde vragenlijst te maken gehad. Tabel 1 geeft daar nader inzicht in. Reden hiervoor was dat enerzijds zoveel mogelijk informatie ingewonnen moest worden, anderzijds moest dit niet leiden tot uitzonderlijk lange vragenlijsten. Dit laatste zou immers de kans op non-response alleen maar vergroten. Uit het oogpunt van betrouwbaarheid is deze wijze van dataverzameling overigens wel problematisch te noemen. Het betrouwbaar vaststellen van organisatiekenmerken is immers gebaat bij zoveel mogelijk respondenten per school (cf. Rowan, Raudenbusch & Kang, 1989). Duidelijk is dat ook hier een probleem ligt.

Bij de verwerking zijn die gegevens die afkomstig zijn van meerdere respondenten binnen een school geaggregeerd. Veelal betekent dit dat de gemiddelde score van deze docenten gebruikt is als score van een sectie. Dit geldt ook voor die variabelen die oorspronkelijk betrekking hadden op kenmerken van de docent (ervaring, prestatiegerichtheid). Reden hiervoor was dat het technisch niet mogelijk was de individuele docentgegevens te koppelen aan de leerlinggegevens. In feite gaat het hier dus om kenmerken op schoolniveau. Een uitzondering op deze wijze van dataverwerking vormen de schalen die betrekking hebben op het concept consensus.

Tabel 1. Overzicht concepten en variabelen.

Concept (variabelen)	no. items	alpha	$\bar{x}$	schaal breedte	parameter waarde	info bron
Beleid						
sectieregels	17	0.84	2.53	1-4	0.82	VL
gemeenschappelijke proefwerken	4	0.94	6.02*	nvt	0.54	VL
planning	8	0.91	2.05	1-4	0.33	VL
schoolwerkplan	3	0.73	1.23	1-3	0.52	VL
Overleg en samenwerking						
frequentie overleg	1	nvt	14.83**	nvt	0.38	VL
aard formeel overleg	14	0.83	2.54	1-4	0.50	VL
aard informeel	14	0.79	2.73	1-4	0.59	D
cohesie	11	0.85	2.97	1-4	0.20	D
samenwerkingsklimaat	11	0.82	2.73	1-4	0.87	D
Consensus						
homogeniteit tav taken/functie school	14	0.74	0.27	0-2.31***	0.73	A
homogeniteit tav doelen vakonderwijs	10	0.81	1.70	0-5.77***	0.92	A
homogeniteit in lesgedrag	12	0.85	1.79	0-5.77***	0.85	A
homogeniteit tav leerinhouden	5	0.77	6.35	nvt	0.73	A
Besluitvorming						
invloed docent	14	0.86	2.67	1-4	0.40	A
invloed vaksectie	14	0.86	2.88	1-4	0.96	A
Schoolleiding						
controle schoolleiding	17	0.92	1.43	1-4	0.88	VL
verticale communicatie	17	0.91	1.67	1-4	0.79	VL
schoolregels	17	0.90	1.98	1-4	0.63	VL
Invloed schoolleiding	14	0.78	1.39	1-4	0.52	VL
Vaksectieleider						
rol vaksectieleider	3	0.79	2.15	1-4	0.66	D
invloed vaksectieleider	14	0.94	1.55	1-4	0.63	A

* = aantal gemiddelde proefwerken per leerjaar
** = aantal uren overleg op jaarbasis
*** = gebaseerd op minimale\maximale spreiding
VL = informatie ingewonnen bij vaksectieleider
D = informatie ingewonnen bij docenten
A = informatie ingewonnen bij alle betrokkenen

Van belang is hier dat docenten konden aangeven in welke mate zij het eens waren met uitspraken die betrekking hadden op de taken en functie van de school, de doelen van het vakonderwijs, hun wijze van lesgeven en de gewenste vakinhouden. Als sectiemaat is hier gebruik gemaakt van de gemiddelde spreiding in de antwoorden van docenten per item. Een lage score van een sectie op deze variabele betekent dus een hoge mate van consensus.

Analyseprocedure

Voor de beantwoording van de eerste -beschrijvende- vraag is vooral gebruik gemaakt van de gemiddelde scores op de afzonderlijke schalen. Het gaat hier zowel om gemiddelden berekend over secties als over de totale groep van docenten.

Voor de tweede vraag is gebruik gemaakt van een multi-level analyse. Deze analyse is

uitgevoerd met behulp van het programma VARCL (Longford, 1986). Met dit programma kunnen statistische problemen die optreden bij de analyse van gegevens op meerdere niveaus het hoofd geboden worden (cf. Aitkin & Longford, 1986; Bosker, 1990; Goldstein, 1987). In dit onderzoek is van een dergelijke hiërarchische structuur sprake; er zijn zowel gegevens op het niveau van leerlingen (leerprestaties) als op schoolniveau (school- en vaksectiekenmerken) verzameld. Het programma VARCL verdeelt in de eerste plaats de variantie in de onafhankelijke variabele in respectievelijk variantie op leerlingniveau en variantie op schoolniveau, waarbij rekening gehouden wordt met het feit dat van bepaalde scholen meer leerlingen bij de steekproef betrokken zijn dan van andere. Vervolgens wordt op individueel niveau (in dit geval prestaties van leerlingen op een wiskundetoets) een regressiemodel geschat, waarbij rekening gehouden wordt met het gegeven dat leerprestaties van leerlingen binnen een school niet onafhankelijk van elkaar zijn (cf. Bosker, 1990). De gebruikte analyse-strategie is daarbij als volgt. Begonnen is met een model zonder voorspellende variabelen (model 0). Deze stap is met name belangrijk om vast te stellen hoeveel variantie de verschillende niveaus verklaren. Vervolgens zijn in de eerste stap de leerling-covariaten ingebracht (model 1) en in de tweede stap de covariaten op schoolniveau (model 2). In de laatste stap zijn tenslotte de verklarende variabelen (model 3) ingebracht. Bij de analyse zijn daarbij bij elke hogere stap die variabelen opgenomen die in de voorgaande stap een significante bijdrage leverden aan het verklaren van leerprestaties. Het significantieniveau van de te hanteren significantietoetsen is gezien het relatief geringe aantal scholen (n=39) gesteld op 10%. Daarbij is sprake van een eenzijdige toetsing. Voor de leerlingvariabelen is een significantieniveau van 5% aangehouden. Voor een beter begrip van de resultaten is nog van belang dat de variabelen tot z-scores (gemiddelde nul, variantie een) zijn getransformeerd.

Ter verduidelijking moet hier opgemerkt worden dat bij deze analyse gebruik is gemaakt van de zes samengestelde variabelen en niet van de afzonderlijke indicatoren binnen ieder concept. Dit om de kans op kanskapitalisatie te reduceren. Deze variabelen zijn samengesteld op basis van een LISREL-analyse waarover hier niet verder gerapporteerd wordt. Wel kan aan tabel 1 (onder het kopje parameterwaarden) ontleend worden welke indicator de beste voorspeller is van elk latent concept. Daarbij kan vermeld worden dat alle bijdragen significant zijn op een 5% niveau en dat het model niet essentieel verbeterd kon worden door indicatoren te laden op andere concepten dan waar zij oorspronkelijk onder gedacht zijn.

RESULTATEN

Beschrijvende gegevens

Zoals eerder is opgemerkt, zal, voordat de resultaten met betrekking tot de centrale vraagstelling ter sprake komen, eerst worden ingegaan op de descriptieve resultaten.

De resultaten van het beschrijvende gedeelte van het onderzoek kunnen gedeeltelijk ontleend worden aan de gemiddelden zoals die naar voren komen in de eerder gepresenteerde tabel 1. In deze tabel is een belangrijk gegeven dat de homogeniteit in opvattingen tussen docenten aanzienlijk is. Zo is de gemiddelde spreiding tussen docenten per sectie op de veertien items die de mate van homogeniteit in opvattingen over de taken en functie van de school representeren slechts 0.27. Ook uit de andere schalen die op dit concept betrekking hebben komt eenzelfde beeld naar voren. Als zodanig geven deze resultaten aanleiding om te veronderstellen dat verschillen in opvattingen en gedrag binnen een sectie niet erg groot zijn. Daarnaast blijkt uit deze tabel dat vaksecties wiskunde regelmatig overleggen (circa 14 uur op jaarbasis) en dat secties per leerjaar gemiddeld 6 gemeenschappelijke proefwerken geven.

In Tabel 2 worden nadere gegevens die zijn voortgekomen uit het beschrijvend gedeelte van het onderzoek weergegeven. Deze tabel geeft de mate waarin verschillende actoren binnen de schoolorganisatie invloed hebben op een aantal belangrijke onderwijskundige beslispunten weer. De resultaten weergegeven in deze tabel kunnen als typerend beschouwd worden voor de wijze waarop het vakonderwijs binnen schoolorganisaties gecoördineerd wordt.

Tabel 2. De invloed van verschillende actoren binnen de schoolorganisatie op het (vak)onderwijs (gemiddelden; n = 169).

	school leiding	vaksectie-leider	vaksectie	docent	overige personen en organen
uitgangspunten onderwijs	2.55	1.40	2.19	2.33	1.86
keuze leergang	1.13	1.81	3.77	1.99	1.04
vaststellen leerdoelen	1.51	1.70	3.12	3.24	1.27
leerstofkeuze	1.09	1.69	3.36	2.61	1.05
wijze van lesgeven	1.20	1.41	2.24	3.60	1.11
pedagogische aanpak	1.50	1.33	1.80	3.68	1.27
inhoud proefwerken	1.00	1.35	2.77	3.30	1.04
aantal proefwerken	1.59	1.42	2.70	3.08	1.17
normering proefwerken	1.03	1.32	2.63	3.20	1.03
rapportcijfer	1.40	1.35	2.57	3.16	1.27
inhoud schoolonderzoek	1.24	1.49	3.01	2.72	1.04
hoeveelheid huiswerk	1.21	1.16	1.74	3.79	1.12
vergaderfrequentie	1.63	2.54	2.80	1.71	1.15

Schaalbreedte loopt van 1 (niet of nauwelijks van invloed) tot 4 (zeer grote invloed).

Uit Tabel 2 blijkt dat vaksecties wiskunde vooral belangrijke organen zijn bij de keuze van de leergang, voor het vastleggen van de leerdoelen en de inhoud van het onderwijs en voor de toetsing. Uit de tabel blijkt verder dat de invloed van de schoolleiding en vaksectieleider op de meeste punten te verwaarlozen is en dat de invloed van de individuele docent zich vooral beperkt tot het pedagogisch-didactische terrein. Met andere woorden, de docent beslist in belangrijke mate op welke wijze het onderwijs wordt vormgegeven. Daarnaast kan uit deze tabel afgeleid worden dat de invloed van andere personen en organen binnen de school op het (vak)onderwijs zeer gering is.

Vooruitlopend op de conclusies betekent dit een tweetal zaken. Enerzijds dat docenten weliswaar een zekere autonomie hebben, maar dat deze gestalte krijgt binnen een door de vaksectie bepaald kader. Anderzijds dat de rol van leidinggevenden op het onderwijskundig terrein als gering ingeschat moet worden, al kan dit genuanceerd worden door te wijzen op de invloed die de schoolleiding heeft op de uitgangspunten van het wiskunde-onderwijs. Deze conclusies worden ondersteund door de overige resultaten van het onderzoek. De resultaten met betrekking tot regelgeving binnen secties geven aan dat secties wiskunde krachtig het handelen van hun docenten reguleren op punten als de leerdoelen, de inhoud van het onderwijs en de wijze en aard van toetsing. Uit de gegevens over de aard van het overleg tussen docenten blijkt daarnaast dat dit overleg veelal beperkt blijft tot aspecten als de planning van het onderwijs, de realisatie van de planning, de wijze en inhoud van toetsing, het tempo van lesgeven en de inhoud van het onderwijs, terwijl weinig overleg plaatsvindt over aspecten als het didactisch functioneren van docenten en de problemen die docenten in de dagelijkse lespraktijk ondervinden. De resultaten ten aanzien van de rol van de schoolleiding en de vaksectieleider tenslotte wijzen uit dat deze rollen onderwijskundig gezien weinig voorstellen. Zo geven de resultaten aan dat schoolleiders weinig overleggen met (docenten van) vaksecties wiskunde, dat er weinig regels op schoolniveau zijn die het gedrag van docenten aan banden leggen en dat de vaksectieleider door docenten niet of nauwelijks beschouwd wordt als een onderwijskundig leider van de vaksectie.

Effectiviteit

Bij de effectiviteitsvraag is zoals gezegd gebruik gemaakt van het programma VARCL. Daarbij zijn eerst de proporties variantie op leerling- en schoolniveau vastgesteld door een 'leeg model' te specificeren, een model zonder leerling en schoolkenmerken. Vervolgens zijn de covariaten

Tabel 3. Regressie-coëfficienten (en standard errors) voor de regressie van leerlingprestaties op leerling- en schoolvariabelen (aantal leerlingen = 975; aantal scholen = 39).

	model 0	model 1	model 2	model 3
Algemeen gemiddelde	-0.01 (0.07)	-0.01 (0.07)	-0.01 (0.07)	0.01 (0.07)
LEERLINGNIVEAU				
sekse		**-0.36 (0.06)**	**-0.36 (0.06)**	**-0.36 (0.06)**
doublurestatus		0.02 (0.06)		
opleiding vader		0.02 (0.02)		
opleiding moeder		-0.01 (0.02)		
etniciteit		**-0.45 (0.22)**	**-0.45 (0.22)**	**-0.43 (0.22)**
SCHOOLNIVEAU				
denominatie			-0.13 (0.19)	
schooltype			**0.30 (0.15)**	**0.38 (0.15)**
structuur			0.06 (0.15)	
ervaring			**0.11 (0.07)**	0.07 (0.07)
prestatiegerichtheid			**0.09 (0.07)**	**0.12 (0.07)**
VERKLARENDE KENMERKEN				
overleg en samenwerking				**-0.18 (0.10)**
collegiale besluitvorming				0.02 (0.10)
consensus				0.04 (0.07)
beleid				**0.21 (0.10)**
schoolleiding				**0.18 (0.08)**
sectieleider				**-0.10 (0.07)**

Vetgedrukt: significante relaties (leerlingkenmerken op 5% (eenzijdig), schoolkenmerken op 10% niveau (eenzijdig).

op leerlingniveau en schoolniveau ingebracht en tenslotte de verklarende variabelen. De resultaten staan vermeld in Tabel 3. De bespreking van de modellen zal daarbij beginnen met de laatste drie modellen.

Uit de eerste stap (model 1) waarin de covariaten op leerlingniveau in het model ingebracht worden blijkt dat sekse en etniciteit verreweg de belangrijkste voorspellers zijn van wiskunde-prestaties. Geen rol van betekenis spelen het opleidingsniveau van de vader en dat van de moeder. Deze resultaten zijn consistent met de gevonden resultaten van het onderzoek 'Uitgangssituatiemeting eerste fase voortgezet onderwijs' – het onderzoek waaruit de test- en achtergrondgevens zoals gezegd afkomstig zijn – en andere bevindingen op het terrein van wiskunde-prestaties. Uit deze onderzoeken blijkt veelal dat meisjes en allochtone leerlingen significant lager scoren op wiskunde-toetsen, terwijl variabelen die betrekking hebben op het opleidingsniveau een veel mindere rol spelen bij het verklaren van deze prestaties (cf. Bosker, 1991). De invloed van het opleidingsniveau van de ouders verdwijnt zelfs als gekeken wordt binnen een schooltype (cf. Kremers, 1990). Aangezien er in dit onderzoek sprake is van een tweetal schooltypen die erg dicht bij elkaar liggen, kan inderdaad verwacht worden dat de invloed van de opleiding van de ouders gering is.

Uit de tweede stap (model 2) waarin de covariaten op schoolniveau ingebracht worden, blijkt dat veel van deze variabelen een bijdrage leveren aan de leerprestaties. Geen significante bijdrage leveren de denominatie van de school en de schoolstructuur. Ook deze resultaten komen grotendeels overeen met die uit ander onderzoek. Een uitzondering hierop vormen de resultaten met betrekking tot de variabele denominatie van de school. In veel onderzoeken speelt deze variabele wel een belangrijke rol (cf. Scheerens, 1989). De grote bijdrage van de variabele

Tabel 4. Variantiecomponenten.

	model 0	model 1	model 2	model 3
VARIANTIECOMPONENTEN				
leerlingniveau	0.80	0.75	0.75	0.75
schoolniveau	0.20	0.20	0.18	0.12
verklaarde variantie	n.v.t.	5%	2%	6%

schooltype, waarbij HAVO-scholen een substantiële bijdrage leveren aan het prestatieniveau, kan verklaard worden door het gegeven dat deze variabele in feite een weerspiegeling vormt van de wijze waarop de leerlingpopulatie van de desbetreffende schooltypen is samengesteld. Naarmate het schooltype hoger is, zal deze bevolkt worden door leerlingen met een hoger IQ. In dit opzicht kan de bijdrage van deze factor ongetwijfeld juist door dit gegeven verklaard worden. Daarbij moet wel vermeld worden dat hier in zekere zin sprake is van speculatie; er waren immers geen precieze gegevens over bijvoorbeeld het IQ van leerlingen beschikbaar.

Uit de laatste stap (model 3) waarin de verklarende variabelen zijn ingebracht, blijkt dat er een aantal significante relaties te onderkennen zijn tussen de verklarende variabelen en de afhankelijke variabele. Een positieve samenhang is er te onderkennen tussen de mate van beleid binnen vaksecties en leerprestaties, terwijl hetzelfde geldt voor de relatie tussen de mate waarin de schoolleiding actief is op het terrein van het onderwijs en de afhankelijke variabele. Negatieve relaties zijn er daarentegen te onderkennen tussen respectievelijk de mate waarin de vaksectieleider als teamleider optreedt en de mate van overleg en samenwerking binnen vaksecties en leerprestaties. De overige verklarende variabelen hebben geen significante relatie met de afhankelijke variabele. De bespreking van dit laatste model wordt besloten met de opmerking dat de bijdragen van de significante variabelen niet anders worden als de niet-significante variabelen uit het model worden weggelaten. In de volgende paragraaf zal worden ingegaan op de geldigheid en betekenis van deze resultaten.

De belangrijkste informatie over de door de verschillende kenmerken verklaarde variantie kan afgelezen worden in Tabel 4. Daarbij gaat het steeds om de hoeveelheid variantie verklaard door die variabelen die blijkens model 3 een significante relatie hebben met de afhankelijke variabele.

Een belangrijk gegeven in deze tabel (onder model 0) is dat 80% van de totale variantie gelokaliseerd kan worden op het niveau van de leerling en 20% op het niveau van de school. Dit laatste percentage kan als opvallend groot aangemerkt worden, daar in het algemeen dit percentage in de Nederlandse situatie op 11 à 12% wordt geschat (cf. Scheerens, 1989). Uit de gegevens onder model 1 kan afgeleid worden dat slechts 5% van de totale variantie door leerlingkenmerken verklaard wordt. Deze geringe bijdrage kan grotendeels toegeschreven worden aan het feit dat – zoals boven reeds naar voren is gekomen – geen kenmerken van leerlingen die op een directe wijze verband houden met het vaardigheidsniveau van leerlingen – zoals het IQ van de leerling – in de analyses opgenomen zijn. Uit de gegevens onder model 2 blijkt dat de ingebrachte controle-variabelen op schoolniveau 2% van de totale variantie verklaren, terwijl uit het verschil tussen model 2 en model 3 afgeleid kan worden dat de significante 'coördinatie'-variabelen 6% van de totale variantie verklaren.

CONCLUSIES EN DISCUSSIE

Zoals eerder naar voren is gekomen, staat in dit artikel een tweetal vragen centraal. De eerste is de vraag of scholen voor voortgezet onderwijs al dan niet geïntegreerde organisaties zijn, waarbij met name het functioneren van vaksecties centraal staat. De tweede vraag is of kenmer-

ken die hieraan gerelateerd zijn een relatie hebben met de effectiviteit van de schoolorganisatie. In het navolgende zal begonnen worden met de beantwoording van de eerste vraag.

Afgaande op de resultaten is een belangrijke conclusie dat scholen met betrekking tot het curriculaire domein veel hechtere organisaties zijn dan vaak wordt verondersteld. Vooral de gegevens over de wijze waarop vaksecties wiskunde functioneren wijzen op dit feit. Deze vaksecties zijn veelal sterk ontwikkelde organen. Uiteraard liggen hier consequenties voor de wijze waarop scholen getypeerd dienen te worden. In algemene zin kan gesteld worden dat er meer wordt overlegd, er meer afspraken worden gemaakt, de positie van de docent minder autonoom is en de onderwijsopvattingen homogener zijn dan op basis van de in de inleiding beschreven schoolorganisatiemodellen verwacht mag worden. Bij veel van de veronderstellingen die in deze modellen vervat zitten, kunnen dan ook op basis van de hier gevonden resultaten vraagtekens gezet worden. Meer specifiek valt te wijzen op de relatief grote rol van coördinatiemechanismen als standaardisatie van de inhoud en de doelen van het werk. De resultaten geven aan dat bij vaksecties de regelgeving, het (informele) overleg en de besluitvorming zich vooral centreren rondom aspecten die op deze standaardisatie betrekking hebben. Een voorbeeld is de aard van de regelgeving binnen secties. Vaksecties reguleren in hoge mate het gedrag van docenten ten aanzien van de leerdoelen die bereikt moeten worden, de te geven leerstof en de wijze en aard van toetsing. In dit opzicht is het niet verwonderlijk dat in onderzoek naar de consistentie naar schooleffecten in het voortgezet onderwijs gevonden wordt dat de leeropbrengsten van parallelklassen binnen eenzelfde school grotendeels gelijk zijn (cf. Bosker, Kremers & Lugthart, 1991). Vanuit de resultaten van het hier gepresenteerde onderzoek is het aannemelijk dat binnen veel scholen docenten toewerken naar een gezamenlijk afgesproken niveau. Voorzover deze modellen wel een juist beeld geven van het functioneren van scholen moet dit vooral gezocht worden in de wijze waarop de rol van de schoolleiding getypeerd wordt. De resultaten geven aan dat de rol van de schoolleiding binnen het curriculaire domein niet erg sterk ontwikkeld is.

Een belangrijke vraag aangaande het bovenstaande is of de resultaten een waarheidsgetrouw beeld geven van de werkelijke situatie binnen scholen. Eerder is gesteld dat het relatief kleine aantal docenten dat betrokken is geweest bij het vaststellen van de sectie- en organisatiekenmerken een probleem is. Het meest relevant is dit wellicht voor de schalen die betrekking hebben op het concept consensus. Dat door het geringe aantal deelnemende docenten een vertekening kan zijn opgetreden is duidelijk. Een ander mogelijk probleem vloeit voort uit de selectie van de betrokken docenten. Zij vormen steeds een homogene groep in die zin dat het steeds gaat om docenten betrokken bij het MAVO-onderwijs of bij het HAVO-onderwijs. Dit kan betekenen dat verschillen in visie tussen beide groepen onderbelicht zijn gebleven. Daar staat tegenover dat in een verwant onderzoek (Hylkema, 1990) op dit punt met betrekking tot vaksecties wiskunde tot dezelfde conclusie gekomen wordt. Hetzelfde geldt ook ten aanzien van de overige door hem onderzochte kenmerken van deze vaksecties. Er lijkt dus van een bijzondere groep secties in dit onderzoek geen sprake te zijn. Een ander vraag is of het functioneren van vaksecties wiskunde zoals dit in dit onderzoek naar voren komt representatief is voor secties in andere vakgebieden. Als gekeken wordt naar het onderzoek van Hylkema (1990) blijkt dat secties wiskunde meer coördineren dan andere secties. Uit een vergelijking van secties wiskunde met de overige bij dit onderzoek betrokken secties (zoals gezegd secties Engels en geschiedenis) blijkt daarentegen dat secties wiskunde een middenpositie innemen (cf. Witziers, 1992). Deze vraag is dus moeilijk te beantwoorden, temeer daar bij verschillen tussen secties niet alleen de aard van het vakgebied van belang is, maar ook de school waartoe de vaksectie behoort (cf. Witziers, 1992). Tenslotte kan ook gewezen worden op de overeenkomst met de resultaten van het onderzoek naar coördinatie binnen scholen voor voortgezet onderwijs van Hofman en Lugthart (1991), met name als het gaat om aspecten als onderwijskundig leiderschap, overleg en consensus. Voorzover er wel verschillen zijn te constateren, kan dit toegeschreven worden aan het feit dat de wijze van operationalisering van kenmerken in beide onderzoeken wezenlijk verschilt.

Wat betreft de tweede vraag – die naar de relatie tussen coördinatie en leerlingprestaties – geven

de resultaten aan dat er aan ene kant kenmerken van de schoolorganisatie (vaksectieleiderschap en overleg en samenwerking) zijn die een negatief verband vertonen met leerlingprestaties en aan de andere kant kenmerken (mate van schoolleiderschap en mate van beleid binnen secties) die juist een positieve relatie hiermee hebben. Tenslotte zijn er variabelen die geen invloed hebben. Deze diversiteit maakt interpretatie geen gemakkelijke zaak. Nu kan in het algemeen gezegd worden dat de interpretatie van gegevens bij effectiviteitsonderzoek vaak moeilijk is. Problemen die daarbij spelen zijn onder andere de aard van de gebruikte toets (cf. Madaus, 1979), het vak waarop de test van toepassing is (cf. Bosker, 1990), de omvang van de steekproeven (cf. Purkey & Smith, 1984) en het gebruik van een ex-post facto design (cf. Scheerens, 1989). Duidelijk is dat sommige van deze problemen ook voor dit onderzoek relevant zijn. De belangrijkste hiervan zijn wel de problemen die kleven aan de beperkte omvang van de steekproef en aan de aard van het gebruikte onderzoeksopzet die het onmogelijk maakt om eenduidig causale relaties vast te stellen.

Hoezeer deze laatste problematiek relevant is voor dit onderzoek blijkt uit de gesignaleerde negatieve relaties. Deze resultaten zijn duidelijk in tegenspraak met regelmatig naar voren gebrachte kenmerken van effectieve scholen als overleg en samenwerking en onderwijskundig leiderschap in de vorm van 'teacher leaders' (Hallinger, 1990; Smith & Scott, 1990; Wilson & Corcoran, 1988). Een verklaring voor de negatieve uitwerking van overleg en samenwerking binnen vaksecties ligt mogelijk in de eerder naar voren gebrachte stelling van Creemers en Scheerens (1991) dat een teveel aan overleg en samenwerking schadelijk kan uitwerken op de effectiviteit van de schoolorganisatie. Gezien het vele onderzoek dat aangeeft dat docenten veeleer hun prioriteiten leggen bij hun onderwijs (cf. Lortie, 1975; Prick, 1983), waardoor overleg en samenwerking zelden excessieve vormen zullen aannemen, en de resultaten van dit onderzoek deze stelling min of meer onderschrijven, is het echter de vraag of hier een verklaringsgrond ligt. Waarschijnlijker is ons inziens de verklaring dat niet zozeer overleg en samenwerking leiden tot 'slechte' leerlingprestaties als wel dat van een omgekeerde relatie sprake is. Eenzelfde redenering kan daarbij opgesteld worden ten aanzien van de vaksectieleider. Ook hier is het waarschijnlijk dat niet zozeer het optreden van de vaksectieleider verantwoordelijk is voor de 'slechte' leerprestaties, als wel dat een verklaring gezocht moet worden in tegenovergestelde richting. De plausibiliteit van deze redenering kan daarbij onderstreept worden door te wijzen op het eerder naar voren gebrachte feit dat in Nederlands onderzoek op dit terrein regelmatig relaties heel anders uitpakken dan verwacht. Duidelijk is overigens wel dat dergelijke bevindingen het belang van onderzoeksdesigns die meer eenduidig causale uitspraken mogelijk maken nog eens onderstrepen (cf. Scheerens, 1989).

Wat betreft de kenmerken die geen relatie hebben met de effectiviteit van de organisaties is een tweetal conclusies mogelijk. De eerste is dat deze aspecten niet valide gemeten zijn en daarom geen verband hebben met de afhankelijke variabele. Levine en Lezotte (1990) menen dat het meten van aspecten als overleg en samenwerking, consensus en participatie in de besluitvorming met schriftelijke vragenlijsten – zoals in dit onderzoek gebeurd is – te problematisch is om tot een valide bepaling te kunnen komen. Dit is uiteraard mogelijk, maar daar staat echter tegenover dat tussen alle onderzochte kenmerken – zoals verwacht – onderling redelijk hoge correlaties bestaan. Dit betekent dat niet per se aan de constructvaliditeit van deze variabelen getwijfeld hoeft te worden (cf. Swanborn, 1987). In deze zin ligt dan ook eerder de tweede conclusie voor de hand, namelijk dat deze variabelen voor de effectiviteit van de schoolorganisatie geen belangrijke rol spelen. Relevant in deze zijn de opmerkingen van Little (1990) over verschillende vormen van samenwerking. Zij stelt dat er tal van vormen van samenwerking zijn die niet per definitie gunstig uitwerken op de kwaliteit van het onderwijs. Veel van deze vormen omschrijft zij als oppervlakkig en met weinig gevolgen voor de dagelijkse praktijk van docenten. Wat volgens haar wel positief zou werken is een samenwerkingscultuur waarbij docenten een gemeenschappelijke verantwoordelijkheid hebben voor het onderwijs en daarbij duidelijke prioriteiten stellen, zonder daarmee overigens de autonomie van docenten te zeer aan banden te leggen. Toegepast op dit onderzoek kan deze redenering de afwezige relatie van kenmerken als consensus en (participatie in de) besluitvorming met de afhankelijke variabele verklaren.

Verder bewijs voor de stelling van Little kan ontleend worden aan de positieve relatie tussen de mate van beleidsvorming binnen vaksecties en de variabele leerprestaties. Juist deze variabele kan bij uitstek gezien worden als een operationalisering van het bovengenoemde door Little beschreven kenmerk van effectieve schoolorganisaties. Van zaken als afspraken ten aanzien van het onderwijs en gemeenschappelijke proefwerken kan immers gezegd worden dat zij een wezenlijke invloed hebben op het gedrag van docenten, met name vanwege de controlerende werking die hiervan uitgaat. Zo impliceren gemeenschappelijke proefwerken dat er standaarden zijn met betrekking tot de prioriteiten die gelden in het onderwijs en het niveau waarop het onderwijs gegeven moet worden. Dit zal betekenen dat factoren die vaak gezien worden als nadelig voor de effectiviteit van de schoolorganisatie, zoals het nastreven van persoonlijke doeleinden (cf. Boyd & Crowson, 1985) en een grote invloed van leerlingen op het aspiratieniveau van docenten (cf. Rosenholtz, 1985), geen grote rol spelen. Daarnaast kan, aangezien het bij het concept beleid vooral gaat om duidelijke leerstofafspraken (zie boven) en veelvuldig toetsen (gemeenschappelijke proefwerken), geconcludeerd worden, dat ook in scholen voor voortgezet onderwijs de kwaliteit van het onderwijs gebaat is bij veelvuldig in de literatuur genoemde aspecten als:

- een consistente opbouw van het onderwijs door de school over de leerjaren heen, waarbij de onderwijsdoelen van de diverse leerjaren naadloos op elkaar aansluiten;
- het vastleggen van de accenten en prioriteiten die de school/sectie stelt in het onderwijs en de wijze waarop men die wil bereiken;
- een beleid gericht op evaluatie van de prestaties van leerlingen;
- het verzorgen van feedback en correctief onderwijs op basis van deze evaluatie.

De resultaten met betrekking tot de rol van de schoolleiding tenslotte zijn consistent met de noties die veelal naar voren worden gebracht in het kader van de rol van de schoolleider in effectieve organisaties (cf. Leighwood & Montgomery, 1982; Smith & Andrews, 1989). In dit onderzoek is de mate van onderwijskundig leiderschap een belangrijk verschil tussen effectieve en niet-effectieve scholen. Hoewel vanuit dit onderzoek moeilijk precies te achterhalen valt welke kenmerken van schoolleiderschap nu daadwerkelijk van belang zijn, kan op basis van ander onderzoek (Van Vilsteren & Witziers, 1990) wel vermoed worden dat het daarbij vooral gaat om aspecten die onderwijsbeleid binnen de school mogelijk maken en om aspecten die er voor zorg dragen dat de samenwerking binnen secties inderdaad goed verloopt. Te denken valt hier bijvoorbeeld aan het nemen van roostertechnische maatregelen die gemeenschappelijke proefwerken mogelijk maken, het stimuleren van overleg en samenwerking binnen vaksecties en een zekere controle -bijvoorbeeld in de vorm van functioneringsgesprekken- op het werk van vaksecties. Dat voor dit laatste ook een zeker bewijs te vinden is in dit onderzoek, kan opgemaakt worden uit het feit dat het concept schoolleiding in hoge mate geïndiceerd wordt door de variabele controle.

Als overstijgende conclusie kan geformuleerd worden dat de effectiviteit van de schoolorganisatie inderdaad gebaat is bij coördinatie. Deze conclusie kan nader gespecificeerd worden door te stellen dat een effectieve school eerder gebaat is bij meer 'bureaucratische' (beleid, actieve rol van de schoolleiding met betrekking tot het onderwijs) dan bij meer 'organische' vormen van coördinatie (besluitvorming, consensus, overleg en samenwerking). Dat wil ons inziens overigens niet zeggen dat deze laatste aspecten niet van belang zouden zijn voor het functioneren van schoolorganisaties. Beleidsvorming ten aanzien van het onderwijs is immers niet of nauwelijks voorstelbaar zonder aspecten als overleg en een zekere overeenstemming over de doelen en inhouden van het onderwijs. Desondanks kan op basis van dit onderzoek geconcludeerd worden dat deze aspecten eerder als conditioneel gezien moeten worden voor een effectieve schoolorganisatie dan dat ze zelf daadwerkelijk van belang zijn voor het effectief functioneren van de schoolorganisatie. Evident daarbij is dat, gezien de eerder genoemde methodologische bezwaren, deze conclusie een hypothetisch karakter heeft, alhoewel een zekere steun hiervoor wel te ontlenen valt aan de resultaten van ander onderzoek op dit terrein (cf. Hofman & Lugthart, 1991; Hüpscher-Post, 1990).

Gegeven dit hypothetische karakter ligt het daarom voor de hand de in dit onderzoek gevonden aanwijzingen nader te onderzoeken, bijvoorbeeld op andere vakgebieden. Daarbij kan met name voortgebouwd worden op de gevonden resultaten over het functioneren van vaksecties. De resultaten vormen op zijn zwakst gezegd een aanwijzing dat secties voor de kwaliteit van het onderwijs van belang zijn. Aannemelijk is vaksecties ten aanzien van veel aspecten die als wenselijk gezien worden voor de kwaliteit van het onderwijs (zoals overleg over en evaluatie van het onderwijs, de ontwikkeling van onderwijsbeleid) een belangrijke rol spelen. Bovendien moet de rol van deze organen voor docenten ook niet onderschat worden. Immers in gezamenlijkheid wordt veelal beslist en overlegd over het te geven onderwijs. Veelal zal dit betekenen dat het werk van docenten beïnvloed wordt door het team waarvan zij deel uitmaken. In dit opzicht ligt het in de rede belangrijke thema's die spelen in het debat over effectiviteit van scholen – bijvoorbeeld de relatie tussen organisatiekenmerken enerzijds en effectief instructiegedrag en leerprestaties anderzijds – (mede) te situeren op het niveau van de vaksectie.

LITERATUUR

Aitkin, M., & Longford, N. (1986). Statistics Modelling Issues in School Effectiveness Studies. *The Journal of the Royal Statistical Society, 149*, 1-43.

Ashton, P.T. & Webb, R.B. (1986). *Making a difference: Teachers' Sense of Efficacy and Student Achievement*. London: Longman.

Astuto, T., & Clarke, D. (1985). Strength of organizational coupling in the instructionally effective school. *Urban Education, 19*, 331-355.

Bosker, R.J. (1990). *Extra Kansen dankzij de school*? (dissertatie) Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. Tijdschrift voor Onderwijsresearch, 16, 193-205

Bosker, R.J., Kremers, E, & Lugthart, E. (1991). De consistentie van schooleffecten in het voortgezet onderwijs. *Tijdschrijft voor Onderwijsresearch, 16*, 193-205.

Boyd, W.L., & Crowson, R.L. (1985). The changing conception and practice of public school administration. *Review of Research in Education, 9*, 311-373.

Brandsma, H.P., & Stoel, W.G.R. (1986). *Schoolleiders in het voortgezet onderwijs. Een exploratieve studie naar determinanten en effecten van schooleiderskenmerken*. Groningen: RION.

Creemers, B.P.M., & Scheerens J., (1991). Onderwijseffectiviteit: overwegingen voor een programma van onderzoek. *Tijdschrijft voor Onderwijsresearch, 16*, 193-205.

Cuttance, P. (1987). *Modelling variation in effectiveness of schooling*. Edinburgh: Centre for Educational Sociology.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 35*, 15-24.

Firestone, W.A., & Wilson, B.L. (1985). Using bureaucratic and cultural images to improve instruction: The principals' contribution. *Educational Administration Quarterly, 21*, 7-30.

Goldstein, H. (1987). *Multilevel models in educational and social research* London: Charles Griffin & Co.

Grift, W. van der & Akkermans, W. (1991). Onderwijskundig leiderschap en leerlingprestaties in het basisonderwijs. *Tijdschrijft voor Onderwijsresearch, 16*, 244-256.

Hallinger, Ph. (1989). Developing Instructional Leadership Teams in High Schools. In B.P.M. Creemers, T. Peters, & D. Reynolds (Red.), *School Effectiveness and School Improvement* (pp. 319- 330). Amsterdam/Lisse: Swets & Zeitlinger.

Hofman, R.H. & Lugthart, E. (1991). *Interne capaciteit in het voortgezet onderwijs*. Groningen: RION.

Hüpscher-Post, A.D. (1990). *De ene school is de ander niet. Allochtone meisjes in het voortgezet onderwijs*. De Lier: Academisch Boekencentrum ABC.

Hylkema, W.F.S. (1990). *Docenten en hun vaksecties* (dissertatie). Nijmegen: Katholieke Universiteit Nijmegen.

Kieser, A., & Kubicek, H. (1983). *Organisation*. Berlijn/New York: Walter de Gruyter.

Kremers, E.J.J. (Red.). (1990). *Overzicht van leerresultaten aan het eind van de eerste fase voorgezet onderwijs*. Arnhem: CITO.

Leighwood, K.A. & Montgomery, D.J. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research, 52*, 309-339.

Levine, D.K., Lezotte, L.W. (1990). *Unusually efffective schools: a review and analysis of research and practice*. Madison: National Center for Effective Schools Research and Development.

Little, J.W. (1982). Norms of Collegiality and Experimentation: Workplace conditions of School Succes. *American Educational Research Journal, 19*, 325-340.
Little, J.W. (1990). The Persistence of Privacy: Autonomy and Initiative in Teachers' Professional Relations. *Teacher Colleger Record, 91*, 509-536.
Longford, N.T. (1986). *Variance component analysis: manual*. Lancaster: University of Lancaster.
Lortie, D.C. (1975). *School teacher, a sociological study*. Chicago: University of Chicago Press.
Madaus, G.F., Kellaghan, Th., Rakow, E.A., & King, D.J. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational review, 49*, 207-230.
Marx, E.C.H. (1975). *De organisatie van scholengemeenschappen in onderwijskundige optiek*. Groningen: Wolters-Noordhoff.
Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs: Prentice Hall.
Pelkmans, A. & Vrieze, G. (1987). *Meer over coördinatie en management in het voortgezet onderwijs*. Nijmegen:ITS.
Prick, L.G.M. (1983). *Het beroep van leraar: satisfactie en crises in de leraarsloopbaan*. (dissertatie). Amsterdam: Vrije Universiteit.
Rosenholtz, S.J. (1985). Effective schools: Interpreting the evidence. *American Journal of Education, 93*, 352-388.
Rosenholtz, S.J. (1989). *Teachers workplace: the social organization of schools*. New York: Longman.
Rowan, B., Raudenbush, S.W., & Kang, S.J. (1989). *Schoolclimate in secondary schools: A multi-level analysis*. Paper presented to the ESRC International Conference on Multilevel Methods in Educational Research, Edinburgh, August 1989.
Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*. 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.
Siskin, L.S. (1991). Departments as different worlds: Subject subcultures in secondary schools. *Educational Administration Quarterly, 17*, 134-160.
Sleegers, P.J.C., Bergen, Th.C.M. & Giesbers, J.H.G.I. (1992).
School en beleidsvoering, *Pedagogische Studiën, 69*, 162-176.
Smith, S.C., & Scott, J. (1990). *The collaborative school: A work environment for effective instruction*. Eugene: University of Oregon.
Smith, W.F., & Andrews, R.L. (1989). *Instructional leaderschip: How principals make a difference*. Alexandria, V.A.:Association for Supervision and Curriculum Development.
Stoel, W.G.R. (1986). *Schoolkenmerken en het gedrag van leerlingen en docenten in het voortgezet onderwijs*. Groningen: RION.
Swanborn, P.G. (1987). *Methoden van sociaal wetenschappelijk onderzoek*. Meppel: Boom.
Tesser, P. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen: I.T.S.
Tyler, W. (1988). *Schoolorganisation: A sociological perspective*. London: Croom Helm.
Vilsteren, C.A. van (1986). Management van onderwijsleersituatie-overstijgende aktiviteiten in scholen. In F.J. van der Krogt (Red.), *Schoolleiding en management* (pp. 99-109). Lisse: Swets & Zeitlinger.
Vilsteren, C.A. van, Witziers, B. (1990). Vaksecties en het onderwijskunidg leiderschap van schoolleidiers. In J.Imants & W.Weijzen (Red.), *Organisatie van Onderwijsinstellingen* (pp. 61-70). Amsterdam/ Lisse: Swets & Zeitlinger.
Vonk, J.H.C. (1982). *Beginnend leraarschap: het professionele ontwikkelingsproces van beginnende docenten*. (dissertatie). Amsterdam: Vrije Universiteit.
Weick, K.E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*, 1-19.
Wilson, B.L., & Corcoran, T.B. (1988). *Successfull secondary schools: Visions of excellence in American public education*. London (etc): The Falmer Press.
Witziers, B. (1992). *Coördinatie binnen scholen voor voortgezet onderwijs*. (dissertatie), Enschede.

Manuscript ontvangen 26-4-1993
Definitieve versie ontvangen 3-11-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 5, pp. 295-305

Aanpak van kwalificatieproblemen in organisaties: een onderzoek naar strategieën van opleidingskundigen

F.J. van der Krogt, A.A. Vermulst en I.E.C. ter Woerds
*Vakgroep Algemene Pedagogiek, Katholieke Universiteit Nijmegen**

ABSTRACT

Training managers are involved with qualification problems of individuals in organizations increasingly. Qualifications problems arise when discrepancies exist between individuals' qualifications and the required qualifications for tasks and functions.
Which strategies are used by training managers to solve these problems? On theoretical grounds, three strategies are distinguished. With help of empirical data obtained from fifty-nine training managers the theoretical strategies are further explored.

INLEIDING

Met de toename van bedrijfsopleidingen zijn steeds meer opleidingskundigen werkzaam binnen arbeidsorganisaties en bij, meer of minder commerciële, opleidingsbureau's. Een belangrijk onderdeel van hun takenpakket is het (doen) ontwikkelen van cursussen, trainingen en andere leersituaties. De aanleidingen daarvoor liggen in problemen bij het uitvoeren van het werk en in veranderingen van arbeid en functies in organisaties. Dit ontwikkelingswerk kan op vele manieren gebeuren. Vaak worden behoeften- of taakanalyses uitgevoerd die de basis vormen voor het opstellen van een opleidingsprofiel waarna het opleidingsprogramma ontwikkeld wordt. Deze taakgerichte ontwikkelingsstrategie is veelvuldig toegepast (Brandsma, Nijhof & Kamphorst, 1990; Mulder ea, 1988). Maar de verscheidenheid aan problemen en arbeid is groot en we vermoeden dat deze gevolgen heeft voor de ontwikkelingsstrategie. Het ontwikkelen van leersituaties voor werk dat bestaat uit kleine taakonderdelen, taakarbeid zoals bij administratief werk of productiearbeid, zal anders verlopen dan voor professionele arbeid zoals die van artsen of adviseurs. In dit artikel stellen we de vraag aan de orde welke ontwikkelingsstrategie opleidingskundigen hanteren en of ze deze afstemmen op het soort arbeid.

Eerst maken we plausibel dat er verschillende leerontwikkelingsstrategieën te onderscheiden zijn en dat deze wellicht verband houden met het soort arbeid. Daarna wordt het onderzoek besproken waarin leerontwikkelingsstrategieën van opleiders zijn geoperationaliseerd. Er is gezocht naar de dimensies in de strategieen en er is nagegaan of opleiders in verschillende situaties andere strategieen hanteren.

THEORETISCHE ACHTERGROND

Kwalificatiestrategieën

Eén van de voornaamste redenen om in arbeidsorganisaties leersituaties te ontwikkelen is gelegen in aansluitingsproblemen tussen de beschikbare kwalificaties van de medewerkers en de vereiste kwalificaties die nodig zijn om bepaalde taken uit te voeren. Als we kwalificaties opvatten als het geheel van kennis, vaardigheden en houdingen dat iemand in staat stelt ar-

* Adres: Vakgroep Algemene Pedagogiek, Postbus 9103, 6500 HD Nijmegen.

beidsprestaties te verrichten (Van Hoof & Dronkers, 1980; Brandsma, Nijhof & Kamphorst, 1990), dan zijn beschikbare kwalificaties die kwalificaties die bij functievervullers voorhanden zijn, en vereiste kwalificaties die kwalificaties die nodig zijn om de functie te vervullen. De discrepanties tussen beschikbare en vereiste kwalificaties kunnen onder andere ontstaan doordat de veranderingen in het productieproces of in de organisatiestructuur doorwerken in de taken en functies, de vereiste kwalificaties veranderen. Aan de andere kant kan de situatie op de arbeidsmarkt leiden tot het moeten inzetten van onvoldoende gekwalificeerd personeel.

Om de discrepanties te reduceren kan men maatregelen nemen die betrekking hebben op zowel de beschikbare als op de vereiste kwalificaties. In principe zijn er drie kwalificatiestrategieën mogelijk om die aansluitingsproblemen op te lossen (Van der Krogt, 1990, 1991; Warmerdam & Van den Berg, 1992).

Men kan de beschikbare kwalificaties van de medewerkers aanpassen aan de vereiste kwalificaties. Dit gebeurt bijvoorbeeld door personeel met de vereiste kwalificaties te werven of intern over te plaatsen, maar ook door cursussen en trainingen voor het zittend personeel te verzorgen. Dit noemen we de strategie van de aanpassende kwalificering, de kwalificaties van de medewerkers worden aangepast aan de taken die ze moeten vervullen. Bij de tweede kwalificatiestrategie, die van de aanpassende werkorganisatie, wordt het werk zo veranderd dat de vereiste kwalificaties passen bij de beschikbare kwalificaties van de medewerkers. Het werk wordt aangepast aan de kwaliteiten van de zittende medewerkers. De strategie van de wederzijdse aanpassing tenslotte werkt via veranderingen zowel aan de kant van de vereiste als aan de kant van de beschikbare kwalificaties.

Deze drie kwalificatiestrategieën kunnen het best toegelicht worden aan de hand van de strategieën zoals die in verschillende soorten organisaties voorkomen (Mintzberg, 1989; Van der Krogt, 1990, 1991).

De strategie van de aanpassende kwalificering komen we tegen binnen de machinebureaucratie. Dit is een organisatie, zoals verzekeringsmaatschappijen of massaproductiebedrijven, met zwaar bemande stafafdelingen die het werk ontwerpen, plannen en voorbereiden. Zij ontwerpen o.a. de taken en functies van het uitvoerend personeel. Veel van het uitvoerende werk bestaat uit enkele kleine deeltaken. De wijze waarop de taken dienen te worden uitgevoerd, is van te voren door de voorbereidende staf bepaald. Discrepanties tussen de beschikbare en vereiste kwalificaties worden aangepakt met de strategie van de aanpassende kwalificering, dus werving en/of scholing van personeel. Opleiden bestaat vooral uit instrueren en trainen, tijdens cursussen of in de vorm van trainingen op de werkplek. Aanleiding tot het ontwikkelen van dergelijke leersituaties binnen de machinebureaucratie is vooral wanneer blijkt dat het werk niet goed wordt uitgevoerd of wanneer de functies veranderen. Bij de ontwikkeling van de leersituaties spelen staffunctionarissen, opleidingskundigen en managers in de organisatie een belangrijke rol. De uitvoerende medewerkers, voor wie de leeractiviteiten bedoeld zijn, zijn overwegend passief.

De strategie van de aanpassende werkorganisatie kan het best geïllustreerd worden aan de hand van de professionele bureaucratie. In zo'n type organisatie doen professionals het uitvoerende werk (het primaire proces); denk bijvoorbeeld aan algemene ziekenhuizen en andere instellingen voor gezondheidszorg, grote advocatenkantoren, accountantskantoren. Dit werk is wel complex maar is anderzijds toch ook gestandaardiseerd. De kwalificaties hebben de professionals middels (vaak) jarenlange scholing verworven. Leersituaties voor de professionals worden door instanties in het professionele veld (ontwikkel- en opleidingsinstituten gelieerd aan de professie) ontwikkeld en aangeboden. De wetenschappelijke ontwikkelingen in het vakgebied - de kennis en kunde- worden tijdens cursussen, trainingen, conferenties, symposia, workshops gepresenteerd. De professionals nemen aan deze cursussen en trainingen deel om de ontwikkelingen in hun vak bij te houden. Daarnaast vinden er ook leeractiviteiten on the job plaats; professionals oefenen en evalueren, individueel of in mono-disciplinair verband, hun (standaard)aanpak en toetsen nieuw verworven kennis op bruikbaarheid in de eigen werksituatie. De professionals komen dus als het ware met hun nieuw verworven kennis en vaardigheden de organisatie binnen. Om deze te kunnen toepassen wordt er, door de professie en de professio-

nals, op het management van de organisaties druk uitgeoefend hiervoor de voorwaarden te scheppen, en de taken en functies zo te veranderen dat deze passen bij de kwalificaties die de professionals verworven hebben. Het werk wordt dus aangepast aan de kwalificaties, vandaar de term aanpassende werkorganisatie.

De strategie van de wederzijdse aanpassing kunnen we goed illustreren aan het werk in de adhocatische organisatie. In een adhocratie, zoals in ontwikkelingslaboratoria, in onderzoeksinstituten, in adviesbureau's, wordt veel in projectgroepen gewerkt die, naar gelang de opdracht, ad hoc worden samengesteld. Werken in dergelijke projectgroepen is bijna voortdurend leren. Men wordt voor problemen gesteld die nieuwe uitdagingen bevatten, de standaardkennis en kunde van de professionals is niet direct toepasbaar. Het kan zelfs zo zijn dat het probleem zo ingewikkeld is dat het geherdefinieerd moet worden, omdat het met de beschikbare kennis en kunde niet op te lossen is. Tijdens het werken aan het probleem wordt duidelijk hoe er verder gegaan kan worden, de taken worden al doende bepaald en verdeeld. Men kan zich het werken voorstellen als een voortdurend wederzijds aanpassen van probleem en kennis/kunde van de projectgroep. Het werk zelf is in de adhocratie dus een belangrijke leersituatie. Daarnaast zullen de (professionele) medewerkers de ontwikkelingen in hun eigen vakdiscipline bijhouden en formele en informele contacten met collega's onderhouden. Het ontwikkelen van netwerken is heel belangrijk voor deze professionals.

In Figuur 1 zijn de verschillen tussen de drie organisaties en de daar (vermoedelijk) gehanteerde strategieën samengevat.

Leerontwikkelingsstrategieën

Een centraal onderdeel van een kwalificatiestrategie is het verschaffen van leermogelijkheden aan de medewerkers. Voor de medewerkers moeten mogelijkheden bestaan om de benodigde kwalificaties te verwerven. Organisaties bieden uiteenlopende leersituaties aan, die op verschillende manieren tot stand komen. Dit noemen we leerontwikkelingsstrategieën dat wil zeggen: maatregelen en activiteiten in verband met het ontwikkelen en vormgeven van leersituaties die gericht zijn op het reduceren van discrepanties tussen beschikbare en vereiste kwalificaties.

	MACHINE BUREAUCRATIE	PROFESSIONELE BUREAUCRATIE	ADHOCRATIE
SOORTEN ARBEID	taakarbeid . kleine deeltaken . werken volgens voorschriften.	prof. bureaucr. arbeid . ingewikkeld werk . individueel uitvoeren volgens (standaard) methodieken geleerd tijdens opleiding.	adhocratische arbeid . groepen medewerkers werken aan complexe problemen. . tijdens werken worden taken verhelderd en verdeeld tussen medewerkers.
KWALIFICATIE-STRATEGIE	aanpassende kwalificering . beschikbare kwalificaties aanpassen aan eisen van werk. . verwerven kwalificaties nav. eisen die taken stellen. Medewerkers volgen cursussen.	aanpassende werkorganisatie . werk aanpassen aan beschikbare kwalificaties. . verwerven kwalificaties nav. ontwikkelingen die in beroep plaatsvinden. Professionals volgen cursussen over nieuwe methodieken.	wederzijdse aanpassing . werk en beschikbare kwalificaties voortdurend aan elkaar aanpassen. . verwerven kwalificaties nav. problemen die zich in werk voordoen. Leren is werk, werk is leren.

Fig. 1. Kwalificatiestrategieën voor arbeid in drie typen organisaties.

We spreken over 'leer'ontwikkeling – en niet over opleidingsontwikkeling – omdat het een verscheidenheid aan leersituaties kan betreffen en niet alleen opleidingen, cursussen of trainingen. We denken ook aan leersituaties als intervisie, werkoverleg, training op de werkplek, functioneringsgesprekken. Het gaat om het ontwikkelen van leersituaties en leeractiviteiten die medewerkers kunnen gebruiken om zich te kwalificeren, zowel op de werkplek als daarbuiten, speciaal ontwikkeld voor dit probleem of een standaardcursus enzovoorts.

Leersituaties kunnen voor vele soorten functies en taken ontwikkeld worden. Hier worden deze beperkt tot leerontwikkelingsstrategieën voor uitvoerende functies in verschillende soorten organisaties.

In leerontwikkelingsstrategieën is een viertal aspecten te onderscheiden. Deze zijn (1) de aanleiding voor het ontwikkelen van leersituaties, (2) de werkwijze bij het ontwikkelen van leersituaties, (3) de aard van leersituaties en (4) de organisatie van het proces om het leeraanbod te ontwikkelen. Elk van deze aspecten kan op verschillende wijze ingevuld worden.

Bij de aanleiding speelt datgene waardoor de discrepantie tussen vereiste en beschikbare kwalificaties is ontstaan, een belangrijke rol; het betreft vooral de analyse van de situatie en de achtergronden van het kwalificatievraagstuk. De werkwijze omschrijven we als die activiteiten die leiden tot leeractiviteiten waarmee de geconstateerde aansluitingsproblemen opgeheven kunnen worden. Het gaat hierbij vooral om voorbereidende activiteiten die nodig zijn om te inventariseren en te analyseren waar de aansluitingsproblemen zitten en wat deze precies inhouden. Met de aard van leeractiviteiten worden de kenmerken bedoeld waaraan de leeractiviteiten moeten voldoen om de geconstateerde aansluitingsproblemen te reduceren. Bij dit aspect komen zaken aan de orde als het doel van leeractiviteiten, de leervorm en de leersituatie. De organisatie

LEERONTWIKKELINGSSTRATEGIEËN

ASPECTEN	TAAKGERICHT	PROFESSIEGERICHT	PROBLEEMGERICHT
AANLEIDING	Huidige taken niet goed uitgevoerd. Medewerkers beschikken over onvoldoende of onjuiste kwalificaties	Bijhouden van nieuwe ontwikkelingen binnen het vakgebied.	Complex en veranderlijk werk vereist voortdurend nieuwe kwalificaties.
WERKWIJZE	Taak- en functieanalyses bepalen de inhoud van de leeractiviteiten.	Opleidingsinstituten organiseren in opdracht van beroepsverenigingen leeractiviteiten.	Medewerkers analyseren zelf problemen en bijbehorende kwalificaties. Dit is tevens een leeractiviteit.
AARD	Aanleren van taakspecifieke kwalificaties. Doel is taken adequaat uit te kunnen oefenen Vooral cursussen off the job. Grote groepen lerenden.	Aanleren van breed toepasbare kwalificaties. Zowel cursussen off the job (in grote groepen) als leeractiviteiten on the job (werkbesprekingen, evalueren van handelen).	Doel is vergroten van probleemoplossingsstrategieën. Zowel taakspecifieke als taakgenerieke kwalificaties. Vooral leren op de werkplek, in teams.
ORGANISATIE	Staffunctionarissen en management spelen de hoofdrol. Medewerkers spelen passieve rol.	De beroepsgroep van de professionals speelt hoofdrol. Binnen organisatie heeft management een bescheiden rol.	Medewerkers zijn belangrijkste actoren bij ontwikkelen van leeractiviteiten. Zij worden begeleid door coördinator.

Fig. 2. Beschrijving van drie leerontwikkelingsstrategieën naar vier aspecten.

gaat in op de verdeling van taken en verantwoordelijkheden van de actoren (opleidingskundigen, managers, lerenden) bij het ontwikkelen van leeractiviteiten.

We onderscheiden drie leerontwikkelingsstrategieën respectievelijk de taakgerichte, de professiegerichte en de probleemgerichte leerontwikkelingsstrategie. In Figuur 2 is een korte beschrijving van deze strategieën gegeven

Welke strategieën zijn nu het meest geschikt? In de contingentietheorie wordt dit afhankelijk gesteld van de situatie (Mintzberg, 1983, 1989; Scheerens, 1989). Vermoedelijk speelt het soort arbeid een belangrijke rol als het leerontwikkelingsstrategieën betreft.

Het ontwikkelen van leeractiviteiten voor adhocratische arbeid zal anders verlopen dan voor taakarbeid in een machine bureaucratische organisatie. Taakarbeid leent zich voor het opstellen van een gedetailleerde taakanalyse. Deze kan als basis dienen voor het in kaart brengen van de te verwerven kwalificaties. Voor een uitvoerende functie in een adhocratie (bijvoorbeeld organisatie-advieswerk) zal deze werkwijze minder geschikt zijn aangezien deze functie niet uit eenduidig te definiëren en af te bakenen taken bestaat. In zo'n geval zal een andere werkwijze, die beter aansluit bij de aard en de kenmerken van de functie, wenselijk zijn.

We veronderstellen dan dat bij taakarbeid de taakgerichte leerontwikkelingsstrategie zal voorkomen, bij professioneel bureaucratische arbeid een professiegerichte leerontwikkelingsstrategie en dat bij adhocratische arbeid een probleemgerichte leerstrategie aangetroffen wordt.

ONDERZOEKSVRAGEN

In het algemeen willen we weten welke dimensies er in de leerontwikkelingsstrategieën voorkomen en of de toepassing van de strategie varieert naar gelang het soort arbeid.

Alvorens dit te onderzoeken zijn de volgende beperkingen aangebracht.

Bij het ontwikkelen van leeractiviteiten spelen diverse actoren in en rondom organisaties een rol: managers, uitvoerende medewerkers, personeelsfunctionarissen, opleidingskundigen en andere staffunctionarissen. Opleidingskundigen zijn het meest betrokken bij het ontwikkelen van leersituaties. Om deze reden is gekozen voor de opleidingskundige als centrale figuur in het onderzoek: welke aanpak hanteren opleidingskundigen bij kwalifciatieproblemen en in het bijzonder welke dimensies zijn er in de maatregelen van opleidingskundigen te onderscheiden. We noemen dat de leerontwikkelingsdimensies.

Naast de beperking tot de strategieën van opleidingskundigen, is ook gekozen voor een vergelijking tussen twee (in plaats van drie) typen arbeid die naar ons idee ver uiteenlopen: taakarbeid in de bureaucratie en adhocratische arbeid in de adhocratie. We willen nagaan in hoeverre opleidingskundigen kwalificatieproblemen bij verschillende soorten arbeid ook anders aanpakken.

Samenvattend kunnen we de volgende onderzoeksvragen formuleren: zijn er dimensies in de aanpak van kwalificatieproblemen door opleidingskundigen te onderkennen? In aansluiting hierop gaan we na of die aanpak anders is bij de twee typen arbeid.

ONDERZOEKSOPZET

Het onderzoek is uitgevoerd onder 59 opleidingskundigen. De ene helft kreeg een kwalificatieprobleem bij taakarbeid voorgelegd, de andere helft een soortgelijk probleem bij adhocratische arbeid. Hen is door middel van een schriftelijke vragenlijst gevraagd hoe zij dit probleem zouden aanpakken. In de vragen zijn de drie strategieën en vier aspecten verwerkt, zie ook Figuur 2.

Subjecten

De opleidingskundigen zijn geworven via een adressenlijst van een beroepsvereniging voor afgestudeerde onderwijskundigen en adressen van opleidingsbureaus. Verder is gebruik ge-

maakt van de sneeuwbalmethode: opleidingskundigen benaderen en hen vragen naar adressen van collega's. Uiteindelijk is een bestand ontstaan van 80 opleidingskundigen aan wie de door ons gemaakte vragenlijst werd toegestuurd. Er werden 60 vragenlijsten teruggestuurd waarvan er 59 bruikbaar bleken. De hierna gepresenteerde onderzoeksresultaten zijn daarmee gebaseerd op een response van 74%.

Van deze 59 opleidingskundigen bleken 35 personen binnen de eigen arbeidsorganisatie te werken en 24 op een zelfstandig opleidingsbureau. Dit laatste betekent dat de opleidingskundige bezig is met opleiden en leren voor medewerkers uit andere organisaties. Het grootste deel (38) heeft een wetenschappelijke opleiding genoten. Daarnaast hebben 19 opleidingskundigen een HBO-opleiding en 2 opleidingskundigen een MBO-opleiding gevolgd.

Procedure en onderzoeksinstrumenten

Er zijn twee verschillende casebeschrijvingen gemaakt waarin een kwalificatieprobleem werd beschreven.

In de ene case-beschrijving staat een verzekeringsmaatschappij centraal. Het primaire proces in deze maatschappij kenmerkt zich door een organisatie van de arbeid die vergelijkbaar is met die in een machinebureaucratie: het werk van de verzekeringsagenten bestaat uit kleine deeltaken die op een gestandaardiseede manier moeten worden uitgevoerd, ze hebben weinig autonomie en hun functie is als weinig complex te omschrijven. Het management signaleert dat de verzekeringsagenten bij moeilijke vragen (die afwijken van het standaardpakket) van cliënten steeds vaker een beroep doen op stafmedewerkers. Dit heeft langere wachttijden tot gevolg en leidt dus tot een vermindering van de kwaliteit van de dienstverlening.

De andere case-beschrijving gaat over een reclame-adviesbureau. De organisatie van de arbeid in dit bureau heeft overeenkomsten met de organisatie van de arbeid in een adhocratische organisatie: de reclameontwerpers werken in teamverband aan complexe opdrachten, tijdens het werken aan deze opdrachten wordt in onderling overleg bepaald wie welke taken uitvoert en in welke mate deze taken verdeeld worden. In het reclamebureau wordt gesignaleerd dat reclames en advertenties vaker een voor de klant specifieke vorm moeten krijgen en dat de opdrachten steeds diverser van aard worden. Het wordt steeds moeilijker om vragen van cliënten om te zetten in ontwerpen. Door de vragen uit de schriftelijke vragenlijst te beantwoorden, kunnen de respondenten aangeven hoe ze de genoemde kwalificatieproblemen aan gaan pakken.

Zowel de vragenlijst als de case-beschrijvingen zijn aan een kritisch vooronderzoek onderworpen alvorens ze aan de gehele steekproef werden voorgelegd. Eerst zijn proefinterviews gehouden met vier opleidingskundigen. Als resultaat daarvan zijn de case-beschrijvingen en een aantal items uit de vragenlijst bijgesteld. In de daarop volgende pilot-studie is het gehele materiaal voorgelegd aan drie opleidingskundigen met het verzoek het geheel kritisch te evalueren en waar nodig wijzigingen voor te stellen. Dit resulteerde in een definitieve vragenlijst met bijbehorende casebeschrijvingen. De vragenlijst bestaat uit 107 items met 7-punts Likert antwoord schalen variërend van 1 (=klopt helemaal niet) tot en met 7 (=klopt helemaal wel). De taakgerichte leerontwikkelingsstrategie bestaat uit 46 items (2 voor aanleiding, 21 voor werkwijze, 16 voor aard en 7 voor organisatie), de professiegerichte uit 24 (respectievelijk 1, 9, 12 en 2) en de probleemgerichte uit 37 items (respectievelijk 2, 13, 19 en 3).

De 80 opleidingskundigen zijn aselect verdeeld in twee groepen. De ene groep kreeg het kwalificatieprobleem met taakarbeid voorgelegd, de andere groep het probleem met de adhocratische arbeid. Beide groepen moesten dezelfde vragenlijst invullen. Over de case-beschrijving van de verzekeringsmaatschappij zijn van 31 opleidingskundigen de antwoorden verwerkt, over de case-beschrijving van het reclame-adviesbureau 28. Van alle opleidingskundigen werden ook enige achtergrondgegevens verzameld (genoten opleidingen en dergelijke). Geen significante verschillen werden gevonden tussen beide groepen op diverse achtergrondkenmerken.

Voor een meer uitvoerige beschrijving van de case-beschrijvingen, de vragenlijst en de vergelijking op achtergrondkenmerken, zie Keurentjes & Ter Woerds (1992).

Analyse opzet

De antwoorden van alle respondenten op de 107 items van de vragenlijst vormen de input voor de analyses. Met behulp van iteratieve principale factor analyse wordt geprobeerd factoren of dimensies te realiseren die een weerspiegeling zijn van leerontwikkelingsdimensies. Er wordt niet bij voorbaat aangenomen dat de dimensies onafhankelijke concepten zijn. Daarom zal gebruik worden gemaakt van oblique rotaties (OBLIMIN) voor interpretatie van de onderliggende factorstructuur. Items die niet in de factorstructuur thuis horen zullen worden verwijderd. In eerste instantie worden de minst passende items verwijderd maar als eis stellen we dat een item in de uiteindelijke patroonmatrix minstens een gewicht moet hebben van .40 op de factor waarvoor het item geformuleerd is en een gewicht lager dan .20 op de overige factoren. Items die verwijderd zijn zullen in de loop van het analyseproces weer in de analyse worden opgenomen om te kijken of ze nu wel in de factorstructuur passen. Een dergelijke werkwijze leidt tot duidelijk van elkaar onderscheidbare factoren en daarmee tot leerontwikkelingsdimensies die ten opzichte van elkaar afgebakend zijn. Dit komt de validiteit van het meetinstrument ten goede. Nadat de factorstructuur is vastgesteld worden betrouwbaarheden vastgesteld in termen van Cronbach's interne consistentiemaat alpha. Items die de betrouwbaarheid in behoorlijke mate drukken zullen zonodig verwijderd worden.

RESULTATEN

Vier leerontwikkelingsdimensies

Van de 107 geformuleerde items zijn tenslotte 31 items na factor analyses overgebleven.

De overgebleven items zijn vooral afkomstig uit de aspecten werkwijze en aard van de strategie: 27 van de 31. Één item blijkt nog het aspect aanleiding te vertegenwoordigen, terwijl 3 items betrekking hebben op de organisatie. Dat van de aspecten aanleiding en organisatie zo weinig items zijn overgebleven heeft waarschijnlijk te maken met het gering aantal items dat voor deze aspecten al vooraf was geformuleerd. Er zijn vier factoren geconstrueerd waarvan het resulterend factorpatroon vermeld staat in Tabel 1.

De eerste kolom in Tabel 1 geeft de originele nummering weer van de items. Kolom 2 geeft aan van welk aspect de items afkomstig zijn: al=aanleiding, w=werkwijze, ar=aard en or=organisatie). Kolom 3 geeft een verkorte omschrijving van de inhoud van de items, terwijl de kolommen 4, 5, 6 en 7 de gewichten weergeven van de items op elk van de vier factoren. In de laatste regel van de tabel zijn de betrouwbaarheden vermeld.

De vier factoren verklaren in totaal 42.5% van de variantie in de variabelen, wat voor analyses op itemniveau een hoog percentage genoemd kan worden. De factoren blijken onderlinge correlaties te laten zien van -.07 tot .12 wat betekent dat ze als onafhankelijke factoren kunnen worden beschouwd. Een orthogonaal uitgevoerde rotatie (VARIMAX) laat dan ook qua factorstructuur een identiek resultaat zien.

De betrouwbaarheden zijn voor drie van de vier factoren goed te noemen. De eerste drie factoren hebben respectievelijk alpha's van .81, .87 en .81, terwijl de laatste factor een alpha heeft van .69. Er zijn geen items gevonden die een negatieve invloed hebben op de betrouwbaarheid vandaar dat op basis van de betrouwbaarheidsanalyses geen items verwijderd werden.

De factoren verwijzen ons inziens naar vier leerontwikkelingsdimensies. We duiden deze aan met:

1 taakanalytische oriëntatie. (F1).
2 professionele leeraanbod oriëntatie. (F2).
3 leer-werk oriëntatie. (F3).
4 werkaanpassing oriëntatie. (F4).

De taakanalytische oriëntatie geeft aan in welke mate een opleidingskundige een analyse maakt van de taken die moeten worden uitgevoerd. Opleiders met een hoge score op deze dimensie leggen de nadruk op het voortraject, de analyses die nodig zijn alvorens leersituaties te ontwikkelen.

Tabel 1. Factorpatroon van de vier-factoroplossing.

nr	as	verkorte item-omschrijving	F1	F2	F3	F4
8	w	nieuwe taken moeten in kaart worden gebracht	**.75**	.02	-.03	.06
7	w	huidige taken moeten geanalyseerd worden	**.75**	.13	-.12	-.02
9	w	discrepantie tussen huidige en nieuwe taken moet bepaald worden	**.74**	-.11	.07	.02
27	w	taken van werknemers duidelijk beschrijven	**.73**	-.01	.19	.04
14	w	taakanalyse is grondslag voor het ontwerpen van de leeractiviteit	**.70**	.05	-.02	-.07
38	w	bestaande taakomschrijvingen worden geraadpleegd	**.63**	.12	-.03	-.14
64	ar	eerst kwalificaties verwerven voordat nieuwe taken uitgevoerd worden	**.50**	-.05	-.05	-.02
98	or	leeractiviteit in overleg met het management grondig plannen	**.41**	-.02	-.04	.04
81	a	kwalificaties verwerven via cursussen bij eigen beroepsvereniging	-.11	**.86**	-.10	.00
47	w	een beroepsvereniging consulteren	.08	**.84**	-.01	.14
36	w	samenwerken met een beroepsvereniging	.03	**.73**	.09	-.01
43	w	overzicht van leeractiviteiten van de beroepsvereniging raadplegen	.12	**.73**	.01	.20
65	ar	beginnen met het volgen van een cursus bij de eigen beroepsvereniging	.04	**.68**	-.11	.02
22	w	de deskundigheid van de beroepsvereniging is belangrijk	.06	**.61**	-.02	.08
20	w	het bestaan van een beroepsvereniging speelt een belangrijke rol	-.13	**.61**	-.12	.09
95	ar	samen met collega's uit dezelfde branche kwalificaties verwerven	.03	**.44**	.02	.06
2	al	ontwikkelingen binnen eigen vakgebied is aanleiding	.09	**.42**	.09	-.07
57	ar	professionele kennis is belangrijker dan specifieke kwalificaties	-.11	**.40**	.03	-.14
50	ar	leeractiviteit moet samenhangen met complexe vragen van het werk	.03	-.07	**.75**	.03
48	w	het complexe werk van werknemers vormt uitgangspunt	-.19	-.09	**.73**	-.03
101	or	werknemers spelen een actieve rol bij ontwikkeling van leeractiviteiten	-.09	-.12	**.62**	.15
19	w	de mate van complexiteit van het werk is van belang	.00	.06	**.58**	-.08
89	ar	werknemers moeten in kleine groepjes met elkaar reflecteren	-.12	-.11	**.54**	-.03
84	ar	de feitelijke werksituatie is belangrijk voor kwalificatieverwerving	.04	.17	**.49**	.02
21	w	het probleemoplossend vermogen moet worden vergroot	-.01	.20	**.44**	-.20
100	or	leeractiviteit is meer procesbegeleiden	.17	-.07	**.44**	.09
85	ar	kwalificaties verwerven via aanpassing van werkopdrachten	-.05	-.02	.12	**.77**
25	w	de mate waarin werknemers opdrachten aanpassen aan hun kwalificaties	-.02	.14	-.09	**.62**
54	ar	leeractiviteit betekent aanpassen van werkopdrachten aan kwalificaties	.02	.00	-.11	**.58**
11	w	kwaliteiten analyseren en opdrachten daarop afstemmen	-.07	.03	-.01	**.57**
90	ar	kwalificaties verwerven door aanpassing van het werk aan problemen	.04	.04	.10	**.49**
		Cronbach's alpha coëfficiënt	**.81**	**.87**	**.81**	**.69**

De professionele leeraanbod oriëntatie benadrukt de betekenis van de beroepsvereniging en het leeraanbod dat vanuit die kringen wordt aangeboden. De inhoud van de na- en bijscholing is vooral een aangelegenheid van de professie, zo vinden opleidingskundigen met een hoge score op deze dimensie.

De leer-werk oriëntatie brengt de verbondenheid van leren en werken tot uitdrukking. Het kwalificatievraagstuk zou vooral dicht bij de werkplek in samenwerking met de werkenden zelf geanalyseerd en aangepakt moeten worden. Deze aanpak wordt mede ingegeven door de complexiteit van het werk.

De vierde oriëntatie verwijst naar de mate waarin opleidingskundigen mogelijkheden zien om door het werk aan te passen de kwalificatieproblematiek op te lossen. Er zijn diverse elementen van de wederzijdse aanpassing aanwezig.

Tabel 2. Resultaten van de verschiltoetsen.

leerontwikkelings-dimensie		taak-arbeid	adhocratische arbeid	df	t	p
taakanalytische oriëntatie	gem.	5.40	4.62	47	2.58	.013
	s.d.	.91	1.34			
professionele leeraanbod oriëntatie	gem.	3.34	3.98	56	-2.54	.013
	s.d.	1.09	.82			
leer-werk oriëntatie	gem.	5.54	5.51	56	.16	n.s
	s.d.	.78	.77			
werkaanpassing oriëntatie	gem.	3.17	3.41	56	-.78	n.s
	s.d.	1.12	1.15			

n.s = niet significant

Verschillen in strategieën bij de twee soorten arbeid

Met behulp van t-toetsen voor onafhankelijke steekproeven is nagegaan of op de vier leerontwikkelingsdimensies opleidingskundigen bij taakarbeid anders te werk gaan dan bij adhocratische arbeid. De resultaten staan in Tabel 2.

Bij taakarbeid hebben opleidingskundigen significant een andere taakanalytische oriëntatie dan bij adhocratische arbeid (t=2.58, p=.013). De cijfers wijzen erop dat de taakanalytische oriëntatie hoger scoort bij taakarbeid dan bij adhocratische arbeid. Voor de professionele leeraanbod oriëntatie vinden we eveneens een significant verschil (t=-2.54, p=.013) maar nu in omgekeerde richting: bij taakarbeid scoort men gemiddeld lager dan bij adhocratische arbeid. De leer-werk en de werkaanpassingsoriëntatie laten geen significante verschillen zien tussen beide typen van arbeid (t=.16 respectievelijk -.78, beide dus niet significant).

Conclusies en perspectieven

De aanpak van kwalificatieproblemen door opleidingskundigen kan langs vier leerontwikkelingsdimensies beschreven worden. Bovendien blijken opleidingskundigen op uiteenlopende manieren met die kwalificatieproblemen om te gaan en daarbij enigermate rekening te houden met de situatie waarin zich het probleem voordoet. De aanpak van kwalificeringsproblemen bij taakarbeid verloopt anders dan bij adhocratische arbeid.

De vier dimensies in de aanpak van kwalificatieproblemen, die het resultaat zijn van de analyses, zijn als volgt te omschrijven:

a. de taakanalytische oriëntatie, dat wil zeggen de nadruk die opleidingskundigen leggen op de (taak)analyses in het voortraject. Deze oriëntatie, waarin de uit te voeren taken een belangrijk referentiepunt voor de aanpak van het probleem vormt, is bij de onderzochte opleidingskundigen in sterke mate aanwezig. In de opleidingskundige literatuur en technologie is deze opvatting ook het verst ontwikkeld.
b. de professionele leeraanbod oriëntatie, dat wil zeggen de betekenis die opleidingskundigen toekennen aan de professie voor de na- en bijscholing. Het kwalificatieprobleem wordt gerelateerd aan de professie van de uitvoerenden. Daar komen de inhoudelijke impulsen voor de scholing vandaan. Deze oriëntatie is duidelijk minder aanwezig bij de opleidingskundigen.
c. de leer-werk oriëntatie, dat wil zeggen de betekenis van de integratie van werken en leren. Ook hier speelt het werk een belangrijke rol maar op een geheel andere manier dan bij de taakanalytische oriëntatie. Het werk is zo complex dat het minder gemakkelijk is te analyseren en er geen standaard leersituaties voor te maken zijn. Het leren en werken wordt meer

met elkaar verbonden. Hierin is het leren op de werkplek onder andere door zelf de problemen te analyseren, herkenbaar. Dit impliceert ook een veel actievere rol van de werknemers dan bij de taakanalytische oriëntatie.

d. de werkaanpassing oriëntatie, dat wil zeggen de mogelijkheden die opleidingskundigen zien om door middel van het aanpassen van het werk de kwalificatieproblemen aan te pakken en leeractiviteiten te ondernemen. In de taakanalytische en de leer-werk oriëntatie wordt het werk als een gegeven beschouwd. Bij de werkaanpassing wordt het kwalificatieprobleem aan twee kanten aangepakt, zowel de beschikbare als de vereiste kwalificaties staan ter discussie. Door het werk te veranderen wordt er ook geleerd.

In de eerste dimensie is de meest gangbare leerontwikkelingsstrategie, de taak- of functiegerichte, duidelijk herkenbaar. De drie andere oriëntaties laten ook de nieuwe elementen zien namelijk dat veranderingen in het werk mogelijkheden bieden voor de aanpak van kwalificatieproblemen en tevens dat de eigen inbreng van de (lerende) werknemers bij de aanpak van dergelijke problemen van belang is. De verbinding van leer- en ontwikkelingsactiviteiten met de analyse en verandering van de werksituatie blijkt niet alleen in theorie maar ook empirisch herkenbaar in de aanpak van deze opleidingskundigen.

De tweede onderzoeksvraag heeft betrekking op het hanteren van verschillende leerontwikkelingsstrategieën door opleidingskundigen bij verschillende soorten arbeid. Ons vermoeden was dat er weinig of geen verschillen zouden bestaan en dat de taakgerichte leerontwikkelingsstrategie het sterkst vertegenwoordigd zou zijn. Deze verwachting was gebaseerd op de constatering dat in de onderwijskundige literatuur de taakgerichte leerontwikkelingsstrategie sterk vertegenwoordigd is. In deze literatuur, die ook veel in opleidingen voor opleidingskundigen wordt gebruikt, zijn de beide andere strategieën nog niet sterk vertegenwoordigd (zie bijvoorbeeld Robinson & Robinson, 1989; Kessels & Smit, 1989; Thijssen, 1988; Mulder e.a., 1988, Romiszowski, 1989). Uit Tabel 2 blijkt echter dat opleidingskundigen bij beide soorten arbeid verschillend scoren op de taakanalytische oriëntatie en op de professionele leeraanbod oriëntatie. Dit kan het volgende betekenen: opleidingskundigen laten zich bij het ontwikkelen en vormgeven van een leeraanbod niet alleen leiden door de dominante taakgerichte leerontwikkelingsstrategie. Blijkbaar bestaan er verschillende opvattingen over hoe leeractiviteiten ontwikkeld moeten worden; in die opvattingen speelt de (contingente) gedachte dat de situatie mede de aanpak bepaalt een belangrijker rol dan we eerst aannamen.

In dit verband is het goed om ons af te vragen wat we hier gemeten hebben: de orientaties en strategieën zoals opleidingskundigen die in hun hoofd hebben of hoe zij in de praktijk te werk gaan? Dit is een vraag naar de "espoused theory", de wenselijke manier van werken en de "theory-in-use" ofwel de feitelijke manier van werken (Argyris & Schön, 1978). Mensen zijn zich niet altijd bewust van hun "theory-in-use" en de "espoused theory". Tijdens proefinterviews bleek dat opleidingskundigen moeilijk het waarom van hun handelen konden beargumenteren. Ze hadden moeite hun praktijktheorie, het geheel aan veronderstellingen dat aan het (keuze-)gedrag van functionarissen ten grondslag ligt (Van Gageldonk & Meyer, 1988, p. 9) te expliciteren. Verder onderzoek hiernaar kan de professionalisering van opleidingskundigen ten goede komen. Door inzicht in de eigen praktijktheorieën kan men de veronderstellingen die aan het handelen ten grondslag liggen ter discussie stellen. Deze vorm van leren is ons inziens van belang omdat opleidingskundigen hiermee inzicht kijgen in hun eigen leerontwikkelingsstrategie maar ook in andere. Het bewust variëren van de strategie naar gelang de situatie, arbeid en organisatie, wordt daarmee bevorderd.

Voorgaande overwegingen en kanttekeningen geven al aan dat dit onderzoek bedoeld is als aanzet voor verdere, empirisch onderbouwde, studies en theorievorming over de samenhang tussen kwalificering van medewerkers en arbeid in verschillende soorten organisaties. Onderzoek en theorievorming op dit terrein staat nog in de kinderschoenen en onze resultaten beschouwen we, door de beperkte omvang van het onderzoek en de gebruikte meetinstrumenten, als empirisch gefundeerde werkhypothesen voor verdere studies.

LITERATUUR

Argyris, C., & Schön, D.A. (1978). *Organizational learning: A theory of action perspective.* Massachusetts, California: Addison-Wesley Publishing.

Brandsma, T.F., Nijhof, W.J. & Kamphorst, J.C. (1990). *Kwalificatie en curriculum.* Amsterdam/Lisse: Swets & Zeitlinger.

Gageldonk, A. van, & Meyer, R.F. (1988). *Praktijktheorieën van het management in het middelbaar beroepsonderwijs: Reconstructie en analyse van hoofdlijnen.* Leiden: Leids Instituut voor Sociaal Beleidsonderzoek (LISBON).

Hoof, J.J. van & Dronkers, J. (1980). *Onderwijs en arbeidsmarkt: een verkenning van relatie tussen onderwijs, arbeidsmarkt en arbeidssysteem.* Deventer: Van Loghum Slaterus.

Kessels, J.W.M., & Smit, C.A. (1989). *Opleidingskunde: Een bedrijfsgerichte benadering van leerprocessen.* Deventer: Kluwer Bedrijfswetenschappen.

Keurentjes, J.G.M.M., & Woerds, I.E.C. ter. (1992). *Opleidingskundigen en de ontwikkeling van leeractiviteiten. Een onderzoek naar de leerstrategieën die opleidingskundigen toepassen bij een verschillende organisatie van de arbeid.* Nijmegen: Katholieke Universiteit Nijmegen, doctoraalscriptie Vakgroep Algemene Pedagogiek.

Kraayvanger, G., & Onna B. van. (1985). *Arbeid en leren: Bijdragen tot de volwasseneneducatie.* Baarn: Nelissen.

Krogt, F.J. van der. (1990). De leerfunctie in verschillende organisaties. *Opleiding & Ontwikkeling, 11,* 17-24.

Krogt, F.J. van der (1991a). Aansluitingsstrategieën en opleidingsactiviteiten in verschillende organisatietypen. *Tijdschrift voor Arbeidsvraagstukken, 4,* 3-15.

Krogt, F.J. van der. (1991b). Autonomie en leren van professionals. *Opleiding & Ontwikkeling, 9,* 15-18.

Mintzberg, H. (1983). *Structure in Fives: Designing effective organizations.* New Jersey: Prentice-Hall International Editions.

Mintzberg, H. (1989). *Mintzberg on management.* New York: Free Press.

Mulder, M., Akkerman, J.S., & Bentvelsen, N. (1988). *Bedrijfsopleidingen in Nederland.* Den Haag: Instituut voor Onderzoek van het Onderwijs.

Robinson, D.G. & Robinson, J.C. (1989). *Training for impact.* San Francisco: Jossey Bass.

Romiszowsky, A.J. (1989). Ontwikkelen van opleidingen. Instructie: De integratieve benadering. *Handboek Opleiders in Organisaties.* Deventer: Kluwer Bedrijfswetenschappen.

Scheerens, J. (1989). Onderzoek naar schoolorganisatorische effectiviteit en de contingentiebenadering. In: Scheerens, J. & Verhoeven, J.C. (red.) *Schoolorganisatie, beleid en onderwijskwaliteit.* Lisse: Swets & Zeitlinger.

Thijssen, J.G.L. (1988). *Bedrijfsopleidingen als werkterrein.* 's-Gravenhage: VUGA uitgeverij.

Warmerdam, J., & Berg, J. van den. (1986). *Opleidingsactiviteiten in arbeidsorganisaties.* Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Warmerdam, J. & Berg, J. van den (1992). *Scholing van werknemers in veranderende organisaties.* Den Haag/Nijmegen: Ministerie van Sociale Zaken/ITS.

Manuscript ontvangen 14-5-1993
Definitieve versie ontvangen 8-9-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 5, pp. 306-316

Tijdsbesteding van docenten in het voortgezet onderwijs op verschillende manieren gemeten: een validatiestudie

Jurrien Heidelberg, Huub van den Bergh*[1], Hans Kuhlemeier[2] en Jan Hoeks[3]
[1] *Centre for Language and Communication, Universiteit Utrecht*
[2] *Cito, Afdeling Peilingsonderzoek, Arnhem*
[3] *Psychologisch Laboratorium van de Universiteit van Amsterdam*

ABSTRACT

In the current study the convergent validity of three ways to measure time allocation in geography lessons in secondary education is studied: observations, keeping logs and traditional questionnaire. Four geography lessens of twelve teachers in the third grade of intermediate general education were observed. The teachers kept the logs for the same lessons, and afterwards they filled in the questionnaire with the same questions concerning the allocation of time to different aspects of the geography lessons. The convergent validity of the three type of measurements of time allocation is analyzed with respect to four questions: (the convergence) of the mean scores, (the convergence) of the between-teacher-variances, the correlation between measurements, and the reliability of the differences between teachers for observations and logs only. Differences due to type of measurement were assessed. This holds for three oppositions, and concerns as well the mean estimates as the between-teacher-variance estimates. Nevertheless observations and logs showed less differences then either of those with a traditional questionnaire. The reliability of the observations and the logs proved to be dependent on the aspect analyzed.

INLEIDING

In instructie-effectiviteitsonderzoek worden leeruitkomsten in verband gebracht met instructieverschillen (Creemers, 1991; Scheerens, 1989). Aan de meting van de leeruitkomsten (prestaties) wordt vaak veel aandacht besteed. De meetinstrumenten worden (vaak) minitieus gescreend in een vooronderzoek, waarbij ruim aandacht besteed wordt aan de betrouwbaarheid en de dimensionaliteit van de meting. De meting van de verklarende variabelen, de instructie, komt er daarentegen vaak bekaaid vanaf. In de meeste gevallen wordt volstaan met een meting met behulp van een vragenlijst welke éénmaal afgenomen wordt. Hierin wordt de docenten dan gevraagd aan te geven wat zij zoal doen, en hoeveel tijd zij zoal aan verschillende vakonderdelen besteden. Hiermee wordt geen recht gedaan aan de complexiteit van (de meting van) het onderwijsaanbod. Enerzijds worden de psychometrische aspecten verwaarloosd. In welk instructie-effectiviteitsonderzoek wordt bij voorbeeld inzicht gegeven in de stabiliteit van de instructie-metingen? Anderzijds blijkt de convergentie met andere meetmethoden voor dezelfde aspecten van de instructie vaak laag.

Wanneer we ons beperken tot de meting van de daadwerkelijk bestede lestijd aan verschillende aspecten van een vak dan is een drietal recente studies van belang. Otter, Mellenbergh, Blok en Van den Bergh (1992) demonstreerden bij voorbeeld dat verschillende metingen (vragenlijsten, telefonische interviews en paarsgewijze vergelijkingen) van de tijdsbesteding aan het vak Nederlands in het basisonderwijs maar zeer ten dele convergeren; de gemiddelde tijdsbesteding varieert ten gevolge van de meetmethode, en de correlatie tussen de drie meetmethoden laat te wensen over (gemiddeld: r = .43).

* Adres: Centre for Language and Communication, Universiteit Utrecht, Trans 10, 3512 JK Utrecht.

Ook Overmaat (1992) onderzocht de validiteit van vragenlijsten ter meting van (aspecten van) het taalonderwijs op de basisschool. Zij constateerde bij een herhaalde afname van dezelfde vragenlijst zowel verschillen in gemiddelde tijdsbesteding, als een matige correlatie (r = .47) tussen de beide metingen. Ook bij een vergelijking tussen de vragenlijstgegevens met die van door de docent bijgehouden logboeken kwamen duidelijke verschillen naar voren. Niet alleen bleek de (gemiddelde) correlatie over de totaal bestede lestijd laag (r = .11), ook de gemiddelde lestijd bleek afhankelijk van de meetmethode. Overmaat beveelt aan naast vragenlijsten ook logboeken te gebruiken ter meting van (aspecten van) het onderwijsaanbod (p. 118). Hierbij noemt zij acht weken als richtperiode voor het bijhouden van logboeken.

Ook Harskamp en Suhre (1993) concluderen dat vragenlijst- en logboekgegevens voor aspecten van het rekenonderwijs in het basisonderwijs slechts zeer ten dele samenhangen: de metingen van de variabelen door middel van een vragenlijst en een logboekprocedure stemmen niet met elkaar overeen (p. 367). In hun aanbevelingen spreken zij een voorkeur uit voor logboeken. Voor een betrouwbare meting moeten dan wel de tijdsbesteding in veertig rekenlessen verdeeld over veertig schoolweken bijgehouden worden (p. 369).

Zeker zo belangrijk als de psychometrische aspecten van een meting van het onderwijsaanbod zijn de theoretische aspecten. Ook deze worden regelmatig verwaarloosd. Rosenshine en Furst (1973) komen bij voorbeeld tot de conclusie dat van de 73 geanalyseerde (observatie-) instrumenten ter inventarisering (o.a.) docentgedrag er slechts in vijftien gevallen sprake is van enige theoretische fundering. Lugthart, Roeders, Bosker en Bos (1989) concluderen op grond van hun literatuurstudie naar onderwijseffecitiviteit dat er door een gebrek aan theoretische modellen sprake is van een vaagheid van de te meten begrippen, waardoor in het ene onderzoek een zelfde variabele op een andere manier geoperationaliseerd wordt als in een ander onderzoek.

In de onderhavige studie staat de vraag naar de betrouwbaarheid, en daarvan afgeleid de vraag naar de validiteit, van metingen van het onderwijsaanbod in het voortgezet onderwijs centraal. We beperken ons slechts tot één aspect van het onderwijsaanbod van één vak: de tijdsbesteding in de lessen aardrijkskunde. De centrale vraag in deze studie is die naar de convergente validiteit van drie verschillende manieren om de tijdsbesteding in de lessen aardrijkskunde te meten: een éénmalig ingevulde vragenlijst, logboeken ingevuld na iedere aardrijkskundeles en observaties.

Geen enkele methode om de tijdsbesteding te meten is ideaal. Zo heeft een vragenlijst als nadeel dat de gegevens retrospectief van aard zijn, en verstoord kunnen zijn door herinnneringseffecten. Logboeken hebben als nadeel dat ze vaak vlug en onregelmatig bijgehouden worden. En, observaties tenslotte kunnen verstoord worden door invloeden van de observator(en).

Deze studie heeft plaats gevonden in het kader van een peilingsonderzoek naar het niveau van het aardrijkskunde-onderwijs. In dit peilingsonderzoek wordt niet alleen een beschrijving gegeven van de leerprestaties van derdeklassers in het voortgezet onderwijs, maar ook wordt een beschrijving gegeven van de wijze waarop het aardrijkskunde-onderwijs in de derde klas gegeven wordt. Daarnaast wordt in deze peiling van het onderwijsniveau getracht een relatie te leggen tussen de leeruitkomsten en de instructie-variabelen. Vandaar dat de centrale vraag naar de convergente validiteit van de drie meetmethoden voor de tijdsbesteding (de vragenlijst, het logboek en de observaties) opgesplitst kan worden in drie deelvragen. Ten eerste, is een schatting van de gemiddelde tijdsbesteding afhankelijk van de methode? Ten tweede, zijn de schattingen van de varianties (tussen docenten) afhankelijk van de onderzoeksmethode? En, ten derde hoe hoog is de samenhang tussen de metingen van de tijdsbesteding met de verschillende instrumenten.

De eerste twee onderzoeksvragen zijn met name relevant voor onderzoek waarin het beschrijvende element centraal staat (zoals peilingsonderzoek). Dan wordt immers inzicht gegeven in de gemiddelde tijdsbesteding en de verschillen tussen docenten (varianties). De derde en laatste onderzoeksvraag is met name van belang indien men een relatie wil leggen tussen de tijdsbesteding en de leerprestaties.

ONDERZOEKSOPZET

Instrumenten

Het observatieschema is het meest centrale instrument, daar zowel het logboek als de vragenlijst hier een afgeleide van zijn. Met behulp van het observatieschema kan het instructie-gedrag van de docent beschreven worden. Uitgaande van het model van didactisch handelen van De Corte (1981) zijn in het observatieschema acht gedragscategorieën onderscheiden: Uitleg geven, vragen stellen, vragen beantwoorden, leerlingen stellen vragen, het geven van opdrachten, het handhaven van de orde, organisatie en management, en een restcategorie.

Voor twee categorieën, het geven van uitleg en het stellen van vragen is een onderscheid gemaakt of de docent al dan niet gebruik maakt van hulpmiddelen als het bord, kaarten, dia's et cetera. We noemen dit het gebruik van media.

Daarnaast is voor de eerste vijf gedragscategorieën (uitleg geven, vragen stellen, vragen beantwoorden, leerlingen stellen vragen en het geven van opdrachten) een onderscheid gemaakt met betrekking tot het publiek waartoe de docent zich richt: de hele klas (frontaal-klassikaal), een groep (van minimaal twee leerlingen), en individuele leerlingen.

Voor dezelfde inhoudelijke gedragscategorieën is ook een (grof) aardrijkskundig onderscheid in tijdsbesteding gemaakt. Bijgehouden is of de tijd besteed is aan sociaal-geografische dan wel aan fysisch-geografische onderwerpen. Bij sociale geografie staat de mens centraal, die geleid door zijn streven naar welzijn en welvaart, de omgeving inricht en zonodig verandert, terwijl bij fysische geografie natuurlijke verschijnselen van het aardoppervlak centraal staan. Dit observatie-schema is in Figuur 1 schematisch weergegeven.

Zowel het logboek als de vragenlijst zijn een directe afgeleide van dit observatieschema. Ter illustratie is in Figuur 2 de vraag over frontaal-klassikaal lesgeven weergegeven.

	Frontaal		Groeps.		Indiv.	
	sociaal	fysisch	sociaal	fysisch	sociaal	fysisch
Uitleg geven: met media						
Uitleg geven: zonder media						
Vragen stellen: met media						
Vragen stellen: zonder media						
Vragen beantwoorden						
Leerlingen stellen vragen						
Opdrachten geven						
Orde						
Organisatie en management						
Rest						

Fig. 1. Een schematische weergave van het observatieschema.

Hoe heeft u de afgelopen les, uw tijd besteed aan frontaal-klassikaal lesgeven, inhoudelijk uitgesplitst naar fysische en sociale geografie?
Onder frontaal-klassikaal verstaan wij: de docent richt zich tot de gehele klas, tot alle leerlingen, en zijn acties zijn van verbale aard.

Frontaal-klassikaal	Sociaal	Fysisch	
Uitleg geven: met hulpmiddelen	——	——	in minuten
Uitleg geven: zonder hulpmiddelen	——	——	in minuten
Vragen stellen: met hulpmiddelen	——	——	in minuten
Vragen stellen: zonder hulpmiddelen	——	——	in minuten
Vragen beantwoorden	——	——	in minuten
Leerlingen stellen vragen	——	——	in minuten
Opdrachten geven	——	——	in minuten

Fig. 2. Een voorbeeld van de vragen uit het logboek.

De tijdregistratie tijdens de observaties is bijgehouden met een voor dit doel geschreven programma. Het programma maakt het mogelijk gelijktijdig plaatsvindende gebeurtenissen simultaan te registreren door middel van het aanslaan van verschillende toetsen tegelijkertijd. De pc registreerde dan de tijd die aan de met deze toetsen corresponderende categorieën besteed werd. In feite is hier dus een combinatie van 'event-' en 'time-sampling' toegepast.

Het logbock is direct na afloop van de les, onder toezicht van de observator, door de docent ingevuld.

De vragenlijst is na afloop van de observatie-periode aan de docent toegestuurd, met het verzoek deze in te vullen.

Gegevens

Het onderzoek is, zoals gezegd, uitgevoerd in het kader van een peilingsonderzoek naar de inhoud en het niveau van het voortgezet onderwijs voor het vak aardrijkskunde. Aan dit deelonderzoek naar de convergente validiteit van verschillende metingen van de tijdsbesteding hebben in totaal twaalf mavo-docenten meegewerkt.

Bij alle twaalf docenten zijn in één maand vier willekeurige aardrijkskunde lessen aan in één klas geobserveerd. In totaal zijn (12 x 4 =) 48 lessen geobserveerd. Om enig inzicht in de betrouwbaarheid van de observaties te verkrijgen zijn vier lessen van verschillende docenten door twee onafhankelijke observatoren geobserveerd. De overeenstemming tussen hun waarnemingen bleek redelijk [r = .85; N = 4 (lessen) x 11 (onderdelen) = 44].

Na elk van de hierboven genoemde lessen is door de docenten ook een logboek ingevuld. Dit bevatte 23 vragen over de tijdsbesteding in de zojuist gegeven les. Vanzelfsprekend zijn de gestelde vragen overeenkomstig met de geobserveerde onderdelen uit het observatieschema. In totaal zijn (12 x 4 =) 48 logboeken verzameld.

Na afloop van de observatie-periode is een vragenlijst aan de docenten toegestuurd. In deze vragenlijst, die vrijwel gelijk was aan het logboek, is verzocht de vragen met betrekking tot de tijdsbesteding te beantwoorden over de onderzoeksperiode. Alle twaalf docenten hebben de vragenlijst ingevuld geretourneerd.

Analyse

Er is in dit onderzoek sprake van een geneste steekproef. De geobserveerde lessen zijn te beschouwen als een steekproef uit de populatie lessen van een docent. Hetzelfde geldt mutatis mutandis voor de logboekgegevens; deze hebben betrekking op een steekproef van lessen uit de populatie lessen van een docent.

De waarnemingen van verschillende lessen van één docent zullen nauwelijks als onafhankelijke waarnemingen van de tijdsbesteding beschouwd kunnen worden; waarnemingen van verschillende lessen van één docent zullen meer overeenkomsten vertonen dan waarnemingen van lessen van verschillende docenten. Om met deze afhankelijkheid rekening te houden zijn de gegevens geanalyseerd met behulp van een multiniveau-model. Daarnaast heeft een multiniveau-model het voordeel dat niet aan de stringente aannamen van meer traditionele analyse modellen (zoals bij voorbeeld: variantie-analyse) voldaan behoeft te worden.

Stel, Y_{ij1} is de geobserveerde tijdsbesteding voor een onderscheiden aspect in les i ($i = 1, 2, 3, 4$) van docent j ($j = 1, 2, ..., 12$), Y_{ij2} is de met behulp van logboeken geregistreerde tijd voor dit aspect in les i van docent j, en Y_{j3} is de met behulp van de vragenlijst gemeten tijdsbesteding aan dit aspect van docent j. Het te analyseren model kan dan met de volgende drie vergelijkingen weergegeven worden:

$$\begin{aligned} Y_{ij1} &= \beta_1 + e_{ij1} + \mu_{0j1}, \\ Y_{ij2} &= \beta_2 + e_{ij2} + \mu_{0j2}, \\ Y_{j1} &= \beta_3 + \mu_{0j3}. \end{aligned} \qquad (1)$$

In de bovenstaande vergelijkingen representeren β_1, β_2 en β_3 de gemiddelde tijdsbesteding voor respectievelijk de observaties, de logboeken en de vragenlijst. Er worden verschillende residuele scores onderscheiden: met e_{ij1} worden de residuele scores voor les i van docent j aangegeven zoals gemeten met de observaties, en met e_{ij2} de zelfde residuele score voor de logboekmetingen. Met behulp van de vragenlijst, die slechts eenmaal ingevuld is, kunnen vanzelfsprekend geen verschillen tussen lessen geschat worden; een residuele score voor les i van docent j kan derhalve niet geschat worden.

Het tweede type residuele score (de μ's), representeert de residuele score voor docent j. Er worden op het tweede niveau drie residuele scores onderscheiden; voor alle drie de meetinstrumenten één: μ_{0j1}, μ_{0j2} en μ_{0j3}.

Door Y_{ijk} te beschouwen als een multivariate variabele kunnen, met behulp van dummy-variabelen, de drie vergelijkingen als één vergelijking geschreven worden. Hierbij geeft de index k ($k = 1, 2, 3$) de meetmethode aan.

$$Y_{ijk} = O_{ijk} (\beta_1 + e_{ijk} + \mu_{0jk}) + L_{ijk} (\beta_2 + e_{ijk} + \mu_{0jk}) + V_{jk} (\beta_3 + \mu_{0jk}). \qquad (2)$$

In de bovenstaande vergelijking zijn O_{ijk}, L_{ijk} en V_{jk} de dummy-variabelen die aangeven met welk instrument de tijdsbesteding voor het aspect is gemeten, respectievelijk: observaties, logboeken en de vragenlijst. (Dat wil zeggen: de dummy-variabele voor bij voorbeeld O_{ijk} is gelijk aan één indien een score een observatie betreft en nul in alle andere gevallen, et cetera).

Vergelijking van de drie gemiddelden ($\beta_1 * O_{ijk}$, $\beta_2 * L_{ijk}$ en $\beta_3 * V_{jk}$) geeft een antwoord op de eerste onderzoeksvraag, naar de convergentie van de gemiddelde tijdsbesteding tussen de drie meetmethoden, en verschillen tussen de gemiddelden kunnen getoetst worden met behulp van een chi-kwadraat verdeelde toetsingsgrootheid (Goldstein, 1987, p. 29).

We nemen aan dat de residuen op het eerste niveau ($O_{ijk} * e_{ijk}$ en $L_{ijk} * e_{ijk}$) normaal verdeeld zijn met een gemiddelde van nul, en een variantie van respectievelijk $\delta^2_{e,O1}$ en $\delta^2_{e,L1}$. Bovendien nemen wij aan dat zij ongecorreleerd zijn. De variantie van deze residuen geeft inzicht in de tussen-lessen-variantie.

Op het tweede niveau worden de drie residuele scores onderscheiden die de docent karakteriseren ($O_{ijk} * \mu_{0jk}$; $L_{ijk} * \mu_{0jk}$; $V_{jk} * \mu_{0jk}$). Deze residuele scores zijn bij aanname normaal verdeeld met een gemiddelde van nul, en een variantie van respectievelijk δ^2_{O2}, δ^2_{L2} en δ^2_{V2}.

Vanzelfsprekend nemen we aan dat op dit niveau de drie residuen gecorreleerd zijn, zodat we op het tweede niveau de volledige covariantie matrix schatten.

De varianties op niveau 2 geven inzicht in de verschillen tussen docenten in de tijdsbesteding per aspect. En, na toetsing van eventuele verschillen tussen deze schattingen kan de tweede onderzoeksvraag beantwoord worden.

De covarianties op niveau twee kunnen eenvoudig omgerekend worden tot correlaties ($r_{a,b} = \delta_{a,b} / \sqrt{[\delta^2_a * \delta^2_b]}$). Aan de hand van deze correlaties kan een antwoord gegeven worden op de derde onderzoeksvraag naar de samenhang tussen de scores op de drie meetinstrumenten. Idealiter zijn deze correlaties gelijk aan één. Dat wil zeggen, indien de meetinstrumenten convergent valide zijn, dan zouden de correlaties niet significant van één behoren af te wijken.

Tot slot kan met behulp van de niveau 1 en de niveau 2 varianties een schatting gemaakt worden van de betrouwbaarheid van de tijdsbestedingsmeting met behulp van de observaties en de logboeken. Als we de betrouwbaarheid aangeven met ρ, dan kan de betrouwbaarheid geschat worden als $\rho = [\delta^2_\mu / (0.25 * \delta^2_e + \delta^2_\mu)]$ (vgl. Rowan, Raudenbush & Jin King, 1992, p. 213). Hierbij geeft δ^2_e de niveau 1 variantie voor de meetmethode aan, terwijl δ^2_μ de niveau 2 variantie van diezelfde meetmethode aangeeft.

RESULTATEN

Voor de vergelijking van de tijdsbesteding zoals gemeten met de drie methodes zijn elf variabelen geconstrueerd. Het betreft: de totale lestijd die effectief aan aardrijkskunde is besteed (1), de totale lestijd die aan frontaal-klassikaal onderwijs (2) groepsgewijs (3) en individueel onderwijs is besteed (4), de totale tijd waarin gebruik gemaakt is van hulpmiddelen (5), en de tijd waarin onderwijs gegeven is zonder gebruik van media (5), de tijd besteed aan fysisch geografische (7) en sociaal geografische onderwerpen (8), de tijd dat leerlingen actief vragen gesteld hebben (9), en de tijd die verloren is gegaan door orde (10) en organisatie en management zoals het uitdelen van materialen, het klaarzetten van een dia-projector e.d. (11).

De gemiddelden en de standaarddeviaties zijn in Tabel 1 per meetmethode, per onderdeel weergegeven.

Uit de gegevens in Tabel 1 suggereren verschilien tussen de drie meetmethoden. Zo lijken de gemiddelden voor verschillende onderdelen afhankelijk van de meetmethode. Voor het onderdeel 'Frontaal onderwijs' bij voorbeeld verschilt de gemiddelde tijdsbesteding meer dan tien minuten tussen de geobserveerde tijd en de bij de vragenlijst ingevulde tijd, en voor het onderwijs in de sociale geografie bedraagt het verschil tussen de gemiddelde voor de genoemde instrumenten zelfs meer dan twintig minuten. Opvallend is voorts dat bij zowel de observaties

Tabel 1. Gemiddelden (in minuten) en standaarddeviaties (tussen haakjes) per onderdeel per meetmethode.

	observaties	logboeken	vragenlijst
Totaal ak-onderwijs	32.0 (3.3)	41.0 (15.0)	49.0 (26.3)
Frontaal onderwijs	33.4 (7.9)	35.1 (14.7)	43.5 (21.2)
Groepsgewijs onderwijs	2.2 (5.9)	6.4 (12.9)	13.7 (19.7)
Individueel onderwijs	1.3 (2.9)	5.5 (10.7)	10.5 (16.5)
Onderwijs met media	17.8 (8.4)	20.6 (14.4)	35.6 (27.8)
Onderwijs zonder media	9.9 (7.4)	9.2 (12.0)	8.7 (10.7)
Onderwijs soc. geogr.	25.1 (16.8)	32.6 (28.7)	46.6 (31.2)
Onderwijs fys. geogr.	11.7 (16.7)	15.5 (19.3)	27.1 (29.0)
Initiatieven van ll.	2.4 (2.4)	12.0 (12.8)	23.8 (14.4)
Handhaven orde	1.0 (0.9)	3.4 (2.0)	3.8 (1.5)
Organis. en management	2.2 (2.3)	3.0 (2.8)	5.4 (1.9)

Tabel 2. Toetsingsgrootheid (χ^2) voor de vergelijking van de gemiddelden ($\overline{X}$) en varianties (δ^2) tussen de drie meetmethoden ($\chi^2 > 3.84$; df = 1; $p < .05$).

	observaties versus logboeken		observaties versus vragenlijst		logboeken versus vragenlijst	
	$\overline{X}$	δ^2	$\overline{X}$	δ^2	$\overline{X}$	δ^2
Totaal ak-onderwijs	7.1	4.2	22.2	5.9	9.2	4.1
Frontaal onderwijs	1.9	.9	11.1	5.6	7.7	4.7
Groepsgewijs onderwijs	1.6	4.5	4.9	6.1	1.4	2.6
Individueel onderwijs	2.7	2.8	10.0	6.0	6.9	4.2
Onderwijs met media	1.1	2.6	4.0	5.8	3.7	6.0
Onderwijs zonder media	.3	.5	.3	6.1	.2	6.0
Onderwijs soc. geogr.	1.6	2.7	12.7	4.0	8.8	.8
Onderwijs fys. geogr.	1.4	.4	3.9	3.7	3.9	4.1
Initiatieven van ll.	9.3	5.1	28.5	5.9	6.5	.6
Handhaven orde	31.6	.7	57.0	4.5	1.1	1.1
Organis. en management	1.0	1.8	22.3	2.5	14.6	.3

als bij de logboeken de som van diverse gemiddelden (bij voorbeeld: frontaal onderwijs + groepsgewijs onderwijs + individueel onderwijs) een redelijke benadering van de werkelijk tijdsbesteding *zou* kunnen zijn. Bij de tijdsbesteding volgens de vragenlijst daarentegen wordt bij een dergelijke sommatie duidelijk de totaal besteedbare lestijd duidelijk overschreden.

De resultaten van de analyse met betrekking tot de gemiddelden en de varianties van deze elf onderdelen zijn samengevat in Tabel 2. In deze tabel is voor de onderscheiden toetsingen een toetsingsgrootheid weergegeven. Onder H_0 is deze grootheid chi-kwadraat verdeeld met 1 vrijheidsgraad.

We concentreren ons eerst op de vergelijking van de gemiddelden van de drie verschillende meetmethoden. Uit Tabel 2 kunnen we opmaken dat voor twintig van de drieëndertig vergelijkingen er een significant verschil is tussen de gemiddelden van hetzelfde aspect tussen twee meetmethoden (i.e. $\chi^2 > 3.84$ bij 1 vrijheidsgraad, $p < .05$). Deze verschillen zijn niet gelijkelijk verdeeld over de vergelijkingen tussen de meetmethoden. Voor de oppositie observaties versus logboeken is slechts in drie van de elf vergelijkingen een verschil aangetoond, terwijl tien van de elf vergelijkingen tussen observaties en de vragenlijstgemiddelden een significant verschil te zien geven, en bij de vergelijking tussen de logboek- en vragenlijstgemiddelden is zeven van de elf keer een significant verschil aangetoond. Voor slechts één van de elf geanalyseerde aspecten kan geen verschil tussen de drie meetmethoden aangetoond worden: de tijd waarin onderwijs gegeven wordt zonder gebruik van media. Aan de andere kant is er alleen voor de totale tijd besteed aan aardrijkskunde-onderwijs, en de tijd voor initiatieven van de leerlingen een verschil tussen alle drie de gemiddelden.

Ook ten aanzien van de tussen-docenten-varianties zijn er duidelijke verschillen ten gevolge van de meetmethode aangetoond; achttien van de drieëndertig keer moet de nulhypothese van gelijke varianties voor eenzelfde onderdeel gemeten met verschillende methoden verworpen worden. Het aantal maal dat H_0 verworpen moet worden is ook hier niet gelijkelijk verdeeld over de drie vergelijkingen. Zo is er drie keer een verschil aangetoond tussen de observatie- en logboekvarianties, negen keer een verschil tussen de observatie- en vragenlijstvarianties en zes keer een verschil tussen de logboek- en vragenlijstvarianties. Alleen voor het al genoemde

Tabel 3. Boom- en takdiagram van de correlaties in tijdbesteding tussen de drie meetmethodes.

r	Observaties versus Logboek	Observaties versus Vagenlijst	Logboek versus Vragenlijst
1.0	00	0	
.9			
.8	13		
.7	8		5
.6			
.5			3
.4		1	24
.3		9	14
.2	9	7	
.1	000	0	48
.0	9	07	00
−.1	1	4	
−.2			4
−.3		23	
−.4			
−.5			
−.6			
−.7			
−.8			
−.9			
−1.0		0	

aspect 'totale tijd besteed aan aardrijkskunde-onderwijs' is er een verschil in tussen-docenten-varianties tussen alle drie de meetmethoden. En, enkel voor het aspect 'organisatie en management' konden geen verschillen in tussen-docenten-varianties ten gevolge van de meetmethode aangetoond worden.

De derde onderzoeksvraag betreft de covarianties tussen de meetmethoden. Per aspect is er een 3 x 3 covariantiematrix tussen de meetmethoden geschat (zie Vergelijking 1). Ter wille van een meer eenvoudige interpretatie zijn deze covarianties gestandaardiseerd. De resulterende correlaties zijn samengevat in Tabel 3.

In het bovenstaande boom- en takdiagram wordt de boom gevormd door de eerste decimaal, en de tak door de tweede decimaal. Bij de correlaties tussen de observaties en het logboek is dus één keer .29, en drie keer .10 geobserveerd.

De geobserveerde correlaties zijn zoals in Tabel 3 te zien is over het algemeen nogal laag. Er zij acht uitzonderingen. Vijf hiervan betreffen de correlatie tussen observaties en logboekgegevens, twee de correlatie tussen observaties en de vragenlijst en één de correlatie tussen de logboek- en vragenlijstgegevens.

De correlatie tussen de geobserveerde tijdsbesteding en de logboekregistraties is hoog voor: frontaal- en groepsgewijsonderwijs (1.0), fysische geografie (.83), sociale geografie (.81) en de tijd dat leerlingen vragen stellen (.78); bij deze vijf correlaties kan niet aangetoond worden dat zij afwijken van één. Voor de overige zes aspecten is de correlatie laag, en met één uitzondering wijkt deze niet significant van nul af.

Bij de vergelijking van de geobserveerde tijdsbesteding en die volgens de vragenlijst blijkt dat de correlaties variëren van +1.0 (groepsgewijsonderwijs) tot -1.0 (onderwijs zonder gebruik van media). In het algemeen zijn de correlaties geclusterd rond 0.0.

Tabel 4. Boom- en takdiagram van de betrouwbaarheidsschattingen (ρ) voor de observaties en de logboek gegevens.

ρ	observaties	logboeken
1.0	0	
.9		2
.8	44	3779
.7	24	07
.6		
.5		
.4	0	7
.3	13	
.2		
.1		
.0	00	00

De correlaties tussen de logboekgegevens en de vragenlijstgegevens zijn over het algemeen laag positief, met uitzondering van .75 (onderwijs in de fysische geografie) en -.24 (totaal aan aardrijkskunde-onderwijs bestede tijd). Behoudens de genoemde correlatie voor onderwijs in de fysische geografie wijken alle correlaties af van 1.0.

In de bovenstaande tabel wordt een boom- en tak diagram van de betrouwbaarheidsschattingen voor de observaties en de logboekgegevens gepresenteerd. In deze is de betrouwbaarheid van de observaties en de logboeken weergegeven indien één les geobserveerd zou worden. Het is op grond van de betrouwbaarheidsschattingen uit Tabel 4 in één oogopslag duidelijk dat deze nogal varieert afhankelijk van het onderdeel in kwestie. De betrouwbaarheid van de observaties en de logboeken laten in twee gevallen overduidelijk te wensen over, voor: individu gericht onderwijs, en onderwijs zonder media. Ook voor het overige doen de betrouwbaarheidsschattingen voor de observaties en de logboekgegevens nauwelijks voor elkaar onder. De over aspecten gemiddelde betrouwbaarheid verschilt dan ook nauwelijks tussen de beide meetmethoden ($\bar{\rho}$ voor observaties en logboeken is respectievelijk .53 versus .62).

DISCUSSIE

In dit onderzoek is de convergente validiteit van drie methoden om de tijdsbesteding van docenten in de lessen aardrijkskunde in het voortgezet onderwijs te meten geëvalueerd: observaties, tijdregistraties met behulp van logboeken, en een traditionele vragenlijst. In totaal zijn elf verschillende aspecten van de tijdsbesteding onderzocht. Hierbij hebben drie vragen centraal gestaan: de convergentie van de gemiddelde tijdsbesteding (per aspect), de convergentie van de schatting van de tussen-docenten-variantie (per aspect) en de correlaties tussen de metingen met behulp van de drie meetmethoden. Tot slot is voor de observaties en de logboekgegevens ook aandacht aan de betrouwbaarheid van de meting geschonken.

Met betrekking tot de schatting van de gemiddelde tijdsbesteding zijn duidelijke methode-effecten aangetoond. Van convergentie tussen de vragenlijstgegevens enerzijds en de logboekgegevens en observaties anderzijds lijkt nauwelijks sprake. De gemiddelden voor de logboekgegevens en de observaties corresponderen daarentegen redelijk. Op zich is een discrepantie tussen de gemiddelden dit nog niet zorgelijk, aangezien we niet mogen aannemen dat de gegevens op ratio-niveau gerepresenteerd zijn. Hier kan dus sprake zijn van louter een schalingsprobleem. Opmerkelijk is ook dat bij de vragenlijst sommatie over verschillende categorieën een

duidelijke overschrijding van de totaal te besteden lestijd te zien geeft. Kennelijk is lastig om een (precieze) schatting te geven van de tijd die gemiddeld aan een onderdeel wordt besteed.

Ten aanzien van de schattingen van de tussen-docenten-varianties zijn ook aanzienlijke methode-effecten aangetoond. Ook hier is een gebrek aan convergentie tussen enerzijds de vragenlijstgegevens en de logboekgegevens en anderzijds de vragenlijstgegevens en de observaties te constateren. Terwijl de metingen met behulp van de beide laatste methodes (logboekgegevens en observaties) wel enige convergentie te zien geven. Ook dit gegeven op zich is nog niet problematisch. Immers, als we er van uitgaan dat ten hoogste gemeten is op intervalniveau (en, met name dan voor de vragenlijst), dan zijn ook de varianties in principe betekenisloos (Suppes & Zinnes, 1963).

Ook wanneer we de correlaties tussen de meetmethoden in ogenschouw nemen, is het beeld niet al te rooskleurig. De correlaties zijn in veel gevallen laag. Hoewel er met name tussen de observaties en de logboekgegevens enkele duidelijke hoge uitschieters te melden zijn. Alle drie de onderzochte onderdelen tezamen geven o.i. wel enige reden tot zorg; noch de gemiddelden, noch de varianties, noch de covarianties (correlaties) blijken duidelijk te convergeren, en juist het meest frequent gebruikte instrument blijkt duidelijk af te wijken. De verschillen tussen de vragenlijst en de beide andere meetinstrumenten voor de directe tijdsbesteding lijkt niet teruggevoerd worden te kunnen worden tot een schaalprobleem.

Het is opmerkelijk dat de voor instructie-effectiviteitsonderzoek waarschijnlijk belangrijkste variabele , 'de effectieve onderwijstijd', op geen van de onderzochte onderdelen convergeert over methoden.

De betrouwbaarheid van de metingen is onderzocht voor de observaties en de logboekgegevens. Deze lijkt afhankelijk van het aspect in kwestie; voor het ene aspect overheersen de verschillen tussen docenten, terwijl voor een ander aspect de verschillen tussen lessen overheersen. In dit laatste geval is natuurlijk nauwelijks sprake van docent-gebonden verschillen, met een lage betrouwbaarheid als consequentie.

Niet voor alle onderzochte aspecten lijkt het mogelijk om een voor onderzoeksdoeleinden redelijke betrouwbaarheid (van bij voorbeeld .80) te verkrijgen. Uitgaande van parallelle toetsverlenging lijkt het echter voor de meeste onderdelen voldoende om gedurende vier lessen logboeken in te laten vullen, dan wel vijf lessen te observeren.

Een voorlaatste opmerking die wij willen maken betreft de power. Slechts twaalf docenten hebben aan dit onderzoek meegewerkt. De a priori kans om verschillen tussen docenten ten gevolge van de meetmethode aan te kunnen tonen (de power) is derhalve laag. Of, anders gezegd, alleen redelijk grote verschillen zullen een aanleiding geven voor (een significant) verschil tussen de drie meetmethoden. Dat desalniettemin veel van de getoetste effecten significant bleken, lijkt er op te wijzen dat het hier substantiële effecten betreft.

Zowel de schattingen van de gemiddelden als de schatting van de varianties voor de vragenlijst voor de directe tijdsbesteding wijken sterk af van die van de beide andere meetmethoden. De vraag in hoeverre het zinvol is om de directe tijdsbesteding van docenten met een vragenlijst te meten wint derhalve aan belang. Op grond van de resultaten van dit onderzoek zijn wij geneigd vraagtekens te plaatsen voor bij het nut van vragenlijsten voor de meting van de directe tijdsbesteding in het voortgezet onderwijs. Over het nut van gegevens verkregen op basis van een vragenlijst voor de directe tijdsbesteding in de analyse van instructie-prestatie-relaties kunnen wij derhalve niet al te optimistisch zijn.

Conform de bevindingen in het basisonderwijs, waar één docent verschillende vakken geeft, blijkt ook in het voortgezet onderwijs, waar in tegenstelling tot het basisonderwijs één docent slechts één vak (of hooguit enkele vakken) geeft, een vragenlijst niet al te valide tijdsbestedingsgevens op te leveren. Misschien kan de tijdsbesteding beter geëvalueerd worden met behulp van minder precieze vragen, zoals vragen naar het aantal lessen waarin een onderwerp behandeld is. Echter, aan de ene kant missen we dan één van de meest centrale variabelen uit het instructie-

effectiviteitsonderzoek: effectieve lestijd (zie o.a. Scheerens, 1989). En, aan de andere kant zijn er ook onderzoeksresultaten op grond waarvan aan de kwaliteit van een dergelijke meer globale maat getwijfeld kan worden (zie bij voorbeeld: Overmaat, 1992).

LITERATUUR

Corte , E. de (1981). *Beknopte didaxologie.* Groningen: Wolters-Noordhoff.

Creemers, B.P.M. (1991). *Effectieve instructie; Een empirische bijdrage aan de verbetering van het onderwijs in de klas.* Den Haag: SVO (Balans van Onderwijsonderzoek).

Goldstein, H., (1987), *Multilevel models in educational and social research.* London: Charles Griffin & Company ltd.

Harskamp, E. & Suhre, C. (1993). Psychometrische kwaliteiten van meetinstrumenten van het onderwijsaanbod. *Tijdschrift voor Onderwijsresearch, 17,* 355-369.

Lugthart, E., Roeders, P.J.B., Bosker, R.J. & Bos, K.T. (1989). *Effectieve schoolkenmerken in het voortgezet onderwijs: Een literatuuroverzicht.* Groningen: Instituut voor Onderwijsonderzoek.

Otter, M., Mellenbergh, G., Blok, H. & Bergh, H. van den (1992). Measuring time spent on teaching native language. *Journal of Experimental Education, 60,* 281-287.

Overmaat, M. (1992). *De kwaliteit van instrumenten voor het aanbod van taal en lezen.* Amsterdam: SCO.

Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In: Travers, R.M.W. (Ed.), *Second handbook of research on teaching* (122-183). Chicago: Rand McNally College Publishing Company.

Rowan, B., Raudenbush, S.W. & Jin King, S. (1992). School climate in secondary schools. In, S.W. Raudenbush & D. Willms (Eds.), *Schools, classrooms and pupils: International studies of schooling from a multilevel perspective.* New York: Academic Press.

Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten.* Den Haag: SVO (Balans van Onderwijsonderzoek).

Suppes, P. & Zinnes, J.L. (1963). Basic measurement theory. In: R.D. Luce, R.R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I). New York: Wiley.

Manuscript ontvangen 22-6-1993
Definitieve versie ontvangen 21-10-1993

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres
Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen,
tel. 045-762293

Jrg. 17, nr. 2 1993

Inhoudsopgave

Congressen, Symposia, Cursussen, Trainingen 318

Congressen, Symposia, Cursussen, Trainingen

De **Onderwijs Research Dagen 1994** zullen worden georganiseerd van **24 t/m 26 mei 1994** in de Jaarbeurs te Utrecht door de vakgroep Onderwijskunde van de Universiteit Utrecht, onder auspiciën van de VOR. De ORD is georganiseerd rond een negental thema's, waarvan er acht gekoppeld zijn aan de divisies van de VOR. Het negende thema, Basisvorming, is aangedragen door de organiserende vakgroep.

U kunt zich voor de ORD '94 inschrijven middels een formulier dat, evenals de folder en de aanmeldingsformulieren voor presentatievoorstellen, is op te vragen bij:

FBU Congresbureau/Universiteit Utrecht
T.a.v. Mw. Loes Donkers
Heidelberglaan 8
3584 CS Utrecht
Tel.: 030-535044/1642
Fax: 030-533667

De inschrijvingskosten zijn bij betaling vóór 15 april 1994 als volgt:

VOR-leden	*f* 300,–
AiO's tevens VOR-lid	*f* 225,–
niet VOR-leden	*f* 380,–
AiO's niet VOR-leden	*f* 300,–
Studenten	*f* 175,–

De vermelde bedragen zijn inclusief: toegang en deelname aan de ORD, congresbundel, koffie/thee/lunches. De betaling geschiedt middels een na inschrijving toegezonden acceptgirokaart en dient vóór 15 april '94 te zijn voldaan. Wie zich daarna inschrijft betaalt *f* 50,– extra.

De ORD'94 begint op dinsdagavond 24 mei 1994 in het Academiegebouw en eindigt om 17.00 uur op donderdag 26 mei in de Jaarbeurs te Utrecht. De organisatiecommissie hoopt u op de ORD welkom te kunnen heten.

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, p. 321

Mededeling van de uitgever

Jaargang 19 (1994) van het *Tijdschrift voor onderwijsresearch* zal in 4 nummers verschijnen in plaats van in 6. Het aantal bladzijden voor de gehele jaargang blijft echter wel 384. Dit betekent een gemiddelde van 96 bladzijden per nummer.

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, pp. 322-330

Thinking aloud: does it affect regulatory processes in learning?

Marcel V.J. Veenman [1], Jan J. Elshout[2] and Martin G. M. Groen[2]
[1]*Dept. of Educational Psychology, Leiden University* and*
[2]*Dept. of Psychonomics, University of Amsterdam*

ABSTRACT

The objective of this study was to determine to what extent thinking aloud during the performance of a learning task interferes with metacognitive or regulatory processes and learning performances. Forty first-year psychology students performed a discovery learning task in a computer simulated Heat lab while thinking aloud or working silently. All student activities in Heat lab were registered in logfiles on the computer. These logfiles were analyzed on performance parameters relevant to regulatory processes. Two posttests were administered in order to obtain measures of learning. The results show that thinking-aloud subjects do not differ from control subjects with regard to logfile measures of regulatory processes and measures of learning. A significant difference between thinking-aloud and control subjects was found for Time on task. Thinking aloud appeared to slow down performance on the learning task, however without altering the ongoing processes.

INTRODUCTION

The thinking-aloud method is being used frequently in cognitive research nowadays for gathering verbal data during task performance. The aim of the present study is to determine whether thinking aloud concurrent to task performance affects the metacognitive or regulatory processes involved in task performance (and consequently affects the learning outcomes). The issue is a methodological one, not one arising from an educational point of view. Our concern is the validity of the thinking-aloud method for assessing metacognitive skills in learning processes.

Due to an incorrect association with the technique of introspection, the thinking-aloud method has often been regarded as suspect (Ericsson & Simon, 1980). However, introspection requires subjects to observe and interpret the contents of consciousness, whereas thinking aloud merely is the verbalization of task related information in working memory that is attended to during task performance (Elshout & van Leeuwen, 1992). Because thinking aloud occurs simultaneously with the actual task performance, the risk of subjective interpretations is far less.

Even so, the possibility that the activity of verbalization, performed as a secondary task concurrent to the primary task, may interfere with ongoing thought processes is not far-fetched (Breuker, Elshout, van Someren, & Wielinga, 1986; Elshout & van Leeuwen, 1992). Because working memory has a limited capacity, there may be a competition between processes involved in performing the primary task and those involved in performing the secondary task of verbalization (Elshout, 1976; Ericsson & Simon, 1980, 1984). This may be the case if the primary task is extremely demanding or if verbalization requires a lot of attention.

The extent to which verbalization interferes with performance on the primary task depends on the verbal fluency of subjects (Elshout, 1976; Elshout & van Leeuwen, 1992). Training subjects in thinking aloud reduces the cognitive load of verbalization, presumably by automatizing the verbalization process (Breuker et al., 1986). However, verbalization may be very demanding if the mode of information to be verbalized is not semantic or symbolic (Elshout, 1976; Ericsson & Simon, 1984). For instance, the verbalization of figurative information that is hard to put into words, will obviously demand a lot of working memory capacity.

*Adres: Wassenaarseweg 52, 2333 AK Leiden, Tel. 071-27.3463, E-mail. veenman@rulfsw.LeidenUniv.nl

The extent to which performing the primary task makes high demands on working memory depends on the nature of the task. Tasks that require subjects to focus their attention on several memory items simultaneously, or tasks that are performed under time pressure are less suited to the purpose of thinking aloud (Breuker et al., 1986; Elshout & van Leeuwen, 1992). However, for most tasks with no time pressure verbalization only tends to slow down performance on the primary task without changing the course and structure of the cognitive processes involved (Ericsson & Simon, 1980).

Finally, the subject's level of expertise in performing the primary task may affect the process of thinking aloud (Breuker et al., 1986). Expert reasoning proceeds rapidly, leaving little opportunity of verbalizing the contents of working memory. Some processes may even be entirely automatized and inaccessible to verbalization (van Someren, Barnard, & Sandberg, 1993). Novices on the other hand, may need all working memory capacity at the expense of verbalization in order to perform the primary task. The most complete thinking-aloud protocols are obtained from moderately advanced subjects (Elshout & van Leeuwen, 1992). Apparently, the task difficulty relative to the subject's level of expertise determines the grain of analysis in the thinking-aloud method.

Although many experiments have been performed to study the effects of verbalization on the primary task, only few were carefully designed to expose the experimental and control groups to identical conditions except for thinking-aloud instructions (for a review, see Ericsson & Simon, 1980, 1984). Furthermore, in most studies reported by Ericsson and Simon only task performance (e.g. the number of problems solved) and performance speed were compared for subjects thinking aloud vs. silent subjects. Occasionally, more detailed comparisons of thought processes were made. For instance, Newell & Simon (1972) compared the solution paths of subjects discovering proofs in propositional logic. Except for performance speed, no differences were found between thinking-aloud and control groups in these studies if all other conditions were identical.

However, these studies only concerned the cognitive parameters of the problem solving process. In educational research the thinking-aloud method often has been used for assessing the metacognitive or self-regulatory skillfulness of subjects (de Jong & de Bil, 1989; Veenman & Elshout, 1991, 1992; Vermunt, Lodewijks, & Simons, 1986) and their approach to problem solving (Jansweijer, Elshout, & Wielinga, 1990; de Jong & Ferguson-Hessler, 1984; Span & Overtoom-Corsmit, 1986; Veenman, Elshout, & van Hell, 1992). Therefore, a first objective of this study is to determine to which extent regulatory processes in learning are affected by thinking aloud concurrent to performing the learning task. Because regulatory processes are part of the processes competing for working memory capacity, verbalization may or may not interfere with self-regulation. Most likely, thinking aloud will pose a serious threat to regulatory processes only if working memory is overloaded. However, thinking aloud may also encourage subjects to focus on one's own behavior, leading to more reflection on that behavior and, possibly, to a higher investment in control. In that case thinking aloud would lead to an improvement of regulatory processes. For instance, research by van Someren and Elshout (1985) showed that learning was fostered by reflecting on one's own problem solving processes during task performance.

As learning outcomes are highly related to the quality of regulatory processes (Veenman, 1993; Veenman & Elshout, 1992), these outcomes may be affected similarly by thinking aloud during the learning task. Therefore, a second objective of this study is to compare learning outcomes of thinking-aloud vs. control subjects. It is hypothesized that merely thinking aloud during task performance (viz. without interpreting or reflecting on one's thoughts) does not affect regulatory processes and learning outcomes. However, thinking aloud may be expected to stretch the time on task, even if verbalization leaves the cognitive and metacognitive processes unaffected.

METHOD

Overall design

Three groups of subjects were included in this experiment: a thinking-aloud group, a control group working silently, and a reference group. Effects of thinking-aloud were examined by contrasting the thinking-aloud group with the control group. As the control group worked silently, metacognitive skills could not be assessed from thinking-aloud protocols. Therefore, the metacognitive skills of both thinking-aloud and control subjects were assessed from logfiles that contained information about all student activities during task performance (see below). In order to determine which of the performance parameters in these logfiles were relevant to metacognitive or regulatory processes, another group of subjects was included as the reference group. Subjects in the reference group performed the same task while thinking aloud. For these subjects both the thinking-aloud protocols and the logfiles were analyzed and compared for selection of the logfile measures that would adequately predict metacognitive skill assessed from the thinking-aloud protocols. This selection of relevant logfile measures was not based on the data of the thinking-aloud group as to avoid capitalization on chance.

Subjects

Forty first-year psychology students signed up for participation in the experiment. They were randomly assigned to either the thinking-aloud condition or the control condition. Eventually, twenty-one subjects participated in the thinking-aloud condition while nineteen subjects participated in the control condition. The two experimental groups did not differ in sex or age (two-third of the subjects were female while one-third was male; mean age was 22.5 years).

Another twenty-eight subjects from the experiment of Veenman, Elshout and Hoeks (1993) were included in the study as the reference group. For reasons beyond the scope of this article these subjects had been selected from another class of first-year psychology students on the basis of a relatively higher or lower level of intellectual ability. More details are provided by Veenman (1993) en by Veenman et al. (1993).

Learning environment

A computer simulated Heat lab was adapted from Veenman et al. (1993). In this simulated Heat lab several objects of different weights and materials could be moved around and could be heated on a burner. The amount of heat transferred to an object was to be regulated with a time-switch and could be read off a joules-meter. Temperature was measured by attaching a thermometer to an object. Thus, the laboratory contained the required means for examining the relationship between heat and temperature depending on weight and material. In Heat lab subjects were required to solve some broadly defined problems by designing their own experiments. An instruction of a subtask was for instance: "Find out what is the relationship between weight, heat and temperature." Altogether, five subtasks were presented.

Simulations, used as an instructional tool, allow for learning by discovery under restricted realistic conditions. Because of its exploratory nature, learning with simulations involves complex problem solving and inductive reasoning, which puts a high cognitive demand on the student (de Jong, 1991). As this learning task may be expected to put some strain on working memory, it is an appropriate task for studying effects of thinking aloud.

Procedure

All subjects in the present experiment passed through Heat lab. Subjects in the thinking-aloud condition were instructed to think aloud while working in Heat lab, viz. to verbalize their activities and thoughts. They were not encouraged to interpret or reflect upon these activities and thoughts. Whenever they lapsed into silence the experimenter prompted them to continue thinking aloud. All thinking-aloud protocols were tape-recorded and subjects were unaware that these protocols were never transcribed. Obviously, subjects in the control condition did not receive thinking-aloud instructions. Their sessions were not tape-recorded and the experimenter

was unobtrusively present. Except for the thinking-aloud instructions, the treatment in both conditions was identical. All activities performed and measurements taken in Heat lab were registered by the computer in a logfile. Student notes, if any, were preserved in order to be analyzed afterwards. Time on task was not controlled, but was included in the study for post-experimental inspection. A Knowledge test assessing declarative knowledge about heat theory (see also Veenman et al., 1993) was administered in a pretest-posttest design. After completion of the Posttest subjects were asked to solve three problems (Problems) while thinking aloud.

Metacognitive skills

A student's logfile contained information about which block was moved where to, about which thermometer was attached to or detached from which block, about the adjustment of the burner, about starting the burner, about changes in the temperature of a block, about re-reading task instructions, and about requesting help from the computer with regard to operating Heat lab. From these logfiles the complete course of an experiment could be reconstructed for each subject. Therefore, the logfiles could be used for counting activities as well as for some judgemental analyses.

It was not known beforehand which logfile measures would be relevant predictors of metacognitive skill. However, for subjects from the reference group (adopted from Veenman et al., 1993) the quality of their Working Method had been assessed through the analyses of thinking-aloud protocols. Working Method may be regarded as a set of higher-order (metacognitive) processes that regulate and control the cognitive processes in problem solving and learning (Elshout, 1990; Veenman, 1993; Veenman & Elshout, 1992; Veenman, Samarapungavan, van Hout-Wolters, & Beishuizen, 1992). For subjects in the reference group Working Method had been assessed by judging the Heat lab protocols on indications of analyzing the problem statement thoroughly, planning before acting, working systematically according to an action plan (i.e. systematical orderliness), monitoring and checking, the quality of note taking, and elaboration activities such as drawing conclusions. An exhaustive discussion of scoring method and scoring criteria for Working Method is presented elsewhere (Veenman, 1993; Veenman & Elshout, 1992; Veenman et al., 1993). As the logfiles of subjects from the reference group were available too, measures obtained from their logfiles could be related to their Working Method scores obtained through protocol analysis. Hence, it could be determined which of the logfile measures represented metacognitive skills.

Several activities were scored objectively from the logfiles of all subjects, including those from the reference group. 'Burner-on' was scored as the total number of times that the burner was started for heating a block. Burner-on may be regarded as the total number of observations (possibly) taken by a subject. 'No-initT' and 'No-finalT' concerned the number of times a subject failed to measure the initial and final temperature respectively. Not measuring the initial or final temperature may be regarded as an indication of a poor systematical orderliness, because it would lead to incomplete observations. 'Block-abs' referred to the absolute number of blocks heated (including blocks that were heated repeatedly), whereas 'Block-rel' concerned the (relative) number of *different* blocks heated out of the nine possible blocks. The number of blocks heated may be regarded as an indication of the number of data points generated. 'Re-reading' concerned the number of times a subject asked the computer for the text of the task instruction. Re-reading was expected to indicate re-orientation. Finally, 'Help' was scored by the number of times help was requested from the computer with regard to operating Heat lab. A high level of Help may indicate that subjects did not process the introduction on operating Heat lab adequately, possibly due to the demand of reading aloud these instructions in the thinking-aloud condition.

Furthermore, all logfiles were judged on indications of Systematical Orderliness. Two judges performed these analyses together, arguing until agreement was reached. As the logfiles were presented in a randomly mixed order without indications of the subject's identity, the judges were unable to determine to which group (reference group, thinking-aloud group, or control group) the subjects belonged. Systematical Orderliness was assessed on the extent to which

variables (weight and material) were varied systematically and on the extent to which an experiment was 'complete' (that is including the necessary measurements of temperature and generating a sufficient number of data points). Systematical Orderliness was rated on a three-point scale (ranging from 0 to 2) for the five subtasks separately. For each subject a sumscore of Systematical Orderliness was calculated from the scores on the subtasks.

Finally, student notes of subjects in the present experiment were judged on the basis of the Quantity and Quality of those notes. Both Quantity and Quality of note-taking were rated on a three-point scale (ranging from 0 to 2). Unfortunately, the notes of subjects from the reference group were not available for these analyses.

Learning measures

The Knowledge test (see also Veenman et al., 1993) contained a series of 37 statements that had to be evaluated as either right or wrong (or "I don't know"). For instance, a statement was: "A certain temperature indicates a certain amount of heat." Every correct evaluation added one point to the subject's score. Alpha reliabilities were .73 for the Pretest and .72 for the Posttest. Solving the three Problems required a fair amount of reasoning based upon understanding of heat theory. An example:

> "Two blocks, made of the same material but having different weights, both have a temperature of 20°C. The smallest block is heated for 30 seconds until it reached a temperature of 50°C. The largest block is heated for 2 minutes until it reached a temperature of 80°C. What is the ratio of the weights of the two blocks? You may assume that there is no exchange of heat with the environment."

The analyses of the Problems protocols were performed by two judges and focused on the correctness and completeness of reasoning, similar to Veenman et al. (1993). Protocols of subjects from the thinking-aloud condition and from the control condition were presented in a randomly mixed order and the judges were unable to determine the conditions to which subjects belonged. Each problem was rated on a three-point scale, ranging from 0 to 2. A sumscore (with an alpha reliability of .52) was calculated for each subject.

RESULTS

Logfile measures of the reference group

First, the logfile measures of 28 subjects from the reference group (adapted from Veenman et al., 1993) were related to their scores on Working Method and learning measures. As subjects in this experiment were selected on the basis of intellectual ability, correlations had to be corrected for selection following the procedures of Gulliksen (1961). Inspection of the bivariate plots did not indicate violation of the requirement of homoscedasticity. In Table 1 the corrected correlations with scores counted from the logfiles are displayed.

Three logfile scores (Burner-on, No-initT and No-finalT) appeared to correlate fairly high with Working Method. The negative correlations for No-initT and No-finalT were expected as these measures were meant to indicate a *lack* of systematical orderliness. Although correlations of these three scores with learning measures were rather low, they showed an appropriate trend. The other four logfile scores (Blocks-abs, Blocks-rel, Re-reading and Help) correlated low with Working Method, while the correlations with learning measures were extremely low or even negative.

A composite measure ('Composite') was calculated from the three scores that were related to Working Method (Burner-on, No-initT, and No-finalT). First, the three scores were transformed to z-scores. Because No-initT and No-finalT represented lack of systematical orderliness, the signs of these z-scores were reversed. Finally, for each subject a sumscore was calculated. The Composite scores (with an alpha reliability of .43) correlated highly with Working Method (see Table 2).

Similarly, the judgemental scores of Systematical Orderliness (with an alpha reliability of

Table 1. Correlations with logfile counted scores for subjects from the reference group.

	W.M.	Pretest	Posttest	Problems
Burner-on	.49 (†)	.01	.21	.21
No-initT	-.40 (*)	-.35 (*)	-.21	-.18
No-finalT	-.37 (*)	-.28	-.23	-.14
Blocks-abs	.17	-.26	-.06	.01
Blocks-rel	.29	-.08	-.02	.01
Re-reading	.09	-.14	-.11	-.29
Help	-.08	-.18	.00	-.09

Note:All correlations are corrected for selection on the basis of scores on intellectual ability. W.M. means Working Method. (*) means that the uncorrected correlations were significant at a .05 level; (†) means that the uncorrected correlations were significant at a .01 level.

Table 2. Correlations with predictors of Working Method for subjects from the reference group.

	W.M.	Pretest	Posttest	Problems
Composite scores	.62 (†)	.32 (*)	.32 (*)	.26
Syst.Orderliness	.64 (†)	.35 (*)	.41 (*)	.34 (*)

Note: All correlations are corrected for selection on the basis of scores on intellectual ability. W.M. means Working Method. (*) means that the uncorrected correlations were significant at a .05 level; (†) means that the uncorrected correlations were significant at a .01 level.

.64) correlated highly with Working Method (see Table 2). Furthermore, Systematical Orderliness correlated .73 ($p < .01$) with Composite scores (after correction for selection on the basis of intellectual ability this correlation was .72). Both Systematical Orderliness and Composite scores may be regarded as vicarious measures of Working Method (or metacognitive skill).

Comparing thinking-aloud conditions on logfile measures

For subjects in the present experiment a Composite score was calculated from scores on Burner-on, No-initT and No-finalT in accordance with the afore-mentioned procedures. Anovas were performed on all logfile measures from the present experiment with thinking-aloud condition vs. control condition as the between-persons factor. Clearly, none of these Anovas revealed a significant difference between the two conditions (see Table 3). In most cases the means of both conditions were quite balanced. Only for Burner-on a marginally significant difference was established, suggesting that control subjects may have taken more observations than did subjects in the thinking-aloud condition. However, in the scores on the two relevant predictors of Working Method (Composite and Systematical Orderliness) such differences were not present.

Comparing thinking-aloud conditions on learning measures

Although control subjects on the average scored slightly higher on Pretest than subjects in the thinking-aloud condition, an Anova showed that this difference was not significant (see Table 4). Anovas on Posttest and Problems evidently showed that differences in learning performances between the thinking-aloud condition and the control condition were not significant too. Using the Pretest as covariate did not result in stronger effects of thinking-aloud condition vs. control condition ($F(1,37)$ was 0.01 and 0.89 for Posttest and Problems respectively). Finally, an Anova with repeated measures on time of testing (Pretest and Posttest) and thinking-aloud

Table 3. Results of logfile scores for thinking-aloud vs. control subjects.

	Thinking-aloud condition		Control condition		Results Anovas	
	M	*Sd*	*M*	*Sd*	*F*(1,38)	*p*-value
Burner-on	8.67	2.35	11.42	6.74	3.10	.09
No-initT	3.14	3.10	3.90	3.40	0.54	.47
No-finalT	0.33	0.58	1.16	3.11	1.42	.24
Blocks-abs	8.95	1.69	8.84	2.24	0.03	.86
Blocks-rel	6.05	0.97	6.10	1.33	0.02	.88
Re-reading	2.57	2.09	2.00	2.24	0.70	.41
Help	0.19	0.40	0.21	0.42	0.02	.88
Composite scores	0.03	1.02	-0.04	1.93	0.02	.89
Syst.Orderliness	4.14	1.90	3.74	2.45	0.35	.56
Note Quantity	0.76	0.77	0.79	0.86	0.01	.92
Note Quality	0.38	0.67	0.74	0.81	2.33	.14

condition vs. control condition as the between-persons factor showed a significant effect of time of testing ($F(1,38) = 51.40$, $p < .001$). Performances on the Knowledge tests improved significantly between Pretest and Posttest. However, the conditions effect and the interaction effect of conditions with time of testing were not significant ($F(1,38) = 0.40$ and $F(1,38) = 0.21$ respectively).

Time on task

Because it was expected that thinking aloud would slow down task performance (Elshout & van Leeuwen, 1992; Ericsson & Simon, 1980), a one-tail t-test was performed on Time on task contrasting the thinking-aloud condition with the control condition. Subjects in the thinking-aloud condition spent significantly more time in Heat lab than did control subjects ($t = 2.02$, $df = 38$, $p < .03$). Means were 44.33 min. ($sd = 14.78$) for subjects in the thinking-aloud condition and 35.90 min. ($sd = 11.25$) for control subjects. However, the negative correlations with both learning measures ($r = -.22$ for the Posttest and $r = -.20$ for the Problems) indicated that Time on task did not contribute to learning.

Table 4. Results of learning measures for thinking-aloud vs. control subjects.

	Thinking-aloud condition		Control condition		Results Anovas	
	M	*Sd*	*M*	*Sd*	*F*(1,38)	*p*-value
Pretest	19.95	5.30	21.05	3.95	0.54	.47
Posttest	24.38	4.52	24.95	4.20	0.17	.68
Problems	2.19	2.11	1.90	1.60	0.25	.62

DISCUSSION

Ericsson & Simon (1980, 1984) showed that, except for performance speed, other parameters of cognitive performance were not altered by thinking aloud concurrently. The results of this experiment justify the same conclusion for parameters of metacognitive or regulatory processes in learning. Neither for Composite scores and Systematical Orderliness, nor for learning measures, significant differences between the thinking-aloud and control conditions were obtained. The significant difference between the two experimental conditions on Time on task indicates that concurrent verbalization slows down performance on the learning task without, however, altering the ongoing cognitive and metacognitive processes. These results raise no objections to using the thinking-aloud method for assessing metacognitive skills.

The comparison of logfile measures with scores on Working Method adopted from Veenman et al. (1993) shows that some aspects of Working Method may be scored adequately from logfiles. However, further research is needed to cross-validate this selection of performance parameters as predictors of Working Method. Furthermore, the magnitude of the correlations between Composite scores and Systematical Orderliness on the one hand and Working Method on the other hand indicates that logfile measures do not cover *all* aspects of Working Method. For instance, orientation activities such as analyzing the problem statement can not be assessed directly from these logfiles. Similarly, the logfile information does not reflect elaboration activities. However, research by Veenman (1993) showed that different aspects of Working Method are highly correlated (correlations ranging from .51 to .82). This may indicate that some latent aspects of Working Method such as orientation and elaboration activities, are assessed indirectly from the logfiles. Yet, it remains to be verified more fully that these *other* aspects of Working Method are unaffected by thinking aloud concurrent to task performance.

NOTE

1. Address for correspondence: Dept. of Educational Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden. The authors express their gratitude to Christiaan Hamaker for his thoughtful comment.

REFERENCES

Breuker, J. A., Elshout, J. J., Someren, M. W. van, & Wielinga, B. J. (1986). Hardop-denken en protocolanalyse [Thinking aloud and protocol analysis]. *Tijdschrift voor Onderwijsresearch, 11*, 241-254.

Elshout, J. J. (1976). *Karakteristieke moeilijkheden in het denken.* Dissertation. Amsterdam: University of Amsterdam.

Elshout, J. J. (1990). Hogere-orde vaardigheden: discussie [Higher order skills: a discussion]. In M. J. Ippel, & J. J. Elshout (Eds.), *Training van hogere-orde denkprocessen* (pp. 149-155). Lisse: Swets & Zeitlinger.

Elshout, J. J., & Leeuwen, C. van (1992). Analyse van hardopdenk-protocollen: theorie en praktijk [Analysis of thinking-aloud protocols: Theory and practise]. In R. J. Jorna, & J. L. Simons (Eds.), *Kennis in organisaties. Toepassingen en theorie van kennissystemen* (pp. 154-174). Muiderberg: Coutinho.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Gulliksen, H. (1961). *Theory of mental tests.* New York: Wiley.

Jansweijer, W., Elshout, J. J., & Wielinga, B. J. (1990). On the multiplicity of learning to solve problems. In H. Mandl, E. de Corte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction: (Vol. 2.1). Social and cognitive aspects of learning and instruction.* Oxford: Pergamon press.

Jong, F. P. C. M. de, & Bil, M. de (1989). Metacognitie als 'executive control' tijdens tekstbestudering [Metacognition as executive control during learning from text]. In P. Span, E. de Corte, & B. van Hout Wolters (Eds.), *Onderwijsleerprocessen: Strategieën voor de verwerking van informatie* (pp. 71-84). Amsterdam: Swets & Zeitlinger.

Jong, T. de (1991). Learning and instruction with computer simulations. *Education & Computing*, *6*, 215-227.

Jong, T. de, & Ferguson-Hessler, M. G. M. (1984). Strategiegebruik bij het oplossen van problemen in een semantisch rijk domein: electriciteit en magnetisme. [The use of strategy in solving problems in a semantically rich domain: electricity and magnetism]. *Tijdschrift voor Onderwijsresearch, 9*, 3-15.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Someren, M. W. van, Barnard, Y., & Sandberg, J. (1993). *The think aloud method*. Amsterdam: University of Amsterdam.

Someren, M. W. van, & Elshout, J. J. (1985). Het effekt van zelfreflektie op leren probleemoplossen [The effect of self-reflection on learning to solve problems]. In J. G. L. C. Lodewijks, & P. R. J. Simons (Eds.). *Zelfstandig leren* (pp. 110-117). Lisse: Swets & Zeitlinger.

Span, P., & Overtoom-Corsmit, R. (1986). Information processing by intellectually gifted pupils solving mathematical problems. *Educational Studies in Mathematics*, *17*, 273-295.

Veenman, M. V. J. (1993). *Intellectual ability and metacognitive skill: determinants of discovery learning in computerized learning environments*. Dissertation. Amsterdam: University of Amsterdam.

Veenman, M. V. J., & Elshout, J. J. (1991). Intellectual ability and working method as predictors of novice learning. *Learning and Instruction, 1*, 303-317.

Veenman, M. V. J., & Elshout, J. J. (1992). Intelligentie en metacognitieve vaardigheden [Intelligence and metacognitive skills]. *Tijdschrift voor Onderwijsresearch*, *17*, 290-302.

Veenman, M. V. J., Elshout, J. J. & Hell, J. G. van (1992). Using the computer as a help tool during learning by doing. In T. J. Plomp, J. M. Pieters, & A. Feteris (Eds.), *Book of Summaries: European Conference on Educational Technology* (pp. 461-464). Enschede: Twente University.

Veenman, M. V. J., Elshout, J. J., & Hoeks, J. C. J. (1993). Determinants of learning in simulation environments across domains. In D. Towne, T. de Jong, & H. Spada (Eds.). *Simulation-based experiential learning* (pp. 235-249). Berlin: Springer Verlag.

Veenman, M. V. J., Samarapungavan, A., Hout-Wolters, B. H. A. M. van, & Beishuizen, J. J. (1992). De relatie tussen cognitieve en metacognitieve vaardigheden: een introductie [The relation between cognitive and metacognitive skills: An introduction]. *Tijdschrift voor Onderwijsresearch*, *17*, 269-272.

Vermunt, J. D. H. M., Lodewijks, J. G. L. C., & Simons, P. R. J. (1986). Hardop-denken als onderzoeksmethode naar regulatieprocessen bij tekstbestudering [Thinking-aloud as method of research on regulation processes in learning from text]. *Tijdschrift voor Onderwijsresearch*, *11*, 187-202.

Manuscript ontvangen 1-4-1993
Definitieve versie ontvangen 8-12-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, pp. 331-342

De stabiliteit van het schooleffect in de tijd: een analyse op basis van vijf jaar Eindtoets Basisonderwijs van het Cito

H. Blok en J.B. Hoeksma*

ABSTRACT

The school effect refers to the well-known phenomenon that pupils in different schools perform differently. In school effectiveness research it is generally assumed that these differences are considerable and relatively stable over time. These assumptions have been tested in a representative sample of 134 Dutch primary schools. Between 1987 and 1991 these schools yearly took the 'Eindtoets Basisonderwijs', produced by Cito. This test is considered to be the best available operationalization of the primary school curriculum. The size of the school effect was evaluated by the intraclass correlation coefficient (ICC), and the stability by the correlation between the school results in consecutive years, as estimated in a multilevel model. The estimated median ICC amounted to .16, the median stability (lag 1) turned out to be .84. Both estimates were domain dependent. The ICC and the stability were lowest for reading (.13, resp. .80), and highest for arithmetic (.19, resp. .87). The estimates for the third domain (written language production) were in between. Furthermore, there was a distinct relationship between the size and the stability of the school effect on the one hand and the socio-economic status (SES) of pupils on the other. Introduction of SES in the model resulted in a 20 to 45 percent decrease in the ICC and a still greater decrease in the stability. The reduction was greatest for the language domain. This finding suggests that differences between schools in performance of their pupils are strongly related to differences in the intake of pupils.

PROBLEEMSTELLING

Het is een bekend fenomeen dat er tussen scholen verschillen bestaan wat betreft de prestaties van hun leerlingen. Sommige scholen leveren leerlingen af die betere schoolprestaties leveren dan leerlingen van een andere school. Dit fenomeen wordt in het vervolg van deze bijdrage aangeduid als het *schooleffect*. Voor het basisonderwijs zijn betrouwbare schattingen beschikbaar over de omvang van het schooleffect. In een eerdere bijdrage (Blok, 1992) zijn schattingen gepresenteerd, gebaseerd op afnamen van de Eindtoets Basisonderwijs (EB) van het Cito, in een steekproef van 179 scholen die tussen 1987 en 1991 jaarlijks aan de toets hebben deelgenomen. De betrokken steekproef is niet alleen representatief voor de populatie van alle EB-deelnemers, maar waarschijnlijk zelfs voor de populatie van alle basisscholen in Nederland. Uit de schattingen blijkt dat het schooleffect aan het eind van de basisschool weliswaar betrekkelijk klein is, maar toch zeker niet valt te verwaarlozen. De variantie van de schoolgemiddelden bedraagt ongeveer 17 procent van de totale variantie. Het schooleffect blijkt enigszins te variëren over de drie onderscheiden domeinen. Voor rekenen en taal is het schooleffect wat groter dan voor lezen (rekenen: .19, taal: .18, lezen: .13). Voorts lijkt het schooleffect in de loop der onderzochte jaren enigszins te dalen (vanaf .19 in 1987 tot aan .15 in 1991).

Niet alleen de precieze grootte van het schooleffect is een relevant thema, maar ook de stabiliteit ervan in de tijd gezien. Is een school die het ene jaar leerlingen aflevert die hoge

* H. Blok, SCO-Kohnstamm Instituut voor Onderzoek van Opvoeding en Onderwijs van de Universiteit van Amsterdam, Grote Bickersstraat 72, 1013 KS Amsterdam.
J.B. Hoeksma, Vrije Universiteit, Vakgroep Kinder- en Jeugdpsychologie, Van der Boechorststraat 1, 1081 BT Amsterdam.

prestaties realiseren, het volgende jaar ook zo'n school? De volgende redenering laat zien, dat er een aanzienlijke mate van stabiliteit valt te verwachten. Uit eerdere onderzoekingen is steeds duidelijker geworden dat het schooleffect statistisch significant groter is dan nul. Er moeten derhalve factoren werkzaam zijn – andere dan de toevalsfactor – die eraan bijdragen dat er tussen leerlingen van verschillende scholen verschillen in schoolprestaties bestaan. Twee van zulke factoren worden vaak genoemd: de kwaliteit van de leerlingeninstroom (met aspecten als de voorkennis van leerlingen, hun geschiktheid om te leren en de steun die leerlingen van huis uit meekrijgen) en de effectiviteit van het door de school verzorgde onderwijs (met aspecten als de vakbekwaamheid van de leerkrachten en de geschiktheid van de gebruikte leermiddelen). Beide factoren vertonen zelf een aanzienlijke stabiliteit. De kwaliteit van de leerlingeninstroom verandert van het ene op het volgende jaar weinig: scholen die het ene jaar een instroom met weinig kinderen in achterstandssituaties hebben, hebben het volgende jaar een instroom die daarop lijkt. Een zekere stabiliteit is er ook voor de tweede factor: de effectiviteit van het door de school verzorgde onderwijs zal van het ene jaar op het volgende jaar niet erg veranderen.

Empirisch onderzoek naar de temporele stabiliteit van het schooleffect laat nog geen duidelijke uitspraak toe, eenvoudigweg omdat de bevindingen inconsistent zijn. Wij illustreren dit aan de hand van drie studies – die het basisonderwijs (of 'primary education') betreffen – en die uiteenlopende schattingen voor de stabiliteit opleveren.

Mandeville (1988) rapporteert correlaties van .34 tot en met .66 tussen 'school effectiveness indices' voor twee opeenvolgende jaren. Zijn gegevens hebben betrekking op een steekproef van 431 Amerikaanse basisscholen waar metingen verricht zijn in vier afzonderlijke leerjaren, namelijk 'grade 1' tot en met 'grade 4'. De aard van het schoolvak – de analyses zijn afzonderlijk gedaan voor lezen en rekenen – blijkt niet samen te hangen met de hoogte van de correlaties, het leerjaar daarentegen wel. De hoogste correlaties betreffen 'grade 1'. Voor 'grade 2' en de daaropvolgende leerjaren blijken de correlaties steeds lager te worden.

Van Batenburg (1990) schat de stabiliteit van het schooleffect als een correlatie tussen schoolgemiddelden op de EB van het Cito. Die correlaties ('lag 1') variëren van .59 tot en met .68 en ze zijn voor rekenen iets hoger dan voor taal en informatieverwerking (het lezen van teksten, kaarten, tabellen en grafieken). De steekproef omvat de gegevens van 2382 scholen die in de periode van 1986 tot en met 1988 aan de EB deelgenomen hebben. De correlaties kunnen evenwel geattenueerd zijn in verband met een mogelijke onbetrouwbaarheid van het schoolgemiddelde. Een dergelijke onbetrouwbaarheid leidt ertoe dat de gerapporteerde correlaties een onderschatting vormen van de stabiliteit van het schooleffect.

Door toepassing van meerniveau-modellen speelt attenuatie ten gevolge van de onbetrouwbaarheid van het schoolgemiddelde in het onderzoek van Bosker (1991) geen rol. Hij heeft de stabiliteit van het schooleffect – door hem consistentie genoemd – op twee manieren bestudeerd. In de eerste plaats schat hij een correlatie tussen de resultaten van jaargroep 6 en van jaargroep 8 binnen dezelfde school (n is 150 scholen). Deze correlatie bedraagt .58 voor rekenen en .69 voor taal (beide op basis van het zogenoemde 'bruto schooleffect', waarin geen predictoren of covariaten een rol spelen). In de tweede plaats schat hij een correlatie tussen de resultaten van parallelklassen. Deze schatting is gebaseerd op de gegevens van 24 scholen met een parallelgroep (16 scholen met twee jaargroepen 6, 8 scholen met twee jaargroepen 8). De correlatie bedraagt .81 voor rekenen en 1.00 voor taal (beide wederom op basis van het bruto schooleffect). Uit het feit dat de op de tweede manier geschatte correlaties beduidend hoger zijn dan de correlaties die op de eerste manier geschat zijn, leidt de onderzoeker af dat het niet zo zeer leerkracht-factoren zijn die de effectiviteit van de school bepalen, maar met name school gebonden factoren.

In de onderhavige bijdrage worden schattingen gepresenteerd van de stabiliteit van het schooleffect over een periode van vijf jaar, eveneens gebruikmakend van een 'multilevel' model. De schattingen worden gebaseerd op de reeds genoemde EB van het Cito. Deze toets is voor het genoemde oogmerk bijzonder geschikt. Ten eerste is de EB wat betreft de dekking van het basisschoolcurriculum een behoorlijk complete toets. De kerndomeinen taal, lezen en rekenen worden jaar in jaar uit met een opmerkelijke stabiliteit geoperationaliseerd in 180 items. Een

tweede voordeel is de telkenjare hoge betrouwbaarheid van de toets (vanaf .83 tot en met .93 op het niveau van de hoofdonderdelen). Van belang is voorts dat de toets op heel veel scholen gemaakt wordt (in de genoemde jaren is het deelnemerspercentage gegroeid tot meer dan 50 procent van alle basisscholen) en dat de groep van deelnemende scholen bij benadering representatief te achten is voor alle ruim 8000 basisscholen in ons land. Als laatste voordeel valt aan te merken, dat er weinig verloop onder de scholen is. Van de deelnemers in 1987 heeft 89 procent (3390 basisscholen) in alle volgende jaren tot en met 1991 eveneens deelgenomen. Daardoor vormen de EB-gegevensbestanden een betrekkelijk stabiele basis.

Het doel van deze bijdrage reikt in bescheiden mate verder dan het presenteren van schattingen voor de stabiliteit alleen. Tevens komt de vraag aan de orde wat de achtergrond vormt van het schooleffect en de stabiliteit ervan. Voor de op deze vraag betrekking hebbende vervolganalyses zijn twee additionele gegevens beschikbaar.

Het ene extra gegeven is een aanduiding van de achterstandscategorieën waaruit een school haar leerlingen recruteert. Dit gegeven is een indicator voor de eerder genoemde factor 'kwaliteit van de leerlingeninstroom'. Het tweede extra gegeven is het antwoord op de vraag of voor een gegeven school de leerkracht voor groep 8 van jaar tot jaar dezelfde is, dan wel dat het telkens wisselende leerkrachten zijn die groep 8 onder hun hoede hebben. Bosker en Scheerens (1989) spreken het vermoeden uit, dat leerkracht-gebonden factoren belangrijk zijn bij de verklaring van de stabiliteit van het schooleffect. Zij baseren dit vermoeden mede op resultaten uit Amerikaans onderzoek, gerapporteerd door Mandeville (o.c.). Deze rapporteert een mediane correlatie tussen onderwijsresultaten in opeenvolgende kalenderjaren van .44 in condities waarin de leerkrachten van jaar tot jaar dezelfde blijven, maar de leerlinglichtingen wisselen. In condities waarin niet alleen de leerlingen maar ook de leerkrachten wisselen, zakt de correlatie tussen onderwijsresultaten in twee opeenvolgende jaren tot .07. Het op Mandeville's resultaten gebaseerde vermoeden vindt overigens, zoals we zojuist zagen, geen steun bij Bosker (o.c.).

De beide extra gegevens maken het mogelijk de vraag naar de stabiliteit van het schooleffect met twee volgende vragen uit te breiden. De eerste is hoe sterk de *omvang* van het schooleffect samenhangt met de achterstandsindex en met leerkrachtwisselingen. De tweede is in welke mate de *stabiliteit* van het schooleffect gerelateerd is aan de achterstandsindex en aan leerkrachtwisselingen. Met andere woorden, in hoeverre vallen de omvang en de stabiliteit van de verschillen tussen scholen te verklaren uit de achterstandsindex en uit leerkrachtwisselingen?

BESCHRIJVING GEGEVENSBESTANDEN

Zoals gezegd, er zijn 3390 basisscholen die aan alle afnamen vanaf 1987 tot en met 1991 hebben deelgenomen. Uit deze groep van scholen is aselect een steekproef van 200 scholen getrokken, aan wie vervolgens toestemming is verzocht voor het gebruik van hun toetsresultaten. Behalve toestemming is aan deze scholen tevens gevraagd op te geven in hoeverre er wisselingen hebben plaatsgehad wat betreft de leerkracht voor de achtste groep. Een groep van 134 scholen heeft op beide verzoeken positief gereageerd. Hun toetsresultaten vormen de basis van de hier te rapporteren gegevens. Deze groep blijkt – gemeten aan de ruwe totaalscore – in hoge mate representatief voor de populaties van EB-deelnemers in de afzonderlijke jaren (zie Tabel 1). De verschillen in gemiddelden (uitgedrukt als effectgrootte) bedragen ten hoogste een tiende van de standaarddeviatie.

De EB bestaat elk jaar uit 180 items, verdeeld over drie subtoetsen: taal (60 items), rekenen (60 items) en informatieverwerking (60 items). De scores op de beide eerste subtoetsen, voor taal en rekenen, zijn ongewijzigd overgenomen. Maar aangaande de subtoets voor informatieverwerking heeft een aanpassing plaats gevonden. Gebruik is alleen gemaakt van de 40 items die betrekking hebben op het lezen van teksten. De 20 overige items – betreffende het hanteren van informatiebronnen, het kaartlezen en het lezen van tabellen en grafieken – zijn buiten beschouwing gelaten. Door deze aanpassing is de eenduidigheid van het door de derde subtoets bestreken domein toegenomen. Dit domein en deze subtoets worden in het navolgende aange-

Tabel 1. Scoreverdeling op de EB voor de versies 1987 tot en met 1991 (ruwe totaalscore, berekend over alle 180 items, zowel voor de afzonderlijke jaarlijkse populatie van EB-deelnemers als voor de steekproef van 134 scholen wier gegevens de basis voor de analyses vormen (n = aantal deelnemende scholen; = $\bar{x}$ gemiddelde; s.d. = standaarddeviatie; effectgrootte = $(\bar{x}_p - \bar{x}_s)/s.d._p$).

	populatie			steekproef (n = 179)		
versie	n	$\bar{x}_p$	$s.d._p$	$\bar{x}_s$	$s.d._s$	eff.gr.
1987	3801	125.38	16.02	124.73	15.91	.04
1988	4260	128.87	15.37	128.84	13.74	.00
1989	4653	128.54	15.22	129.99	14.11	–.10
1990	4836	125.07	15.71	126.45	14.51	–.09
1991	4956	125.21	14.93	126.76	12.87	–.10

duid als 'lezen'. De schattingen voor de betrouwbaarheid (Cronbach alpha) zijn overwegend hoog, al variëren ze enigszins over de onderscheiden domeinen. De betrouwbaarheid van de rekentoetsen (variërend vanaf .92 tot en met .94) is stelselmatig hoger dan de betrouwbaarheid van de toetsen voor taal (variërend vanaf .87 tot en met .89) en lezen (variërend vanaf .83 tot en met .86).

Wat betreft de validiteit van de toetsen wordt verwezen naar de beide doelenboeken (Cito, 1986; 1990), waarvan de ene betrekking heeft op de versies 1987 tot en met 1990 en de andere op de versies vanaf 1991. Voor taal en lezen is er in de betrokken periode van 1987 tot en met 1991 geen verandering geweest in de omschrijving van de toetsmatrijs. Voor rekenen is de toetsmatrijs met ingang van 1991 enigszins veranderd. De oorspronkelijke subrubriek 'Vraagstukken' is met ingang van 1991 verdwenen. De items uit deze rubriek zijn ondergebracht bij andere rubrieken. Voorts heeft het hoofdrekenen wat meer gewicht gekregen. Bevatte de 1987-versie nog slechts 10 items op het gebied van hoofdrekenen, de 1991-versie bevat er inmiddels 20. Deze verschuiving is gerealiseerd door in de meeste rubrieken wat minder items op te nemen, waardoor het totaal aantal items voor de subtoets rekenen alle jaren constant op 60 is gehouden. De achtergrond van de beschreven veranderingen is het streven de EB een afspiegeling te laten zijn van de communale doelstellingen, de doelstellingen die bijna alle scholen gemeenschappelijk hebben. De veranderingen weerspiegelen de opkomst van het 'realistische reken- en wiskundeonderwijs'.

De *achterstandsindex (AI)* is op de volgende wijze tot stand gekomen. Van het Ministerie voor Onderwijs en Wetenschappen is voor de scholen in de steekproef een opgave verkregen van het aantal 11- tot en met 13-jarigen in de afzonderlijke achterstandscategorieën voor vijf teldata (telkens 16 januari, vanaf 1987 tot en met 1991). De gegevens van het ministerie betreffen een ruimer gedefinieerde groep dan de leerlingen die daadwerkelijk aan de EB hebben deelgenomen. Met name 11-jarigen maar ook sommige 12-jarigen zitten voor een deel nog in groep 7 en nemen derhalve niet aan de EB deel. Het bleek evenwel niet mogelijk deze groep apart te onderscheiden. De opgaven van het ministerie over het totale aantal 11- tot en met 13-jarige leerlingen blijken ongeveer 45 procent hoger te zijn dan de opgaven van het Cito over het aantal in hetzelfde jaar getoetste leerlingen. Uit de ministerie-gegevens is de achterstandsindex berekend door de leerlingaantallen te vermenigvuldigen met hun gewichtscategorie en te wegen naar het totaal aantal leerlingen. Een school zonder leerlingen in enige achterstandscategorie heeft hierbij de waarde 100 gekregen, een school met uitsluitend leerlingen in de hoogste categorie de waarde 190. Tabel 2 geeft enig inzicht in de verdeling van de index. Het blijkt dat de achterstandsindex voor 1987 een afwijkend hoog gemiddelde en spreiding laat zien. Dit wordt veroorzaakt door een verandering van de telsystematiek van het ministerie per augustus 1987. Omdat de intercorrelaties géén afwijkend patroon laten zien, is besloten de variabele

Tabel 2. Intercorrelaties (pmc), gemiddelden en standaarddeviaties voor de Achterstandsindex in 1987 tot en met 1991 voor de steekproef van 134 scholen; onder de onderbroken lijn worden gemiddelden en standaarddeviaties gegeven zoals bepaald in de populatie van alle circa 8400 basisscholen.

	AI87	AI88	AI89	AI90	AI91
Steekproef					
AI87	1.00				
AI88	.84	1.00			
AI89	.87	.92	1.00		
AI90	.85	.88	.91	1.00	
AI91	.86	.89	.93	.92	1.00
gem.	134.20	117.28	118.48	118.40	118.72
s.d.	22.31	11.71	12.75	13.25	13.51
Populatie					
gem.	136.50	118.92	119.21	119.51	119.72
s.d.	23.80	15.36	15.90	16.40	16.87

AI87 toch in de analyses op te nemen. Om problemen als gevolg van de heterogene varianties te vermijden, zijn de vijf AI-variabelen ten behoeve van de in de volgende sectie gerapporteerde analyses lineair getransformeerd naar een schaal met een gemiddelde van 120 en een standaarddeviatie van 15. Men beseffe overigens dat de correlaties 'confounded' zijn. Ze hebben betrekking op deels overlappende leerlinggroepen, daar bijvoorbeeld de 11-jarigen in 1987 ook in 1988 en 1989 nog kunnen meetellen als ze als 12-, respectievelijk 13-jarigen nog steeds dezelfde school bezoeken. De mate van overlap is niet precies aan te geven, maar heeft als bovengrens de hiervoor genoemde 45 procent.

Overigens zijn in Tabel 2 ook de gemiddelden en standaardafwijkingen van AI in de populatie van alle circa 8400 basisscholen opgenomen. Te constateren valt dat de steekproef wat de AI betreft nauwelijks van de populatie afwijkt. Het grootste verschil betreft 1987, waarin het steekproefgemiddelde ruim twee punten lager is dan het populatiegemiddelde (overeenkomend met een effectgrootte van .10).

De variabele *leerkrachtwisseling (LKW)* is op de volgende manier tot stand gekomen. De scholen hebben voor de vijf betrokken afnamejaren aangegeven welke leerkrachten in welke jaren groep 8 onder de hoede hadden. Daaruit is bepaald in hoeverre er leerkrachtwisselingen plaatsgevonden hebben. Van het ene op het volgende jaar blijken betrekkelijk weinig wisselingen plaats te vinden. Het percentage leerkrachtwisselingen (LKW = 1) wordt tussen haakjes vermeld: 1987/1988 (17 procent), 1988/1989 (25 procent), 1989/1990 (20 procent), 1990/1991 (24 procent).

ANALYSE

Het databestand bevat gegevens van 13.106 leerlingen, verdeeld over 134 scholen die vijf achtereenvolgende jaren aan de EB hebben deelgenomen. Dit suggereert een hiërarchische structuur met drie niveaus, te weten leerling, school en afnamejaar. Het probleem is evenwel dat het jaar van afname niet zinvol als een 'random' effect gemodelleerd kan worden. Kernoverweging hierbij is dat de EB jaarlijks uit andere items bestaat. De verschillen tussen de jaren zijn daardoor arbitrair en niet te interpreteren als een effect van tijd. Een bijkomend probleem is dat de hiërarchische structuur tussen school en afnamejaar niet vastligt: scholen zijn niet genest in afnamejaren, noch omgekeerd. Er is derhalve geen doorslaggevend argument om of de school of

het afnamejaar als het hoogste niveau te beschouwen. Onze conclusie is dat we te maken hebben met 'random' gegevens op twee niveaus (leerling, school), waarbij we de verschillen tussen de afnamejaren als 'fixed' opvatten.

In het basismodel is de score van de leerlingen als volgt ontleed:

$$y_{ijk} = \beta_{jk} + e_{ijk}. \quad (1)$$

Hierin representeert

y_{ijk} de toetsscore (voor rekenen, taal of lezen) van leerling i in school j in jaar k;
β_{jk} de bijdrage van school j in jaar k;
e_{ijk} het leerlingspecifieke residu, met variantie $\sigma^2(e_k)$.

De schoolbijdrage β_{jk} wordt gesplitst in een 'fixed' en een 'random' component, namelijk als volgt:

$$\beta_{jk} = \beta_k + u_{jk}, \quad (2)$$

waarbij

β_k het jaargemiddelde ('fixed');
u_{jk} de bijdrage van de school als afwijkingsscore, met variantie $\sigma^2(u_k)$.

Combinatie van Vergelijkingen 1 en 2 leidt tot

$$y_{ijk} = \beta_k + u_{jk} + e_{ijk}. \quad (3)$$

In Vergelijking 3 representeert de eerste term het jaargemiddelde, de tweede term de bijdrage van de school (geschreven als afwijking van het jaargemiddelde) en de derde term het leerlingspecifieke residu (geschreven als afwijking van het jaargemiddelde en de schoolbijdrage). Het model wordt gecompleteerd met de opname van de covarianties $\sigma(u_{jk},u_{jk'})$, die vrijgelaten worden vanuit de veronderstelling dat de verschillen tussen scholen een zekere mate van stabiliteit zullen bezitten.

Men kan zich het geformuleerde model voorstellen als een vijftal twee-niveaumodellen (waarbij de leerlingen genest zijn in de scholen) die gelijktijdig worden geanalyseerd. Voor elk afnamejaar zijn er schoolspecifieke bijdragen (u_{jk}), die over de jaren heen gecorreleerd kunnen zijn.

Op grond van de variantie-covariantiematrix van de schoolspecifieke bijdragen $\sigma(u_{jk},u_{jk'})$ kan de intraklasse correlatiecoëfficiënt ρ in jaar k als volgt worden bepaald:

$$\rho_k = \sigma^2(u_k) / (\sigma^2(u_k) + \sigma^2(e_k)). \quad (4)$$

De correlatie $r_{kk'}$, tussen de schooleffecten tussen de jaren k en k' wordt verkregen via:

$$r_{kk'} = \sigma(u_{jk},u_{jk'}) / (\sigma^2(u_k) * \sigma^2(u_{k'}))^{1/2}. \quad (5)$$

Aangaande de coëfficiënten ρ_k en $r_{kk'}$ zijn diverse statistische toetsen mogelijk. Getoetst kan bijvoorbeeld worden of de intraklasse correlatiecoëfficiënt ρ_k in de loop van de tijd gelijk gebleven is. Daartoe wordt het model van Vergelijking 3 opnieuw geschat, met als toegevoegde restricties dat de varianties op het lagere niveau aan elkaar gelijk zijn ($\sigma^2(e_k) = \sigma^2(e)$), evenals de varianties op het hogere niveau ($\sigma^2(u_k) = \sigma^2(u)$). De chikwadraat verdeelde 'loglikelihood' van dit model wordt vergeleken met die van het model zonder deze restricties. Het verschil volgt – onder de nulhypothese – eveneens een chikwadraat verdeling met vier vrijheidsgraden (cor-

responderend met een verschil van vier te schatten parameters). Merk op dat standaardisatie van de toetsscores (y_{ijk}) hierbij noodzakelijk is, omdat de varianties van de toetsen van jaar tot jaar enigszins verschillen, al zijn de betrouwbaarheden per toets van jaar tot jaar erg constant. De toetsen bevatten immers elk jaar andere items.

Indien de nulhypothese van gelijke intraklasse correlaties niet wordt verworpen, kan vervolgens worden getoetst of de correlaties $r_{kk'}$ tussen de schooleffecten aan elkaar gelijk zijn. Daartoe worden de covarianties $\sigma(u_{jk},u_{jk'})$ aan elkaar gelijk gesteld. De toetsingsgrootheid (het verschil tussen de 'loglikelihood' van het model zonder en het model met deze restrictie) is chikwadraat verdeeld met tien vrijheidsgraden.

Ter beantwoording van de beide vervolgvragen – naar de effecten van AI en LKW op de omvang en de stabiliteit van de verschillen tussen scholen – kan Vergelijking 1 worden uitgebreid tot:

$$y_{ijk} = \beta_{jk} + e_{ijk} + \gamma_1 AI_{jk} \qquad (6)$$

of tot:

$$y_{ijk} = \beta_{jk} + e_{ijk} + \gamma_2 LKW_{jk} . \qquad (7)$$

Indien AI (of LKW) een effect heeft op de grootte van de verschillen tussen scholen, zal opname van deze term leiden tot een reductie van de variantie tussen scholen ($\sigma^2(u_k)$) en bijgevolg van de intraklasse correlaties (ρ_k). De omvang van de reductie is een directe indicatie voor de sterkte van het effect van AI (of LKW). Als AI (of LKW) inderdaad een dergelijk effect heeft, dan zullen ook de covarianties $\sigma(u_{jk},u_{jk'})$ onder Vergelijking 6 (of 7) kleiner zijn dan die onder Vergelijking 3. De omvang van de reductie geeft aan, in welke mate de stabiliteit van de verschillen tussen scholen aan AI (of LKW) kan worden toegeschreven.

Alle modelspecificaties en -schattingen zijn uitgevoerd met het programma ML3 (Prosser, Rasbash & Goldstein, 1991). Hierbij hebben zich geen convergentieproblemen voorgedaan.

RESULTATEN

De stabiliteit van de verschillen tussen scholen

Tabel 3 geeft de resulterende intraklasse correlaties voor de drie vakgebieden, geschat volgens Vergelijking 4. In het rechter gedeelte van de tabel zijn de schattingen ontleed door middel van 'mean polishing' (Hoaglin, Mosteller & Tukey, 1983). De variantie van de intraklasse correlaties wordt door de randeffecten met 89 procent gereduceerd. Aangezien de residuen voorts klein zijn en geen systematisch patroon vertonen, mag worden geconcludeerd dat de randeffecten in Tabel 4 een goede beschrijving van de gegevens bieden. De 'overall' intraklasse correlatie bedraagt .16, met kleine effecten voor de vakgebieden en de afnamejaren. Deze resultaten komen nagenoeg overeen met de eerder gerapporteerde schattingen (Blok, o.c.). Dit viel overigens te verwachten, aangezien de onderhavige steekproef van 134 scholen een subset is van de steekproef van 179 scholen die de basis vormen voor de eerdere schattingen.

Er is getoetst of de intraklasse correlaties van jaar tot jaar gelijk gebleven zijn. Voor lezen en rekenen wordt deze veronderstelling niet verworpen, maar voor taal wel (Tabel 4). Vanwege de overzichtelijkheid van de presentatie is besloten deze restrictie in volgende analyses voor *geen* van de drie vakken op te leggen.

De correlaties tussen de schoolspecifieke bijdragen voor de afzonderlijke jaren, die de stabiliteit ervan representeren, zijn hoog (Tabel 5). De mediane correlaties bedragen voor lezen, taal en rekenen respectievelijk .80, .84 en .87. De beide eerste 'lag 1' correlaties (namelijk 87/88 en 88/89) zijn binnen alle drie de vakgebieden de hoogste, terwijl de 'lag 4' correlaties enigszins lager uitvallen. Maar in zijn algemeenheid is er onder invloed van de tijd geen sterke afname te zien van de sterkte van de verbanden. Toetsing van de veronderstelling dat de stabiliteiten van jaar tot jaar gelijk zijn, is echter niet zinvol aangezien de intraklasse correlaties variabel zijn.

Tabel 3. De geschatte intraklasse correlaties, voor vijf jaren en drie domeinen (lezen, taal, rekenen); voor en na ontleding in een lineair additief model.

	voor ontleding			na ontleding			
jaar	lezen	taal	rek.	lezen	taal	rek.	jaareffect
1987	.15	.20	.23	−.01	.00	.01	.03
1988	.12	.14	.20	.00	−.02	.02	−.01
1989	.13	.20	.20	−.01	.02	.00	.01
1990	.13	.18	.18	.00	.01	−.01	.00
1991	.12	.14	.16	.01	−.01	−.01	−.02
domeineffect				−.03	.01	.03	.16[a]

[a] universeel gemiddelde

Tabel 4. Toetsingresultaten van de hypothese $\rho_k = \rho$ voor lezen, taal en rekenen (toetsingsgrootheid, aantal vrijheidsgraden, rechter overschrijdingskans en de onder de nulhypothese geschatte ρ.

domein	χ^2	d.f.	p	$\hat{\rho}$
lezen	2.3	4	.68	.13
taal	14.2	4	<.01	.17
rekenen	8.6	4	.07	.19

Tabel 5. De geschatte correlaties tussen de schoolspecifieke effecten ($r_{kk'}$; k vanaf 1987 tot en met 1991) voor lezen, taal en rekenen; n = 134 scholen.

		domein		
'lag'	jaren	lezen	taal	rekenen
'lag' 1	1987/88	.86	.95	.93
	1988/89	.85	.92	.92
	1989/90	.79	.79	.80
	1990/91	.75	.85	.87
'lag' 2	1987/89	.76	.84	.89
	1988/90	.80	.88	.89
	1989/91	.81	.82	.81
'lag' 3	1987/90	.78	.80	.85
	1988/91	.80	.84	.86
'lag' 4	1987/91	.75	.80	.81
mediaan		.80	.84	.87

Tabel 6. Het effect van het opnemen van de Achterstandsindex AI als 'fixed' term: toetsingsresultaten voor lezen, taal en rekenen (parameterschatting γ, standaardfout, toetsingsgrootheid, aantal vrijheidsgraden en rechter overschrijdingskans).

domein	$\hat{\gamma}$	s.e.	χ^2	d.f.	p
lezen	−.012	.0013	58.1	1	<.001
taal	−.012	.0015	46.3	1	<.001
rekenen	−.010	.0016	27.0	1	<.001

Effect van de achterstandsindex AI

De achterstandsindex heeft een aanzienlijk effect op de verschillen tussen scholen (Tabel 6). Opname van AI als een 'fixed' parameter in het model van Vergelijking 3 leidt voor alle drie de vakgebieden tot een statistisch significante verbetering van de modelpassing. Voorts blijken de geschatte parameterwaarden groot in vergelijking met hun geschatte standaardfout. De richting van het effect is conform de verwachting: een toename van AI leidt tot een verlaging van de gemiddelde schoolscore. Voor lezen correspondeert de verlaging per 10 punten op de AI-schaal met 1,0 punt (ofwel items correct) op de prestatieschalen. Voor taal en rekenen bedraagt de corresponderende verlaging respectievelijk 1,4 en 1,7 punten. Daar het bereik van de AI-schaal omstreeks 90 punten bedraagt, gaapt er een kloof van 9 tot 15 punten tussen scholen met de laagste AI-score en scholen met de hoogste AI-score.

In Tabel 7 is aangegeven wat het effect van opname van AI is op de omvang van de verschillen tussen scholen. Er blijkt sprake van een aanzienlijke reductie van de intraklassecorrelaties. De grootste reductie vindt plaats bij lezen (de mediane reductie bedraagt 34 procent, berekend over vijf intraklasse correlaties). De overeenkomstige percentages voor taal en rekenen zijn 31, respectievelijk 21 procent. Opvallend is de relatief grote reductie die 1989 in alle drie de vakgebieden laat zien.

Ook de stabiliteit van de verschillen tussen scholen neemt af door de opname van AI als 'fixed' parameter. Voor lezen, taal en rekenen bedraagt de afname (uitgedrukt als een mediaan percentage, berekend over tien covariantietermen) 47, respectievelijk 37 en 24 procent. Ondanks de reductie van de covarianties blijven de stabiliteiten tamelijk hoog (namelijk voor lezen: .66; taal: .75; rekenen: .83; wederom uitgedrukt als mediaan), omdat ook de varianties van de schoolspecifieke bijdragen afnemen. Deze gereduceerde stabiliteiten hebben betrekking

Tabel 7. De intraklasse correlaties ρ, na de opname van γAI in het model, inclusief de reductie in procenten (vergelijk Tabel 3).

	lezen		taal		rekenen	
jaar	$\hat{\rho}$	red.(%)	$\hat{\rho}$	red.(%)	$\hat{\rho}$	red.(%)
1987	.11	23	.16	20	.20	13
1988	.07	39	.09	38	.15	24
1989	.07	50	.12	41	.15	27
1990	.08	37	.13	28	.14	21
1991	.08	31	.10	31	.14	15
mediaan	.08	34	.12	31	.15	21

Tabel 8. Het effect van het opnemen van de leerkrachtwisseling LKW als 'fixed' term: toetsingsresultaten voor lezen, taal en rekenen (parameterschatting γ, standaardfout, toetsingsgrootheid en rechter overschrijdingskans).

domein	$\hat{\gamma}$	s.e.	χ^2	d.f.	p
lezen	.095	.0326	7.9	1	<.01
taal	.060	.0321	3.3	1	.07
rekenen	.088	.0324	7.0	1	<.01

op de stabiliteit van dàt deel van de verschillen tussen scholen dat *niet* aan de invloed van AI toe te schrijven is.

Effect van leerkrachtwisseling LKW

Opname van LKW als 'fixed' parameter in het basismodel van Vergelijking 3 leidt alleen voor lezen en rekenen tot een statistisch significante verbetering van de modelpassing (Tabel 8). Voorts blijken de geschatte parameters klein te zijn in relatie tot hun geschatte standaardfout. Het effect van wel of geen leerkrachtwisseling (LKW is een dichotome variabele) op het schoolgemiddelde bedraagt .06 tot .10 maal de standaarddeviatie van de toetsscores. Per toets komt dit neer op een ongeveer één punt hogere score voor scholen waar de leerkracht van het ene op het volgende jaar gewisseld is.

Hoe klein het effect van LKW is, blijkt eveneens uit het reducerende effect op de intraklasse correlaties, de covarianties en de stabiliteiten. De intraklasse correlaties worden met nul tot vier procent gereduceerd, de covarianties en de stabiliteiten met maximaal twee procent.

DISCUSSIE EN CONCLUSIE

In de probleemstelling is de verwachting uitgesproken, dat verschillen tussen basisscholen in de prestaties van hun leerlingen een aanzienlijke stabiliteit zouden vertonen. Deze verwachting was gebaseerd op de veronderstelling dat de belangrijkste oorzakelijke factoren voor deze verschillen – de kwaliteit van de leerlingeninstroom en de effectiviteit van het door de school verzorgde onderwijs – van het ene op het andere jaar weinig veranderen. De resultaten van onze analyses kunnen, naar nu gebleken is, beschouwd worden als een bevestiging van deze verwachting. De correlaties tussen de schoolspecifieke bijdragen van het ene op het volgende jaar ('lag' 1) variëren van .75 tot en met .95, met een mediane waarde van .87. De correlaties over een langere periode (tot en met vier jaar) zijn nauwelijks lager. Ze variëren van .75 tot en met .89, met een mediane waarde van .81. Met andere woorden, zelfs op een termijn van een handvol jaren blijken basisscholen behoorlijk stabiel te zijn wat betreft hun schooleffect.

Voor een juiste interpretatie realisere men zich dat de verschillen tussen scholen in de tijd weliswaar tamelijk stabiel zijn, maar de omvang van de verschillen is niet groot, zoals blijkt uit de geschatte intraklasse correlaties. De variantie van de schoolspecifieke bijdragen bedraagt – samengenomen over alle jaren en vakgebieden – niet meer dan 16 procent van de totale variantie.

De hier gebleken stabiliteiten zijn aanzienlijk hoger dan de eerder gerapporteerde (Van Batenburg, o.c.; Mandeville, o.c.). Vooral het verschil met Van Batenburgs resultaten is opvallend, omdat deze eveneens op de EB gebaseerd zijn. Het verschil moet berusten op een verschil in het statistische model. Berekenen wij namelijk in navolging van Van Batenburg de 'lag 1' correlaties tussen de geobserveerde schoolgemiddelden, dan blijken deze een bereik te hebben

– vanaf .50 tot en met .74 – dat nauwelijks afwijkt van het bereik (vanaf .59 tot en met .68) dat Van Batenburg rapporteerde. In het door ons geformuleerde 'multilevel' model wordt 'random' variantie gesplitst in een tussen-scholen gedeelte en een tussen-leerlingen gedeelte. De geschatte stabiliteiten zijn uitsluitend op het tussen-scholen gedeelte gebaseerd. Van Batenburg baseert zijn analyse evenwel op schoolgemiddelden waarvan beide variantiebronnen deel uitmaken. Het gevolg is dat zijn stabiliteiten niet alleen betrekking hebben op de tussen-scholen variantie, maar ook op de tussen-leerlingen variantie. En omdat de leerlingen van jaar tot jaar wisselen, zullen deze stabiliteiten dan ook lager uitvallen.

Een rechtstreekse vergelijking tussen onze resultaten en die van Bosker is misleidend door verschillen in de operationalisatie van het schooleffect (bij ons gebaseerd op de resultaten van achtste groepers van verschillende scholen; bij Bosker gebaseerd op de resultaten van leerlingen in groep 6 en groep 8 van dezelfde school). Boskers schattingen voor de consistentie van de verschillen tussen scholen zijn daardoor mede beïnvloed door de variantie tussen de leerjaren en derhalve onvergelijkbaar met onze schattingen. Naarmate de leerjaarvariantie groter is, neemt de consistentie van de verschillen tussen scholen namelijk af. In zijn onderzoek blijkt de variantie tussen de leerjaren voor rekenen groter te zijn dan voor taal. Dit hangt vermoedelijk samen met de leerplanopbouw van de beide vakken, die zodanig is dat het verschil in onderwijsaanbod tussen groep 6 en groep 8 voor rekenen groter is dan voor taal. Dit kan de verklaring zijn voor de afwijkend hoge consistentie die Bosker voor taal rapporteert in vergelijking met rekenen.

Keren we terug naar de eigen resultaten, dan wijzen we nog op de kleine verschillen die er tussen de drie vakgebieden bestaan wat betreft de omvang èn de stabiliteit van de verschillen tussen scholen. Over het geheel genomen zijn beide voor lezen enigszins lager dan voor taal, terwijl deze op hun beurt weer iets lager zijn dan voor rekenen (de mediane intraklasse correlaties voor lezen, taal en rekenen bedragen .13, respectievelijk .17 en .19; de mediane stabiliteiten voor lezen, taal en rekenen bedragen .80, respectievelijk .84 en .87). Men kan deze verschillen zien als het gevolg van het feit dat rekenen en taal (zoals geoperationaliseerd in de EB) veel exclusiever dan lezen vooral op school – en minder buiten school – beoefend worden. Scholen kunnen zich, met andere woorden, voor rekenen en taal beter profileren dan voor lezen, waar zich een grotere buitenschoolse invloed gelden laat.

De achterstandsindex AI – indicator voor de kwaliteit van de leerlingeninstroom – blijkt een forse bijdrage te leveren aan het grootte van de verschillen tussen scholen. De verschillen tussen scholen nemen met circa een derde af, na verdiscontering van het effect van AI op de prestatiematen. De afname is voor lezen en taal beduidend groter dan voor rekenen (respectievelijk 34 en 31 versus 21 procent). Dit verschil hangt waarschijnlijk samen met het feit dat allochtone leerlingen – wier aanwezigheid een sterk effect heeft op de AI-score van een school, ze wegen immers voor 1.9 – voor taal en lezen een grotere achterstand hebben dan voor rekenen. Uit door Uiterwijk (1990) vermelde gegevens over de EB-1987 valt voor Turkse en Marokkaanse leerlingen een achterstand te becijferen van 1.3 à 1.4 maal de standaarddeviatie (voor taal en lezen), respectievelijk 1.0 à 1.1 maal de standaarddeviatie (voor rekenen).

Niet alleen aan de omvang, ook aan de stabiliteit van de verschillen tussen scholen levert AI blijkens de resultaten een grote bijdrage. De covarianties tussen de jaarspecifieke schooleffecten nemen met 24 procent (bij rekenen) tot 47 procent (bij lezen) af. Dit resultaat werpt een interessant licht op de positie van lezen. Enerzijds blijken de verschillen tussen scholen wat betreft lezen relatief het kleinst. Maar diezelfde verschillen blijken relatief het sterkst samen te hangen met AI. Het netto resultaat voor lezen, na 'aftrek' van het effect van AI, is een relatief kleine tussen-scholenvariantie (de intraklasse correlatie bedraagt .08) en een relatief lage stabiliteit (de stabiliteit bedraagt .66). Het netto resultaat voor rekenen is omgekeerd: relatief grote verschillen tussen scholen (de intraklasse correlatie bedraagt .15) en een relatief grote stabiliteit van deze verschillen (de stabiliteit bedraagt .83). Taal neemt een tussenpositie in, die samen zou kunnen hangen met het gemengde karakter van de toets.

In wezen is AI een betrekkelijk grove indicator voor de kwaliteit van de instroom. Immers, niet àlle allochtone leerlingen en arbeiderskinderen verkeren in een nadelige positie, zo min als alle andere leerlingen in een gunstige positie verkeren. Bovendien kon AI slechts op school-

niveau worden gemeten, hetgeen minder adequaat lijkt dan een meting op het individuele niveau. Eigenlijk is het verbazingwekkend dat met zo'n gebrekkige indicator al zo'n 20 tot 45 procent van de verschillen tussen scholen te verklaren is. Zou men over ruimere mogelijkheden beschikken bij het vaststellen van de kwaliteit van de instroom – waardoor de voorkennis en de leergeschiktheid van leerlingen rechtstreeks en op het niveau van het individu te bepalen is, en niet afgeleid moet worden uit de sociaal-economische positie van de schoolbevolking – dan valt te verwachten dat deze groep van factoren nog sterker gerelateerd blijkt te zijn aan de schoolresultaten dan nu al het geval is.

Desalniettemin zijn er vermoedelijk nog andere factoren – niet behorend tot de kwaliteit van de instroom – die een effect hebben op de grootte en de stabiliteit van de verschillen tussen scholen. Over de aard van deze factoren kunnen we slechts speculeren. De enige bijdrage die onze studie in dit opzicht heeft opgeleverd, is dat de leerkracht een onbelangrijke factor lijkt. De variabele LKW blijkt niet of nauwelijks in verband te staan met verschillen tussen scholen. Introductie van LKW heeft verwaarloosbare effecten op de grootte en de stabiliteit van de verschillen tussen scholen. Dit resultaat weerspreekt de bevinding van Mandeville (o.c.) dat leerkrachtgebonden factoren een belangrijke bijdrage leveren aan de stabiliteit van de verschillen tussen scholen. Het geeft daarentegen steun aan Boskers (o.c.) speculatie dat schoolgebonden factoren belangrijker zijn dan leerkrachtgebonden factoren.

NOTEN

Dit onderzoek werd mogelijk gemaakt door een subsidie van het Instituut voor Onderzoek van het Onderwijs (SVO). Het Instituut voor Toetsontwikkeling (Cito) en het Ministerie van Onderwijs en Wetenschappen zijn we dankbaar voor de beschikbaar gestelde gegevens.
De anonieme 'reviewers' zijn wij erkentelijk voor hun commentaar.

LITERATUUR

Van Batenburg, Th.A. (1990). Variatie in schoolgemiddelden op de Cito-eindtoets basisonderwijs. *Tijdschrift voor Onderwijsresearch, 15,* 362-369.

Blok, H. (1992). De grootte van het schooleffect in het basisonderwijs; een analyse op basis van vijf jaar Eindtoets Basisonderwijs. *Tijdschrift voor Onderwijsresearch, 17,* 343-354.

Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch, 16,* 206-218.

Bosker, R.J. & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness research. *International Journal of Educational Research, 13,* 721-740.

Cito (1986). *Eindtoets Basisonderwijs; doelenboek.* Arnhem: Cito.

Cito (1990). *Eindtoets Basisonderwijs; doelenboek.* Arnhem: Cito.

Hoaglin, D.C., Mosteller, F. & Tukey, W.J. (1983). *Understanding robust and exploratory data analysis.* New York: John Wiley.

Mandeville, G.K. (1988). School effectiveness indices revisited: Cross-year stability. *Journal of Educational Measurement, 25,* 349-356.

Prosser, R., Rasbash, J. & Goldstein, H. (1991). *ML3 Software for three-level analysis.* London: Institute of Education, University of London.

Uiterwijk, J.H. (1990). *Item- en testbias in de Eindtoets Basisonderwijs 1987.* Arnhem: Cito.

Manuscript ontvangen 23-2-1993
Definitieve versie ontvangen 25-11-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, pp. 343-354

De invloed van contingente feedback op attributies en prestaties in de klas

Frans Siero en Jan Pieter van Oudenhoven
*Rijksuniversiteit Groningen, Vakgroep Psychologie**

ABSTRACT

The topic of this study is the impact of three contingent feedback strategies by teachers on students' performance. Contingency means that the feedback strongly corresponds with task behaviour that can be controlled by the students. Students received individualized feedback about their performance during a series of five lessons. Within this contingent feedback structure, three strategies were applied which were assumed to enhance the perceived controllability of the task situation: (1) an increase in the perceptibility of the contingency between feedback and task behavior; (2) explicit reference to effort as part of the feedback, and (3) goal setting. As predicted, the three contingent feedback strategies had a positive effect on perceived controllability, and led to better task performance.

INLEIDING

Het probleem dat leerlingen 'onder hun kunnen' functioneren vormt de achtergrond van de hier beschreven studie. Centraal staat het beoordelingsproces in de klas. Leerlingen ontvangen tijdens het uitvoeren van taken bij voortduring evaluatieve reacties. De goedkeuring of afkeuring door de leerkracht kan de leerling stimuleren maar ook ontmoedigen in zijn poging bij een volgende taak de aanwijzingen van de leerkracht te volgen en zijn prestaties te verbeteren. Voor een effectieve 'sturing' van het leerproces is het essentieel dat de leerkracht zich bij het vaststellen van zijn oordeel richt op die gedragsaspecten die beïnvloedbaar zijn, zoals de inzet van de leerling. Er is dan sprake van contingente oordelen; dat wil zeggen dat de oordelen covariëren met door de leerling beïnvloedbaar taakgedrag. Basisgedachte in deze studie is dat de leerkracht via de contingente beoordeling de leerling kan stimuleren tot meer inzet en betere prestaties. Kernvariabele is de 'waargenomen controle'; dit is de subjectieve contingentie die de leerling aan de taak- en beoordelingssituatie ontleent. Een hoge waargenomen controle houdt in dat men de oorzaak voor een (beoordeelde) prestatie vooral legt bij factoren die door de persoon zelf te beïnvloeden zijn (zoals inzet) en in geringe mate bij factoren die zich aan zijn controle onttrekken (zoals bekwaamheid, toeval en de moeilijkheid van de taak).

De wijze waarop de leerkracht tot een oordeel over een prestatie komt bepaalt in belangrijke mate hoe de leerling het oordeel waarneemt en interpreteert. Bij de in dit onderzoek toegepaste *geïndividualiseerde* vorm van beoordeling kiest de leerkracht de eerdere verrichtingen van de individuele leerling als criterium voor het vaststellen van een oordeel: verbetering van taakresultaten leidt tot een positief oordeel, verslechtering leidt tot een negatief of minder positief oordeel. Door deze *intrapersoonlijke* vergelijking verdwijnen verschillen in capaciteiten naar de achtergrond. Deze wijze van beoordeling leidt ertoe dat de oordelen over de prestaties (bij gelijke moeilijkheidsgraad) vooral door inzet bepaald worden. Vergeleken met een meer klassikale benadering waarin de prestaties van medeleerlingen de beoordelingsnorm mede bepalen, is de contingentie tussen oordeel en taakgedrag dus groter.

Oordelen zijn aan schommelingen onderhevig al naar gelang een leerling zijn inspanning varieert. Contingente beoordeling leidt er niet alleen toe dat de uitkomsten van een leerling (i.e.,

* Adres: F.W. Siero, Vakgroep Psychologie, Rijksuniversiteit Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen.

de oordelen die hij ontvangt), variëren, maar ook dat deze uitkomsten zelden sterk zullen overeenstemmen met de uitkomsten van zijn medeleerlingen. Contingentie tussen oordelen en taakgedrag geeft leerlingen dus uitkomstinformatie met een lage temporele consistentie (een wisselend patroon) en neutrale consensus (weinig overeenkomst met de uitkomstinformatie van medeleerlingen). De combinatie van deze beide principes maakt het waarschijnlijk dat de persoon het oordeel toeschrijft aan een instabiele oorzaak, zoals inzet en toeval (vgl. Forsyth & McMillan, 1981a; Hewstone & Jaspers, 1983, 1987; Weiner, 1982, 1985), maar door de covariatie tussen oordeel en inzet zal vooral inzet (zijnde een instabiele, interne èn *beïnvloedbare* factor) als oorzaak gezien worden.

Attributies naar beïnvloedbare oorzaken leiden tot hoge verwachtingen van succes en vervolgens tot betere prestaties (vgl. Forsyth & McMillan, 1981b). Deze auteurs vonden dat personen hun kans op een positieve uitkomst in de toekomst hoger inschatten naar de mate waarin zij hun taakuitkomsten toeschrijven aan beïnvloedbare oorzaken zoals inzet. Zij wijzen tevens op een positieve relatie tussen beïnvloedbaarheid en de affectieve reacties van een persoon op succes en falen: onafhankelijk van de uitkomst blijken personen tevredener te zijn met het oordeel naarmate ze de situatie als meer beïnvloedbaar waarnemen. Toeschrijving aan beïnvloedbare oorzaken gaat dus gepaard met een hogere uitkomstverwachting en met een hogere uitkomstwaardering. Beide elementen, uitkomstverwachting en uitkomstwaardering, zouden een persoon moeten motiveren zich sterker in te spannen en betere prestaties te leveren.

De controle die een persoon afleidt uit de taak- en beoordelingssituatie, kan op uiteenlopende wijze worden beïnvloed. In dit onderzoek zijn we daartoe uitgegaan van een beoordelingssituatie waarin verschillen in capaciteiten naar de achtergrond zijn geplaatst en het oordeel gebaseerd is op verschuivingen in prestaties. Aan deze geïndividualiseerde wijze van beoordeling zijn andere elementen toegevoegd waarvan we verwachtten dat ze de relatie tussen inzet en oordeel extra zouden benadrukken. Op grond van praktische overwegingen zijn drie strategieën toegepast: het extra zichtbaar presenteren van de contingentie tussen taakgedrag en oordeel, het gericht bijsturen van causale reacties en het gebruik van taakdoelen.

Zichtbaarheid

De vraag is hoe sterk de contingentie tussen taakgedrag en oordeel moet worden gepresenteerd om door de leerling te worden opgemerkt. Uit verschillende studies (bijv. Alloy & Tabachnik, 1984; Crocker, 1981; Kelley, 1973, pag. 120; Nisbett & Ross, 1980) blijkt, dat allerlei omstandigheden kunnen verhinderen dat een objectief aanwezige covariatie tussen gedrag en uitkomsten ook daadwerkelijk wordt waargenomen. De tijdsduur tussen het gedrag en de uitkomst kan te groot zijn. Ook kan er sprake zijn van een cognitieve barrière om contingenties waar te nemen. Een persoon kan bijvoorbeeld gewend zijn een vast verklaringschema toe te passen (zoals 'ik haal onvoldoendes omdat ik te stom ben') waardoor hij geen covariatie verwacht tussen inzet en prestaties. Het is daarom van belang dat de inzet van de leerling een duidelijk herkenbare component van het oordeel is. De eerste strategie houdt dan ook in dat we de zichtbaarheid van de contingentie tussen taakgedrag en oordeel versterken. Het is te verwachten dat de leerlingen in sterkere mate gestimuleerd worden zich in te spannen; de inzetverhoging zou moeten resulteren in betere taakprestaties.

Attributiesturing

Men kan ook op een meer directe wijze trachten de causale percepties te beïnvloeden door concreet te wijzen op de rol van inzet bij het tot stand komen van de prestatie. Aan het gunstige of ongunstige oordeel wordt dan een opmerking toegevoegd dat er kennelijk met een voldoende, dan wel te geringe inzet aan de taak is gewerkt. Deze attributiegerichte informatie stimuleert de persoon om aan zichzelf meer controle toe te schrijven. De hogere waargenomen controle motiveert vervolgens tot betere prestaties.

Stellen van taakdoelen

Bij de beide hierboven uitgewerkte beoordelingsstrategieën was het oogmerk de leerlingen via

ingrepen in de *presentatie van de beoordeling* de ervaring te geven dat zij controle kunnen uitoefenen op de succes- en faaluitkomsten. Het is ook mogelijk het controlekarakter van de *taaksituatie* te verhogen en wel door de persoon te richten op een na te streven taakdoel. Uit de literatuur blijkt, dat het stellen van doelen tot betere taakresultaten kan leiden (zie voor een overzicht: Locke & Latham, 1990; Mento, Steel & Karren, 1987). De uitwerking is optimaal als aan een aantal voorwaarden wordt voldaan. Zo is het noodzakelijk dat het stellen van een doel vergezeld gaat van informatie over de te leveren prestatie, zodat een persoon zijn gedrag kan wijzigen als hij het doel niet bereikt heeft (Becker, 1978). Doelen mogen hoog zijn maar moeten wel bereikbaar zijn; dat wil zeggen, de persoon moet over voldoende capaciteiten beschikken (Locke & Latham, 1990; Garland, 1983). Het is ook belangrijk dat men zijn inspanning kan richten op een zo specifiek mogelijk doel. Hoe specifieker het doel, des te sterker het stellen van een taakdoel zou werken (vgl. Jackson & Zedeck, 1982).

Het taakdoel stellen werd ingepast in de eerder beschreven geïndividualiseerde beoordelingsstructuur. Daardoor wordt redelijk voldaan aan de gestelde eisen met betrekking tot feedback, bereikbaarheid en specificiteit. Het (opgelegde) doel is niets anders dan de presentatie van een specifieke mate van vooruitgang die de persoon opgedragen wordt. Hij kan voor zichzelf bepalen of hij het doel gehaald heeft door het te vergelijken met het beoordelingsresultaat. De taakdoelen zijn op de leerlingen afgestemd. Iedere persoon beschikt dus over voldoende mogelijkheden om het gestelde doel te realiseren. We verwachten dat de aanwezigheid van een realiseerbaar doel de taaksituatie een meer beïnvloedbaar karakter zal geven en de persoon zal motiveren zich sterker in te spannen en betere prestaties te leveren (Bandura & Wood, 1989). Over de rol van waargenomen controle in dit proces verschaft de nogal behavioristisch georiënteerde literatuur over taakdoelen geen directe aanwijzingen. Dit onderzoek geeft de mogelijkheid de invloed van taakdoelen op de waargenomen controle na te gaan.

Richting van het oordeel

In het algemeen blijkt men successen vooral op te vatten als een gevolg van 'interne factoren' (eigen bekwaamheid en inzet) en falen vooral toe te schrijven aan een te moeilijke taak en toevallige factoren ('attributievertekening': Weiner, 1982, 1985; Zuckerman, 1979). In deze studie heeft succes betrekking op vooruitgang in prestaties en falen betrekking op achteruitgang in prestaties.

Deze koppeling tussen het oordeel en individuele verschuivingen in taakgedrag zou ertoe moeten leiden dat men ook voor negatieve, in dit geval beïnvloedbare, uitkomsten meer verantwoordelijkheid neemt en zijn falen met inzet in verband brengt. Het falen zou men meer aan inzet moeten toeschrijven naar de mate dat de relatie tussen oordelen en eigen taakgedrag in de beoordelingscondities sterker benadrukt wordt.

Hypothesen

Er werden vier beoordelingscondities toegepast. In de eerste conditie werden oordelen over de vooruitgang en achteruitgang in prestaties verstrekt; de oordelen waren gebaseerd op reële prestaties. Men ontving in deze conditie een oordeel dat door de vergelijking met eerdere prestaties, voor alle leerlingen in gelijke mate contingent was met het taakgedrag. In de tweede conditie werd de beoordeling op dezelfde wijze verstrekt, maar werd de zichtbaarheid van de contingentie vergroot. In de derde conditie werd aan deze versterkte contingentie een vorm van attributiesturing toegevoegd. In de vierde conditie, tenslotte, werd het stellen van taakdoelen in combinatie met de derde conditie toegepast. In de vier beoordelingscondities werd de relatie tussen het oordeel en het taakgedrag in steeds sterkere mate tot uitdrukking gebracht. De volgende hypotheses werden getoetst:

1. Het sterk zichtbaar presenteren van de contingentie tussen taakgedrag en oordeel leidt tot een hogere waargenomen controle (hyp. 1a) en betere taakresultaten (hyp. 1b): conditie II versus conditie I.
2. De toevoeging van attributiesturing leidt eveneens tot een hogere waargenomen controle (hyp. 2a) en betere taakresultaten (hyp. 2b): conditie III versus conditie II.

3. Het versterkt presenteren van de contingentie en de attributiesturing in combinatie met het stellen van taakdoelen leidt tot een verbetering van taakresultaten: conditie IV versus conditie III.

METHODE

Proefpersonen en taak

Aan het onderzoek deden 296 leerlingen mee van 13 willekeurig gekozen basisscholen, die alle een klassikaal beoordelingssysteem hanteerden. De 13 scholen werden op willekeurige wijze over vier beoordelingscondities verdeeld. In de condities I, II, III en IV zaten respectievelijk 76, 79, 63 en 78 leerlingen, allen uit jaargroep 7.

De leerlingen werkten gedurende vijf lessen aan een betrekkelijk nieuw onderwerp dat we ook elders gebruikten (Siero, Van Oudenhoven & Veen, 1983), namelijk een lespakket over de sociale geschiedenis van Noord-Amerika. De lessen werden gegeven door een voor dit doel aangestelde leerkracht. Zij waren alle op dezelfde wijze opgebouwd en werden, met uitzondering van de vijfde les, afgesloten met een meer-keuze toets. De lessen, die elk ongeveer vijftig minuten duurden, werden binnen een periode van 2 weken gegeven.

Beoordelingscondities

Zoals boven is aangegeven, verschillen de vier condities, met uitzondering van de vierde conditie (taakdoel stellen), in de wijze waarop de individuele vooruitgang of achteruitgang in prestaties aan de leerling werd gepresenteerd. De verandering in prestaties, uitgedrukt in punten, was gebaseerd op een vergelijking van de score op de laatst afgenomen toets met de score op de vorige toets of, als dat mogelijk was, de twee voorafgaande toetsen. Hierbij is het uiteraard van belang, dat de toetsen, wat de moeilijkheidsgraad betreft, onderling vergelijkbaar zijn. Daartoe werd bij de berekening van de totaalscore voor elk van de toetsen de moeilijkheidsgraad van de items meegewogen. In conditie I werd de range van punten voor de vooruitgang en achteruitgang, betrekkelijk willekeurig, gelegd tussen -15 en +15.[1]

De leerlingen kregen de beoordeling aangeboden in de vorm van een rapportje waarop men na elke toets kon aflezen in hoeverre men vooruit of achteruit was gegaan. De verschuiving in prestaties werd uitgedrukt in punten die werden aangegeven op een horizontale balk. De punten werden met behulp van labels ('sterk achteruit ..... 'sterk vooruit') van een betekenis voorzien. In figuur 1 is als voorbeeld het rapportje weergegeven voor de beoordeling van de verandering in prestaties van les 1 naar les 2.

In *conditie II* werd de zichtbaarheid van de contingentie tussen oordeel en taakgedrag versterkt door het aantal punten, dat bij een bepaalde vooruitgang of achteruitgang werd verstrekt, met een factor drie te vermenigvuldigen. Voor leerlingen in deze (en de volgende) conditie(s) kon het aantal punten dus variëren van -45 tot +45 (vgl. conditie I: van -15 tot +15). In het

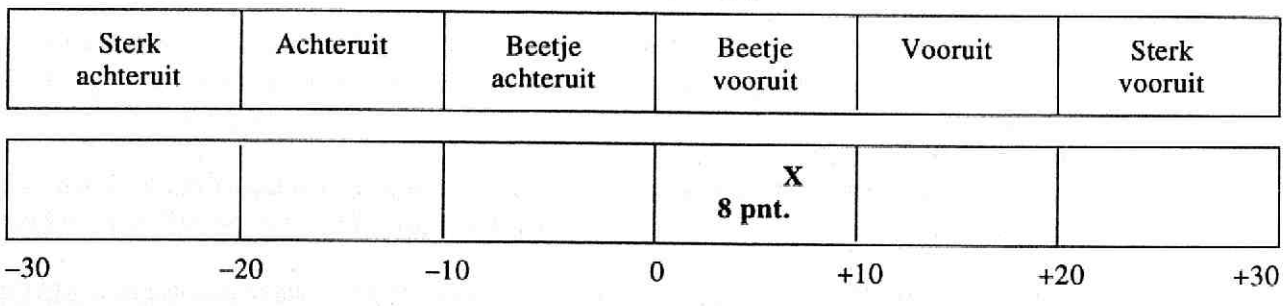


Fig. 1. Voorbeeld van een feedbackformulier (i.e. prestatieverschuiving van les 1 naar les 2).

rapportje, hierboven als voorbeeld gegeven, ontving een leerling (in de eerste conditie) voor een bepaalde vooruitgang bijvoorbeeld +8 punten ('beetje vooruit'). Eenzelfde vooruitgang werd in de tweede conditie 'beloond' met +24 punten ('sterk vooruit'). In het rapportje is binnen alle vier condities als maximale grens steeds -30 en +30 genomen. Als in het geval van een extreme verschuiving het aantal punten groter dan +30 of kleiner dan -30 was, werden de punten buiten de balk genoteerd.

In *conditie III* werd de voor- en achteruitgang, identiek aan de tweede conditie, versterkt weergegeven. Bovendien werd de leerling voorgehouden dat hij door zijn inspanning de prestaties kon beïnvloeden. Daartoe werd op het rapport een schriftelijke opmerking geplaatst waarvan de formulering verwees naar inzet als een plausibele causale factor (bijv. 'wat beter je best doen' bij een achteruitgang in prestaties). Naast alle elementen van de derde conditie werd in de *conditie IV* een taakdoel opgelegd. De leerkracht deelde aan iedere leerling mee, dat deze volgens hem een bepaald aantal punten méér moest kunnen halen op de toets aan het eind van de betreffende les. Aan het rapport was een extra pagina toegevoegd waarop, naast deze mededeling, het gestelde doel met behulp van een voor kinderen begrijpbare afbeelding zichtbaar was gemaakt. Om te vermijden dat leerlingen met een te hoog gesteld doel werden geconfronteerd, kozen we de hoogte van het taakdoel steeds op aselecte basis tussen de 10 en 20 punten vooruitgang; deze range bleek gemiddeld genomen haalbaar op grond van een inspectie van de feitelijke verschuivingen in prestaties bij Siero e.a. (1983).

Afhankelijke variabelen

Waargenomen controle

De leerlingen werd gevraagd in hoeverre bekwaamheid, inzet, toeval en taakmoeilijkheid een rol hadden gespeeld bij het behalen van het resultaat (c.q. de voor- of achteruitgang). Alleen de toeschrijving naar inzet en toeval worden opgevat als indicatoren voor de waargenomen controle. Voor de volledigheid en aansluiting op de literatuur zijn bekwaamheid en taakmoeilijkheid eveneens in de vragenlijst opgenomen.

De vier attributiefactoren werden in zes paren aangeboden, telkens in een andere volgorde. Bij ieder paar vormden de twee causale factoren de uiteinden van een vierpunts schaal. Op deze schaal kon men aangeven in hoeverre de ene dan wel de andere causale factor aan het oordeel had bijgedragen. Boven elk paar van causale factoren stond, afhankelijk van de geleverde prestatie (vooruit of achteruit gegaan) als aanhef 'ik ging vooruit, omdat ...' òf 'ik ging achteruit, omdat ...'. Op basis van de scores op de zes vierpunts schalen werd voor elk van de vier causale factoren een percentagescore berekend. Uitgaande van een verdeling van vijf 'causale' punten tussen elk paar per schaal (1 - 4 of 2 - 3 of 3 - 2 of 4 - 1) kan de score van een causale factor variëren tussen 3 (driemaal geringste causale bijdrage) en 12 (driemaal hoogste causale bijdrage). Deling door de som van de scores op de zes schalen (een constante, namelijk 30) levert de uiteindelijke percentagescore op. Deze kon dus variëren tussen 10 en 40 procent.

Prestaties

De tweede groep van afhankelijke variabelen betrof de prestaties op de vier toetsen. Iedere toets bestond uit een twintigtal meerkeuze items over de lesstof van de betreffende les. Alleen de derde en vierde toets zijn als effectvariabele van belang, omdat zij zijn afgenomen na de invoering van de beoordelingscondities.

Procedure

In de *eerste les* werd kort het onderwerp van de lessen geschetst en aangekondigd, dat er tijdens de vijf lessen regelmatig schriftelijk vragen zouden worden gesteld over de lesstof en de prestaties. Na afloop van de les werd een meer-keuze toets over de stof afgenomen. De leerkracht vermeldde bij de afname van deze en de daarop volgende toetsen steeds nadrukkelijk dat ze van ongeveer gelijke moeilijkheidsgraad waren.

Aan het begin van de *tweede les* vond een bespreking plaats van de moeilijkste items van de eerste toets. Er werd medegedeeld, dat iedere leerling tijdens de derde les een gecombineerd

oordeel zou ontvangen over de resultaten van de eerste en de nog af te nemen tweede toets. Om de leerlingen alvast te laten wennen aan de betrekkelijk complexe beoordelingsvorm werd deze alvast toegelicht. Op het eind van de les vond de afname van de tweede toets plaats.

De *derde en vierde les* hadden eenzelfde opbouw. Na de bespreking van de voorgaande toets vond een instructie plaats over de invulling van de attributielijst en werd de toelichting op de vooruitgangs- en achteruitgangsbeoordeling kort herhaald. De attributielijst en het rapportje werden rondgedeeld. Vervolgens werden de leerlingen verzocht in conditie I en II het aantal punten vooruitgang of achteruitgang en in conditie III en IV bovendien de opmerking over de geleverde inzet in te vullen op de eerste pagina van de attributielijst en gevraagd het oordeel goed in zich op te laten nemen alvorens de vragen te beantwoorden. In conditie IV was aan het rapport een extra pagina toegevoegd waarop het te behalen aantal punten vooruitgang was aangegeven. De leerkracht legde uit dat hij dit aantal had bepaald aan de hand van de toetsresultaten. In feite was de hoogte van het doel aselect bepaald. De les werd afgesloten met een toets over de besproken leerstof.

De procedure tijdens de *vijfde les* verliep grotendeels zoals beschreven voor les 3 en les 4. Omdat feedback na de vijfde les niet mogelijk was, werd in deze les afgezien van een toets over de besproken lesstof en werd geen taakdoel opgelegd. De vijf lessen, die ieder in totaal ongeveer 75 minuten duurden, werden binnen een periode van twee weken met tussenpozen van twee dagen gegeven. Er was, naast de leerkracht, steeds een tweede persoon aanwezig. Deze had tot taak na te gaan in hoeverre de leerkracht zich bij de realisatie van de beoordelingscondities, de afname van de vragenlijsten en het uitvoeren van de lessen hield aan het draaiboek.

Analyse

Het design omvat twee tussen-groepen factoren, namelijk het oordeel (2 niveau's: vooruitgang of achteruitgang) en de contingentiecondities (4 niveau's). Om de effecten van elke beoordelingsstrategie te bepalen, toetsten we steeds het verschil tussen twee opeenvolgende condities: conditie *II versus I* voor het effect van de zichtbaarheid van de contingente beoordeling, conditie *III versus II* voor het effect van attributiesturing en conditie *IV versus III* voor het effect van taakdoelen. De analyses werden afzonderlijk verricht op de gegevens na het eerste en na het tweede oordeel.

Op de attributiegegevens werden MANOVA's uitgevoerd.[2] Voor de bepaling van de effecten op de prestaties (de meer-keuzetoetsen) werd gebruik gemaakt van covariantieanalyses met als covariaten de prestaties op de eerste en tweede toets. Covariantieanalyse leidt tot schattingen van de effecten van de condities na correctie voor initiële verschillen in prestaties.

RESULTATEN

Zichtbaarheid (vergelijking conditie I en II)

In de eerste hypothese (1a) werd voorspeld dat het sterk zichtbaar presenteren van de contingente beoordeling de leerlingen ertoe zou brengen, aan zich zelf meer controle toe te schrijven (meer inzet- en minder toevalattributies). De gemiddelden in Tabel 1 en Tabel 2 en de multivariate en univariate F-toetsen in Tabel 3 laten zien dat de causale reacties op inzet en toeval deze hypothese ondersteunen.

Na het *eerste oordeel* schreven de leerlingen hun uitkomsten bij sterke zichtbaarheid, in vergelijking met de eerste conditie (zwakke zichtbaarheid), minder toe aan *toeval*; er zijn geen verschillen tussen de beide condities op de toeschrijving naar inzet.

Na het *tweede* oordeel zijn de causale reacties ook wat inzet betreft in de voorspelde richting: leerlingen aan wie de contingente beoordeling sterk zichtbaar werd gepresenteerd, beschouwden, in vergelijking met de leerlingen bij wie de contingentie van de oordelen zwakker zichtbaar was, hun uitkomsten als *meer door inzet* en *minder door toeval* bepaald. De gemiddelde waarden zijn weergegeven in Tabel 2. De effecten op inzet en toeval worden ook multivariaat ondersteund.

Tabel 1. Attributies van causaliteit na het eerste oordeel (les 3) onder condities van zwak (I) en sterk zichtbare contingentie (II), attributiesturing (III) en taakdoel stellen (IV).

Beoordelingscondities	Attributie van uitkomst naar			
	Inzet	Toeval	Bekwaamheid	Taakmoeilijkheid
Zichtbaarheid				
Zwak	23.7	26.4	21.6	28.1
Sterk	23.9	24.4	23.9	27.8
Attributiesturing				
Geen sturing	23.9	24.4	23.9	27.8
Wel sturing	25.3	22.8	25.5	26.4
Taakdoelen				
Geen doelen	25.3	22.8	25.5	26.4
Wel doelen	26.7	21.4	24.9	27.0
Richting oordeel				
Achteruit	19.8	25.8	24.6	29.8
Vooruit	30.0	21.6	23.4	25.0

Noot: Range van de attributiescores: minimaal 10, maximaal 40.

Tabel 2. Attributies van causaliteit na het tweede oordeel (les 4) onder condities van zwak (I) en sterk zichtbare contingentie (II), attributiesturing (III) en taakdoel stellen (IV).

Beoordelingscondities	Attributie van uitkomst naar			
	Inzet	Toeval	Bekwaamheid	Taakmoeilijkheid
Zichtbaarheid				
Zwak	23.9	26.7	22.1	27.3
Sterk	25.9	24.5	22.9	26.7
Attributiesturing				
Geen sturing	25.9	24.5	22.9	26.7
Wel sturing	26.3	25.2	22.8	25.7
Taakdoelen				
Geen doelen	26.3	25.2	22.8	25.7
Wel doelen	27.8	22.9	23.1	26.2
Richting oordeel				
Achteruit	21.4	26.7	22.8	29.1
Vooruit	30.6	23.0	22.6	23.8

Noot: Range van de attributiescores: minimaal 10, maximaal 40.

Er is tevens sprake van een niet voorspeld effect op de factor bekwaamheid: na het eerste oordeel schreef men zijn uitkomst bij grote zichtbaarheid, meer dan bij zwakke zichtbaarheid, toe aan bekwaamheid (M = 23.9 versus M = 21.6). Deze verschillen waren afwezig na het tweede oordeel.

Attributiesturing (vergelijking conditie II en III)

De toevoeging van attributiesturing aan de contingente beoordeling zou tot uitdrukking moeten

Tabel 3. Effecten van richting van het oordeel, zichtbaarheid, attributiesturing en taakdoelen op causale attributies (F-toetsen).

Bron	MANOVA[a]	ANOVA[b]			
		Inzet	Toeval	Bekwaamheid	Taakmoeilijkheid
Na eerste oordeel					
Richting oordeel	63.34***	183.93***	26.91***	1.85	45.88**
Beoordelingscondities					
Zichtbaarheid	5.91***	2.45	11.29***	10.83***	1.85
Attributiesturing	7.50***	7.34**	14.10***	8.65**	2.92
Taakdoelen	5.37***	6.89**	12.63***	3.37	< 1.0
Na tweede oordeel					
Richting oordeel	46.42***	128.37***	17.90***	< 1.0	60.85***
Beoordelingscondities					
Zichtbaarheid	4.30**	7.82**	9.37**	1.57	1.07
Attributiesturing	3.13*	6.63*	5.95*	< 1.0	1.01
Taakdoelen	4.13**	5.51*	11.00***	1.35	< 1.0

*** $p < .001$; ** $p < .01$; * $p < .05$

Noot: Er zijn geen significante interacties tussen de richting van het oordeel en de afzonderlijke beoordelingscondities.
[a] Multivariate toets op drie van de vier attributievariabelen; vrijheidsgraden: 3 en 286.
[b] Vrijheidsgraden bij ANOVA: 1 en 288.

komen in een stijging van de waargenomen controle (hypothese 2a). De resultaten na het eerste en tweede oordeel wijzen in deze richting. Zoals Tabel 1 en 2 laten zien, leidde de toevoeging van attributiesturing aan de contingente beoordeling, zowel na het *eerste* als na het *tweede oordeel*, tot meer toeschrijving aan de factor *inzet*.

Wat de toeschrijving naar de factor toeval betreft, zijn de resultaten iets minder duidelijk. Na het *eerste oordeel* reageerde men zoals voorspeld: de factor *toeval* werd als een *minder* belangrijke oorzaak opgevat als aan het oordeel attributiesturing was toegevoegd. Na het *tweede oordeel* waren de reacties tegengesteld aan de voorspelling.

De vermelde verschillen in toeschrijving van het oordeel aan inzet en toeval tussen de twee condities blijken ook multivariaat significant te zijn. Er was ook nu sprake van een niet voorspeld effect op de factor bekwaamheid: onder invloed van de attributiesturing beschouwden de leerlingen, meer dan in de conditie zonder attributiesturing, hun bekwaamheid als een belangrijke oorzaak van hun vooruitgang of achteruitgang.

Taakdoelen (vergelijking conditie III en IV)

In de inleiding is vooral ingegaan op de te verwachten gedragseffecten van taakdoelen. Er waren geen redenen om bij voorbaat te voorspellen dat het opleggen van taakdoelen ook directe cognitieve consequenties zou hebben. De duidelijke verschillen tussen de derde en vierde conditie (attributiesturing versus attributiesturing plus taakdoelen) op de inzet- en toevalattributies wijzen echter wel in deze richting.

In vergelijking met de conditie III (geen taakdoelen) blijken de leerlingen in de taakdoelconditie (IV), zowel na het *eerste* als na het *tweede oordeel*, hun resultaten *meer aan inzet* en *minder aan toeval* toe te schrijven. In Tabel 1 en 2 zijn de gemiddelde waarden weergegeven. Zoals Tabel 3 laat zien, zijn de verschillen zowel univariaat als multivariaat significant.

Tabel 4. Attributies van causaliteit naar inzet, uitgesplitst naar beoordelingscondities en richting van het oordeel.

Richting oordeel	Beoordelingscondities			
	Zwakke zichtbaarheid	Sterke zichtbaarheid	Attributie-sturing	Taakdoel-stelling
Na eerste oordeel				
Achteruit	17.3	18.7	20.3	23.0
Vooruit	30.2	29.1	30.2	30.4
Na tweede oordeel				
Achteruit	18.0	21.9	21.4	24.1
Vooruit	29.7	29.8	31.1	31.5

Noot: Range van de attributiescores: minimaal 10, maximaal 40.

Richting van het oordeel

Is de toeschrijving van het falen aan de bestede inzet te beïnvloeden door de contingentie tussen oordeel en prestaties te vergroten? Een trend-analyse waarin het cumulatieve effect van de vier in contingentie oplopende condities op de causale attributies werd getoetst, laat inderdaad een lineaire stijging van inzet-attributies zien van conditie I naar conditie IV die binnen de achteruit-conditie anders verloopt dan binnen de vooruit-conditie. Het effect komt het sterkst naar voren na het eerste oordeel: F (1, 288) = 2.36 ($p < .01$).

De gemiddelde attributiescores voor inzet in Tabel 4 laten zien dat de leerlingen na *het eerste oordeel* hun achteruitgang meer naar inzet toeschreven naarmate de contingentie in steeds sterkere mate van conditie I naar conditie IV benadrukt werd (F (1, 288) = 14.09, $p < .001$). Leerlingen die vooruitgingen, schrijven hun uitkomst binnen alle vier contingentie-condities in even sterke mate toe aan inzet (F (1, 288) < 1.00).

De inzetattributies na het tweede oordeel laten een sterk gelijkend resultaat zien. Hoewel in de voorspelde richting, is het effect, tweezijdig getoetst, niet significant, namelijk F (1, 288) = 2.38 ($p < .12$). Ook nu werd de toeschrijving van de achteruitgang naar inzet beïnvloed door de mate van contingentie (hoe meer contingentie, des te meer attributie naar inzet: namelijk 18.0, 21.9, 21.4, en 24.1; F (1, 288) = 9.90, $p < .01$). Na een vooruitgangsoordeel was deze lineaire relatie afwezig (namelijk 29.7, 29.8, 31.1, en 31.5; F (1, 288) = 2.35, niet significant).

Ondanks de stimulerende werking van de contingente beoordeling op de toeschrijving van de achteruitgang aan inzet, waren de leerlingen ook in deze geïndividualiseerde beoordelingscontext onderhevig aan de zogenaamde attributievertekening: leerlingen die vooruitgingen, schreven hun succes zowel na het eerste oordeel (zie Tabel 1) als na het tweede oordeel (zie Tabel 2) *meer aan inzet* en *minder aan toeval en taakmoeilijkheid* toe. Opvallend is het identieke patroon van causale reacties na beide oordelen: steeds hetzelfde contrast tussen succes en falen op inzet, toeval en taakmoeilijkheid en geen differentiële attributies naar bekwaamheid. In Tabel 3 zijn de toetsgegevens voor de univariate en multivariate effecten weergegeven. Zowel de multivariate toets op de attributievariabelen na het eerste oordeel als die na het tweede oordeel ondersteunen de univariate effecten.

Prestaties

Elk van de drie strategieën – zichtbaarheid van de contingentie, attributiesturing en taakdoelen – zou de leerling moeten aanzetten tot betere taakresultaten (hypothesen 1b, 2b en 3). Na het *eerste* oordeel bleken de drie beoordelingsstrategieën inderdaad tot een verbetering van taakresultaten te leiden. In Tabel 5 zijn de oorspronkelijke en aangepaste gemiddelden per conditie

Tabel 5. Gemiddelde toetsscores, uitgesplitst naar beoordelingsconditie.

	Toets					
	Oorspronkelijke scores[a]				Aangepaste scores[b]	
Beoordelingsconditie	1	2	3	4	3	4
Zwakke zichtbaarheid	7.77	0.09	−1.50	3.00	−4.67	0.04
Sterke zichtbaarheid	2.78	−1.78	0.35	−0.48	−0.19	−0.92
Attributiesturing	3.01	0.27	1.54	−2.76	0.02	−3.24
Taakdoelstelling	−7.93	-6.10	−0.69	−0.90	4.53	2.97

[a] Gemiddelde toetsscores, per toets getransformeerd tot afwijkingsscores met een standaardafwijking gelijk aan 25.
[b] Via covariantie-analyse aangepast voor initiële verschillen op toets 1 en toets 2.

weergegeven. De aangepaste scores betreffen de prestaties op toets 3 en toets 4 nadat via covariantie-analyse gecorrigeerd is voor initiële verschillen op toets 1 en toets 2.

Na het eerste oordeel (toets 3 in Tabel 5, onder 'aangepaste scores') zette de sterk zichtbare contingentie, meer dan de zwak zichtbare contingentie, de leerlingen aan tot betere prestaties: F (1, 286) = 5.80, $p < .05$. De toevoeging van attributiesturing aan de sterk zichtbare contingente beoordeling resulteerde eveneens in een beter taakresultaat (F (1, 286) = 4.92, $p < .05$). Tenslotte, de leerlingen aan wie, naast attributiesturing, ook nog een taakdoel werd opgelegd, werden het sterkste gestimuleerd; zij waren in staat prestaties te leveren die nog eens uitstegen boven die van de leerlingen binnen de derde conditie (attributiesturing): F (1, 286) = 7.58, $p < .01$. Dit krachtige cumulatieve effect van de vier condities ontbreekt na het *tweede* oordeel (toets 4 onder 'aangepaste scores' in Tabel 5).

DISCUSSIE

Heeft vergroting van de controle binnen de taak- en beoordelingssituatie een stimulerende werking op het taakgedrag van leerlingen? Deze vraag stond centraal in het hier gerapporteerde onderzoek. De hier toegepaste controle-vergrotende maatregelen blijken over het geheel genomen succesrijk zijn. Het sterker zichtbaar maken van de contingentie tussen taakgedrag en oordeel leidt, ook bij herhaalde toepassing, tot een krachtig besef van controle. Ook het versterken van de attributies naar inzet en het opleggen van taakdoelen hebben een controle-versterkende betekenis voor de leerling. Leerlingen blijken onder invloed van de drie hier toegepaste strategieën ter verhoging van de waargenomen controle (zichtbaarheid, attributiesturing en taakdoelen) de door hen geleverde inzet inderdaad op te vatten als een belangrijke determinant van de oordelen. Wat prestaties betreft zien we eveneens een verbetering onder invloed van de drie contingentiestrategieën.

De verhoging van de zichtbaarheid en de attributiesturing leidde initieel ook tot een, niet voorspelde, verhoogde toeschrijving naar bekwaamheid, een stabiele interne factor. Op zich is dat niet zo verwonderlijk. Leerlingen zijn gewend zich met elkaar te vergelijken en gewend oordelen over hun prestaties in verband te brengen met hun bekwaamheden. Het ontvangen van voortgangsfeedback zal dan in eerste instantie ook gevoelens van competentie of incompetentie oproepen: "Ik ben lekker veel meer vooruitgegaan dan jij, dus ik ben beter dan jij".

In deze studie leggen leerlingen de verklaring voor hun vooruitgang meer bij de geleverde inzet en verklaren hun achteruitgang vooral vanuit toeval en de moeilijkheid van de taak. Dit

patroon vertoont gelijkenis met het patroon dat we binnen ons klassikale beoordelingsexperiment (Siero et al., 1983) aantroffen tussen de hoog en laag presterende leerlingen. Opvallend is dat ook binnen deze geïndividualiseerde beoordelingsstructuur, een taakstructuur waarmee juist wordt beoogd de verantwoordelijkheid voor ongunstige uitkomsten te vergroten, de leerlingen minder verantwoordelijkheid nemen voor negatieve dan voor positieve resultaten. Kennelijk is de gunstige dan wel ongunstige aard van het oordeel een dermate dominant gegeven dat deze tendens om negatieve uitkomsten extern te attribueren ondanks de voortgangsbeoordeling optreedt.

Tegelijkertijd blijkt het echter mogelijk leerlingen meer te oriënteren op hun inzet als een verklarende factor voor negatieve resultaten door het beïnvloedbare karakter van het oordeel extra te benadrukken. In dit onderzoek werd 'het falen' het sterkst met inzet in verband gebracht als de drie strategieën, zichtbaarheid van de contingentie tussen oordeel en prestaties, attributiesturing en het opleggen van taakdoelen, in combinatie werden toegepast. Ook in de praktijk is het zeer wel mogelijk om deze strategieën met succes tegelijkertijd toe te passen mits de beoordeling gebaseerd is op de vooruitgang en achteruitgang van de individuele leerling.

De vergroting van de controle laat over het algemeen vooral in het begin een sterk effect zien zowel op waargenomen controle als op prestaties. Wat betreft de waargenomen controle, is het met name de attributiesturing in combinatie met de contingente beoordeling die na het tweede oordeel minder effectief blijkt te zijn. Na het tweede oordeel schat men zijn controle weliswaar hoger in, maar dit komt alleen tot uitdrukking in de attributies naar inzet. De beide andere strategieën, zichtbaarheid van de contingentie en taakdoelen, blijken de leerlingen wel, zowel na het eerste als na het tweede oordeel, ertoe aan te zetten aan zichzelf meer controle op de uitkomsten toe te schrijven.

De drie controle-strategieën hadden alleen in het begin een systematische verbetering van prestaties tot gevolg. Deze invloed was niet aanwijsbaar bij de prestaties van de leerlingen op de toets na het *tweede* oordeel. Wellicht hebben de door ons toegepaste ingrepen door hun voor de leerlingen wat ongewoon krachtige aard slechts een korte-termijneffect teweeggebracht. Het is voor een leerling moeilijk om na een zeer grote prestatieverbetering zo te presteren dat wederom sprake is van een forse vooruitgang. Dit kan leiden tot een verstoring van de beoogde contingentie. Binnen de hier toegepaste beoordelingsprocedure is deze mogelijk opgetreden complicatie onvermijdelijk.

Dit betekent echter niet dat controleverhogende maatregelen niet langere tijd werkzaam zouden kunnen zijn. Dat dit wel mogelijk was, bleek namelijk uit een vier maanden durend veldexperiment (Van Oudenhoven, Siero, Veen & Withag, 1983) dat plaats had onder leerlingen uit jaargroep 5. Drie condities werden toegepast: de gebruikelijke klassikale beoordeling, een conditie waarin voortgangsbeoordeling werd toegepast en een conditie waarin de voortgangsbeoordeling gecombineerd werd met individueel aangepaste oefeningen. In de twee condities met voortgangsbeoordeling waren de prestaties beter, maar deze effecten consolideerden zich vooral in de laatste, vierde maand van het onderzoek. De langere uitwerking van de voortgangsbeoordeling op de prestaties vloeide voort uit het feit dat de contingentie 'verpakt' was in bestaande lesprogramma's en de ingreep daardoor gebaseerd was op kleinere subjectieve vooruitgangsstappen. Overigens is de studie van Van Oudenhoven e.a. (1983) tevens een voorbeeld van hoe de hier kunstmatig onderscheiden beoordelings- en contingentiestrategieën via integratie in bestaande spellingonderwijsmethoden kunnen worden geïmplementeerd.

NOTEN

1. Op basis van de verschuivingen in prestaties in Siero e.a. (1983) is een voor alle leerlingen identieke lineaire transformatie toegepast, zodanig dat de vooruitgang/achteruitgang alleen in incidentele gevallen buiten de range (in conditie I tussen -15 en +15) zou vallen.
2. Omdat de vierde attributievariabele vanwege het gebruik van percentagescores afleidbaar was uit de drie andere variabelen, heeft de multivariate toets steeds betrekking op drie van de vier attributievariabelen.

LITERATUUR

Alloy, L., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review, 91*, 112-149.

Bandura, A., & Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of Personality and Social Psychology, 56*, 805-814.

Becker, L.J. (1978). Joint effect of feedback and goal setting on performance: A field study of residential energy conservation. *Journal of Applied Psychology, 63*, 428-433.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90*, 2, 272-292.

Forsyth, D.R. & McMillan, J.H. (1981a). The attributional cube and reactions to educational outcomes. *Journal of Educational Psychology, 73*, 5, 632-641.

Forsyth, D.R. & McMillan, J.H. (1981b). Attributions, affect, and expectations: a test of Weiner's three-dimensional model. *Journal of Educational Psychology, 73*, 3, 393-403.

Garland, H. (1983). Influence of ability, assigned goals, and normative information on personal goals and performance: A challenge to the goal attainability assumption. *Journal of Applied Psychology, 68*, 1, 20-30.

Hewstone, M. & Jaspers, J. (1983). A re-examination of the roles of consensus, consistency and distinctiveness: Kelley's cube revisited. *British Journal of Social Psychology, 22*, 41-50.

Hewstone, M. & Jaspers, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology, 53*, 663-672.

Jackson, S.E. & Zedeck, S. (1982). Explaining performance variability: Contributions of goal setting, task characteristics, and evaluative contexts. *Journal of Applied Psychology, 67*, 759-768.

Kelley, H.H. (1973). The processes of causal attribution. *American Psychologist*, 107-128.

Locke, E.A.& Latham, G.P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, New Jersey: Prentice Hall.

Mento, A., Steel, R., & Karren, R. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966-1984. *Organizational Behavior and Human Decision Processes, 39*, 52-83.

Nisbett, R.E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, New Jersey: Prentice-Hall.

Oudenhoven, J.P. van, Siero, F.W., Veen, P. & Withag, J. (1983). Effects of individualised feedback and instruction on effort attributions, ability attributions and spelling achievement. *Educational Studies, 9*, 105-113.

Siero, F.W., Oudenhoven, J.P. van & Veen, P. (1983). Causale attributies na oordelen over prestaties: een veldexperiment in het onderwijs. *Gedrag - tijdschrift voor psychologie, 11*, 4, 190-204.

Weiner, B. (1982). An attributional based theory of motivation and emotion: Focus, range, and issues. In N.T. Feather (Ed.) *Expectations and Actions: Expectancy-Value Models in Psychology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 163-203.

Weiner, B. (1985). An attributional theory of achievement motivation. *Psychological Review, 92*, 548-573.

Zuckerman, M. (1979). Attributions of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality, 47*, 245-287.

Manuscript ontvangen 16-7-1993
Definitieve versie ontvangen 30-11-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, pp. 355-368

Opzet en effecten van een preventieve sociale-vaardigheidstraining voor jonge adolescenten

Jan O. Bijstra, Heleen P. van der Kooi en Henk T. van der Molen

ABSTRACT

In this article, the principles of a preventive social skills training for early adolescents and the results of a first evaluation study are discussed. The training which was based on cognitive-social-learning principles, was run on three secondary schools. Pre- and posttest results of 71 13-15-year-olds who had followed the training, were compared with those of 315 youngsters who had not been involved in the training. Participants showed an increase of knowledge about social situations and they became more active in such situations. Furthermore, content and structure of the training was positively evaluated. However, the training did not result in a decrease of anxiety in social situations. This latter result is explained in terms of the relatively little attention which has been paid in the training to the presence of irrational thoughts. Another issue in the discussion is concerned with the way in which adolescents are recruited for the training.

INLEIDING

Uit onderzoek naar het vóórkomen van psychosociale problemen bij adolescenten blijkt dat de meerderheid van de jongeren zich zonder veel problemen ontwikkelt. Er is echter een kleine groep jongeren (ruwweg geschat op 10 à 20 procent) waarvan de ontwikkeling gekenmerkt wordt door stagnatie en steeds groter wordende problemen (Rutter, 1980; Petersen, 1988). Uit de literatuur blijkt dat deze problemen worden veroorzaakt door een complex van factoren (Rutter, 1989). Een belangrijke bepalende factor in dit geheel is het sociale-vaardigheidsniveau van jongeren.

Onderzoekers hebben aangetoond dat er een verband bestaat tussen het ontbreken van adequate sociale vaardigheden op jonge leeftijd en het bestaan van allerlei problematiek op latere leeftijd. Cowen, Pederson, Babigian, Izzo en Trost (1973) vonden bijvoorbeeld dat sociaal geïsoleerde kinderen op latere leeftijd verhoudingsgewijs vaak in aanraking kwamen met sociale zorg- en gezondheidsvoorzieningen. Ringrose en Nijenhuis (1986) geven een overzicht van onderzoek waaruit blijkt dat kinderen met terugtrek- en vermijdingsgedrag te maken kunnen krijgen met problemen op school (aanpassingsproblemen, slechte schoolprestaties, voortijdige schoolverlating), een vertraagde intellectuele ontwikkeling, jeugddelinquentie en psychiatrische problematiek in de volwassenheid. Bartels (1986) suggereert dat gedragsmoeilijkheden en delinquent gedrag mede worden veroorzaakt door een tekort aan sociale vaardigheden.

Voor jongeren die al duidelijk probleemgedrag vertonen, zijn er verscheidene corrigerende interventieprogramma's in omloop (bijv. Bartels, 1986; Van der Zee, Van der Molen & Van der Beek, 1989; Van Bilsen, Jonkers, Schuurman & Swager, 1991). De vraag rijst echter of de hierboven geschetste problematiek niet voorkómen kan worden. Juist vanuit preventief oogpunt lijkt het gewenst om jongeren al aan het begin van de adolescentie in de gelegenheid te stellen een sociale-vaardigheidstraining te volgen (Meichenbaum, 1985). Immers, het proces van ontwikkeling naar zelfstandigheid waar jongeren in de adolescentie mee te maken krijgen, doet een groot beroep op allerlei sociale vaardigheden.

Adres: Dr. J.O. Bijstra, Rijksuniversiteit Groningen, Vakgroep Psychologie, Grote Kruisstraat 2/1, 9712 TS Groningen.
Noot: Dit onderzoek werd via het onderzoeksprogramma "Determinanten van Gezondheid" mogelijk gemaakt door een subsidie van het Ministerie van WVC.

Vanuit dat oogmerk is aan de Vakgroep Psychologie van de Rijksuniversiteit Groningen een preventieve sociale-vaardigheidstraining voor adolescenten ontwikkeld. Deze training is bestemd voor normaal functionerende jongeren, dus jongeren die (nog) niet tot een duidelijk definieerbare probleemgroep behoren. Uiteraard bevinden zich binnen deze "normale" groep zowel jongeren die in sociaal opzicht minder adequaat zijn als jongeren die over uitstekende sociale vaardigheden beschikken. In beginsel is de training bestemd voor sociaal vaardige en minder vaardige jongeren, maar daarbij kan wel worden opgemerkt dat jongeren die aan de negatieve kant van het sociale-vaardigheidsspectrum zitten, een noodzakelijker doelgroep vormen dan zij die reeds goede sociale vaardigheden hebben. Echter, vanuit het idee dat een sociale-vaardigheidstraining in principe voor iedereen leerzame aspecten kan bevatten, kunnen ook sociaal vaardigen baat hebben bij de training.

In dit artikel worden eerst de theoretische achtergronden van de sociale-vaardigheidstraining kort uiteengezet en vervolgens wordt aangegeven hoe de training is ontwikkeld. Tot slot wordt ingegaan op een eerste onderzoek naar de effecten van de training.

Doelstellingen van de training

Het hoofddoel van de training is het bevorderen van sociale vaardigheden van adolescenten. Dit hoofddoel is opgedeeld in drie subdoelen (vgl. Van der Molen, 1985):

- Kennisdoel: het vergroten van kennis en inzicht in sociale vaardigheden.
- Gedragsdoel: het uitbreiden van het repertoire aan sociale vaardigheden.
- Belevingsdoel: het verminderen van spanning in sociale situaties.

Theoretische achtergronden

De training sluit aan bij een binnen de vakgroep Psychologie van de Rijksuniversiteit Groningen ontwikkelde sociale-vaardigheidstraining voor verlegen mensen (Van der Molen, 1982; 1984; 1985). De methode die bij de uitvoering van deze training wordt gehanteerd, is gebaseerd op cognitief-sociaal-leertheoretische principes (Bandura, 1977; Goldstein 1973, 1981; Ladd & Mize, 1983). De elementen uit deze methode zijn instructie, modeling, oefenen, feedback en transfer.

Instructie

De instructie bestaat uit een beperkte hoeveelheid informatie over inhoud, doel en toepassingsmogelijkheden van de betreffende vaardigheid. Elke vaardigheid is opgesplitst in leerpunten (Goldstein, 1981) waardoor deze gemakkelijker onthouden en gereproduceerd kan worden.

Modeling

De vaardigheid wordt gedemonstreerd met behulp van goede en foute videovoorbeelden. Op deze wijze kan aan de deelnemer het verschil tussen effectief en niet-effectief handelen duidelijk worden gemaakt. Het goede voorbeeld vervult hierbij een modelfunctie.

Oefenen

Het oefenen van de vaardigheden gebeurt door middel van rollenspelen. Het belang van deze oefenvorm is ook binnen diverse andere settings benadrukt (Ivey, 1971; Goldstein, 1973; Beekers, 1982; Van der Molen, 1984; 1985).

Feedback

Om het leereffect tijdens de rollenspelen te optimaliseren, is het van belang dat er feedback wordt gegeven op het toepassen van de vaardigheden in de rollenspelen.

Transfer

Goldstein (1973) noemt als belangrijkste element van een sociale-vaardigheidstraining de transfer. De deelnemers moeten de vaardigheden niet alleen in de rollenspelen oefenen maar deze juist in het dagelijks leven leren toepassen. Het geven van "huiswerkopdrachten" is een goede manier om deze transfer te bevorderen.

Uitwerking van de theorie naar de training

De besproken elementen zijn alle in de sociale-vaardigheidstraining voor adolescenten opgenomen. Ten behoeve van de instructie is voor de deelnemers aan de training een klapper geschreven (Van der Kooi, 1992). In afzonderlijke lessen worden het doel, de inhoud en de toepassingsmogelijkheden van de sociale vaardigheden uitgelegd. Bovendien zijn de vaardigheden opgedeeld in een aantal leerpunten. De vaardigheid die centraal staat, wordt tijdens de bijeenkomst kort toegelicht aan de hand van een onderwijsleergesprek.

Voor de realisatie van het onderdeel modeling is gebruik gemaakt van bestaande videobanden (Van der Molen, 1982; Van der Zee, Van der Molen & Van der Beek, 1989). De eerste oefening met een vaardigheid bestaat uit een zogenaamde droogzwemoefening. Droogzwemoefeningen zijn groepsoefeningen waarbij in een speelse vorm wordt kennisgemaakt met de nieuw aan te leren vaardigheid. Vervolgens wordt de nieuwe vaardigheid geoefend in een rollenspel. De feedback op het rollenspel wordt door de begeleider van de groep gegeven, eventueel aangevuld met feedback door de trainingsdeelnemers. Bij het geven van feedback wordt eerst zo concreet mogelijk besproken wat goed ging. Vervolgens worden fouten, bijvoorbeeld het niet of niet effectief toepassen van de vaardigheid, besproken. Om de transfer van de vaardigheden te bewerkstelligen, zijn in de klapper huiswerkopdrachten opgenomen die telkens aan het begin van de volgende bijeenkomst worden besproken.

Opbouw van de training

De training heeft een duidelijke structuur aangezien de elementen uit de cognitief-sociale-leertheorie telkens in dezelfde volgorde terugkomen. Deze vaste structuur kent verschillende voordelen. In de eerste plaats biedt de structuur veiligheid; de deelnemers weten wat er gaat gebeuren en zij weten wat er van hen wordt verwacht. In de tweede plaats biedt de structuur houvast aan de begeleider(s). In de derde plaats bevordert de vaste structuur de overdraagbaarheid van de training.

Inhoud van de training

Wanneer twee mensen met elkaar communiceren, kan een onderscheid worden gemaakt tussen de "ontvanger" en de "zender". Analoog is een onderscheid te maken tussen ontvanger- of luistervaardigheden en zender- of assertieve vaardigheden. Beide categorieën sociale vaardigheden zijn in de training opgenomen. Luistervaardigheden zijn opgenomen om deelnemers meer greep te laten krijgen op het gespreksproces: "actief" luisteren betekent namelijk niet dat men stil blijft, maar dat men op adequate wijze reageert op hetgeen iemand anders vertelt. Bovendien draagt het oefenen van luistervaardigheden bij tot het empathisch vermogen van de deelnemers. Assertieve vaardigheden zijn opgenomen om deelnemers te leren initiatieven te tonen en om hen te leren adequaat om te gaan met lastige situaties (Van der Molen, 1984).

De luistervaardigheden die in de training aan de orde komen zijn: aandachtgevend gedrag, vragen stellen, samenvatten en ingaan op gevoel. De assertieve vaardigheden bestaan uit: zelf iets vertellen, een praatje beginnen, iets vragen, "nee"-zeggen, kritiek geven en kritiek krijgen. De diverse vaardigheden worden besproken aan de hand van allerlei dagelijkse sociale situaties waar jongeren mee te maken (kunnen) krijgen. Daarmee komt dus het gehele scala van sociale relaties aan bod (ouders, vrienden, klasgenoten, docenten etc.).

Begeleidershouding

Om de training op de juiste manier over te brengen, wordt van de begeleider een bepaalde houding verlangd. Deze houding kenmerkt zich door de aspecten acceptatie, directiviteit, constructiviteit en flexibiliteit (Lang & Van der Molen, 1992). Met acceptatie wordt bedoeld dat de begeleider de gevoelens en eigenaardigheden van de deelnemers accepteert. Wanneer de deelnemers zich geaccepteerd weten, voelen zij zich veilig, worden minder onzeker en gaan dus minder defensief reageren. Dit is juist in nieuwe situaties zoals in een sociale-vaardigheidstraining, van belang.

Onder directiviteit wordt verstaan: het duidelijk richting geven aan de activiteiten tijdens de

training en het doelgericht werken aan een bepaalde taak. Doelgerichtheid en duidelijkheid zijn alleen te bereiken wanneer er volgens een gestructureerde manier wordt gewerkt. Het draaiboek van de training waarin het programma grotendeels vastligt, biedt hiervoor een goed aanknopingspunt. Een constructieve houding kenmerkt zich door een merkbare, positieve waardering voor datgene wat de deelnemers aan bruikbare opmerkingen inbrengen en wat zij goed doen in hun praktische oefeningen. Dit houdingsaspect is van belang omdat deelnemers vaak geneigd zijn datgene wat goed gaat, vanzelfsprekend te vinden en alleen de aandacht te richten op wat niet goed gaat. Flexibiliteit heeft als begrip een brede betekenis. Het betekent zowel "buigzaam" en "lenig" als ook "gemakkelijk" en "gedwee". Als onderdeel van de begeleidershouding wordt met flexibiliteit bedoeld: geestelijke lenigheid van de begeleider om met deelnemers te werken.

Begeleidersvaardigheden

De boven beschreven begeleidershouding is in concreet gedrag te vertalen door het toepassen van een aantal begeleidersvaardigheden: het doelgericht begeleiden van een onderwijsleergesprek, het doelgericht presenteren en nabespreken van een videovoorbeeld, het doelgericht structureren en nabespreken van rollenspelen en het doelgericht begeleiden van groepsoefeningen.

In het nu volgende gedeelte zullen de resultaten worden besproken van een onderzoek waarin jonge adolescenten (13 - 15 jaar) volgens bovenstaande principes een preventieve sociale-vaardigheidstraining hebben gevolgd. Deze groep jongeren (de trainingsgroep) werd vergeleken met een groep jongeren die geen training had gevolgd (de controlegroep). In de eerste plaats werd bestudeerd of de training effectief was geweest voor de totale groep jongeren. De verwachting was dat de trainingsgroep zou profiteren van de sociale-vaardigheidstraining. In de tweede plaats werd bestudeerd of de training effectief was geweest voor subgroepen jongeren en of sommige subgroepen meer van de training zouden profiteren dan andere subgroepen. Met betrekking tot de bepaling van de subgroepen ging het om de onafhankelijke variabelen sociale-vaardigheidsniveau, geslacht, leeftijd en opleiding. De verwachting was dat alle trainingssubgroepen zouden profiteren van de training. Ten aanzien van de vraag of sommige subgroepen meer van de training zouden profiteren dan andere subgroepen werd alleen voor de variabele sociale-vaardigheidsniveau een hypothese geformuleerd: de verwachting was dat jongeren met de meest adequate vaardigheden het minst zouden profiteren aangezien bij hen weinig verbetering meer viel te verwachten. Ten aanzien van de variabelen geslacht, leeftijd en opleiding was er geen reden te veronderstellen dat sommige subgroepen meer van de training zouden profiteren dan andere subgroepen.

Het onderzoek zoals hier besproken, is ingebed in een groter onderzoek naar de relatie tussen sociale vaardigheden en verschillende aspecten van psychosociaal functioneren van jongeren en mogelijke effecten van een sociale-vaardigheidstraining op deze aspecten (zie Bosma, Bijstra, Jackson, Van der Kooi & Van der Molen, 1991; Bijstra, Van der Kooi, Bosma, Jackson & Van der Molen, 1993).

METHODE

Proefpersonen

Een groep van 660 13 t/m 15-jarige jongeren – afkomstig van vier middelbare scholen in en rond Groningen – was betrokken bij de voormeting. Een groep van 78 jongeren – de trainingsgroep – meldde zich vervolgens aan voor deelname aan de sociale-vaardigheidstraining. Deze jongeren waren allen afkomstig van drie van de vier scholen (zie ook "Procedure"). Uiteindelijk werd gebruik gemaakt van de gegevens van 71 jongeren aangezien drie voortijdig stopten met de training en van vier slechts zeer onvolledige nameting-gegevens beschikbaar waren.

De controlegroep werd gevormd door de jongeren die zich niet aanmeldden voor de training, maar van wie wel nameting-gegevens beschikbaar waren. Deze groep bestond uit 323 jongeren. Uiteindelijk werd gebruik gemaakt van de gegevens van 315 jongeren omdat de nameting-

Tabel 1. Enkele biografische gegevens over de jongeren in de trainings- en controlegroep.

	trainingsgroep		controlegroep	
	N	%	N	%
meisjes	40	56	188	59
jongens	31	44	127	41
totaal	71	100	315	100
13-jarigen	56	79	161	51
14-jarigen	11	15	112	36
15-jarigen	4	6	42	13
totaal	71	100	315	100
mavo	7	10	94	30
havo	3	4	64	20
havo/vwo	37	52	107	34
vwo	24	34	50	16
totaal	71	100	315	100

gegevens ook hier in enkele gevallen zeer onvolledig waren. In Tabel 1 zijn enkele gegevens vermeld van de twee groepen jongeren. Uit de tabel blijkt dat het aantal meisjes en jongens in beide groepen procentueel gezien vrijwel gelijk was en dat de trainingsgroep gemiddeld jonger was dan de controlegroep (13,3 tegen 13,6 jaar). Verder bleek de trainingsgroep vooral te bestaan uit jongeren van de twee hoogste opleidingsniveaus (86%). Tot slot dient te worden opgemerkt dat één school haar leerlingen uit een minder hoog sociaal-economisch gebied betrekt dan de andere drie scholen; juist deze school leverde alle mavo- en havo-trainingsdeelnemers.

Materiaal

In het onderzoek werd gebruik gemaakt van drie vragenlijsten: 1) een zelf-geconstrueerde kennistoets, 2) een aangepaste adolescentenversie van de Schaal voor Interpersoonlijk Gedrag, de SIG (Arrindell, De Groot & Walburg, 1984) en 3) een evaluatieformulier.

De kennistoets werd gebruikt om na te gaan of er sprake was van kennisvermeerdering ten aanzien van sociale situaties (kennisdoel). Er werden twee parallelversies ontwikkeld: één werd onmiddellijk voor en één direct na de training afgenomen. De twee versies bestaan ieder uit 20 drie-keuzevragen. Een voorbeeld van een vraag is: "Als iemand je in de rede valt, dan....." a) "zeg je dat diegene stil moet zijn"; b) "vraag je diegene of hij/zij je niet in de rede wil vallen"; c) "zeg je niets en laat diegene aan het woord". De alpha-coëfficiënten van beide versies waren aan de lage kant, namelijk .58 voor de voormetingversie en .62 voor de nametingversie.

De SIG werd gebruikt om na te gaan of de training een positief effect had op de mate van gedrag in sociale situaties (gedragsdoel) en op de door de deelnemers ervaren spanning in dergelijke situaties (belevingsdoel). De SIG is opgebouwd uit vier subschalen die verschillende typen sociale situaties reflecteren: 1) kritiek geven, weigeren: situaties waarin het adequaat is negatieve gevoelens te uiten, 2) kritiek ontvangen, aandacht en hulp vragen: situaties waarin het adequaat is eigen onzekerheid en eigen ontoereikendheid te uiten, 3) contact leggen, mening geven: situaties waarin het adequaat is jezelf kenbaar te maken en 4) complimenten, waardering geven en ontvangen: situaties waarin het adequaat is positieve gevoelens te uiten. Respondenten dienen aan te geven hoe vaak zij zich in sociale situaties begeven (frequentiedimensie) en in hoeverre zij dergelijke situaties als "eng" ervaren (spanningsdimensie). Iedere subschaal levert dus een frequentie- en een spanningsscore op. De scores op de frequentiedimensie van de SIG werden in dit onderzoek gebruikt als subjectieve gedragsmaat aangezien het om door de jonge-

ren zelf gerapporteerd gedrag gaat. De scores op de spanningsdimensie van de SIG werden gebruikt als belevingsmaat aangezien zij weergeven hoe jongeren verschillende sociale situaties beleven.

De SIG werd tevens gebruikt voor het bepalen van de onafhankelijke variabele sociale-vaardigheidsniveau. Arrindell et al. suggereren dat het mogelijk is zogenaamde sociale-vaardigheidsprofielen te schetsen door de dimensies frequentie en spanning met elkaar te combineren. Wanneer beide dimensies ruimtelijk worden weergegeven door middel van een assenkruis kunnen er theoretisch vier subgroepen worden onderscheiden: 1) subassertieven: begeven zich weinig in sociale situaties en ervaren veel spanning in dergelijke situaties; 2) onverschilligen: begeven zich weinig in sociale situaties, maar ervaren ook weinig spanning in dergelijke situaties; 3) assertieven: begeven zich frequent in sociale situaties en ervaren weinig spanning in dergelijke situaties 4) angstige uitvoerders: begeven zich frequent in sociale situaties, maar ervaren wel veel spanning in dergelijke situaties.

Aangezien de SIG van oorsprong een vragenlijst voor volwassenen is, werden enkele veranderingen aangebracht. Deze veranderingen betroffen hoofdzakelijk de formulering van de items. Bovendien werden enkele items weggelaten omdat deze geen betrekking hadden op de leefwereld van de jongeren. Tot slot werd de spanningsdimensie van een vijfpuntsschaal veranderd in een vierpuntsschaal aangezien op basis van pilotonderzoek de indruk bestond dat sommige jongeren wat problemen hadden met de differentiatie tussen de oorspronkelijke vijf antwoordcategorieën. Een voorbeeld van een item is: "Een gesprekje beginnen met een jongen of meisje die je niet kent" (spanningsdimensie: "helemaal niet eng", "een beetje eng", "nogal eng", "heel eng"; frequentiedimensie: "nooit", "zelden", "soms", "meestal", "altijd"). In dit onderzoek varieerden de alpha coëfficiënten van de subschalen tijdens voor- en nameting tussen .75 en .97.

Het evaluatieformulier tenslotte werd gebruikt om na te gaan wat de deelnemers vonden van de opzet en het verloop van de training. Het formulier bestaat uit verschillende beoordelingsschalen die betrekking hebben op aspecten als de klapper, de huiswerkopdrachten, de videobanden, het aantal bijeenkomsten en de lengte van de bijeenkomsten.

Procedure

Sociale-vaardigheidstraining

De inhoud en vorm van de training is reeds uitvoerig besproken. Op deze plaats kan daar nog aan worden toegevoegd dat deelname vrijwillig was en dat de training op school werd gegeven aan zes groepjes van 12 à 13 jongeren. Er waren tien wekelijkse bijeenkomsten van anderhalf à twee uur. De zes trainingsgroepen werden begeleid door tweetallen psychologiestudenten die gezamenlijk waren getraind door één van de onderzoekers.

Metingen

Zoals opgemerkt, is het onderzoek ingebed in een bredere studie. In het kader van deze studie bestond de oorspronkelijke opzet van het hier besproken onderzoek uit vier meetmomenten die als volgt waren gepland. Een half jaar voor het begin van de training vond de eerste voormeting plaats (T1). Tijdens deze voormeting werd bij een groep jongeren een uitgebreide survey afgenomen. Deze survey bevatte de SIG en verder enkele vragenlijsten die betrekking hadden op aspecten als welbevinden, zelfwaardering en coping. Tijdens de meting die een half jaar na de training plaatsvond (T4), werd dezelfde survey bij de gehele groep jongeren nogmaals afgenomen. De SIG-gegevens van beide surveys konden dus worden gebruikt ter vergelijking van trainings- en controlegroep.

Naast deze twee metingen, werden bij de trainingsgroep ook nog twee andere metingen verricht die directer gericht waren op de evaluatie van de training. Deze beide metingen vonden onmiddellijk voor (T2) en na (T3) de training plaats en bestonden uit een afname van de kennistoets en (nogmaals) de SIG. Bovendien kreeg de trainingsgroep direct na de training het evaluatieformulier voorgelegd. De tijdens deze metingen verzamelde gegevens konden worden vergeleken met de gegevens van een zogenaamde wachtlijst-controlegroep. Deze wachtlijst-controlegroep had zich eveneens aangemeld voor de training, maar begon uit organisatorische

overwegingen later dan de andere trainingsgroepen (de deelnemers van de trainingsgroep kwamen van drie scholen terwijl de deelnemers van de wachtlijst-controlegroep afkomstig waren van de vierde school).

Met behulp van deze vier metingen zouden de resultaten op adequate wijze kunnen worden geanalyseerd. Echter, op het moment dat de wachtlijst-controlegroep met de training zou beginnen, bleek het aantal deelnemers door allerlei niet te voorziene omstandigheden zodanig afgenomen dat statistische vergelijking met de trainingsgroep niet meer mogelijk was. Dit had met name consequenties voor de analyse van de resultaten van de kennistoets: de gegevens van de trainingsgroep konden niet meer worden afgezet tegen die van een groep jongeren die geen training had gevolgd.

In de hierna te bespreken resultaten worden dus de kennistoets-scores van de trainingsgroep op de twee meetmomenten T2 en T3 met elkaar vergeleken. Met betrekking tot de SIG kunnen de resultaten van de trainingsgroep wel worden afgezet tegen een controlegroep omdat daarvoor de twee meetmomenten T1 en T4 kunnen worden gebruikt.

RESULTATEN

Voordat kon worden bestudeerd of er sprake was van trainingseffecten, diende eerst inhoud te worden gegeven aan de onafhankelijke variabele sociale-vaardigheidsniveau. Zoals opgemerkt, werd dit gedaan met behulp van de SIG. Door middel van een clusteranalyse op de voormetinggegevens van de SIG werd nagegaan in hoeverre de vier door Arrindell et al. (1984) genoemde subgroepen in dit onderzoek aanwezig waren (zie voor een uitgebreidere bespreking Bijstra, Bosma & Jackson, in druk). Het bleek dat drie van de vier subgroepen werden teruggevonden: een subassertieve subgroep (lage frequentie, hoge spanning; n = 75), een onverschillige subgroep (lage frequentie, lage spanning; n = 185) en een assertieve subgroep (hoge frequentie, lage spanning; n = 126). De trainingsgroep werd gevormd door 20 subassertieven (28%), 27 onverschilligen (38%) en 24 assertieven (34%) en de controlegroep door 55 subassertieven (18%), 158 onverschilligen (50%) en 102 assertieven (32%). Uit de percentages blijkt dat er relatief veel subassertieven deelnamen aan de training.

Kennistoets

In Tabel 2 staan voor de totale groep en de subgroepen de gemiddelden en standaarddeviaties weergegeven van de voor- en nametingversie van de kennistoets (in verband met de geringe aantallen zijn bij de variabele opleiding mavo en havo tot één subgroep en havo/vwo en vwo tot een tweede subgroep samengevoegd; hetzelfde geldt voor 13-jarigen enerzijds en 14- en 15-jarigen anderzijds; deze samenvoegingen zijn bij alle verdere analyses gehanteerd). Op de volgende wijze kan globaal enige waarde worden toegekend aan deze gemiddelden: aangezien de kennistoets uit 20 vragen bestaat en drie antwoordmogelijkheden heeft, wordt het aantal vragen dat op basis van kans goed wordt beantwoord, bepaald op 7. Een gemiddelde van 8,3 levert vervolgens het cijfer 1 op, een gemiddelde van 9,6 het cijfer 2 etc. De grens tussen voldoende en onvoldoende ligt daarmee op 14,2 (een gemiddelde van 13,5 levert het cijfer 5 op en een gemiddelde van 14,8 het cijfer 6). Uit Tabel 2 blijkt dat de subassertieve subgroep op basis van deze cijfertoekenning als enige groep een onvoldoende behaalde op de voormeting.

Met behulp van gepaarde t-toetsen werd bestudeerd of de trainingsgroep na de training beter presteerde op de kennistoets dan voor de training. Voor alle duidelijkheid zij nogmaals opgemerkt dat de beoogde controlegroep was weggevallen. Uit Tabel 2 blijkt dat zowel de totale groep als de subgroepen vooruitgingen van voor- naar nameting. De vooruitgang was het minst duidelijk voor de subassertieven: de jongeren die gezien hun sociale-vaardigheidsniveau waarschijnlijk het meest behoefte hebben aan een training, profiteerden het minst. Zij behaalden overigens wel een voldoende op de nameting.

Met behulp van manova's met herhaalde metingen werd vervolgens voor iedere onafhankelijke variabele bestudeerd of sommige subgroepen in de trainingsconditie meer vooruitgingen

Tabel 2. t-toetsanalyses: vergelijking van de gemiddelden van de totale trainingsgroep en de subgroepen op twee meetmomenten; manova's met herhaalde metingen: vergelijking van de subgroepen onderling op twee meetmomenten.

KENNISTOETS	t (df) =	p <	gem. T2	s.d. T2	gem. T3	s.d. T3
totale groep	4.54 (71)	.001	14.9	2.8	16.3	2.5
subassertieven	1.86 (19)	.10	13.6	2.8	14.7	2.8
onverschilligen	3.00 (26)	.01	14.6	2.6	16.1	2.2
assertieven	2.84 (24)	.01	16.2	2.8	17.8	1.7
subgroep x meetmoment: F (2,69) = .16, ns.						
meisjes	2.16 (40)	.05	14.7	2.6	15.6	2.6
jongens	4.50 (30)	.001	15.1	3.1	17.3	2.1
subgroep x meetmoment: F (1,70) = 4.66, p < .05.						
13-jarigen	3.68 (57)	.01	15.0	2.8	16.3	2.5
14+15-jarigen	3.37 (13)	.01	14.6	2.7	16.2	2.7
subgroep x meetmoment: F (1,70) = .11, ns.						
mavo+havo	3.74 (9)	.01	15.5	2.2	17.0	1.8
havo/vwo+vwo	3.93 (61)	.001	14.8	2.9	16.2	2.6
subgroep x meetmoment: F (1,70) = .01, ns.						

dan andere subgroepen. Uit Tabel 2 blijkt dat dat alleen het geval was voor de variabele geslacht. Jongens gingen significant meer vooruit dan meisjes: F (1,70) = 4.66, p < .05.

SIG

In de Tabellen 3 en 4 staan de gemiddelden en standaarddeviaties weergegeven van de trainings- en controlegroepen op de voor- en nameting van de SIG. Ter vermijding van een al te grote hoeveelheid cijfers, zijn de gemiddelden op de vier subschalen gesommeerd tot een frequentie-totaalscore en een spanning-totaalscore (de berekeningen zijn echter wel gebaseerd op de gemiddelden van de verschillende subschalen). Aangezien nog geen normgroepen beschikbaar zijn, kan op basis van deze gemiddelden een indruk worden verkregen van het sociale-vaardigheidsniveau van de subgroepen.

Tabel 3 laat zien dat er op de frequentiedimensie een verschil was tussen de gemiddelden van de subassertieven en onverschilligen enerzijds en dat van de assertieven anderzijds. Wanneer de gemiddelden op de voormeting worden "teruggerekend" naar de antwoordcategorieën, blijkt dat de subassertieven en onverschilligen respectievelijk 2,6 en 2,7 scoorden (tussen "zelden" en "soms") en de assertieven 3,4 (tussen "soms" en "meestal"). Uit Tabel 4 blijkt dat er op de spanningsdimensie een verschil was tussen het gemiddelde van de subassertieven enerzijds en de gemiddelden van de onverschilligen en de assertieven anderzijds. Teruggerekend naar de antwoordcategorieën, scoorden de subassertieven 2,2 (tussen "een beetje eng" en "nogal eng") en de onverschilligen en assertieven respectievelijk 1,5 en 1,3 (tussen "helemaal niet eng" en "een beetje eng").

Tabel 5 laat de resultaten zien van de vergelijkingen tussen de voor- en nametingscores van de trainingsgroepen en die van de controlegroepen. Deze analyses werden gedaan met behulp van manova's met herhaalde metingen. Het blijkt dat de training een effect had op de frequentie waarmee jongeren zeggen zich in sociale situaties te begeven: voor de totale groep en voor bijna alle subgroepen nam de frequentie van de trainingsjongeren van voor- naar nameting significant meer toe dan die van de controlejongeren. De frequentie van de assertieve en mavo+havo-

Tabel 3. Gemiddelden en standaarddeviaties op de frequentiedimensie van de SIG.

FREQUENTIE	gem. (s.d.) T1 training	gem. (s.d.) T1 controle	gem. (s.d.) T4 training	gem. (s.d.) T4 controle
totale groep	136.5 (22.1)	134.7 (22.7)	150.9 (25.6)	142.5 (23.2)
subassertieven	127.7 (13.3)	125.4 (14.9)	141.9 (22.8)	137.3 (18.6)
onverschilligen	123.3 (16.7)	121.8 (14.5)	143.3 (24.9)	134.1 (21.7)
assertieven	158.8 (15.3)	159.8 (14.4)	167.1 (21.2)	158.3 (19.6)
meisjes	135.1 (21.7)	133.7 (21.7)	151.9 (22.6)	144.0 (22.7)
jongens	138.4 (22.7)	136.3 (24.1)	149.7 (29.4)	140.3 (23.9)
13-jarigen	136.0 (23.0)	131.2 (21.7)	151.7 (26.2)	141.3 (23.1)
14+15-jarigen	138.3 (18.8)	138.3 (23.2)	148.2 (24.0)	143.8 (23.4)
mavo+havo	139.8 (20.1)	134.7 (24.3)	149.2 (23.9)	140.4 (25.5)
havo/vwo+vwo	136.0 (22.5)	134.7 (20.9)	151.2 (26.0)	144.6 (20.6)

Tabel 4. Gemiddelden en standaarddeviaties op de spanningsdimensie van de SIG.

SPANNING	gem. (s.d.) T1 training	gem. (s.d.) T1 controle	gem. (s.d.) T4 training	gem. (s.d.) T4 controle
totale groep	77.0 (20.1)	71.6 (18.3)	74.1 (18.4)	69.5 (17.1)
subassertieven	103.8 (6.6)	102.8 (10.9)	87.0 (17.6)	87.6 (18.0)
onverschilligen	70.9 (11.5)	67.7 (11.0)	74.4 (16.5)	68.2 (14.8)
assertieven	61.7 (11.7)	60.9 (11.1)	63.2 (13.8)	61.7 (12.4)
meisjes	79.6 (19.3)	74.1 (17.6)	75.0 (17.9)	70.6 (16.0)
jongens	73.7 (21.0)	67.9 (18.8)	73.1 (19.1)	67.9 (18.4)
13-jarigen	77.6 (20.1)	72.9 (18.6)	74.5 (19.1)	70.8 (17.9)
14+15-jarigen	75.0 (20.7)	70.2 (18.0)	72.7 (15.7)	68.1 (16.1)
mavo+havo	72.5 (16.0)	71.4 (19.2)	69.1 (17.4)	68.8 (17.4)
havo/vwo+vwo	77.8 (20.7)	71.8 (17.5)	75.0 (18.5)	70.2 (16.8)

trainingsjongeren nam niet significant meer toe dan die van de betreffende controlejongeren ondanks het feit dat de gemiddelden van de trainingsjongeren meer stegen dan die van de controlejongeren (zie tabel 3). Uit tabel 3 blijkt verder ook dat er op de nameting nog steeds een duidelijk verschil was tussen de gemiddelden van de subassertieven en de onverschilligen enerzijds en het gemiddelde van de assertieven anderzijds.

Vervolgens werd met behulp van manova's met herhaalde metingen voor iedere onafhankelijke variabele bestudeerd of de ene subgroep meer vooruitging van voor- naar nameting dan de andere subgroep. Uit tabel 5 wordt duidelijk dat dit bij geen enkele variabele het geval was.

Tabel 5 laat verder zien dat de training geen positief effect had op de spanning van jongeren: noch voor de totale groep noch voor de subgroepen nam de spanning van de trainingsjongeren significant meer af van voor- naar nameting dan die van de controlejongeren. Weliswaar was er

Tabel 5. Manova's met herhaalde metingen (conditie x meetmoment): vergelijking van trainingsjongeren met controlejongeren op twee meetmomenten; manova's met herhaalde metingen (conditie x subgroep x meetmoment): vergelijking van trainings- met controlesubgroepen onderling op twee meetmomenten (alleen voor de frequentiedimensie).

SIG	FREQUENTIE		SPANNING	
	F (df) =	p <	F (df) =	p <
conditie x meetmoment				
totale groep	5.89 (4,381)	.001	.61 (4,381)	ns
conditie x meetmoment				
subassertieven	3.06 (4,70)	.05	.66 (4,70)	ns
onverschilligen	4.85 (4,180)	.001	1.30 (4,180)	ns
assertieven	1.69 (4,121)	ns	.21 (4,121)	ns
conditie x subgroep x meetmoment	.58 (8,756)	ns		
conditie x meetmoment				
meisjes	3.29 (4,223)	.05	.96 (4,223)	ns
jongens	6.15 (4,153)	.001	1.03 (4,153)	ns
conditie x subgroep x meetmoment	.69 (4,379)	ns		
conditie x meetmoment				
13-jarigen	6.08 (4,212)	.001	2.04 (4,212)	ns
14+15-jarigen	2.72 (4,164)	.05	3.09 (4,164)	.05
conditie x subgroep x meetmoment	1.02 (4,379)	ns		
conditie x meetmoment				
mavo+havo	1.84 (4,163)	ns	.51 (4,163)	ns
havo/vwo+vwo	4.69 (4,213)	.001	1.06 (4,213)	ns
conditie x subgroep x meetmoment	.23 (4,379)	ns		

bij de 14+15-jarigen sprake van een significant verschil, maar dit werd veroorzaakt door het gegeven dat de spanning van de trainingsgroep op één van de subschalen *steeg*. Aangezien er geen trainingseffect werd geconstateerd, werd vergelijking van de subgroepen onderling achterwege gelaten.

Evaluatieformulier

In Tabel 6 staat vermeld op welke wijze jongeren het evaluatieformulier invulden dat hen na afloop van de training werd aangeboden. Jongeren bleken in het algemeen positief over de training: de training werd als duidelijk, leerzaam en zinvol ervaren en de organisatorische kanten werden als goed beoordeeld. Wel vonden nogal wat jongeren het huiswerk en de rollenspelen gemakkelijk (resp. 90% en 36%). Bovendien vond 55% van de deelnemers de duur van de training te lang en 23% het tempo te laag. Tot slot beoordeelde 39% van de jongeren de videobanden als eentonig.

DISCUSSIE

Voor de in dit artikel beschreven preventieve sociale-vaardigheidstraining voor jonge adolescenten waren drie doelen geformuleerd: een kennisdoel, een gedragsdoel en een belevingsdoel (Van der Molen, 1985). De vraag is nu in hoeverre deze drie doelen zijn verwezenlijkt.

Tabel 6. Meningen van jongeren over verschillende aspecten van de training uitgedrukt in percentages.

HUISWERK		
makkelijk 90	moeilijk 10	heel erg moeilijk 0
niet leerzaam 18	leerzaam 77	heel erg leerzaam 5
onduidelijk 3	duidelijk 79	heel erg duidelijk 18
VIDEO-BANDEN		
niet leerzaam 6	leerzaam 79	heel erg leerzaam 15
onduidelijk 6	duidelijk 60	heel erg duidelijk 34
eentonig 39	afwisselend 49	heel afwisselend 12
ROLLENSPELEN		
makkelijk 63	moeilijk 37	heel erg moeilijk 0
niet leerzaam 11	leerzaam 77	heel erg leerzaam 12
onduidelijk 4	duidelijk 84	heel erg duidelijk 12
niet zinvol 19	zinvol 59	heel erg zinvol 22
SAMENSTELLING VAN DE GROEP		
te groot 1	goed 98	te klein 1
DUUR VAN DE BIJEENKOMSTEN		
te lang 55	goed 45	te kort 0
FREQUENTIE VAN DE BIJEENKOMSTEN		
te veel 1	goed 96	te weinig 3
PERSOONLIJKE AANDACHT TIJDENS DE TRAINING		
te veel 3	goed 94	te weinig 3
TEMPO VAN DE TRAINING		
te hoog 1	goed 76	te laag 23
ORGANISATIE VAN DE TRAINING		
slecht 0	redelijk 27	goed 73

Ten aanzien van het kennisdoel bleek dat de trainingsgroep (totale groep èn alle subgroepen) na de training significant beter presteerde op de kennistoets dan voor de training. De conclusie luidt dat de sociale-vaardigheidstraining heeft bijgedragen aan de toename van kennis over sociale vaardigheden. Aangezien de scores op de kennistoets van de trainingsgroep echter niet konden worden afgezet tegen die van een controlegroep, is deze conclusie niet hard. Daarbij moet tevens worden opgemerkt dat tijdens voor- en nameting gebruik werd gemaakt van twee parallelversies: de vraag is hoe parallel deze versies waren. Uit de resultaten kwam verder naar voren dat jongens significant meer vooruitgingen van voor- naar nameting dan meisjes. Op grond van de beschikbare gegevens is hiervoor geen duidelijke verklaring te geven. Wellicht vormden de jongens een zeer gemotiveerde of "slimme" groep.

Ten aanzien van het gedragsdoel bleek dat de frequentiescore van de trainingsgroep op de SIG significant meer was toegenomen van voor- naar nameting dan die van de controlegroep. Dat gold voor de totale groep en voor alle subgroepen met uitzondering van de assertieven en de mavo+havo-jongeren. De conclusie luidt dat de sociale-vaardigheidstraining ertoe heeft geleid dat jongeren zich frequenter in sociale situaties begeven.

Dat de frequentiescore van de assertieven niet significant toenam, is niet verwonderlijk aangezien bij deze subgroep niet veel vooruitgang meer viel te verwachten. Overigens steeg het

gemiddelde van de assertieve trainingsgroep wel meer van voor- naar nameting dan dat van de assertieve controlegroep. Deze stijging was zelfs nog zodanig dat de drie subgroepen onderling (subassertief, onverschillig, assertief) niet significant verschilden in de mate waarin zij vooruitgingen van voor- naar nameting.

Dat de frequentiescore van de mavo+havo-jongeren niet significant toenam, is minder eenduidig te verklaren. Gezien de algemene, gestructureerde opzet van de training, lijkt schoolniveau geen rol te spelen. Mogelijk speelt de factor motivatie wel een rol: mavo- en havoscholieren bleken bijvoorbeeld veel minder gemotiveerd tot deelname aan de training dan havo/vwo- en vwo-scholieren. Ook het verschil in sociaal economisch milieu tussen de mavo+havo-subgroep en de havo/vwo+vwo-subgroep kan een rol hebben gespeeld: zoals opgemerkt, leverde de school die haar leerlingen uit een minder hoog sociaal-economisch gebied betrekt, alle mavo+havo-trainingsdeelnemers. Ook nu moet echter weer worden opgemerkt dat het gemiddelde van de mavo+havo-jongeren in de trainingsgroep nog zodanig steeg dat de twee subgroepen niet significant verschilden in de mate van vooruitgang van voor- naar nameting.

Bij het gebruik van de frequentiedimensie als gedragsmaat moet verder worden opgemerkt dat deze niet als een objectieve maat voor gedragsverandering kan worden gezien omdat het om door de deelnemers zelf gerapporteerd gedrag gaat. Deelnemers geven wel aan dat ze zich meer in sociale situaties begeven, maar het is niet duidelijk òf ze dat ook daadwerkelijk doen, en zo ja, hòe. Meer objectieve gedragsmaten zoals rollenspeltoetsen (Hersen & Bellack, 1977) of interviews met ouders/leraren over het gedrag van jongeren, bleken echter praktisch niet haalbaar in dit onderzoek. Bovendien kleven aan deze methodes eveneens bezwaren (Bellack, 1979).

Ten aanzien van het belevingsdoel bleek dat de spanningsscore van de trainingsgroep op de SIG niet meer was afgenomen van voor- naar nameting dan die van de controlegroep. De conclusie luidt dat de training niet effectief is geweest in het doen afnemen van spanning in sociale situaties. Met name voor de subassertieve groep was dit resultaat teleurstellend aangezien deze jongeren aangaven relatief veel spanning te ervaren (omdat de assertieve en onverschillige jongeren tijdens de voormeting al een laag spanningsniveau hadden, viel bij hen niet zoveel afname van spanning meer te verwachten). De oorzaak kan zijn dat het belevingsaspect in de training te weinig aandacht heeft gekregen. Het gegeven dat de training is gebaseerd op sociaal-leertheoretische principes (Bandura, 1977; Goldstein 1973, 1981; Ladd & Mize, 1983), lijkt zich dan ook te vertalen in de resultaten: jongeren begeven zich wel frequenter in sociale situaties, maar er is geen sprake van een spanningsafname.

Samenvattend kan dus worden geconcludeerd dat twee van de drie doelen (het kennisdoel – zij het met enige reserve – en het gedragsdoel) zijn verwezenlijkt. Met betrekking tot beide doelen kan nu nog de vraag worden gesteld of de bereikte vooruitgang wel de moeite waard was, met name voor jongeren die gezien hun sociale-vaardigheidsniveau het meest behoefte hadden aan een training. Ten aanzien van het niet-verwezenlijkte belevingsdoel dient de vraag zich aan op welke wijze de training kan worden verbeterd.

Of de bereikte vooruitgang de moeite waard was, is een moeilijk te beantwoorden vraag: er zijn namelijk geen duidelijke criteria voorhanden op grond waarvan kan worden bepaald of een bepaalde mate van vooruitgang al dan niet "de moeite waard" is. Feit is dat de training effect had, maar dat de kennis- en frequentiescores van de subassertieven en onverschilligen ook na de training nog steeds verschilden van die van de assertieven. De vraag is echter of het wel reëel is op korte termijn (een half jaar na de training) zodanig grote veranderingen te verwachten dat er geen enkel verschil meer bestaat tussen de drie groepen.

Met betrekking tot de vraag op welke wijze de training kan worden verbeterd, kan het volgende worden opgemerkt. Op basis van eerste indrukken uit gesprekjes met trainingsdeelnemers bestaat de indruk dat de geringe spanningsafname een gevolg zou kunnen zijn van de aanwezigheid van irrationele denkpatronen en negatieve denkgewoonten ("Dat doe je toch niet, zomaar op een meisje afstappen en een praatje beginnen"). De training zou dan ook kunnen worden verbeterd door nadrukkelijker in te spelen op dergelijke denkpatronen en gewoonten. Daarbij kan een benadering als de rationeel emotieve therapie (Diekstra & Dassen, 1976) als aanknopingspunt dienen.

Tot slot rest de vraag welke procedure voor werving van trainingsdeelnemers de meest geschikte is. In dit onderzoek werd gekozen voor de procedure "vrijwillige deelname voor iedereen". Deze procedure had enkele duidelijke voordelen. De training werd gepresenteerd als een "Cursus Communicatie" voor iedere belangstellende, kende als zodanig een lage drempel en bracht weinig kans op stigmatisering van de deelnemers met zich mee. Aan de procedure kleefden echter ook enkele problemen. In de eerste plaats was het aantal deelnemers aan de training een relatief klein gedeelte van het potentiële aantal en in de tweede plaats trad zelfselectie op door ondervertegenwoordiging van havo+mavo-scholieren en 14+15-jarigen (interessant is overigens dat het aantal subassertieven oververtegenwoordigd was in de trainingsgroep). Het is verder mogelijk dat deze procedure de resultaten in gunstige zin heeft vertekend aangezien de trainingsgroep waarschijnlijk uit relatief gemotiveerde jongeren bestond.

De meeste van deze problemen kunnen worden voorkómen wanneer de training wordt gegeven tijdens een aantal lesuren. Aangezien de training daarmee een zeker verplichtend karakter krijgt, wordt een grotere groep jongeren bereikt en worden zelfselectieproblemen vermeden. Bovendien is er geen sprake van een drempel en bestaat er weinig kans op stigmatisering. Een groot praktisch probleem is echter dat van de scholen een grote tijdsinvestering wordt gevraagd. Een derde mogelijkheid tenslotte is om alleen sociaal minder vaardige jongeren een training aan te bieden (bijvoorbeeld in het kader van de leerlingbegeleidende activiteiten). Bij deze vorm is de drempel echter weer hoog en is de kans op stigmatisering van de trainingsdeelnemers groot.

Kortom, de drie procedures kennen alle voor- en nadelen; redenen als doelstelling en praktische mogelijkheden van een in de training geïnteresseerde school moeten uiteindelijk de doorslag geven bij de keuze voor een bepaalde procedure. Tot slot kan nog worden opgemerkt dat in de nabije toekomst verder zal worden gewerkt aan de in dit onderzoek besproken training. Er ligt een drietal plannen in het verschiet: er zal een aangepaste versie worden getest waarin explicieter aandacht wordt besteed aan de aanwezigheid van irrationele denkpatronen en negatieve denkgewoonten; bij de werving van deelnemers voor de training zal ook gebruik worden gemaakt van de twee andere procedures; de training zal worden uitgeprobeerd bij andere schooltypes (met name lbo).

LITERATUUR

Arrindell, W.A., Groot P.M. de, & Walburg, J.A. (1984). *De Schaal voor Interpersoonlijk Gedrag: SIG. Handleiding.* Lisse: Swets & Zeitlinger.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs: Prentice Hall.

Bartels, A.J. (1986). *Sociale-vaardigheidstraining voor probleemjongeren.* Lisse: Swets & Zeitlinger.

Beekers, M. (1982). *Interpersoonlijke vaardigheidstherapieën voor kansarmen.* Lisse: Swets & Zeitlinger.

Bellack, A.S. (1979). Behavioral assessment of social skills. In: Bellack, A.S. & Hersen, M. (red.). *Research and practice in social skills training.* New York: Plenum Press.

Bilsen, H.P.J.G. van, Jonkers, J., Schuurman, C.M.A. & Swager, H.D. (1991). Geïntegreerde behandeling van sociale-vaardigheidstekorten bij kinderen uit het speciaal onderwijs. *Kind en Adolescent, 12,* 78-86.

Bosma, H.A., Bijstra, J.O., Jackson, S., Kooi, H.P. van der, & Molen, H.T. van der (1991). Ingrijpen in ontwikkeling: onderzoek onder adolescenten. *Jeugdonderzoek, 4,* 10-13.

Bijstra, J.O., Kooi, H.P. van der, Bosma, H.A., Jackson, S. & Molen, H.T. van der (1993). Zelfwaardering en sociale vaardigheid bij adolescenten. Relaties tussen begrippen en effecten van een preventieve sociale-vaardigheidstraining. *Kind en Adolescent, 14,* 161-172.

Bijstra, J.O., Bosma, H.A. & Jackson, S. The relationship between social skills and psycho-social functioning in early adolescence. *Personality and Individual Differences,* in druk.

Cowen, E.L., Pederson, A., Babigian, H., Izzo, L.D. & Trost, M.A. (1973). Long term follow-up of early detected vulnerable children. *Journal of Consulting and Clinical Psychology, 41,* 438-446.

Diekstra, R.F.W. & Dassen, W.F.M. (1976). *Inleiding tot de rationele therapie.* Lisse: Swets & Zeitlinger.

Goldstein, A.P. (1981). *Psychological skill training: the structured learning technique.* New York: Pergamon Press.

Goldstein, A.P. (1973). *Strucured learning therapy: toward a psychotherapy for the poor.* New York: Academic Press.

Hersen, M. & Bellack, A.S. (1977). Assessment of social skills. In: Ciminero, A.R., Calhoun, K.S. & Adams, H.E. (red.). *Handbook of behavioural assessment.* New York: Wiley.
Ivey, A.E. (1971). *Microcounseling: innovations in interview training.* Springfield: Thomas.
Kooi, H.P. van der, (1992). *Draaiboek voor de cursus communicatie voor adolescenten.* Interne publicatie: Vakgroep Psychologie, Rijksuniversiteit Groningen.
Ladd, G.W. & Mize, J. (1983). A cognitive-social learning model of social skill training. *Psychological Review, 90,* 127-157.
Lang, G. & Molen, H.T. van der, (1992). *Methodiek van gesprekstraining.* Baarn: Nelissen.
Meichenbaum, D. (1985). *Stress inoculation training.* New York, Pergamon Press.
Molen, H.T. van der, (1982). *Kommunikatieve vaardigheden (beeld & geluid); instructieprogramma.* Groningen: Rijksuniversiteit Groningen.
Molen, H.T. van der, (1984). *Aan verlegenheid valt iets te doen: een cursus in plaats van een therapie.* Deventer: Van Loghum Slaterus.
Molen, H.T. van der, (1985). *Hulp als onderwijs; effecten van cursussen voor verlegen mensen.* Groningen: Wolters-Noordhoff.
Petersen, A.C. (1988). Adolescent development. *Annual Review Psychology, 39,* 583-607.
Ringrose, H.J. & Nijenhuis, E.J. (1986). *Bang zijn voor andere kinderen: omgang en therapie met sociaal onhandige kinderen.* Groningen: Wolters-Noordhoff.
Rutter, M. (1980). *Changing youth in a changing society.* Cambridge: Harvard University Press.
Rutter, M. (1989). Pathways from childhood to adult life. *Journal of Child Psychology and Psychiatry, 30,* 23-51.
Zee, S.A.M. van der, Molen, H.T. van der & Beek, D.T. van der, (1989). *Sociale vaardigheden voor zwakbegaafde jongeren: praktijkboek Goldsteintraining.* Deventer: Van Loghum Slaterus.

Manuscript ontvangen 9-8-1993
Definitieve versie ontvangen 25-11-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, pp. 369-379

The influence of parents' expectations on school careers

Anna A. van der Hoeven-van Doornum[1], Marinus J.M. Voeten[2] en Paul Jungbluth[1]
[1] *Institute for Applied Social Sciences, Nijmegen**
[2] *Department of Educational Sciences, University of Nijmegen, The Netherlands*

ABSTRACT

In a previous study it was found that teacher expectations in the last two years of primary education influence school careers of pupils. Children may enter higher or lower levels of secondary education depending on the aspiration levels their teachers set for them. Aspiration levels work out as a form of teacher effectiveness. The aim of the present study was to investigate the influence of parental expectations on the school careers of children, using the data from the previous longitudinal study. Analyses were performed on data from a cohort of 701 pupils (10-12 year old) who were followed during their last two years of primary education. As indicator for parental expectations the level of secondary education wanted for the child was used, assessed by a questionnaire during the first wave of measurements.

It was shown that the expectations of parents influence the school careers of children, taking into account prior learning achievement. High expectations of parents have a positive effect on learning achievement and on the principal's advice for level of secondary education. In this sense parents' expectations work out as a form of 'parental effectiveness'. SES-related differences between the advices for level of secondary education could be explained by parental and teacher expectations.

INTRODUCTION

Parents' expectations for their children's academic achievement and occupational career form an important source of influence for academic performance and choices made in a child's school career. According to Lee, Bryk, and Smith (1993) the major influences of parents on their children's learning involve "enforcing normative standards concerning education and exhibiting specific behaviors directly connected to learning" (o.c., p. 190). More specifically, they state, that "parental expectations for their children's achievement and the importance that parents place on education are positively and strongly related to academic performance" (o.c., p. 190). These authors review research on effective secondary schools. But the influence of parental expectations, which may be seen as an aspect of parental support, concerns elementary as well as secondary education. Parental expectations, may be a cause as well as a consequence of children's achievement. Seginer (1983) reviews the research from these two perspectives. She developed a model for the course of parents' educational expectations. Hossler and Stage (1992) present a short review of research on postsecondary aspirations. They conclude that parental expectations play an important role in the predisposition stage of student college choice.

The relationship of parental educational expectations with student achievement and with attained level of schooling may be considered a well-established fact, for students of various ages, young children as well as adolescents. Despite variations in definitions of parental expectations and academic achievement, empirical studies generally support this statement (Seginer, 1983). Operational definitions for parental expectations include years of schooling, type of secondary school wanted for child (Peaker, 1967), attained level of schooling aspired (Hossler & Stage, 1992), occupational expectations, and predicted academic performance. Also the chil-

* Adres: Institute for Applied Social Sciences, P.O. Box 9048, 6500 KJ Nijmegen.

dren's perception of their parents' expectations and encouragement has been used in research (e.g. Carpenter & Fleishman, 1987). The correlations found in the studies reviewed by Seginer range from .00 to .79. Carpenter and Fleishman (1987) found a correlation of .45 between perceived parental encouragement to continue education and the student's actual decision after graduating from high school. The Plowden report on primary schools in England even concluded that parental attitudes (including aspirations for the child) "can account for more of the variation in children's school achievement than either the variation in home circumstances or the variation in schools" (Peaker, 1967, p. 181). In a study concerning Mexican American high school seniors it was found that personal aspirations of students were a much stronger correlate of achievement than the Spanish-language background (Buriel & Cardoza, 1988). Mothers' aspirations had no significant effect apart from the students' own aspirations.

This selection of empirical results might suffice to demonstrate that parents' educational expectations form an important variable for research on school achievement and educational attainment. What is less clear, is how this variable relates to other variables in a model of status attainment. The purpose of this research is to investigate the role of parental expectations in connection with teacher expectations in the choice process for type of secondary education after completing primary education (in the Netherlands the children are by then approximately at the age of 11 or 12).

Parental expectations are one way in which family background influences children's academic achievement. Expectations and, more generally, parental involvement in children's school learning, seem to correlate with socio-economic status even when academic performance is taken into account (Baker & Stevenson, 1986; Carpenter & Fleishman, 1987; Keith et al., 1986). From this one may argue that parents' expectations might help to elucidate the well-known fact that parents' socio-economic status influences the child's academic achievement and educational attainment. These effects of SES can partly be explained by factors related to cultural differences in the social background of pupils and by factors related to school characteristics (Meijnen, 1984; Tesser, 1986; De Graaf, 1987; Leseman, 1989; Bosker, 1990; Van der Hoeven-van Doornum, 1990; Van der Hoeven-van Doornum, Voeten & Jungbluth, 1990a; Tesser, 1991). Recently it was shown that aspiration levels teachers set for their pupils also contribute to the explanation of these SES effects (Van der Hoeven-van Doornum, Voeten & Jungbluth, 1989, 1990b).

Maintaining high achievement standards and high expectations for student achievement belong to the characteristics of effective teachers and schools (Lee et al., 1993). Likewise parental involvement in school learning and parents' educational expectations have their impact on children's school career. This point has been well recognized or – one might say – rediscovered in school improvement research. The active role the family has to play in the education system in recent years receives more attention in research on school effectiveness and school improvement. Families too have a curriculum (Cremin, 1976) and – in the words of Bloom (1981) – a teaching style as well. According to Cooper (1991), parents are critical in supporting education, in preparing to live in modern society and in helping to advocate improvements in education. The school reform movement considers parents as educative partners which cannot be passive observers of their children in the school scene (Melnick, 1991). Following the new insights, policy-makers in western countries have made a shift from compensatory programs in the schools for disadvantaged children to parent partnerships programs before and during the school years (Wallace & Walberg, 1991).

Some parents are more effective than others in influencing their children's school learning. It may be that high-SES parents are more effective than low-SES parents. In raising their children parents express their expectations in many implicit and explicit ways, partly depending upon their social and cultural background. In daily activities – like playing, reading, talking with their children – parents may make high demands, thereby stimulating the cognitive development of children (Rupp, 1969; Tesser, 1991). When parents' behaviour is consistent, there will be a cumulative effect of parents' expectations on the child's behaviour, performance and attitudes. In this study, however, we concentrate on one particular crucial moment in a child's school

career, the transition from elementary to secondary education. Choice for type of secondary education is very important for the child's later life opportunities.

Building upon our earlier work on inequalities in educational opportunities and the influence of teacher expectations (Van der Hoeven-van Doornum, 1990; Van der Hoeven-van Doornum et al., 1990b), we included parental expectations in our model along with teacher expectations. We were specially interested in the extent to which the type of secondary school parents want for their children, is related to learning achievement, to aspiration levels defined by teachers, and to advised type of secondary education.

THEORETICAL FRAMEWORK

In agreement with Finn's definition (1972), parents' educational expectations involve a conscious or unconscious evaluation of the child's academic competence as well as a form of anticipation: the parents anticipate that the child will perform in a way consistent with their evaluation. Finn's analysis of the concept implies three elements: the parents' aspirations – e.g. what they ideally hope for – an estimation of reality based on the parents' knowledge, and standards of achievement maintained for the child. The next elaboration of these three elements draws heavily upon Seginer's model (1983) of the course of parents' expectations and upon our earlier work about teacher expectations.

Parents' own aspirations, possibly based on dissatisfaction resulting from own life experiences or based on the value parents place on a good education, will greatly determine the expectations they have for their children. For instance, the desire to "get ahead" might stimulate the cognitive development of some lower-class children. Another way in which educational aspirations are related to SES concerns the knowledge or lack of knowledge about the range of educational possibilities and the contingencies of the educational system. Deficits in low-SES parents' knowledge may set bounds to their aspirations.

Estimation of reality, i.e. the prediction of a child's performance, will mainly be based on the feedback that parents receive from the school, for instance in the form of school grades. Expectations may be "corrected' during a child's school career by this feedback. After some school years aspirations might become better tuned to school realities. This implies that parents' expectations to a certain extent have become influenced by actual school achievement. The effect of school feedback on educational expectations might differ for lower-, middle-, and upper-class children; perhaps lower-class parents are less familiar with criteria defined by schools than middle- and upper-class parents.

The standards of achievement are the (implicit) criteria by which parents evaluate their children's school achievement. These standards will be based upon the parents' appraisal of the child's ability and behaviour, on the one hand, and their notions concerning children in general and school demands, on the other (Seginer, 1983, p. 11). These standards defined by parents partly determine the kind of feedback children receive on their behaviour. Parents may differ in the measure by which they evaluate their child's learning achievement as good or bad. But, nevertheless, they may also have absolute goals expressed as an explicit preference for a certain type of secondary education for their children. (see Van der Hoeven-van Doornum et al., 1990b for similar concepts applied to teachers). In this study we focus on the effects of this latter type of parental aspiration levels.

From the foregoing analysis it will be clear that parental educational expectations will depend on SES as well as on the child's record of school achievement. The three antecedents of parents' expectations mentioned above, namely parents' own aspirations, school feedback and parental knowledge may explain the dependence of expectations on SES and previous school achievement. The next question is how expectations can affect achievement. Expectations of parents may work out in the same way as aspiration levels teachers define for individual students. We expect a positive effect on educational achievement as well as on the level of secondary education considered for the child.

Seginer (1983) included in her review three factors that may mediate the effects of parents' educational expectations on children's school achievement. Expectations may affect performance by desirable goals parents define for the children (cf. the absolute and relative goals of teachers as described in Van der Hoeven-van Doornum et al., 1990b), and by achievement supporting behaviours associated with high expectations, like showing interest in the child's learning. A second mechanism considered by Seginer is differential reinforcement: rewarding behaviours that conform to expectations, and punishing behaviours that don't. A third explanation is that expectations of parents affect achievement through building up the children's own educational aspirations. The child's self-concept and the activation of motivational processes will be central constructs in this line of thought.

These three mediating mechanisms apply to teacher expectations as well, and they certainly do not contradict each other; all three might be true at the same time. There is not enough empirical evidence, yet, to support these explanations. But they at least make the notion of expectations affecting achievement plausible.

Teacher expectations

Aspiration levels teachers define for their pupils are mainly based on perceived learning performance and social background of the pupils. Aspiration levels, as considered here, are general goals the teacher expects to attain with a pupil. As a consequence of their expectations, teachers will, to a certain extent, achieve these goals in the end. It is to be noted that we are not referring here to the specific contents of the language or the mathematics curriculum a pupil has to master. A distinction between two kinds of aspiration levels must be made: there are absolute and relative goals (Van der Hoeven-van Doornum, 1990; Van der Hoeven-van Doornum et al., 1990b). Absolute goals refer to what the teacher expects to be ultimately attainable for a pupil, such as the level of secondary education. Relative goals are related to daily instructional and educational processes, and are indicated as attainable cognitive goals. Examples of cognitive goals are 'high demands for this pupil' and 'strong emphasis on reading and mathematics rather than practical work'. It is not necessarily the case that aiming at high cognitive goals is coupled with advising a high level of secondary education.

Van der Hoeven-van Doornum (1990) showed that teacher expectations are related to the social background of the pupils as well as to achievement and advised type of secondary education (see also Van der Hoeven-van Doornum et al., 1990b). Teachers' perceptions of the social background and the learning ability of their pupils as well as anticipated achievement levels were shown to be associated with the SES of the pupils. These teacher perceptions and expectations appeared to have (indirect and direct) effects on achievement and attainment, the latter measured by the type of secondary education recommended by the school principal. In these analyses pupils' aptitude and prior learning achievement were controlled for. It was concluded that SES effects on achievement and attainment can partly be explained by teacher expectations.

Parents' expectations are, as shown above, also a likely mediator of SES effects on achievement and attainment. What will happen when parents' and teacher expectations are considered in combination? We expect a positive association between parents' and teacher expectations because they are both based on educational achievement as well as on (perceived) social background. When prior achievement and SES are accounted for, there still might be a positive association between parents' and teacher expectations, because of common antecedents not included in our model or because parents succeed in communicating their expectations to teachers. When high expectations are accompanied with appropriate learning support, positive effects on achievement may be expected. The question is then whether parents' expectations affect school achievement and attainment *in addition to* effects of teacher expectations. But, one could also argue that effects of teacher expectations on achievement and attainment decline when parents' expectations are taken into account. According to Hossler and Stage (1992), in American education teachers appear to have very little influence on college choice of high school students, while they could conclude from their study that parents' expectations exerted

strong positive effects. At the elementary school level we would expect stronger influences of teachers on student aspirations.

Hypotheses

We hypothesized that educational expectations of parents influence both the learning performances of their children and the expectations of the teacher. Also, as argued above, we hypothesised that parents' expectations mediate the effects of SES on learning achievement, and on advised type of secondary education. As a consequence of what was stated above, that parental expectations may be a cause as well as the consequence of children's achievement, a model was developed which includes an effect of prior achievement on parental expectations, and effects of the latter on (later) learning achievement, teachers' expectations and advice for secondary education. From such a model, we can assess to what extent parental expectations affect the process of choosing a type of secondary education. We then can also assess to what extent the effects of SES on school careers can be explained by parental expectations.

METHOD

The study was carried out with 10-12 year old pupils and their parents. A cohort of pupils from a sample of 53 primary schools in the eastern and southern part of The Netherlands was longitudinally followed from 7th through 8th grade.[1] Included in the sample were all pupils from these schools who were by september 1986 in grade 7. For a full description of the sample we refer to Van der Hoeven-van Doornum (1990).

Information on the socio-economic status and parental expectations was obtained from the parents by a questionnaire (february 1987), while the pupils were still in the 7th grade. A SES-index was constructed from the parents' education and occupation (Van der Hoeven-van Doornum, o.c., p. 94-97). Parental expectations were measured by the type of secondary education parents wanted for their child. In the sequel we will call this variable the parents' choice, scored on a scale from 10 (junior vocational education) to 40 (pre-university education). The question was not asked for each of the parents separately, so we cannot differentiate between mother's choice and father's choice; therefore we label it as parents' choice.

An extensive questionnaire on aspiration levels was filled out by the teacher for each individual pupil. This was done while the pupil was in grade 7 and again while the pupil was in grade 8. We here use only the latter data (december 1987). The variable 'relative goals' was measured by a Likert-type scale consisting of 9 items; the scale scores ranged from 1 (low aspiration level) to 5 (high aspiration level). The internal consistency (Cronbach's alpha) was .86. The items and the scale construction were reported in Van der Hoeven-van Doornum (o.c.). Anticipated type of secondary education (the teachers' absolute goals) was scored on the same scale as the parents' choice. June 1988 the advised type of secondary education for each pupil was obtained from the teachers, also scored on the scale from 10 to 40.

An intelligence test (Van Boxtel, Snijders & Welten, 1982) was administered in september 1987. Learning achievement was measured september 1986 and again september 1987 by standardized tests for reading comprehension and mathematics (cf. Van der Hoeven-van Doornum, 1990 for a description of these tests). Finally, february 1988 scores were obtained on the *Cito Eindtoets*, an achievement test many schools in the Netherlands use to obtain information for advising the appropriate type of secondary school for a child.

Model specification

Point of departure was a school career model tested in a previous study (Van der Hoeven-van Doornum, 1990). In this study the effects of SES, aptitude (IQ), learning achievement and teacher expectations on advised type of secondary education were analyzed, using LISREL. SES and IQ were considered correlated exogenous variables. Achievement in mathematics and reading were included for two consecutive years (7th grade and 8th grade). From this initial model a new model was developed to study the effects of parental expectations.

Achievement measured upon entering the 7th grade was, in addition to the exogenous variables, used in the model as a determinant of parents' expectations. In turn, parents' expectations were seen as antecedent to achievement in the 8th grade, measured at the beginning as well as the end of the school year. Parents' expectations were also supposed to affect both types of teacher expectations, relative goals as well as absolute goals. In addition, a direct effect of parents' expectations on advised type of secondary education was entered in the model.

The effect of introducing parents' expectations in our school career model was evaluated by comparing models with and without parents' expectations.

RESULTS

Descriptive statistics

Table 1 presents the descriptive statistics of the variables used in the model. Data were available for 701 pupils. Pupils with a missing value on one or more tests were excluded from the analysis. On the question about preferred type of secondary education, twenty-two percent of the parents (N=157) indicated to have no preference yet for a particular type of secondary education for their child. Therefore Table 1 presents the descriptive statistics for two groups: the pupils whose parents indicated a certain type of secondary education for their child (N=544) and the group of pupils whose parents could not indicate a preference at the time asked. Compared with the latter group (N=157), the scores of the first group of pupils (N=544) were significantly higher ($p < .001$) on all variables. The parents who said to have no preference apparently were lower-SES parents with lower-achieving children.

So two indicators for parents' expectations can be distinguished: whether or not parents could indicate a preference for their child's level of secondary education, and the parents' choice for level of secondary education. This same distinction was made in the Plowden report (Peaker, 1967). We considered the first indicator too coarse; therefore we decided to restrict the analyses to the subsample of pupils whose parents did indicate a preference (N=544).

Table 1. Descriptive statistics for the variables in the school career model, separately for two sub-samples of pupils differing by existence of explicit parental preferences for the child's level of secondary education.

		Explicit preference N=544		Not yet N=157		
	variable	mean	sd	mean	sd	range
grade 7	SES	51.08	10.13	46.88	9.02	39-77
begin grade 7	mathematics score	23.83	7.78	19.49	7.08	1-40
begin grade 7	reading score	25.00	5.23	21.78	5.81	1-36
grade 7	parents' choice	26.40	8.97	-	-	10-40
begin grade 8	IQ score	13.20	2.32	11.71	2.73	1-20
begin grade 8	mathematics score	12.09	5.35	8.96	4.65	1-24
begin grade 8	reading score	12.22	3.33	10.39	3.38	1-24
begin grade 8	relative goals	3.70	.68	3.27	.72	1-5
begin grade 8	absolute goals	24.25	9.59	17.87	8.40	10-40
end grade 8	Cito score (math. & reading)	536.11	9.95	528.60	11.55	500-550
end grade 8	advice sec. ed.	24.57	9.97	17.36	8.98	10-40

Table 2. Correlations between observed variables (N=544).

	1	2	3	4	5	6	7	8	9	10	11
1. SES	1.00										
2. grade 7 math. score	.20	1.00									
3. grade 7 reading score	.32	.50	1.00								
4. grade 7 parents' choice	.48	.38	50	1.00							
5. IQ score	.36	.52	.68	50	1.00						
6. grade 8 math. score	.36	.70	.59	.50	.71	1.00					
7. grade 8 reading score	.27	.48	.66	.49	.59	.58	1.00				
8. grade 8 relative goals	.22	.47	.55	.41	.52	.52	.51	1.00			
9. grade 8 absolute goals	.45	.56	.64	.67	.70	.70	.63	.60	1.00		
10. grade 8 Cito score	.39	.59	.75	59	.77	.78	.63	.61	.77	1.00	
11. grade 8 advice	.48	.52	.65	.71	.72	.70	.62	.56	.91	.79	1.00

Table 2 presents the correlations between the variables for the subsample used in the analyses. As might be expected, all correlations are positive. There are moderate to high correlations of parents' choice with all the other variables.

Test of the school career model

To answer the research question we applied the Lisrel procedure to the covariance matrix. Two models were fitted: an initial model without the parents' choice, and a model in which the parents' choice was included.

As expected, the initial model fitted the data well: $chi^2 = 13.57$, df = 12, p = .33, AGF = .98. The variances of the Cito score and of the advice at the end of grade 8 were quite well explained by the model (78% and 86% respectively). Table 3 shows the standardized direct effects of the variables. About 71% of the total effects of SES on advice could be explained through learning achievements and teachers' aspiration levels.

The new model with the inclusion of parental expectations also fitted the data. But a few modifications were made to remove nonsignificant effects. It appeared that only one of the hypothesized relations of parents' choice with the other variables was not significant: there was no direct relation with the relative cognitive goals reported by the teachers. The final model is

Table 3. Initial model for the influence of teachers' expectations on school careers. Direct effects.

	ses	iq	math.7 score	read.7 score	math.8 score	read.8 score	rel. goals	abs. goals	Cito score
grade 7									
math.	–	.52							
reading	.08	.66							
grade 8									
math.	.11	.43	.45	–					
reading	–	.23	.15	.42					
rel. goals	–	.12	.14	.23	.12	.16			
abs. goals	.18	.23	.09	.13	.23	.18			
Cito score	–	.19	–	.26	.31	–	.09	.20	
advice	.07	.08	–	–	–	–	–	.70	.17

$Chi^2 = 13.57$, df = 12, p = .33, AGF = .98, N = 544

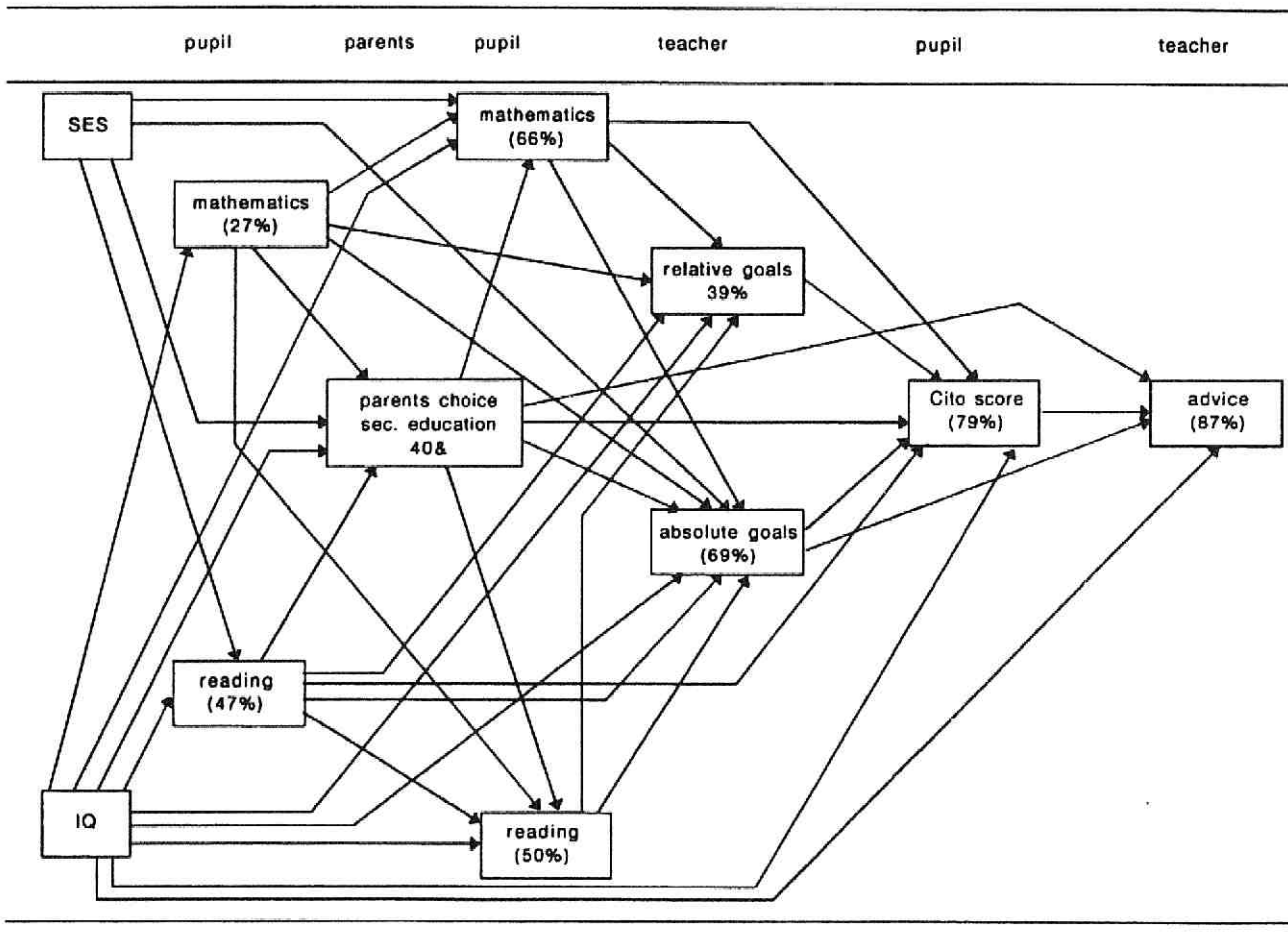


Fig. 1. Model for parents' choice for secondary education, aspiration levels, learning achievement and advice.

depicted in Figure 1. The fit of this model was very satisfying: chi^2= 17.94, df=14, p=.21, AGF=.97. The parents' choice could be explained by SES as well as IQ and prior learning achievement, in agreement with our hypotheses. These variables explained 40% of the variance in the parents' choice. There was only a marginal increase though of the explained variance of Cito score and advice.

The effects of parental expectations

The estimated direct effects are presented in Table 4. These may be compared with the direct effects of the initial model presented in Table 3. As hypothesized parents' expectations had a significant effect on learning achievement, though the effects were not strong. But, nevertheless, for reading achievement the effect was of about the same magnitude as the effect of IQ. Note that these are effects after controlling for, among others, previous learning achievement.

By adding the choice of parents to the model the direct effect of SES on advice for level of secondary education at the end of primary education (grade 8) disappeared completely. Also, a substantial decrease (44%) occurred with respect to the direct effect of SES on absolute goals. In agreement with our hypotheses, the combination of parents' expectations and teacher expectations could explain SES-related differences in advice for secondary education, while controlling for learning achievement. In addition, the parental expectations mediated the SES-effect on the teacher's absolute goals expressed at the beginning of grade 8. In contrast the relations of IQ and learning achievement with other variables (including advice) were only marginally affected by taking the parental expectations into account.

Table 4. Final model for the influence of parents' expectations on school careers. Direct effects.

	ses	iq	math. score	read. score	parents choice	math. score	read. score	rel. goals	abs. goals	Cito score
grade 7										
math.	–	.52								
reading	.08	.66								
par. choice	.33	.17	.12	.22						
grade 8										
math.	.08	.41	.44	–	.09					
reading	–	.18	.13	.39	.15					
rel. goals	–	.12	.14	.23	–	.12	.16			
abs. goals	.10	.21	.08	.09	.27	.20	.14			
Cito score	–	.19	–	.25	.08	.31	–	.09	.16	
advice	–	.09	–	–	.15	–	–	–	.64	.15

$Chi^2 = 17.94$, df = 14, p = .21, AGF = .97, N = 544

As to the relation between parental expectations and teacher expectations the following may be observed. As noted before parents' choice was not related to the relative cognitive goals defined by the teacher, but there was an effect on the teacher's absolute goals. Now 69% of the variance in teachers' absolute goals was explained by the model. Adding parents' choice lead to an increase of 5%. The effects of teacher expectations on the Cito score and especially on advice, were only marginally affected by including parental expectations in the model.

The effect of parents' choice on the Cito score was small, but statistically significant. Parental expectations and teacher expectations both had an independent effect on the Cito score. Finally, the direct effect of parents' choice on the ultimate advice for level of secondary education was of the same magnitude as the effect of the Cito score, and larger than the effect of IQ.

CONCLUSIONS AND DISCUSSION

It was our aim to investigate the role of parents' educational expectations regarding choices made in a child's school career. In particular we were interested in the influence of parents' expectations, in connection with teacher expectations, in the process of choosing the type of secondary education for a child. Parents' as well as teacher expectations were assumed to be factors that mediate the influence of SES on learning achievement and on the principal's advice for type of secondary education. The results generally confirmed our hypotheses.

We have shown that, taking into account differences between children with respect to aptitude and previous learning achievement, parents' expectations influence both the learning performances of the child and the expectations of the teacher. In addition, parents' expectations mediate the effects of socio-economic status on teachers' aspiration levels, and on test scores and advice at the end of primary education.

The analysis was restricted to those parents who reported an explicit preference. A substantial percentage of the parents indicated that they had not made up their mind, yet. This was especially the case for lower-SES parents with low-achieving children. A possible explanation is, that these parents postpone their choice, because they experience a discrepancy between idealistic and realistic expectations. Another interpretation might be that lower-SES parents are more inclined to leave school career decisions to the teacher, "who knows best". We fitted our school career model also to the complete sample with parents' expectations represented as a dichotomous variable indicating whether or not the parents reported an explicit preference for

their child's secondary education. This coarse indicator appeared to be not strong enough to make a difference. Probably the parents who had no explicit preference at the time we asked them, must be situated at the lower end of a scale for parental expectations.

We did not find a relation between the parents' choice and the relative cognitive goals the teacher sets for the child. In our previous study is was shown that these relative goals are related to the teachers' appraisal of the children's ability as well as to their appraisal of the children's home environment. It may be the case that the relative cognitive goals reflect the teacher's psychological insights and their conceptions of the goals of primary education. The effect of this variable on end-of-the-year achievement is only weak.

One could say that aspiration levels set by parents work out as a form of 'parental effectiveness'. The influence of parents' expectations is not restricted to learning achievements of their children, they have also an impact on the aspiration levels teachers define for the children in the form of level of secondary education, deemed appropriate for the child. It appears that school careers are determined by a mechanism of synergism: positive characteristics of school functioning going hand in hand with positive characteristics of the home environment. High expectations of parents partly may enable high aspirations of the teacher for the child. In addition, teacher aspiration levels are also influenced by other aspects of the home environment, which in the model is indicated by the direct effect of socio-economic status of parents on, what we called, the absolute goals of the teacher. A note of caution must be added though. Including the parental expectations in the model brought about only a small increase in the explained variances of Cito score and advice for secondary education. Further to asses the effects of parental expectations on school careers, the research should be extended to the children's career in secondary education. We might hypothesize that mechanisms of synergism between parents, teachers and students also exist in secondary education.

In the school reform movement parents and teachers are considered educative partners. As a consequence of this new insight, researchers and policy-makers in western countries have made a shift from compensatory programs for disadvantaged children to programs for families at home and for clubs or groups after school hours with social and cultural activities. The main target of those programs is to alter the social and cultural background of lower-SES children. It seems likely that parents and children from middle and upper-class families will benefit more from measures taken in educational or school reform movements and intervention programs, than parents and children of lower-income families will do. This 'Matthew effect' of the educationally rich getting richer has been observed in many U.S. studies (Walberg & Tsai, 1983). From the point of view of unequal opportunities due to socio-economic status, an implication of our results is that researchers and policy-makers must take into account the synergism between teachers and parents. Otherwise, the problem of unequal opportunities will only be retaylored: a new dress made of an old tissue. The school reform movement has to be more than a deficit approach in disguise.

NOTE

1. In the Dutch educational system the counting of grades starts at Kindergarten level. The 7th and 8th grade correspond to the fifth and the sixth year of primary education.

REFERENCES

Baker, D.P., & Stevenson, D.L. (1986). Mothers' strategies for children's school achievement: Managing the transition to high school. *Sociology of Education, 59,* 156-166.

Bloom, B.S. (1981). *All our children learning: A primer for parents, teachers and other educators.* New York: McGraw-Hill.

Bosker, R.J. (1990). *Extra kansen dankzij de school?* Dissertation. Nijmegen: ITS.

Boxtel, H.W. van, Snijders, J.Th., & Welten, V.J. (1982). *Interesse Intelligentie Schoolvorderingen*. Groningen: Wolters-Noordhoff.

Buriel, R., & Cardoza, D. (1988). Sociocultural correlates of achievement among three generations of Mexican American high school seniors. *American Educational Research Journal, 25*, 177-192.

Carpenter, P.G., & Fleishman, J.A. (1987). Linking intentions and behavior: Australian students' college plans and college attendance. *American Educational Research Journal, 24*, 79-105.

Cooper, B.S. (1991). Parent choice and school involvement: perspectives and dilemmas in The United States and Great Britain. *International Journal of Educational Research, 15*, 235-250.

Cremin, L.A. (1976). *Public education.* New York: Basic Books.

Finn, J.D. (1972). Expectations and the educational environment. *Review of Educational Research, 42*, 387-410.

Graaf, P.M. de (1987). *De invloed van financiële en culturele hulpbronnen in onderwijsloopbanen*. Nijmegen: ITS.

Hoeven-van Doornum, A.A. van der (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen*. Dissertatie. Nijmegen: ITS.

Hoeven-van Doornum, A. A. van der, Voeten, M.J.M., & Jungbluth, P. (1989). The effect of aspiration levels set by teachers for their pupils on learning achievement. In B. Creemers, T. Peters & D. Reynolds (Eds.), *School Effectiveness and School Improvement* (pp. 231-239). Lisse: Swets & Zeitlinger.

Hoeven-van Doornum A.A. van der, Voeten M.J.M. & Jungbluth, P. (1990a). De invloed van klas- en schoolkenmerken op leerprestaties en vervolgadvies in het basisonderwijs. In: C.A.C. Klaassen & P.L.M. Jungbluth (Eds.), *Onderwijs Research dagen 1990. Onderwijs en Samenleving*. Nijmegen: ITS.

Hoeven-van Doornum A.A. van der, Voeten M.J.M. & Jungbluth, P. (1990b). Het effect van leerkrachtverwachtingen en streefniveaus op schoolloopbanen. *Tijdschrift voor Onderwijs Research, 15*, 23-42.

Hossler, D. & Stage F.K. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. *American Educational Research Journal, 29*, 425-451.

Keith, T.Z., Reimers, Th.M., Fehrmann, P.G., Pottebaum, S.M., & Aubey, L.W. (1986). Parental involvement, homework, and TV time: Direct and indirect effects on high school achievement. *Journal of Educational Psychology, 78*, 373-380.

Lee, V.E., Bryk, A.S., & Smith, J.B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (Ed.), *Review of Research in Education, vol. 19* (pp. 171-267). Washington, D.C.: American Educational Research Association.

Leseman, P.P.M. (1989). *Structurele en pedagogische determinanten van schoolloopbanen*. Rotterdam: Rotterdamse Schooladviesdienst. Project Onderwijs en Sociaal Milieu.

Meijnen, G.W. (1984). *Van zes tot twaalf. Een longitudinaal onderzoek naar de milieu- en schooleffekten van loopbanen in het lager onderwijs*. 's-Gravenhage: SVO.

Melnick, C.R.(1991). Parents and teachers as educative partners. Introduction. *International Journal of Educational Research, 15*, 125-130.

Peaker, G.F. (1967). The regression analyses of the national survey. In Plowden Report, *Children and their primary schools, Volume 2: Research and surveys* (pp. 179-221). London: Her Majesty's Stationery Office.

Rupp, J.C. (1969). Opvoeding tot schoolweerbaarheid. Groningen: Wolters-Noordhoff.

Seginer, R. (1983). Parents' educational expectations and children's academic achievements: A literature review. *Merrill-Palmer Quarterly, 29*, 1-23.

Tesser, P. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen: ITS.

Tesser, P., L. Mulder & M.P.C. van der Werf (1991). *De eerste fase van de longitudinale OVB-onderzoeken: het leerlingenonderzoek*. Nijmegen/ Groningen: ITS/ RION.

Walberg, H.J. & Tsai, S.L. (1983). Matthew effects in education. *American Educational Research Journal, 20*, 359-373.

Wallace, T. & Walberg, H.J.(1991). Parental partnerships for learning. *International Journal of Educational Research, 15*, 131-145.

White, K.R., Taylor, M.J. & Moss, V.D. (1992). Does research support claims about the benefits of involving parents in early intervention programs? *Review of Educational Research, 62*, 91-125.

Manuscript ontvangen 13-7-1993
Definitieve versie ontvangen 30-11-1993

Tijdschrift voor Onderwijsresearch, 18 (1993), nr. 6, pp. 380-381

Boekbespreking

E. Wagenaar (1993): *Tweetaligheid in het aanvangsonderwijs. Een onderzoek naar de effecten van tweetalig kleuteronderwijs op de schoolloopbaan van Marokkaanse kinderen.* Universiteit van Amsterdam, dissertatie, 206 pp.

Omtrent de wenselijke vormen van opvang van allochtone leerlingen in het onderwijs lopen de meningen uiteen. Voor een niet gering deel zijn deze meningsverschillen terug te voeren op uiteenlopende doelstellingen. Toen in de zestiger jaren de eerste groepen Zuideuropese leerlingen hun intrede in het onderwijs deden, werden zij aangeduid als kinderen van gastarbeiders. In de veronderstelling dat deze na verloop van tijd terug zouden keren naar hun moederland, leek het wenselijk de kinderen daar ook op voor te bereiden. Onderhoud, en eventueel zelfs verrijking, van hun moedertaal leek destijds dus een voor de hand liggende onderwijsdoelstelling. Met het verstrijken van de tijd bleek de 'terugkeer'-optie een weinig realistische. Permanente vestiging kreeg de overhand. Daardoor begon de legitimering van onderwijs in de eigen (moeder-)taal af te brokkelen.

De meningsvorming omtrent de opvang in het onderwijs van allochtone leerlingen begint vervolgens een meandervormige structuur aan te nemen. De 'onderdompelingsthese' is één stroom: leerlingen dienen uitsluitend in het Nederlands aangesproken en onderwezen te worden. Onderhoud van en instructie in de moedertaal is uit den boze. Het komt de verwerving van de tweede taal (lees: Nederlands) niet ten goede. De 'drempelhypothese' van Cummings is een andere: een zekere vaardigheid in de moedertaal is noodzakelijk om met succes een tweede taal te kunnen leren. Dan is er nog een stroom van sociaal-psychologische origine: onderwijs in de moedertaal draagt bij aan een positief zelfbeeld van leerlingen, hetgeen zowel hun schoolcarrière als hun sociaal functioneren ten goede komt. Zonder volledigheid te pretenderen dient tenslotte ook nog een cultuur-politieke invalshoek genoemd te worden: onderwijs in de moedertaal dient niet op zijn maatschappelijke functionaliteit beoordeeld te worden (wat draagt het bij aan integratie van individuen en groepen), maar op zijn culturele waarde. Friezen stellen prijs op het behoud van onderricht in hun eigen taal, als wezenlijk element van hun cultuur. Waarom worden rechten die dit volksdeel heeft verworven, niet toegezegd aan andere culturele groepen binnen onze landsgrenzen? In het in 1992 verschenen rapport van de Commissie Allochtone Leerlingen in het Onderwijs (de zgn. Commissie Van Kemenade) wordt deze cultuur-politieke optie verdedigd.

Wagenaar legitimeert de uitgangspunten van haar proefschrift in het snijvlak van meerdere opties: "Het meest zinvol lijkt een transitiemodel [...], waarbij van de thuistaal van de kinderen gebruik wordt gemaakt om drie redenen: omdat een hoog beheersingsniveau in de eerste taal voor de algemene (cognitieve, linguïstische en emotionele) ontwikkeling van het kind van belang wordt geacht, om op korte termijn een optimaal onderwijsrendement voor het kind mogelijk te maken, en om de ontwikkeling van de tweede-taalvaardigheid te stimuleren." (p. 65).

De kern van haar onderzoek is een experiment waarbij op een basisschool in Amsterdam een etnisch-homogene kleutergroep voor Marokkaanse kinderen wordt gevormd die vanaf hun entree in de basisschool tweetalig kleuteronderwijs krijgen. Het Marokkaans Arabisch is voertaal gedurende de ochtenden (15 uur) en het Nederlands in de middagen (8 uur). In groep drie wordt de verhouding omgedraaid. Het curriculum dient overeen te komen met een Nederlandstalige kleuterklas.

Het design van haar onderzoek is zo opgezet dat er sprake is van een beginmeting op vierjarige leeftijd, een meting op zesjarige leeftijd, aan het eind van de kleuterperiode (direct na de interventie) en een meting op achtjarige leeftijd, aan het einde van groep vier. Gemeten worden het algemeen cognitief ontwikkelingsniveau, de taalvaardigheid in het Nederlands, de taalvaardigheid in het Marokkaans Arabisch, de schoolvorderingen en de sociaal-emotionele ontwikkeling. De 30 leerlingen van de proefschool zijn gematcht met 30 leerlingen uit twee scholen uit de

omgeving die aldus de controlegroep representeren. Voorts is er op tamelijk ondoorzichtige gronden nog een referentiegroep gecreëerd van 25 Nederlandstalige kinderen en hun ouders, om zicht te krijgen op eventuele verschillen en overeenkomsten tussen allochtone en autochtone kleuters bij de aanvang van het basisonderwijs. Door de weinig doordachte wijze waarop deze groep is samengesteld, kan ze de beoogde functies niet vervullen en de auteur besteedt er in haar verslag dan ook terecht weinig aandacht aan.

De resultaten kunnen als volgt worden samengevat. Op zesjarige leeftijd beheersen de leerlingen uit de proefgroep het Marokkaans Arabisch beter dan de leerlingen uit de controlegroep. Andere verschillen in de talige en cognitieve domeinen treden niet op. Wel worden de leerlingen uit de proefgroep positiever beoordeeld door hun leerkracht dan de controlegroep-leerlingen door hun leerkracht waar het de sociaal-emotionele ontwikkeling betreft. Dat dit een dubieus resultaat is omdat het hier zeer goed een beoordelaarseffect zou kunnen betreffen, vermeldt de auteur nergens.

Op achtjarige leeftijd zijn ook de verschillen in beheersing van de moedertaal en in sociaal-emotioneel functioneren verdwenen. Opvallend is vervolgens dat op dit meetmoment maar liefst zes kinderen uit de proefgroep naar het speciaal onderwijs zijn verdwenen, tegen één uit de controlegroep. De auteur schenkt aan deze leerlingen bijzondere aandacht waar het om hun individuele kenmerken gaat. Zij vergelijkt echter niet meer op match-niveau waardoor duidelijk zou kunnen worden of het hier wellicht om verschillen in verwijzingsbeleid tussen twee scholen gaat.

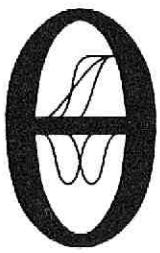
Daarmee komen we bij het hoofdpunt van kritiek. Behalve verschillen in instructietaal kunnen er tal van andere verschillen tussen de proefschool en de twee controlescholen zijn opgetreden. Verschillen die wellicht niet te onderdrukken waren, maar dan toch op zijn minst systematisch beschreven hadden moeten worden. Niets daarvan, zodat de lezer met de vraag blijft zitten: wat is nu eigenlijk precies geëvalueerd? Ongetwijfeld verschillen in instructietaal, maar misschien ook verschillen in leermateriaal en didactische aanpak? Af en toe licht de auteur een tipje van de sluier op, bijvoorbeeld wanneer zij beschrijft hoe de allochtone leidster in de proefklas het moeilijk heeft met de Nederlandse aanpak in het kleuteronderwijs. Ook bleek ze veel lager geschoold te zijn dan haar Nederlandse collega's (dus ook die van de controlescholen).

Kortom, het onderzoeksverslag vertoont nogal wat tekortkomingen met betrekking tot het design. De statistische analyses daarentegen, zijn naar behoren uitgevoerd. Ook het literatuuroverzicht is informatief, up to date, goed gestructureerd en zeer goed leesbaar. Twee landelijke dagbladen kopten op de dag na de verdediging van het proefschrift (vrij vertaald): "Tweetalig onderwijs is succesvol", respectievelijk "Tweetalig onderwijs heeft geen effect". In termen van Kuhn is dit kennelijk geen cruciaal experiment geweest.

G.W. Meijnen
Universiteit van Amsterdam
SCO-Kohnstamm Instituut
Grote Bickersstraat 72
1013 KS Amsterdam

Cito Instituut voor Toetsontwikkeling

PSYCHOMETRIE
IN DE
PRAKTIJK



Redactie
T.J.H.M. Eggen
P.F. Sanders

Psychometrie in de praktijk

- de nieuwste ontwikkelingen uitgewerkt en geïllustreerd met voorbeelden uit de psychometrische praktijk.

Psychometrie in de praktijk

- enige Nederlandstalige boek over klassieke testtheorie, generaliseerbaarheidstheorie en itemresponstheorie
- voor studenten sociale wetenschappenmet kennis van kansrekening en statistiek op inleidend sociaal wetenschappelijk universitair niveau en voor toets- en test-deskundigen
- uitvoerige behandeling van onderwerpen als equivaleren van toetsen, vraag-onzuiverheid, meten van veranderingen, optimaal samenstellen van toetsen, overeenstemming tussen beoordelaars en rapporteren van toetsresultaten.

Psychometrie in de praktijk
554 pagina's, genaaid gebonden, *f* 60,00
bestellen bij Cito, Postbus 1034, dienst Communicatie en verkoop, 6801 MG Arnhem, telefoon (085) 521 590

Cito

BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres
Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen,
tel. 045-762293

Jrg. 17, nr. 3 1993

Inhoudsopgave

Onderwijsonderzoekers in Europa verenigen zich 384

Onderwijsonderzoekers in Europa verenigen zich

Nijs Lagerweij

De inwoners van het kleine Zwitserse stadje Aarau zullen er niet veel van gemerkt hebben. Toch was het een opmerkelijke gebeurtenis: vertegenwoordigers van 15 landen in Europa namen het principe besluit tot oprichting van de "European Educational Research Association". Het komend jaar wordt gezamenlijk gewerkt aan het opstellen en het aanvaarden van de statuten. Men hoopt gereed te zijn voor de tweejaarlijkse ministers conferentie in 1995. Centrale opgave is de positie en de mogelijkheden van onderwijsonderzoek in Europees verband te versterken. Hierna een korte impressie van de besprekingen in Aarau, die ik bijwoonde als vertegenwoordiger van de VOR (Vereniging voor Onderwijsresearch in Nederland) .

Initiatief

De ontmoeting vond plaats in een tweedaagse conferentie van 14 tot 15 oktober 1993 in Aarau in Zwitserland. De Zwitser Armin Gretler had in juni 1992 op de ECER (European Conference on Educational Research) in Enschede toegezegd een bijeenkomst voor te bereiden in zijn land voor vertegenwoordigers van organisaties van onderwijsonderzoekers. Hij heeft daartoe een tekst voor een statuut opgesteld, waarin een Europese samenwerking wordt geregeld tussen organisaties voor onderwijsonderzoekers. Op zijn uitnodiging werd gretig ingegaan. Vanuit vijftien landen kwamen personen bijeen om de tekst van het statuut te becommentariëren. Ook andere landen gaven te verstaan belangstelling te hebben, maar konden dit keer nog geen vertegenwoordiger sturen.

Eerst werd vastgesteld dat er bij alle aanwezigen in principe bereidheid bestond een dergelijke vorm van samenwerking in het leven te roepen. Het eerste initiatief, dat vooral door de Nederlander Tjeerd Plomp in Enschede was genomen, bleek allerwegen instemming te verkrijgen. In het bijzonder werd op samenwerking aangedrongen door een vertegenwoordiger van de OECD uit Parijs, Albert Tuynman, en door een vertegenwoordiger vanuit de Raad van Europa, Michael Vorbeck. Op Europees niveau is er dringend behoefte aan een soort overkoepelende organisatie voor de vele groepen die zich in Europa bezighouden met onderzoek op het gebied van onderwijs. De Europese gemeenschap heeft inmiddels in een nieuwe nota richtlijnen voor activiteiten opgesteld op het gebied van onderwijs en vorming. Dit zogenaamde Ruberti-plan verscheen in mei 1993 en bevat onder meer voorstellen voor programma's als Comett, Erasmus, Iris, Lingua en Tempus. Door het vormen van netwerken, door uitwisseling en door gezamenlijke projecten hoopt men in Europa tot intensieve vormen van samenwerking te komen. Daarbij is ook bescheiden enige aandacht besteed aan onderwijsonderzoek op Europees niveau.

Reputatie

De positie van onderwijsonderzoek in het kader van Europese activiteiten blijkt nogal omstreden. In feite heeft onderwijsonderzoek in ogen van de politici een weinig gunstige reputatie. Vooral van die zijde valt enorm veel negatieve kritiek te vernemen. Men vindt het onderzoek te zeer versnipperd. Er zou sprake zijn van een vergaarbak van veel klein grut zonder betekenis voor de ontwikkeling van onderwijsbeleid. Er zou ook te vaak in verschillende landen hetzelfde onderzocht worden hetgeen geldverspilling zou opleveren. Bovendien zouden de onderzoekers voor buitenstaanders nauwelijks toegankelijk rapporteren over hun werkzaamheden.

Dergelijke opvattingen of waardeoordelen verklaren waarom de relevantie van het werk van onderwijsonderzoekers laag wordt ingeschat. Dit is een teneur die in veel landen aanwezig blijkt. Ook op Europees niveau is er aarzeling gelden te steken in onderwijsonderzoek. Er bestaat steeds minder bereidheid fondsen vrij te maken. In Nederland is de ondermijning van de positie van SVO in de besprekingen over de Wet op de Onderwijsverzorging daarvan een pijnlijke illustratie.

Dat het ook anders kan uitpakken gaf Tuynman aan vanuit zijn betrokkenheid bij een OECD

project. Daar loopt al enkele jaren een project waarin men de onderwijsuitgaven van de verschillende lidstaten vergelijkt. Onlangs werd door dat project duidelijk dat Nederland relatief veel geld stopt in het hoger en universitair onderwijs. Ongetwijfeld heeft Ritzen daarin argumenten gevonden om het zogenaamde stapelen van opleidingen tegen te gaan.

Gemeenschappelijk belang

Bovenstaande situatie betekent alle reden voor een gemeenschappelijk overleg op Europees niveau om de sterke en zwakke kanten van onderwijsonderzoek te verkennen en wegen ter verbetering te ontwikkelen. Artikel 2 van het statuut voor een Europese samenwerking bevat dan ook als doelstelling: "The aim and the objectives of the Association are to promote research in education in Europe and thus to promote educational science and to improve education."

De problemen voor een dergelijke vorm van samenwerking bleken groter dan zo op het oog te verwachten is. Niet geheel duidelijk is wat zoal tot onderwijsonderzoek ("educational research") gerekend kan worden. In het bijzonder de koppeling met "development" roept vragen op betreffende de afgrenzing van onderzoek en ontwikkelingswerk.

Eveneens problematisch is de kwestie wie lid kunnen worden van de Association. Moeten dat alleen nationale organisaties van onderwijsonderzoekers zijn, of kunnen ook individuele lidmaatschappen voorkomen? Welke landen kunnen toetreden, of anders gezegd waar houdt Europa op en beginnen andere werelddelen? (Het antwoord luidt dat Europa ophoudt bij de grens van Siberië.)

Nog onvoldoende helder is welke activiteiten door de Association ontplooid kunnen worden. Hoe aantrekkelijk zijn die voor de potentiële deelnemers? Wie betaalt de kosten voor een lidmaatschap en hoe hoog moeten die zijn? Zal er ook een gemeenschappelijk tijdschrift voor Europees onderwijsonderzoek komen? Er heeft zich inmiddels al een uitgever bereid verklaard met een nieuw blad op de markt te komen.

Vooral van Franse zijde werd opgemerkt dat het nog steeds geen vanzelfsprekende zaak is om het Engels tot voertaal te maken in Europa. Er zijn ook andere landen die daarbij vraagtekens hebben.

Afspraken

Deze tweede ontmoeting stond in het teken van voorzichtig aftasten hoe ver men tot overeenstemming kan komen over doelen en werkwijzen van een Europees samenwerkingsverband. Zo gauw er sprake is van internationaal overleg is er ook de noodzaak van behoedzaam manoeuvreren. Culturele en politieke verschillen tussen landen maken dat overeenstemming over samenwerking niet via een leien dakje tot stand komt.

Om meer zicht te krijgen op de opinies omtrent de nieuwe organisatie is besloten de statuten hier en daar op grond van de gevoerde discussies te herzien. In januari 1993 krijgen alle deelnemers een nieuwe tekst toegezonden, die in eigen kring besproken kan worden. Het commentaar wordt weer verzameld door Gretler. Die zal op grond van de reacties met vijf andere personen (resp. uit Tsjechië, Zweden, Engeland, Italië en Frankrijk) een voorbereidende bespreking houden in april 1993. Dat overigens alleen als er ernstige meningsverschillen manifest geworden zijn. In elk geval moet op 18 juni 1993 de definitieve afrondende besluitvorming met alle betrokken vertegenwoordigers plaats vinden. Het is mogelijk dat in september 1994 er een (Europees) ECER94 verbonden wordt met de jaarlijkse conferentie van de British Educational Research Association (BERA).